Aspects Of Copulas And Goodness-Of-Fit

by

Tchilabalo Abozou Kpanzou

Assignment presented in partial fulfilment of the requirements for the degree of

Master of Commerce at Stellenbosch University

Supervisor:

Prof. Tertius De Wet

December 2008

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 15 December 2008

Abstract

The goodness-of-fit of a statistical model describes how well it fits a set of observations. Measures of goodness-of-fit typically summarize the discrepancy between observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing, for example to test for normality, to test whether two samples are drawn from identical distributions, or whether outcome frequencies follow a specified distribution. Goodness-of-fit for copulas is a special case of the more general problem of testing multivariate models, but is complicated due to the difficulty of specifying marginal distributions.

In this thesis, the goodness-of-fit test statistics for general distributions and the tests for copulas are investigated, but prior to that an understanding of copulas and their properties is developed. In fact copulas are useful tools for understanding relationships among multivariate variables, and are important tools for describing the dependence structure between random variables. Several univariate, bivariate and multivariate test statistics are investigated, the emphasis being on tests for normality. Among goodness-of-fit tests for copulas, tests based on the probability integral transform, Rosenblatt's transformation, as well as some dimension reduction techniques are considered. Bootstrap procedures are also described. Simulation studies are conducted to first compare the power of rejection of the null hypothesis of the Clayton copula by four different test statistics under the alternative of the Gumbel-Hougaard copula, and also to compare the power of rejection of the null hypothesis of the Gumbel-Hougaard copula under the alternative of the Clayton copula. An application of the described techniques is made to a practical data set.

Uittreksel

Die passing van 'n statistiese model beskryf hoe goed die model pas op 'n stel data. Maatstawwe van passing gee gewoonlik die afwyking tussen waargenome waardes en die waardes wat verwag word onder die model ter sprake. Sodanige maatstawwe kan gebruik word in statistiese hipoteses, byvoorbeeld in toetse vir normaliteit, om te toets of twee steekproewe uit dieselfde verdeling kom of om te toets of gegewe frekwensies ooreenkom met 'n bepaalde verdeling. Passingstoetse vir copulas is 'n spesiale geval van die meer algemene probleem om te toets vir 'n meerveranderlike model, maar word bemoeilik deur die nodigheid van spesifisering van marginale verdelings.

In hierdie tesis word passingstoetse vir algemene verdelings asook vir copulas, ondersoek. Vooraf word daar egter eers aandag gegee aan die verstaan van copulas en hul eienskappe. Copulas is baie geskik om die verwantskappe tussen veranderlikes te verstaan en is belangrike gereedskap om die afhanklikheid tussen stogastiese veranderlikes te beskryf. 'n Verskeidenheid van eenveranderlike-, tweeveranderlike- en meerverandelike toets statistieke word ondersoek, met die klem op toetse vir normaliteit. As passingstoetse gebaseer op copulas, word aandag gegee aan die waarskynlikheidsintegraal transformasie, Rosenblatt se transformasie asook sekere dimensie verminderingstegnieke. Tegnieke gebaseer op die skoenlus word ook beskou. Simulasie studies word gebruik om die onderskeidingsvermoë van vier statistieke te bepaal vir die toets van die Clayton copula teenoor die Gumbel-Hougaard copula, asook die omgekeerde. Die tegnieke wat beskryf is word dan, ter illustrasie, toegepas op 'n praktiese datastel.

Acknowledgements

- I am very grateful to my supervisor, Prof T. De Wet, for his guidance throughout this project. I would also like to thank him for securing the financial assistance which made it possible for me to undertake this study.
- I would like to thank Dr P.J.U. Van Deventer for the description of the data.
- I am very grateful to Prof N.J. Le Roux, for helping me develop my skills in R/S-PLUS programming.
- I would like to express my sincere appreciation to my colleague and friend, Mr A.M. La Grange, who was never too busy to help.
- Finally I would like to thank my parents for their continued love, support and encouragement throughout the years, and for helping me to keep courageous and strong although I am far from them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The objective of statistics is to extract information from data in order to better explain the situations that these data portray. In other words, it is to describe a certain real phenomenon using data from that phenomenon. In fact the objects of the real world can not be described in such a complete and exact way that can form the basis of an exact theory. In order to carry out statistical inference, many techniques have been developed through the years. These techniques include point estimation, interval estimation and hypothesis testing, and are usually based on limited samples. Once a description of a real phenomenon is made through a model, some rules are needed in order to establish the correspondence between the idealized model and the real world. A statistical problem encountered in many areas of research is the need to assess whether a sample of observations comes from a specified distribution. Typically such situations are known as goodness-of-fit (GOF) problems.

The objective of this thesis is to investigate some goodness-of-fit techniques and applications to copulas. The goodness-of-fit of a statistical model describes how well it fits a set of observations. Measures of goodness-of-fit typically summarize the discrepancy between observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing, for example to test for normality, to test whether two samples are drawn from identical distributions, or whether outcome frequencies follow a specified distribution. Whether data are univariate or multivariate, continuous or categorical, researchers are interested in determining if the observed data differ from the expected data. A measure of how well the null, hypothesized or expected distribution fits the observed data, underlies the basic concept in the area of GOF statistics. The basic reasoning underlying most statistical hypothesis tests can be summarized as follows:

1. Choose a test statistic T, whose distribution is known when the null hypothesis is true;

2. Use the distribution of T to calculate the probability p of observing a value of T more extreme than its observed value, given that the null hypothesis is true;

3. Given a significance level α, reject the null hypothesis if $p < \alpha$.

In fact a hypothesis test requires formulation of null and alternative hypotheses. The confidence level of the test is then defined as the probability of not rejecting the null hypothesis given it is true, and the power of the test is the probability of rejecting the null hypothesis given the alternative is true. In this thesis, we describe several test statistics, and we investigate the goodness-of-fit tests for copulas. But prior to that, we give an overview of copulas and some properties. The outline of the thesis is as follows.

In Chapters 2 and 3 we describe copulas and we provide some important properties. In fact, the study of the relationship between two or more random variables remains an important problem in statistical inference, and copulas proved to constitute a convenient way to express joint distributions. In Chapter 4 we provide ways to generate copulas when we are given a data set.

The regression analysis is a statistical technique intensively used to measure the degree of relationship between two or more variables. In Chapter 5 we discuss an alternative way of looking at regression analysis by using copulas.

As described in [35], goodness-of-fit tests can be put into two classes.

1. The first class of tests divides the range of the data into disjoint bins; the number of observations falling in each bin is compared to the expected number under the hypothesized distribution. These tests can be used for both discrete and continuous distributions although they are most natural for discrete distributions as the definition of the bins tends to be less arbitrary for discrete distributions than it is for continuous distributions.

2. The second class of tests are used almost exclusively for testing continuous distributions. For these tests, empirical distribution functions of the data are compared to the hypothesized distribution functions. The test statistics for these tests are based either on some measure of distance between the two distributions, or on a measure of correlation between them.

In Chapter 6 we describe several goodness-of-fit test statistics, and we look at the asymptotic behavior of some of them. There are in fact many situations in statistics where we want to test whether a particular distribution fits our observed data. In some cases, these tests are informal; for example in linear regression modeling, a statistician usually examines diagnostic plots (or other procedures) that allow him to determine whether the particular model assumptions (for example normality and/or independence) are satisfied. However, in other cases where the form of the model has more significance, statisticians tend to rely more and more on formal hypothesis testing. We concentrate on these methods in that chapter. The univariate, the bivariate and the multivariate tests are investigated. In many cases, it is not easy to obtain true p-values and so computer-based methods are used. One of these methods is the bootstrap. Therefore some bootstrap goodness-of-fit methods are also considered. In Chapter 7 we discuss goodness-of-fit tests for copulas. In fact, goodness-of-fit testing for copulas recently emerged as a challenging inferential problem and some approaches have been proposed in the literature. We also conduct simulation studies to test the null hypothesis of the Clayton copula against that of the Gumbel-Hougaard copulas, and vice versa, through four test statistics. An application of the previously mentioned methods is made to a practical data set in Chapter 8. The data consist of partials of a carillon of bells in the University library of the Catholic University of Leuven in Belgium. The bells were founded in years 1928 and 1983 by respectively Gillett & Johnston and Eisjbouts.

Chapter 2

Notion of Copulas and Some Examples

The study of the relationship between two or more random variables remains an important problem in statistical inference, and copulas proved to constitute a convenient way to express joint distributions. Copulas provide one of the most widely used tools to study multivariate outcomes, giving a useful tool to assist in the process of model building. In this chapter we define copulas and we give some examples.

2.1 Notion of Copulas

The term Copula comes from the Latin noun which means "a link, tie, bond" (see [41]) referring to joining together. With this meaning, a copula is defined as a function that joins multivariate distribution functions to their one-dimensional marginal distribution functions. It is a multivariate distribution function defined on the unit n-cube $[0,1]^n$, with uniformly distributed marginals. Before we give a formal definition of a copula, let us define the H-volume of an n-box.

Definition 2.1. Let S_1, S_2, ..., S_n be nonempty subsets of $\overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ is the extended real line $[-\infty, \infty]$, and let H be an n-dimensional real function whose domain, $\operatorname{Dom} H$, is given by $\operatorname{Dom} H = S_1 \times S_2 \times ... \times S_n$. Let $B = [a, b]$ be an n-box whose vertices are all in $\operatorname{Dom} H$. The H-volume of B is given by

$$V_H(B) = \sum_{c \in B} \operatorname{Sign}(c) H(c),$$

where $\mathrm{Sign}(c)$ is given by

$$\mathrm{Sign}(c) = \begin{cases} 1 & \text{if } c_k = a_k, \text{ for an even number of } k\text{'s} \\ -1 & \text{if } c_k = a_k, \text{ for an odd number of } k\text{'s} \end{cases}$$

(Notice that $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots,\ b_n)$, $c = (c_1, c_2, \ldots,\ c_n)$, and $B = [a, b] = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$ is well defined if $a_k < b_k$ for all k).

Now let us give a formal definition of a copula.

Definition 2.2. An n-dimensional copula is a function $C : [0, 1]^n \to [0, 1]$, with the following properties:

1. C is grounded. This means that for every $u = (u_1, u_2, \ldots,\ u_n) \in [0, 1]^n$, $C(u) = 0$ if at least one coordinate u_i is zero, $i = 1, 2, \ldots,\ n$,
2. C is n-increasing. This means that for every $u \in [0, 1]^n$ and $v \in [0, 1]^n$ such that $u \leq v$, the C-volume $V_C([u, v])$ of the box $[u, v]$ is non-negative,
3. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$, for all $u_i \in [0, 1]$, $i = 1, 2, \ldots, n$.

For $n = 2$ this definition is reduced to the one following, which is easy to deal with.

Definition 2.3. A two-dimensional (bivariate) copula is a function $C : [0, 1]^2 \to [0, 1]$, with the following properties:

1. C is grounded: for all u, $v \in [0, 1]$, $C(u, 0) = 0$ and $C(0, v) = 0$
2. C is 2-increasing: for all u_1,u_2,v_1,$v_2 \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

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

3. For all u,$v \in [0, 1]$, $C(u, 1) = u$ and $C(1, v) = v$.

Remark 2.1. The word copula was first employed in a mathematical or statistical sense by Sklar (1959). (See [41].) Copulas have recently become popular in financial and insurance applications.

2.2 Examples of Copulas

In this section we give some examples of copulas.

Example 2.1. Marshal-Olkin family (1967).

If $\alpha, \beta \in [0,1]$, then the function $C_{\alpha,\beta} : [0,1]^2 \to [0,1]$, defined by

$$C_{\alpha,\beta}(u,v) = \min(u^{1-\alpha}v, uv^{1-\beta}),$$

is a bivariate copula function. This two-parameter family of copulas is the Marshal-Olkin family (see [37]).

Example 2.2. Bivariate Pareto copula.

This copula is defined by the following formula (see [19])

$$C_\alpha(u,v) = u + v - 1 + [(1-u)^{-\frac{1}{\alpha}} + (1-v)^{-\frac{1}{\alpha}}]^{-\alpha},$$

where α is a parameter ($\alpha \in \mathbb{R} \setminus \{0\}$).

Example 2.3. Farlie-Gumbel-Morgenstern family.

If $\theta \in [-1,1]$, then the function C_θ defined on $[0,1]^2$ by

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v),$$

is a one-parameter bivariate copula. This family is known as the Farlie-Gumbel-Morgenstern family (see [46]).

Example 2.4. Cuadras-Augé family of copulas.

Let $\theta \in [0,1]$. The function C_θ defined by

$$C_\theta(u,v) = [\min(u,v)]^\theta [uv]^{1-\theta} = \begin{cases} uv^{1-\theta}, & \text{if } u \le v \\ u^{1-\theta}v, & \text{if } u \ge v \end{cases},$$

is a copula function. This family is known as the Cuadras-Augé family of copulas (see [41], page 15).

Example 2.5. The Frechet and Mardia family of copulas.

The Frechet and Mardia copula is defined by

$$C(u,v) = \theta_1 \min\{u,v\} + (1-\theta_1-\theta_2)uv + \theta_2 \max\{u+v-1,0\},$$

where $\theta_1, \theta_2 \in [0,1]$ and $\theta_1 + \theta_2 \leq 1$.

Example 2.6. The Rodriquez-Lellena and Ubena-Flores family of copulas.

A copula from this family has the form

$$C(u,v) = uv + f(u)g(v),$$

where

1. $f(0) = f(1) = g(0) = g(1) = 0$
2. f and g are absolutely continuous and
3. $\min\{\alpha\delta, \beta\gamma\} \geq 1$,

where

$$\alpha = \inf\{f'(u) : u \in A\} < 0,$$

$$\beta = \sup\{f'(u) : u \in A\} > 0,$$

$$\gamma = \inf\{g'(v) : v \in B\} < 0,$$

$$\delta = \sup\{g'(v) : v \in B\} > 0,$$

with

$$A = \{u \in [0,1] : f'(u) \text{ exists}\}$$

and

$$B = \{v \in [0,1] : g'(v) \text{ exists}\}.$$

Example 2.7. The Gaussian Copula.

This copula is simply derived from a multivariate Gaussian distribution function Φ_Σ with mean zero and correlation matrix Σ by transforming the marginals by the inverse of the standard normal distribution function Φ. It is given by (see [39])

$$C(x_1, x_2, \ldots, x_d) = \Phi_\Sigma(\Phi^{-1}(x_1), \Phi^{-1}(x_2), \ldots, \Phi^{-1}(x_d)).$$

A bivariate Gaussian copula is defined by

$$\begin{aligned} C(u,v;\rho) &= \Phi_\rho(\Phi^{-1}(u),\Phi^{-1}(v)) \\ &= \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-(s^2-2\rho st+t^2)}{2(1-\rho^2)}\right\} dsdt, \end{aligned}$$

where Φ denotes the distribution function of the univariate standard normal distribution and $\Phi_\rho(.,.)$ denotes the distribution function of the bivariate standard normal distribution with correlation parameter ρ such that $-1<\rho<1$.

Note that

$$\lim_{\rho\to+1} C(u,v;\rho) = \min\{u,v\},$$

$$\lim_{\rho\to-1} C(u,v;\rho) = \max\{u+v-1,0\}$$

and

$$C(u,v;0) = u.v$$

for $(u,v)\in[0,1]^2$.

Example 2.8. The t-Copula.

The t-copula is derived in the same way as the Gaussian copula. Given a multivariate centered t-distribution function $t_{\Sigma,\nu}$ with correlation matrix Σ, ν degrees of freedom and with marginal distribution function t_ν, this copula is given by (see [39])

$$C(x_1,x_2,\ldots,x_d) = t_{\Sigma,\nu}(t_\nu^{-1}(x_1),t_\nu^{-1}(x_2),\ldots,t_\nu^{-1}(x_d)).$$

A bivariate t_ν-copula is defined by

$$C(u,v;\nu,\rho) = \int_{-\infty}^{F_\nu^{-1}(u)}\int_{-\infty}^{F_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{1+\frac{(s^2-2\rho st+t^2)}{\nu(1-\rho^2)}\right\}^{\frac{-(\nu+2)}{2}} dsdt,$$

where F_ν denotes the univariate distribution function of a t-distribution with ν degrees of freedom, and the parameters are ν and ρ such that $\nu\in\mathbb{N}$ and $-1<\rho<1$.

2.3 Bivariate Extreme Value Copulas

This family of copulas is obtained by using bivariate extreme value distributions. A bivariate extreme value copula has the form

$$C_A(u,v) = \exp\left[\log(uv)A\left\{\frac{\log(u)}{\log(v)}\right\}\right],$$

where the dependence function, A, defined on $[0,1]$ is convex and such that

$$\max(t, 1-t) \leq A(t) \leq 1, \text{ for all } t \in [0,1].$$

The most common parametric models of bivariate extreme value copulas are given in Table 2.1.

Model	$A_\theta(t)$	$C_{A_\theta}(u,v)$
Gumbel[28]	$\theta t^2 - \theta t + 1$, $\theta \in (0,1)$	$uv\exp(-\theta\frac{\log(u)\log(v)}{\log(uv)})$
Gumbel-Hougaard	$[t^{\frac{1}{1-\theta}} + (1-t)^{\frac{1}{1-\theta}}]^{1-\theta}$, $\theta \in (0,1)$	$\exp\left\{-[\lvert\log(u)\rvert^{\frac{1}{1-\theta}} + \lvert\log(v)\rvert^{\frac{1}{1-\theta}}]^{1-\theta}\right\}$
Galambos	$1-[t^{-\theta} + (1-t)^{-\theta}]^{-\frac{1}{\theta}}$, $\theta \in (0,\infty)$	$uv\exp\left\{(\lvert\log(u)\rvert^{-\theta} + \lvert\log(v)\rvert^{-\theta})^{-\frac{1}{\theta}}\right\}$
Generalized Marshall-Olkin[38]	$\max\{1-\theta_1 t, 1-\theta_2(1-t)\}$, $(\theta_1,\theta_2) \in (0,1)^2$	$u^{1-\theta_1}v^{1-\theta_2}\min(u^{\theta_1}, v^{\theta_2})$

Table 2.1: Families of bivariate extreme value copulas

2.4 Archimedean Copulas

Definition 2.4. A copula is an Archimedean copula if it can be expressed in the form

$$C_\phi(u_1,u_2,\ldots,u_n) = \phi^{-1}\{\phi(u_1) + \phi(u_2) + \ldots + \phi(u_n)\},$$

where $\phi : [0,1] \to [0,\infty)$ is a bijection such that $\phi(1) = 0$ and

$$(-1)^i \tfrac{d^i}{dx^i}\phi^{-1}(x) > 0,\ i \in \mathbb{N} \text{ (see [21]).}$$

ϕ is called the generator of the copula C_ϕ.

One key characteristic of Archimedean copulas is the fact that all the information about the n-dimensional dependence structure is contained in a univariate generator ϕ. So the Archimedean representation allows the study of a multivariate copula to be reduced to a single univariate function.

Some important families of Archimedean copulas are given in Table 2.2.

Family	Generator $\phi(t)$	Bivariate copula $C_\phi(u,v)$
Independence	$-\log(t)$	uv
Clayton[5], Cook-Johnson[6], Oakes[43]	$\frac{t^{-\alpha}-1}{\alpha}$, $\alpha \in (0,\infty)$	$(u^{-\alpha}+v^{-\alpha}-1)^{-\frac{1}{\alpha}}$
Gumbel[27], Hougaard[29]	$(-\log(t))^\alpha$, $\alpha \in [1,\infty)$	$\exp\left\{-[(-\log(u))^\alpha+(-\log(v))^\alpha]^{\frac{1}{\alpha}}\right\}$
Frank[17]	$\log(\frac{e^{\alpha t}-1}{e^\alpha-1})$, $\alpha \in R \setminus \{0\}$	$\frac{1}{\alpha}\log\left\{1+\frac{(e^{\alpha u}-1)(e^{\alpha v}-1)}{e^\alpha-1}\right\}$

Table 2.2: Families of bivariate Archimedean copulas

Chapter 3

Some Properties of Copulas

3.1 Sklar's Theorem

The importance of copulas in statistics is described in Sklar's theorem. In this sense, this theorem is considered as the central theorem of copula theory.

Theorem 3.1. (Sklar). See [41], page 17.

Let H be an n-dimensional distribution function with marginals $F_1, F_2, \ldots, F_n$. Then there exists an n-copula C such that for all $x_1, x_2, \ldots, x_n \in \overline{\mathbb{R}}$,

$$H(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \tag{3.1}$$

Conversely, if C is an n-copula and $F_1, F_2, \ldots, F_n$ are distribution functions, then the function H defined by Equation (3.1) is an n-dimensional distribution with marginals $F_1, F_2, \ldots, F_n$. Furthermore, if the marginals are all continuous, then C is unique. Otherwise C is uniquely determined on $\operatorname{Ran} F_1 \times \operatorname{Ran} F_2 \times \ldots \times \operatorname{Ran} F_n$, where $\operatorname{Ran} F_i$ is the range of the function F_i.

For $n = 2$, we have the corresponding theorem in two dimensions.

Theorem 3.2. (Sklar in two dimensions)

Let H be a joint distribution function with the marginals F and G. There exists a copula C such that for all x and y in $\overline{\mathbb{R}}$,

$$H(x, y) = C(F(x), G(y)). \tag{3.2}$$

If F and G are continuous, then the copula C is unique; otherwise it is uniquely determined on $\operatorname{Ran} F \times \operatorname{Ran} G$. Conversely, if C is a copula, and F, G are distribution functions, then the function H defined by Equation (3.2) is a distribution function with marginals F and G (see [41], page 18).

With this important theorem we see that the copula function is one of the most useful tools for dealing with multivariate distribution functions with given or known univariate marginals.

We now focus on bivariate copulas.

3.2 Continuity, Differentiability and Invariance

Theorem 3.3. (Continuity)

Let C be a bivariate copula. Then for all $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 < u_2$ and $v_1 \leq v_2$,

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

which means that C is uniformly continuous in its domain (see [41]).

Proof. Let $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 < u_2$ and $v_1 \leq v_2$. Let γ_1 be the track passing through the points (u_1, v_1) and (u_2, v_1), and let γ_2 be a track passing through the points (u_2, v_1) and (u_2, v_2). There exist copulas C_{γ_1} and C_{γ_2} such that

$$C(u_1, v_1) = C_{\gamma_1}(u_1, v_1),\ C(u_2, v_2) = C_{\gamma_2}(u_2, v_2),\ C(u_2, v_1) = C_{\gamma_1}(u_2, v_1) = C_{\gamma_2}(u_2, v_1)$$

Therefore,

$$\begin{aligned}
|C(u_2, v_2) - C(u_1, v_1)| &\leq |C(u_2, v_2) - C(u_2, v_1)| + |C(u_2, v_1) - C(u_1, v_1)| \\
&= |C_{\gamma_2}(u_2, v_2) - C_{\gamma_2}(u_2, v_1)| + |C_{\gamma_1}(u_2, v_1) - C_{\gamma_1}(u_1, v_1)| \\
&\leq |v_2 - v_1| + |u_2 - u_1|,
\end{aligned}$$

the last inequality following from the fact that copulas satisfy Lipschitz's condition (see Lemma 6.1.9 in Schweizer and Sklar [48]). $\square$

Theorem 3.4. (Differentiability)

Let C be a bivariate copula. For any $v \in [0,1]$, the partial derivative $\frac{\partial C}{\partial u}(u, v)$ exists for almost all $u \in [0,1]$, and for such v and u,

$$0 \leq \tfrac{\partial C}{\partial u}(u, v) \leq 1.$$

Similarly, for any $u \in [0,1]$, the partial derivative $\frac{\partial C}{\partial v}(u,v)$ exists for almost all $v \in [0,1]$, and for such u and v,

$$0 \leq \frac{\partial C}{\partial v}(u,v) \leq 1.$$

Furthermore, the functions $u \mapsto \frac{\partial C}{\partial v}(u,v)$ and $v \mapsto \frac{\partial C}{\partial u}(u,v)$ are well-defined and non-decreasing almost everywhere on $[0,1]$ (see [41]).

Theorem 3.5. (Invariance)

Copulas are invariant under strictly monotone transformations of the random variables.

Proof. Let X_1 and X_2 be continuously distributed random variables with copula C, and let T_1, T_2 be strictly increasing transformation functions. Our aim is to prove that $T_1(X_1)$ and $T_2(X_2)$ have the same copula as X_1 and X_2. Let F_1 and F_2 be distribution functions of X_1 and X_2 respectively, and let T_1^{-1} and T_2^{-1} be the inverse functions of T_1 and T_2 respectively. Let G_1 and G_2 be the distribution functions of $T_1(X_1)$ and $T_2(X_2)$ respectively, and let C_T be the copula for $T_1(X_1)$ and $T_2(X_2)$. We have for $i \in \{1,2\}$,

$$\begin{aligned} G_i(x_i) &= P[T_i(X_i) \leq x_i] \\ &= P[X_i \leq T_i^{-1}(x_i)] \\ &= F_i(T_i^{-1}(x_i)). \end{aligned}$$

Therefore,

$$\begin{aligned} C_T(G_1(x_1), G_2(x_2)) &= P[T_1(X_1) \leq x_1, T_2(X_2) \leq x_2] \\ &= P[X_1 \leq T_1^{-1}(x_1), X_2 \leq T_2^{-1}(x_2)] \\ &= C(F_1(T_1^{-1}(x_1)), F_2(T_2^{-1}(x_2))) \\ &= C(G_1(x_1), G_2(x_2)). \end{aligned}$$

Hence $C_T = C$ in $[0,1]^2$, which means that copulas are invariant under strictly increasing transformations of random variables. Similarly one can verify that copulas are invariant under strictly decreasing transformations of random variables. □

3.3 Frechet-Hoeffding Bounds

Theorem 3.6. For every copula C and every $(u,v) \in [0,1]^2$,

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

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

and

$$M(u,v) = \min(u,v)$$

are themselves copulas (see [41], Theorem 2.2.3, page 11).

Proof. Let C be a bivariate copula. Let X and Y be random variables with copula C. Let F and G be distribution functions of X and Y respectively, and let H be the joint distribution function. We have

$$P[X \leq x, Y \leq y] \leq P[X \leq x],$$

and

$$P[X \leq x, Y \leq y] \leq P[Y \leq y],$$

so

$$P[X \leq x, Y \leq y] \leq \min(P[X \leq x], P[Y \leq y])$$

Moreover,

$$P[X \leq x, Y \leq y] = P[X \leq x] + P[Y \leq y] + P[X > x, Y > y] - 1.$$

Since

$$P[X > x, Y > y] \geq 0,$$

we have

$$P[X \leq x] + P[Y \leq y] - 1 \leq P[X \leq x] + P[Y \leq y] + P[X > x, Y > y] - 1,$$

which means that,

$$P[X \leq x] + P[Y \leq y] - 1 \leq P[X \leq x, Y \leq y].$$

Therefore,

$$\max(P[X \leq x] + P[Y \leq y] - 1, 0) \leq P[X \leq x, Y \leq y].$$

It follows that,

$$\max(F(x) + G(y) - 1, 0) \leq H(x,y) \leq \min(F(x), G(y)),$$

for all x and y, hence

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

The proof that $W(u,v) = \max(u+v-1,0)$ and $M(u,v) = \min(u,v)$ are copulas can be found in [56]. □

3.4 Copulas and Association

This section contains different ways in which copulas can be used in the study of dependence between random variables.

3.4.1 Kendall's Tau

Kendall's tau measure of a pair (X,Y), distributed according to H, is defined as the difference between the probabilities of concordance and discordance for two independent pairs (X_1,Y_1) and (X_2,Y_2) each with distribution H; that is

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \tag{3.3}$$

3.4.2 Spearman's Rho

Let (X_1,Y_1), (X_2,Y_2) and (X_3,Y_3) be three independent random vectors, copies of a random vector (X,Y), with a common joint distribution function H. Spearman's rho associated with (X,Y), distributed according to H, is defined by

$$\rho = 3P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]. \tag{3.4}$$

Remark 3.1. If C is the copula associated with (X,Y), distributed according to H, then Kendall's tau and Spearman's rho can be written in the forms (see [46]):

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1, \tag{3.5}$$

$$\rho = 12\int_0^1\int_0^1 (C(u,v) - uv)dudv. \tag{3.6}$$

3.4.3 Schweizer and Wolff's Sigma

If we replace the function, $(u,v) \mapsto C(u,v) - uv$, in Equation (3.6) by its absolute value, then we obtain Schweizer and Wolff's Sigma given by (see [46])

$$\sigma = 12 \int_0^1 \int_0^1 |C(u,v) - uv| du dv. \tag{3.7}$$

3.5 Tail Dependence

Definition 3.1. Let X and Y be random variables with distribution functions F and G respectively. Let $U = F(X)$ and $V = G(Y)$.

The coefficient of upper tail dependence is defined as

$$\lambda_U = \lim_{u \to 1^-} P[V > u | U > u], \tag{3.8}$$

provided this limit exists ($\lambda_U \in [0,1]$).

The coefficient of lower tail dependence is defined as

$$\lambda_L = \lim_{u \to 0^+} P[V \leq u | U \leq u], \tag{3.9}$$

provided this limit exists ($\lambda_L \in [0,1]$).

Interpretation 3.1. The coefficients λ_U and λ_L are interpreted as follow:

1. If $\lambda_U = 0$, then X and Y are independent in the upper tail.
2. If $\lambda_U \in (0,1]$, then X and Y are dependent in the upper tail.
3. If $\lambda_L = 0$, then X and Y are independent in the lower tail.
4. If $\lambda_L \in (0,1]$, then X and Y are dependent in the lower tail.

Proposition 3.7. Let C be a copula associated with (X,Y).

If

$$\lim_{u \to 1^-} \left(\frac{1 - 2u + C(u,u)}{1-u}\right)$$

and

$$\lim_{u \to 0^+} \left(\frac{C(u,u)}{u}\right)$$

exist, then λ_U and λ_L are given by

$$\lambda_U = \lim_{u \to 1^-} \left(\frac{1 - 2u + C(u,u)}{1-u}\right)$$

and

$$\lambda_L = \lim_{u \to 0^+} \left(\frac{C(u,u)}{u}\right).$$

Remark 3.2. We now find λ_U and λ_L for the Archimedean copulas.

Let C be an Archimedean copula generated by ϕ, i.e, $C(u,v) = \phi^{-1}(\phi(u) + \phi(v))$.

Using [1]l'Hopital's rule and the fact that

$$(\phi^{-1})'(y) = \frac{1}{\phi'(\phi^{-1}(y))},$$

λ_U and λ_L are given by

$$\lambda_U = 2 - 2 \lim_{u \to 1^-} \frac{\phi'(u)}{\phi'(\phi^{-1}(2\phi(u)))},$$

$$\lambda_L = 2 \lim_{u \to 0^+} \frac{\phi'(u)}{\phi'(\phi^{-1}(2\phi(u)))}.$$

3.6 Methods of Generating Copulas

In this section we present some methods of constructing bivariate copulas. We particularly focus on two illustrations: the Marshall-Olkin Bivariate Exponential family and the Bivariate Pareto Model. To start, let us define the survival function and the survival copula.

Definition 3.2. For a pair (X, Y) of random variables with joint distribution function H, the joint survival function is defined by

$$\overline{H}(x,y) = P[X > x, Y > y]. \tag{3.10}$$

The marginals of $\overline{H}$ are the functions $\overline{H}(-\infty, y)$ and $\overline{H}(x, -\infty)$ which are univariate survival functions $\overline{F}$ and $\overline{G}$, where F and G are the distribution functions of X and Y respectively.

[1]L'Hopital's rule: Let c be either a finite number or ∞.

If $\lim_{x \to c} f(x) = 0$ and $\lim_{x \to c} g(x) = 0,$ then $\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}.$

Definition 3.3. If C is a copula for X and Y, then the survival copula of X and Y is the function $\widehat{C} : [0,1]^2 \to [0,1]$, given by (see [41], page 32)

$$\widehat{C}(u,v) = u + v - 1 + C(1-u, 1-v). \tag{3.11}$$

Also, if $\overline{C}$ is the joint survival function for two uniform $(0,1)$ random variables U and V whose joint distribution function is the copula C, then we have (see [41], page 33)

$$\overline{C}(u,v) = 1 - u - v + C(u,v) = \widehat{C}(1-u, 1-v). \tag{3.12}$$

3.6.1 The Inversion Method

Let H be a bivariate distribution function with continuous marginals F and G. A copula C can be constructed by using Sklar's Theorem through the relation

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)). \tag{3.13}$$

Using the survival function $\overline{H}$, we can also construct a survival copula by the relation

$$\widehat{C}(u,v) = \overline{H}(\overline{F}^{-1}(u), \overline{G}^{-1}(v)), \tag{3.14}$$

where $\overline{F}$ and $\overline{G}$ are taken as in Definition 3.2.

Let us now use this method to construct the Marshal-Olkin Bivariate Exponential family and the Bivariate Pareto Model.

Example 3.1. We consider a two-component system such as a two engine aircraft. The components are subject to "Shocks", which are always "fatal" to one or both of the components. For example one of the two aircraft engines may fail, or both of them could be destroyed simultaneously. Let X and Y denote the lifetimes of the components 1 and 2, respectively. The survival function $\overline{H}$ is given by

$$\overline{H}(x,y) = P[X > x, Y > y],$$

the probability that the component 1 survives beyond time x and that the component 2 survives beyond time y. The "Shocks" to the two components are assumed to form three independent Poisson processes with (positive) parameters λ_1, λ_2 and λ_{12}, depending on whether the shock kills only component 1, only component 2, or both the two components simultaneously. The times Z_1, Z_2 and Z_{12} of occurrence of these three shocks are independent exponential random variables with parameters λ_1, λ_2 and λ_{12}, respectively. So we have

$$X = \min(Z_1, Z_{12}),$$

$$Y = \min(Z_2, Z_{12}),$$

and then for all nonnegative numbers x and y,

$$\begin{aligned} \overline{H}(x,y) &= P[Z_1 > x]P[Z_2 > y]P[Z_{12} > \max(x,y)] &(3.15)\\ &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)\}. &(3.16)\end{aligned}$$

The marginal survival functions are

$$\overline{F}(x) = \exp\{-(\lambda_1 + \lambda_{12})x\}$$

and

$$\overline{G}(y) = \exp\{-(\lambda_2 + \lambda_{12})y\};$$

and then X and Y are exponential random variables with parameters $\lambda_1 + \lambda_{12}$ and $\lambda_2 + \lambda_{12}$, respectively. To construct the survival copula $\widehat{C}$, let us first express $\overline{H}(x,y)$ in terms of $\overline{F}(x)$ and $\overline{G}(y)$. Using the relation

$$\max(x,y) = x + y - \min(x,y),$$

we get

$$\begin{aligned} \overline{H}(x,y) &= \exp\{-(\lambda_1 + \lambda_{12})x - (\lambda_2 + \lambda_{12})y + \lambda_{12}\min(x,y)\}\\ &= \overline{F}(x)\overline{G}(y)\min\{\exp(\lambda_{12}x), \exp(\lambda_{12}y)\}.\end{aligned}$$

Now we set

$$\overline{F}(x) = u,$$

$$\overline{G}(y) = v,$$

$$\alpha = \frac{\lambda_{12}}{\lambda_1 + \lambda_{12}}$$

and

$$\beta = \frac{\lambda_{12}}{\lambda_2 + \lambda_{12}}.$$

Then the previous relation gives us

$$\begin{aligned}\widehat{C}(u,v) &= uv\min(u^{-\alpha}, v^{-\beta}) \\ &= \min(u^{1-\alpha}v, uv^{1-\beta}). \end{aligned} \tag{3.17}$$

This leads to a two-parameter family of copulas given by

$$\begin{aligned}C_{\alpha,\beta}(u,v) &= \min(u^{1-\alpha}v, uv^{1-\beta}) \\ &= \begin{cases} u^{1-\alpha}v, & if\ u^{\alpha} \geq v^{\beta} \\ uv^{1-\beta}, & if\ u^{\alpha} \leq v^{\beta} \end{cases}\end{aligned} \tag{3.18}$$

This family is the Marshall-Olkin family of copulas. It is also known as the Generalized Cuadras-Augé family of copulas.

Example 3.2. Bivariate Pareto Model.

Here we consider a random variable X that, given a risk classification parameter γ, can be modeled as an exponential distribution; that is (see [19])

$$P[X \leq x|\gamma] = 1 - e^{-\gamma x}.$$

If γ has a gamma distribution, then the marginal distribution of X is Pareto. That is, if γ is gamma (α,λ) then

$$F(x) = 1 - (1 + \frac{x}{\lambda})^{-\alpha}.$$

Now suppose, conditional on the risk class γ, that X_1 and X_2 are independent and identically distributed. Assuming that they come from the same risk class γ, induces a dependency. The joint distribution is

$$F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2] \tag{3.19}$$

$$= 1 - (1 + \frac{x_1}{\lambda})^{-\alpha} - (1 + \frac{x_2}{\lambda})^{-\alpha} + (1 + \frac{x_1 + x_2}{\lambda})^{-\alpha} \tag{3.20}$$

$$= F_1(x_1) + F_2(x_2) - 1 + [(1 - F_1(x_1))^{-\frac{1}{\alpha}} + (1 - F_2(x_2))^{-\frac{1}{\alpha}}]^{-\alpha}. \tag{3.21}$$

This yields the copulas function

$$C(u,v) = u + v - 1 + [(1-u)^{-\frac{1}{\alpha}} + (1-v)^{-\frac{1}{\alpha}}]^{-\alpha}. \tag{3.22}$$

3.6.2 A Way to Generate Archimedean Copulas

An Archimedean copula is known once one knows its generator. Therefore, to generate it, we just need to construct its generator. Genest and Riviest (1993) provided a procedure for identifying an Archimedean copula (see [19]). To start, let us assume that we have available a random sample of bivariate observations, $(X_{11}, X_{21}), (X_{12}, X_{22}), \ldots, (X_{1n}, X_{2n})$. Assume that the distribution function has an Archimedean copula C_ϕ. Our aim is to identify the form of ϕ. We consider an intermediate pseudo-observation Z_i (defined in 2.a below), that has distribution function

$$K(z) = P[Z_i \leq z].$$

Genest and Riviest (1993) (see [19]) showed that K is related to an Archimedean copula through the relation

$$K(z) = z - \frac{\phi(z)}{\phi'(z)}.$$

To identify ϕ, we use the following algorithm:

Algorithm 3.1. Generating an Archimedean copula.

1. Estimate Kendall's correlation coefficient using the usual estimate

$$\tau_n = \tbinom{n}{2}^{-1} \sum_{i<j} \operatorname{Sign}[(X_{1i} - X_{1j})(X_{2i} - X_{2j})]. \tag{3.23}$$

2. Construct a nonparametric estimate of K as follows:
   a. define the pseudo-observations

$$Z_i = \frac{\{\text{number of } (X_{1j}, X_{2j}) \text{ such that } X_{1j} < X_{1i} \text{ and } X_{2j} < X_{2i}\}}{n-1}, \tag{3.24}$$

   b. construct the estimate K_n of K as $K_n(z) =$ proportion of $Z_i's \leq z$.

3. Since K has to satisfy the relation

$$K(z) = z - \frac{\phi(z)}{\phi'(z)}, \tag{3.25}$$

   we obtain an estimate ϕ_n of ϕ, by solving the equation

$$z - \frac{\phi_n(z)}{\phi_n'(z)} = K_n(z). \tag{3.26}$$

Remark 3.3. Some other methods of constructing copulas are illustrated in [41]. There are geometric methods (see examples in [41], pages 59 to 86) and algebraic methods (see examples in [41], pages 89 to 99).

Chapter 4

Estimation of Copulas

An estimation approach is proposed for models for a multivariate response with covariates when each of the parameters (either univariate or a dependence parameter) of the model can be associated with a marginal distribution. In this chapter we give three ways to estimate a copula. We also discuss confidence bands and asymptotic theory.

4.1 Methods of Estimating Copulas

To start, let us make the following assumptions and notations. We assume that the copula we have to estimate belongs to a family $\{C(.,\theta),\ \theta \in \Theta\}$, where Θ is the space of parameters. Consider a copula-based parametric model for the random vector $Y = (Y_1, Y_2, \ldots, Y_d)$, with cumulative distribution function,

$$F(y; \alpha_1, \alpha_2, \ldots, \alpha_d; \theta) = C(F_1(y_1; \alpha_1), F_2(y_2; \alpha_2), \ldots, F_d(y_d; \alpha_d); \theta),$$

where $F_1, F_2, \ldots, F_d$ are univariate cumulative distribution functions with respective parameters $\alpha_1, \alpha_2, \ldots, \alpha_d$. We assume that C has density c (mixed derivatives of order d), and by f_j we denote the marginal probability density of Y_j, for $j \in \{1, 2, \ldots, d\}$. Then Y has the density (see [32]):

$$f(y; \alpha_1, \ldots, \alpha_d; \theta) = c(F_1(y_1; \alpha_1), F_2(y_2; \alpha_2), \ldots, F_d(y_d; \alpha_d); \theta) \prod_{j=1}^{d} f_j(y_j; \alpha_j) \tag{4.1}$$

For a sample of size n with observed random vectors $Y_1, Y_2, \ldots, Y_n$, we consider the d log-likelihood functions for the univariate marginals,

$$L_j(\alpha_j) = \sum_{i=1}^{n} \log f_j(y_{ij}; \alpha_j), j = 1, 2, \ldots, d, \tag{4.2}$$

and the log-likelihood function for the joint distribution,

$$L(\alpha_1, \alpha_2, \ldots, \alpha_d; \theta) = \sum_{i=1}^{n} \log f(y_i; \alpha_1, \ldots, \alpha_d; \theta). \tag{4.3}$$

Once one estimates the parameter θ, one has an estimate of the copula.

4.1.1 The Inference Method for Marginals

The inference function for marginals (IFM) method consists of doing d separate optimizations of the univariate likelihoods, followed by an optimization of the multivariate likelihood as a function of the dependence parameter vector. It consists of the following two steps:

1. the log-likelihoods $L_1(\alpha_1), L_2(\alpha_2), \ldots, L_d(\alpha_d)$, of the d univariate marginals are separately maximized to get estimates $\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_d$ of $\alpha_1, \alpha_2, \ldots, \alpha_d$, respectively,
2. the function $L(\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_d; \theta)$ is maximized over θ to get an estimate $\widehat{\theta}$ of θ.

That is, under regularity conditions, $(\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_d, \widehat{\theta})$ is the solution of

$$\left(\frac{\partial L_1}{\partial \alpha_1}, \frac{\partial L_2}{\partial \alpha_2}, \ldots, \frac{\partial L_d}{\partial \alpha_d}, \frac{\partial L}{\partial \theta} \right) = \underline{0}'. \tag{4.4}$$

The IFM method is useful for models with the closure property of parameters associated with or being expressed in lower-dimensional marginals (see [32]).

4.1.2 The Maximum Likelihood Method

This method obtains estimates $\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_d, \widehat{\theta}$, by solving the equation

$$\left(\frac{\partial L}{\partial \alpha_1}, \frac{\partial L}{\partial \alpha_2}, \ldots, \frac{\partial L}{\partial \alpha_d}, \frac{\partial L}{\partial \theta} \right) = \underline{0}' \tag{4.5}$$

simultaneously. Contrast this with Equation (4.4). An example of the bivariate case can be found in [19], page 14.

4.1.3 The Empirical Copula Function

Here we give a non parametric method for getting a bivariate copula. Consider a sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ of iid copies of a random vector (X, Y). The bivariate empirical distribution function (see [14], page 182) associated with (X, Y) is

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \le x, Y_i \le y\}},$$

with marginals

$$F_n(x) = H_n(x, +\infty) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \le x\}}$$

and

$$G_n(y) = H_n(+\infty, y) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \le y\}},$$

where I_A is the indicator function of the set A.

Then (see [56]) the empirical copula function is given by

$$C_n(u, v) = H_n(F_n^{-1}(u), G_n^{-1}(v)) \tag{4.6}$$

$$= \frac{1}{n} \sum_{k=1}^{n} I_{\{X_k \le F_n^{-1}(u), Y_k \le G_n^{-1}(v)\}}. \tag{4.7}$$

Nelsen (see [41], page 219) defined this copula as

$$C_n(\frac{i}{n}, \frac{j}{n}) = \frac{\text{number of pairs } (x, y) \text{ in the sample with } x \le x_{(i)}, y \le y_{(j)}}{n}, \tag{4.8}$$

where $x_{(i)}$ and $y_{(j)}$, $1 \le i, j \le n$, denote the order statistics of the sample.

Note that the empirical copula function based on $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, is the same as that based on uniform $[0, 1]$ random variables $(U_1, V_1), (U_2, V_2), \ldots, (U_n, V_n)$, where $U_i = F(X_i)$ and $V_i = G(Y_i)$, $i \in \{1, 2, \ldots, n\}$ (see [56]).

4.1.4 Estimating Archimedean Copulas

The following method was proposed by Genest and Rivest [24].

Consider a sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, which are iid copies of (X, Y), and assume that the copula C associated with (X, Y) is Archimedean with parameter α. To construct an estimate of α, Genest and Rivest [24] used the observed value of Kendall's tau. In fact, for Archimedean copulas, Kendall's tau can be conveniently computed via the identity

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt. \tag{4.9}$$

Let us consider the usual estimate of Kendall's tau given by (see [19])

$$\widehat{\tau} = \binom{n}{2}^{-1} \sum_{i<j} \mathrm{Sign}[(X_i - X_j)(Y_i - Y_j)]. \tag{4.10}$$

Since τ is expressed in terms of ϕ (Equation (4.9)), and ϕ is a function of α, an estimate $\widehat{\alpha}$ of α is obtained by solving the equation

$$\widehat{\tau} = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt, \tag{4.11}$$

for α.

4.2 Asymptotic Theory

In this section we present asymptotic results associated with the methods of estimating copula parameters. We present the iid case and an approach of dealing with covariates.

4.2.1 Independent and Identically Distributed Case

Here we assume that the regularity conditions for asymptotic maximum likelihood theory hold for the multivariate model as well as for all its marginals.

Let $\eta = (\alpha_1, \alpha_2, \ldots, \alpha_d; \theta)$ be the row vector of parameters and let Ψ be the row vector of inference functions of the same dimension as η. Let $Y, Y_1, Y_2, \ldots, Y_n$, be iid with density $f(.; \eta)$.

Suppose that the estimator $\widehat{\eta} = (\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_d; \widehat{\theta})$ is given by

$$\sum_{i=1}^{n} \Psi(Y_i, \widehat{\eta}) = 0,$$

and let $\frac{\partial \Psi^T}{\partial \eta}$ be the matrix with (j,k) components $\frac{\partial \Psi_j(y,\eta)}{\partial \eta_k}$. Joe and Xu [32] showed that the asymptotic covariance matrix of $n^{\frac{1}{2}}(\widehat{\eta} - \eta)^T$, called the Godambe information matrix, is

$$V = D_\Psi^{-1} M_\Psi (D_\Psi^{-1})^T, \tag{4.12}$$

where

$$D_\Psi = E\left[\frac{\partial \Psi^T(Y, \eta)}{\partial \eta}\right],$$

and

$$M_\Psi = E\left[\Psi^T(Y, \eta)\Psi(Y, \eta)\right].$$

4.2.2 Inclusion of Covariates

Here we assume that we have independent, non-identically distributed random vectors Y_i, $i = 1, 2, \ldots, n$, with densities $f_i(.; \alpha)$, where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d, \theta)$. In order to include covariates we assume that

$$\alpha_j = a_j(x, \gamma_j), \; j = 1, 2, \ldots, d,$$

and

$$\theta = t(x, \gamma_{d+1}),$$

where $a_1, a_2, \ldots, a_d, t$ are link functions. Instead of $f(y; \alpha_1, \alpha_2, \ldots, \alpha_d, \theta)$ in the case without covariates, we now consider the density

$$\begin{aligned} f_{Y|x}(y|x; \gamma) &= f(y; a_1(x, \gamma_1), a_2(x, \gamma_2), \ldots, a_d(x, \gamma_d), t(x, \gamma_{d+1})) \\ &= c(F_1(y_1; \alpha), F_2(y_2; \alpha), \ldots, F_n(y_n; \alpha)) \prod_{i=1}^{n} f_i(y_i; \alpha), \end{aligned} \tag{4.13}$$

where F_i is the marginal distribution function of Y_i, $i = 1, 2, \ldots, n$, and

$$\alpha = (a_1(x, \gamma_1), a_2(x, \gamma_2), \ldots, a_d(x, \gamma_d), t(x, \gamma_{d+1})).$$

The estimate

$$\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2, \ldots, \widehat{\gamma}_d, \widehat{\gamma}_{d+1})$$

of

$$\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_d, \gamma_{d+1}),$$

is obtained by the maximum likelihood method under the following conditions (see [32]):

1. mixed derivatives of Ψ of first and second order are dominated by integrable functions,
2. products of these derivatives are uniformly integrable,
3. the link functions are twice continuously differentiable with first and second order derivatives bounded away from zero,
4. covariates are uniformly bounded, the sample covariance matrix of the covariates is strictly positive definite,
5. a Lindeberg-Feller type condition holds.

If all these conditions hold then the asymptotic normality result has the form (see [32])

$$n^{-\frac{1}{2}} V_n^{-\frac{1}{2}} (\widehat{\gamma} - \gamma)^T \xrightarrow{d} N(0, I),$$

where

$$V_n = D_n^{-1} M_n (D_n^{-1})^T,$$

with

$$D_n = n^{-1} \sum_{i=1}^{n} E\left[\frac{\partial \Psi^T(Y_i, \gamma)}{\partial \gamma}\right]$$

and

$$M_n = n^{-1} \sum_{i=1}^{n} E\left[\Psi^T(Y_i, \gamma) \Psi(Y_i, \gamma)\right].$$

Note that this approach allows to extend asymptotic theory to the case of random vectors with covariates.

Remark 4.1. This result can also be extended to random covariates. See [32].

Chapter 5

Copula and Regression Analysis

In this chapter we discuss an alternative way of looking at regression analysis by using copulas. In one of his papers, Sungur [53] defined the copula regression function and provided its basic properties. All the material used here can be found in [53].

Definition 5.1. Let (U, V) be a random pair with uniform marginals on $[0, 1]$ and copula C. The copula regression function of V on U, denoted by $r_C(u)$, is defined by

$$r_C(u) = E_C[V|U = u].$$

Some properties of the copula regression function are described in the following theorems.

Theorem 5.1. We have the following properties.

1. If $C^0(u, v) = uv$ then $r_{C^0}(u) = 1/2$
2. If $C^+(u, v) = \min\{u, v\}$ then $r_{C^+}(u) = u$
3. If $C^-(u, v) = \max\{u, v\}$ then $r_{C^-}(u) = 1 - u$

Now define by $C_u(v)$ the conditional distribution function of V given $U = u$, i.e.

$$C_u(v) = P(V \leq v|U = u) = \frac{\partial C(u, v)}{\partial u}.$$

Theorem 5.2. We have the following properties.

1.
$$r_C(u) = 1 - \int_0^1 \left[C_{u_0}(v) + \sum_{l=1}^{n-1} \frac{C_{u_0}^{(l)}(v)}{l!}(u - u_0)^l + \frac{C_{u_r}^{(n)}(v)}{n!}(u - u_r)^n \right] dv,$$

where

$$C_{u_0}^{(l)}(v) = \frac{\partial^l C_u(v)}{\partial u^l}|_{u=u_0}$$

and u_r is an interval joining u and u_0.

2.

$$r_C(u) \geq r(1 - C_u(r)), \text{ for any } r \in (0, 1];$$

3.

$$E(V) = \int_0^1 r_C(u)du = \frac{1}{2};$$

4.

$$\rho_C = 3\left\{1 - 4\int_0^1 \left[\int_0^u r_C(w)dw\right] du\right\},$$

where ρ_C is the Pearson's correlation.

Sungur [53] looked at linear and non linear copula regression functions.

5.1 Linear Copula Regression Functions

The class of copulas with linear copula regression functions is defined by Sungur [53] as

$$\zeta_L = \left\{C: \ 1 - \int_0^1 \frac{\partial C(u, v)}{\partial u} du = \alpha + \beta u\right\}.$$

The following result is given.

Theorem 5.3. A copula has a linear regression function, i.e. $C \in \zeta_L$, if and only if

$$r_C(u) = \alpha + (1 - 2\alpha)u,$$

or

$$r_C(u) = \frac{1 - \beta}{2} + \beta u.$$

From this result, we can observe that the special relationship between the slope and intercept parameters for the linear copula regression functions provides a way of testing for linearity. Moreover, for a linear copula regression function, the coefficient (e.g slope and intercept) will be related to the Pearson correlation as shown in the following theorem.

Theorem 5.4. If $C \in \zeta_L$, then for Pearson correlation

$$\rho_C = 1 - 2\alpha$$

and

$$r_C(u) = \frac{1-\rho_C}{2} + \rho_C u.$$

Therefore, we can observe the strength of a linear relationship by checking the intercept. Note that Sungur [53] started his investigation with two examples: Farlie-Gumbel-Morgenstern family (Example 2.3) and Frechet and Mardia (Example 2.5) family.

5.2 Non Linear Copula Regression Functions

Sungur [53] considered two examples: the Rodriquez-Lallena and Ubena-Flores family (Example 2.6), and the Cuadras-Augé family (Example 2.4). For the first class, he proved the following two results.

Theorem 5.5. The Rodriquez-Lallena and Ubena-Flores family, and the Cuadras-Augé family have a linear copula regression function if and only if the function f in Example 2.6 satisfies the equation

$$f(u) = 6u(1-u)\int_0^1 f(u)du.$$

Theorem 5.6. For the Rodriquez-Lallena and Ubena-Flores family, and the Cuadras-Augé family,

$$f(u) = \beta\left[\frac{1}{2}u - \int_0^u r_C(w)dw\right],$$

where

$$\beta = 12\rho_C^{-1}\int_0^1 f(w)dw,$$

and f defined as in Example 2.6.

By Theorem 5.6, Sungur showed how one can form a class of copulas with polynomial regression functions. From the examples he has provided, he deduced that the functional form of the regression line depends on the joint behavior, determined by the copula, and marginal behavior, shaped by the marginal distribution functions. The real problem from an application point of view is whether it is possible to separately transform each of the variables to achieve linearity in regression. The answer is “yes” if $C \in \zeta_L$, but in the case where C is not in ζ_L, Sungur [53] provided two approaches to solve this problem. See [53] for these approaches.

5.3 Relationship Between Level Curves and Copulas

The dependence structure for a bivariate distribution can be represented by the concept of copulas and so the effect of the dependence can be separated from the effect of the marginal distributions. The idea is to use the concept of quantile curves in order to study the dependence structure (given by the concept of copulas) of a bivariate distribution function. Before we give results that link copulas to level curves, we give the following definitions (see [1]).

Definition 5.2. Let $\mathbf{X} = (X, Y)$ be a random vector under regularity conditions. Let (x, y) be a point in $\mathbb{R}^2$. Denote by $F_\varepsilon(x, y)$ the accumulated probability in the quadrant defined by the direction ε, i.e.

$$F_\varepsilon(x, y) = P\{X\Delta_{\varepsilon_1}x, Y\Delta_{\varepsilon_2}y\},$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2)$ with $\varepsilon_i \in \{-1, +1\},\ i = 1, 2$ denote four directions in $\mathbb{R}^2$, Δ_- and Δ_+ are the inequalities "$\leq$" and "$\geq$", respectively. We write $\Delta_{\varepsilon_i} = \Delta_-$ when $\varepsilon_i = -1$ and $\Delta_{\varepsilon_i} = \Delta_+$ when $\varepsilon_i = +1$.

Definition 5.3. Let $\mathbf{X} = (X, Y)$ be a random vector under the regularity conditions, and let $p \in [0, 1]$. We define the bivariate quantile set or quantile curve for the direction ε, denoted by $Q_X(p, \varepsilon)$, as

$$Q_X(p, \varepsilon) = \{(x, y) \in \mathbb{R}^2 : F_\varepsilon(x, y) = p\}.$$

By the previous definition, for each $p \in [0, 1]$, we have four quantile curves, where each one can be described by an equation.

Definition 5.4. Let $\mathbf{X} = (X, Y)$ be a random vector and let $p \in [\frac{1}{2}, 1]$. Then we define the central region, denoted $\Omega_X(p)$, as

$$\Omega_X(p) = \{(x, y) \in \mathbb{R}^2 : F_\varepsilon(x, y) < p, \forall\varepsilon\}.$$

Definition 5.5. Let $\mathbf{X} = (X, Y)$ be a random vector and let $p \in (0, 1)$. We define the lateral region with order p in the direction ε, denoted $L_X(p, \varepsilon)$, as

$$L_X(p, \varepsilon) = \{(x, y) \in \mathbb{R}^2 : F_\varepsilon(x, y) > p\}.$$

From this, the quantile curves have been described (see [1]) in a parametric form by expressing them by means of the quantiles for the conditional distributions $[Y|X \leq x]$ and $[Y|X \geq x]$ as follows:

$$Q_X(p, \varepsilon_{--}) \to \{(Q_X(u), Q_{Y|X \leq Q_X(u)}(p/u)) : u > p\},$$

$$Q_X(p, \varepsilon_{+-}) \to \{(Q_X(u), Q_{Y|X \geq Q_X(u)}(p/(1-u))) : u < 1-p\},$$

$$Q_X(p, \varepsilon_{-+}) \to \{(Q_X(u), Q_{Y|X \leq Q_X(u)}(1-p/u)) : u > p\},$$

and

$$Q_X(p, \varepsilon_{++}) \to \{(Q_X(u), Q_{Y|X \geq Q_X(u)}(1-p/(1-u))) : u < 1-p\}.$$

A similar parametric form can be obtained if the marginal variables X and Y are interchanged.

The following results show that the accumulated probability in the central region and the lateral regions defined previously depends on the copula of the underlying random variable $\mathbf{X}$.

Theorem 5.7. Let $\mathbf{X} = (X, Y)$ be a random vector under regularity conditions with copula C. Then the accumulated probability in the central region $P\{\mathbf{X} \in \Omega_{\mathbf{X}}(p)\}$ depends solely on the copula.

Theorem 5.8. Let $\mathbf{X} = (X, Y)$ be a random vector under regularity conditions with copula C. Then the accumulated probability in the lateral region with order p in the direction ε, $P\{\mathbf{X} \in L_{\mathbf{X}}(p, \varepsilon)\}$, depends solely on the copula.

Belzunce, Castano, Olveira-Cervantes and Suarez-Llornz (see [1]) analyzed the case of independence through the corollary below.

Corollary 5.9. Let $\mathbf{X} = (X, Y)$ be a random vector under regularity conditions with independent components. Then

$$P\{\mathbf{X} \in L_{\mathbf{X}}(p, \varepsilon)\} = 1 - p + \ln(p),$$

for all direction $\varepsilon,\ p \in [0, 1]$. In addition, for $p > \frac{1}{2}$ it holds that

$$P\{\mathbf{X} \in \Omega_{\mathbf{X}}(p)\} = 4p(1 + \ln(p)) - 3.$$

They also compared the accumulated probability in the lateral regions for a general bivariate distribution with the corresponding probabilities for a bivariate distribution with independent components, and they applied the previous results to an independence test for bivariate distribution. See [1] for more about their work.

Chapter 6

A Review of Goodness-of-Fit Test Statistics

Given a random sample, the idea of goodness-of-fit test is to see whether this sample comes from a particular distribution. In this chapter we review some formal goodness-of-fit testing methods. Bootstrap procedures for goodness-of-fit are also discussed.

6.1 Univariate Test Statistics

6.1.1 Univariate Test Statistics for General Distributions

a. Chi-Square Type Test Statistics

The chi-square goodness-of-fit test is a special type of test, which applies when the possible outcomes are partitioned into a finite number of categories. Given a random sample, in order to carry out this test, two values are involved: an observed value, which is the frequency of a category from the sample, and the expected frequency, which is calculated based upon the claimed distribution. Moreover, chi-square tests can be used for both discrete and continuous distributions.

Pearson's Chi-Square Test Statistic

In [44], Pearson proposed the test statistic T_{χ^2} given by

$$T_{\chi^2} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \tag{6.1}$$

where k is the number of categories, O_i and E_i are the observed and expected frequencies for category i, $(1 \leq i \leq k)$. He showed that under certain conditions the distribution of T_{χ^2} can be approximated by a χ^2-distribution with $k-1$ degrees of freedom.

Apart from Pearson's chi-square test statistic, there are many other measures of fit with distributions that, under certain conditions, may be approximated by a χ^2-distribution. Consider a random sample and divide the range of the sample into k disjoint bins. As previously, let O_i and E_i be the observed and expected frequencies for bin i, $(1 \leq i \leq k)$. The following chi-square test statistics are given.

Modified Chi-Square Test Statistic

This statistic is based on the difference of frequencies and is affected by small observed frequencies. It is given by (see [42])

$$T_{\chi^2(M)} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{O_i}. \tag{6.2}$$

Freeman-Tukey Statistic

It is based on the difference of frequencies, but not affected by small observed or expected frequencies. This statistic is given by (see [18])

$$T_{FT} = 4\sum_{i=1}^{k} (\sqrt{O_i} - \sqrt{E_i})^2. \tag{6.3}$$

Log-Likelihood Ratio Statistic

This statistic is based on the ratio of frequencies and is affected by small expected frequencies. It is given by (see [59])

$$T_{G^2} = 2\sum_{i=1}^{k} O_i \ln(\frac{O_i}{E_i}). \tag{6.4}$$

Modified Log-Likelihood Ratio Statistic

This statistic is based on the ratio of frequencies and is affected by small observed frequencies. It is given by (see [36])

$$T_{G^2(M)} = 2\sum_{i=1}^{k} E_i \ln(\frac{E_i}{O_i}). \tag{6.5}$$

Remark 6.1. Cressie and Read [8] identified the similarities of the previous chi-square test statistics and proposed the general Power-Divergence statistic, T_{PD}, given by

$$T_{PD} = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{k} O_i \left[(\frac{O_i}{E_i})^{\lambda} - 1 \right], \tag{6.6}$$

where $k,\ O_i,\ E_i$ are as before. The value assigned to the coefficient λ is used to formulate the particular test statistic as follows (see [51]):

1. If $\lambda = -\frac{1}{2}$, T_{PD} is equivalent to T_{FT},
2. If $\lambda = 1$, T_{PD} is equivalent to T_{χ^2},
3. If $\lambda = -2$, T_{PD} is equivalent to $T_{\chi^2(M)}$,
4. If $\lambda \to -1$, T_{PD} approaches T_{G^2},
5. If $\lambda \to 0$, T_{PD} approaches $T_{G^2(M)}$.

b. Kolmogorov-Smirnov Test Statistics

One of the simplest ways to determine a difference between the true, but unknown, distribution and the continuous null distribution is the use of the Kolmogorov-Smirnov test statistics. Given a random sample $X_1, X_2, \ldots, X_n$, generated by the cumulative distribution function F, consider the null hypothesis

$$H : F(x) = F_X(x, \theta).$$

We first define the Kolmogorov-Smirnov test statistic in the case where θ is known, and then we discuss the case where θ has to be estimated.

i. The parameter θ is known

Description of The Test

Let θ be fixed at some value θ_0. In this case $F_X(x, \theta_0)$ is fully specified and the hypothesis $H_0 : F(x) = F_X(x, \theta_0)$ is simple. Denote $F_X(x, \theta_0)$ by $F_X(x)$. The Kolmogorov-Smirnov test statistic is given by

$$K_n = \sqrt{n} \sup_{-\infty<x<\infty} |F_n(x) - F_X(x)|, \tag{6.7}$$

where F_n is the sample (empirical) cumulative distribution function.

Asymptotic Distribution of K_n

In order to give the asymptotic distribution of K_n, we first define the Brownian bridge. Define the process

$$B_n(x) = \sqrt{n}(G_n(x) - x),$$

as a random function on the interval $[0, 1]$, where G_n is the empirical distribution function associated with the uniform distribution, and consider a finite number of points $x_1, \ldots, x_k \in [0, 1]$. By the Multivariate Central Limit Theorem, we have (see [35])

$$(B_n(x_1), \ldots, B_n(x_k))^T \xrightarrow{D} N_k(0, C),$$

where $N_k(0, C)$ is the k-variate normal distribution with mean vector 0 and covariance matrix C which (i, j) element is defined by

$$C(i, j) = \min(x_i, x_j) - x_i x_j.$$

A random function B on $[0, 1]$ such that the random vector $(B(x_1), \ldots, B(x_k))$ has the previous limit for any finite number of points $x_1, \ldots, x_k \in [0, 1]$, is called a Brownian bridge.

Now we discuss the asymptotic result of the Kolmogorov-Smirnov statistic in the case where the distribution function F_X is independent of θ. Using the probability integral transform $U =$

$F_X(X)$, under H_0 we can write K_n in the form

$$K_n = \sup_{0 \le u \le 1} |B_n(u)|,$$

and so we have the following asymptotic representation (see [35]).

$$K_n \xrightarrow{D} \sup_{0 \le u \le 1} |B(u)| = K. \tag{6.8}$$

The limiting distribution can be found as

$$P(K > u) = 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp\left(-2j^2 u^2\right).$$

ii. The parameter θ is unknown

Description of the Test

In the case where θ is unknown, it can be replaced by an estimate, say $\widehat{\theta}$, and then the Kolmogorov-Smirnov test statistic is defined by

$$\widehat{K}_n = \sqrt{n} \sup_{-\infty < x < \infty} |F_n(x) - F_X(x, \widehat{\theta})|. \tag{6.9}$$

Before we describe the asymptotic result of $\widehat{K}_n$, we first investigate the estimated empirical process. This investigation is very important and will be used for some further test statistics. We follow the discussion as in Shorack and Wellner [50] (pages 228-237).

Estimated Empirical Process

Suppose we want to test if the sample $X_1, \ldots, X_n$ comes from a distribution function $F_X(., \theta)$. Without loss of generality, we assume that $X_1, \ldots, X_n$ comes from $F_X(., (\theta, \gamma))$, for some pair $(\theta, \gamma) = (\theta_1, \ldots, \theta_J, \gamma_1, \ldots, \gamma_K) \in \mathbb{R}^{J+K}$, and we consider testing the hypothesis $H_0 : \gamma = 0$. Let $\widehat{F}_n(x) = F_X(x, (\widehat{\theta}_n, 0))$ for some estimate $\widehat{\theta}_n$ of θ, and consider the processes

$$U_n(F) = \sqrt{n}(F_n - F)$$

and

$$\widehat{B}_n = \sqrt{n}(F_n - \widehat{F}_n),$$

where $\widehat{F}_n$ denotes the distribution function when θ is estimated, and F_n is the empirical distribution function. Denoting $F_X(.,(\theta,\gamma))$ by $F_{\theta,\gamma}$, $\widehat{B}_n$ can be written in terms of U_n as

$$\widehat{B}_n = U_n(F_{\theta,\gamma/\sqrt{n}}) - \sqrt{n}(\widehat{F}_n - F_{\theta,\gamma/\sqrt{n}}). \tag{6.10}$$

The Taylor expansion of $\sqrt{n}(\widehat{F}_n - F_{\theta,\gamma/\sqrt{n}})$ about $(\theta,0)$ is given by

$$\begin{aligned}\sqrt{n}(\widehat{F}_n - F_{\theta,\gamma/\sqrt{n}}) &= \sqrt{n}(F_{\widehat{\theta}_n,0} - F_{\theta,\gamma/\sqrt{n}}) \\ &\doteq \sum_{j=1}^{J} \sqrt{n}(\widehat{\theta}_{nj} - \theta_j)\frac{\partial F_{\theta,\gamma}}{\partial \theta_j}|_{(\theta,0)} + \sum_{k=1}^{K} \sqrt{n}(0 - \frac{\gamma_k}{\sqrt{n}})\frac{\partial F_{\theta,\gamma}}{\partial \gamma_k}|_{(\theta,0)} \\ &= \sum_{j=1}^{J} \sqrt{n}(\widehat{\theta}_{nj} - \theta_j)\frac{\partial F_{\theta,\gamma}}{\partial \theta_j}|_{(\theta,0)} - \sum_{k=1}^{K} \gamma_k\frac{\partial F_{\theta,\gamma}}{\partial \gamma_k}|_{(\theta,0)},\end{aligned}$$

provided sufficient regularity is assumed for the partial derivatives of $F_{\theta,\gamma}$ to behave nicely. Let

$$\widehat{U}(F_{\theta,0}) \equiv U(F_{\theta,0}) - \sum_{j=1}^{J} \sqrt{n}(\widehat{\theta}_{nj} - \theta_j)\frac{\partial F_{\theta,\gamma}}{\partial \theta_j}|_{(\theta,0)} + \sum_{k=1}^{K} \gamma_k\frac{\partial F_{\theta,\gamma}}{\partial \gamma_k}|_{(\theta,0)}, \tag{6.11}$$

where U is the Brownian bridge. We will show that under some regularity conditions, $\widehat{B}_n$ has the same asymptotic behavior as $\widehat{U}(F_{\theta,0})$. Suppose that the family $F_{\theta,\gamma}$ and the sequence of estimators $\widehat{\theta}_n$ of θ are regular in the following sense.

1. The first-order Taylor series approximation

$$\|F_{\theta',\gamma} - F_{\theta,0} - \sum_{j=1}^{J}(\theta'_j - \theta_j)\frac{\partial F_{\theta,\gamma}}{\partial \theta_j}|_{(\theta,0)} + \sum_{k=1}^{K} \gamma_k\frac{\partial F_{\theta,\gamma}}{\partial \gamma_k}|_{(\theta,0)}\| = o\left(\sum_{j=1}^{J}(\theta'_j - \theta_j)^2 + \sum_{k=1}^{K}\gamma_k^2\right), \tag{6.12}$$

   in a neighborhood of $(\theta,0)$, holds with partial derivatives being uniformly bounded in x.

2. For $j = 1,\ldots,J$, we have

$$Z_{nj} = \sqrt{n}(\widehat{\theta}_{nj} - \theta_j) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_j(\xi_{ni}) + o_p(1), \tag{6.13}$$

   where $\xi_{ni} \equiv \widehat{F}_n(X_i)$, and the h_j's are such that $E(h_j(\xi)) = 0$ and $\mathrm{Var}(h_j(\xi)) = \sigma_j^2$, $j = 1,2,\ldots,J$.

Then $\widehat{B}_n$ satisfies

$$\|\widehat{B}_n - \widehat{U}(F_{\theta,0})\| \xrightarrow{P} 0, \tag{6.14}$$

with $\widehat{U}(F_{\theta,0})$ as in Equation (6.11), and $\|.\|$ denoting the L_2 norm. We write $\widehat{U}$ as

$$\widehat{U}(F_{\theta,0}) \doteq U(F_{\theta,0}) - \sum_{j=1}^{J} Z_{nj} F_j + \sum_{k=1}^{K} \gamma_k G_k \tag{6.15}$$

with

$$F_j \equiv \frac{\partial F_{\theta,\gamma}}{\partial \theta_j}|_{(\theta,0)}, \ G_k \equiv \frac{\partial F_{\theta,\gamma}}{\partial \gamma_k}|_{(\theta,0)}, \text{ and } Z_{nj} \equiv \sqrt{n}(\theta_{nj} - \theta_j) \xrightarrow{D} Z_j = \int_0^1 h_j(s) dU(s).$$

It can be shown that the vector of Z_j's and the Brownian bridge U are jointly normal with 0 mean,

$$\mathrm{Cov}(Z_j, Z_{j'}) = \int_0^1 h_j(s) h_{j'}(s) ds \text{ and } \mathrm{Cov}(Z_j, U(t)) = \int_0^1 h_j(s) ds \tag{6.16}$$

for h_j's given in Equation (6.13).

In fact, from Equations (6.10) and (6.11), we have

$$\begin{aligned} \|\widehat{B}_n - \widehat{U}(F_{\theta,0})\| &\doteq \|U_n(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,0})\| \\ &= \|U_n(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,\gamma/\sqrt{n}}) + U(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,0})\| \\ &\le \|U_n(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,\gamma/\sqrt{n}})\| + \|U(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,0})\|. \end{aligned}$$

But from Equation (6.12), we have

$$\|U(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,0})\| \xrightarrow{P} 0.$$

Moreover, since $U_n(F)$ converges to the Brownian bridge $U(F)$, we have

$$\|U_n(F_{\theta,\gamma/\sqrt{n}}) - U(F_{\theta,\gamma/\sqrt{n}})\| \xrightarrow{P} 0.$$

Therefore, we conclude that

$$\|\widehat{B}_n - \widehat{U}(F_{\theta,0})\| \xrightarrow{P} 0.$$

Let us now show that $\widehat{U}(F_{\theta,0})$ is written in the form (6.15). From Equations (6.10), (6.12) and (6.13), we have

$$\begin{aligned}\widehat{U}(F_{\theta,0}) &= U(F_{\theta,0}) - \sum_{j=1}^{J} \sqrt{n}(\widehat{\theta}_{nj} - \theta_j)F_j + \sum_{k=1}^{K} \gamma_k G_k \\ &= U(F_{\theta,0}) - \sum_{j=1}^{J} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(\xi_{ni})\right] F_j + \sum_{k=1}^{k} \gamma_k G_k + o_p(1),\end{aligned}$$

with $\xi_{ni} \equiv F_n(X_i)$, and so we can write

$$\widehat{U}(F_{\theta,0}) \doteq U(F_{\theta,0}) - \sum_{j=1}^{J} Z_j F_j + \sum_{k=1}^{K} \gamma_k G_k,$$

where Z_j's are in the form

$$Z_j = \int_0^1 h_j(s) dU(s),$$

and

$$\mathrm{Cov}(Z_j, Z_j') = \int_0^1 h_j(s) h_{j'}(s) ds$$

(see Theorem 3.1.2 of [50]).

Now consider the process

$$\widehat{U}_n(t) = \sqrt{n}(\widehat{G}_n(t) - t),\ 0 \leq t \leq 1,$$

where $\widehat{G}_n$ denotes the empirical distribution function of the $\widehat{\xi}_{ni}$'s defined by

$$\widehat{\xi}_{ni} \equiv \widehat{F}_n(X_i) = F_X(X_i, (\widehat{\theta}_n, 0)).$$

Suppose that $F_X(., (\theta, 0)) = F(., -\theta)$ for some distribution function F, and that $\widehat{\theta}_n$ is the maximum likelihood estimate of θ. Then F_1 is $-f_X$, where f_X is the density function associated with F_X. Define the function h by

$$h(t) = -\frac{f'(F^{-1}(t))}{I f(F^{-1}(t))},$$

with

$$I = \int_{-\infty}^{\infty} (f'/f)^2 dF$$

denoting the Fisher information. We have (see [50])

$$\|\widehat{U}_n - \widehat{U}\| \xrightarrow{P} 0, \text{ as } n \to \infty, \tag{6.17}$$

where

$$\widehat{U} = U + Zf(F^{-1}(t)); \tag{6.18}$$

the variable Z, arising as the limit of random variables of the type $\int_0^1 h dU_n$, is distributed such that $E(Z) = 0$,

$$\begin{aligned}
\mathrm{Var}(Z) &= \int_0^1 h^2(t)dt \\
&= \frac{1}{I^2}\int_0^1 \frac{f'^2(F^{-1}(t))}{f^2(F^{-1}(t))}dt \\
&= \frac{1}{I^2}\int_{-\infty}^{\infty} \frac{f'^2(x)}{f^2(x)}dF(x) \text{ (by making change of variable } F^{-1}(t) = x) \\
&= I/I^2 = I^{-1},
\end{aligned}$$

and jointly distributed with U such that

$$\begin{aligned}
\mathrm{Cov}(Z, U(t)) &= \int_0^t h(s)ds \\
&= -\frac{1}{I}\int_0^t \frac{f'(F^{-1}(s))}{f(F^{-1}(s))}ds \\
&= -\frac{1}{I}\int_{-\infty}^{F^{-1}(t)} f'(x)dx \\
&= -\frac{1}{I}\int_{-\infty}^{F^{-1}(t)} df(x) \\
&= -\frac{f(F^{-1}(t))}{I}.
\end{aligned}$$

The covariance function of $\widehat{U}$ is given by

$$\begin{aligned}
K_{\widehat{U}}(s,t) &= \mathrm{Cov}(U(s) + Zf(F^{-1}(s)), U(t) + Zf(F^{-1}(t))) \\
&= \mathrm{Cov}(U(s), U(t)) + \mathrm{Cov}(U(s), Zf(F^{-1}(t))) + \mathrm{Cov}(Zf(F^{-1}(s)), U(t)) \\
&\qquad + \mathrm{Cov}(Zf(F^{-1}(s)), Zf(F^{-1}(t))) \\
&= K_U(s,t) + f(F^{-1}(t))\,\mathrm{Cov}(U(s), Z) + f(F^{-1}(s))\,\mathrm{Cov}(Z, U(t)) \\
&\qquad + f(F^{-1}(s))f(F^{-1}(t))\,\mathrm{Var}(Z) \\
&= K_U(s,t) - f(F^{-1}(t))f(F^{-1}(s))/I - f(F^{-1}(s))f(F^{-1}(t))/I \\
&\qquad + f(F^{-1}(s))f(F^{-1}(t))/I \\
&= K_U(s,t) - f(F^{-1}(s))f(F^{-1}(t))/I \\
&= \min(s,t) - st - \phi(s)\phi(t),
\end{aligned}$$

where $\phi(t) = -f(F^{-1}(t))/\sqrt{I}$.

Asymptotic Behavior of $\widehat{K}_n$

The Kolmogorov-Smirnov test statistic $\widehat{K}_n$ can be written in the form

$$\widehat{K}_n = \sup_{0\le t\le 1} |\widehat{U}_n|,$$

and so we have the following asymptotic result.

$$\widehat{K}_n \xrightarrow{D} \sup_{0\le t\le 1} |\widehat{U}(t)|.$$

c. Generalized Cramer-Von Mises Test Statistic

Let $X_1, \ldots, X_n$ be an i.i.d. random sample from a distribution function F, and let F_n denote the corresponding empirical distribution function. In this section we discuss Cramer-Von Mises test statistics by first looking at the case where the parameter θ is known, and then moving to the case where θ is unknown.

i. The parameter θ is known

Description of the Test

Assume θ is fixed at some value θ_0, and let $F_{\theta_0} = F_0$. The generalized Cramer-Von Mises test statistic for testing the null hypothesis $H_0 : F(x) = F_0(x)$, is given by (see [45])

$$W_n^2(\Psi) = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 \Psi(F_0(x)) dF_0(x), \tag{6.19}$$

where Ψ is a suitably chosen weight function. By means of the probability transformation $U = F_X(X)$, W_n^2 can be written in the form

$$W_n^2 = n \int_0^1 [G_n(u) - u]^2 \Psi(u) du, \tag{6.20}$$

where G_n is the empirical distribution function of the transformed data. Thus, in terms of the empirical process $B_n(u) = \sqrt{n}(G_n(u) - u)$, we have

$$W_n^2 = \int_0^1 B_n^2(u) \Psi(u) du. \tag{6.21}$$

Asymptotic Distribution of W_n^2 with $\Psi \equiv 1$

With $\Psi \equiv 1$, we have

$$W_n^2 = \int_0^1 B_n^2(u) du, \tag{6.22}$$

and so the following asymptotic result is obtained.

$$W_n^2 \xrightarrow{D} \int_0^1 B^2(u) du = W^2. \tag{6.23}$$

One can show that the principal component decomposition of the kernel $K(s,t) = \min(s,t) - st$ is

$$K(s,t) = \sum_{j=1}^{\infty} \lambda f_j(s) f_j(t),$$

where $\lambda_j = \frac{1}{j^2\pi^2}$ and $f_j(t) = \sqrt{2}\sin(j\pi t)$, $j = 1, 2, \ldots$ are respectively the eigenvalues and the corresponding eigenfunctions of the covariance matrix. From this, W^2 is written

$$W^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2\pi^2},$$

where

$$Z_j = \sqrt{2}j\pi \int_0^1 \sin(j\pi u)B(u)du, \; j = 1, 2, \ldots, \tag{6.24}$$

are i.i.d. $N(0,1)$-distributed. Therefore, Equation (6.23) becomes

$$W_n^2 \xrightarrow{D} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2\pi^2} \tag{6.25}$$

i. The parameter θ is unknown

Description of the Test

In the case where θ is unknown, we estimate it by $\widehat{\theta}_n$ and we define the generalized Cramer-Von Mises statistic by

$$\widehat{W}_n^2(\Psi) = n \int_{-\infty}^{\infty} [F_n(x) - \widehat{F}_n(x))]^2 \Psi(\widehat{F}_n(x)) d\widehat{F}_n(x), \tag{6.26}$$

where $\widehat{F}_n = F_{\widehat{\theta}_n}$.

Asymptotic Result for $\widehat{W}_n^2(\Psi)$

Under regularity conditions, we have

$$\widehat{W}_n^2(\Psi) \xrightarrow{D} \widehat{W}^2(\Psi) \equiv \int_0^1 \widehat{U}^2(t)\Psi(t)dt, \tag{6.27}$$

for a process $\widehat{U}$ defined as in (6.18).

From this, the simplified Cramer-Von Mises statistic (case $\Psi \equiv 1$), given by

$$\widehat{W}_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - \widehat{F}_n(x))]^2 d\widehat{F}_n(x),$$

satisfies

$$\widehat{W}_n^2 \xrightarrow{D} \widehat{W}^2 \equiv \int_0^1 \widehat{U}^2(t)dt, \tag{6.28}$$

where $\widehat{U}$ is a normal process with the covariance function $K_{\widehat{U}}(s,t)$ in the form (see [50])

$$K_{\widehat{U}}(s,t) = \min(s,t) - st - \sum_{i=1}^{m} \phi_i(s)\phi_i(t),\ 0 \le s,t \le 1,$$

for some functions $\phi_1, \dots, \phi_n$.

d. Anderson-Darling Test Statistic

It is a particular case of the Cramer-Von Mises statistic, obtained with $\Psi(u) = \frac{1}{u(1-u)}$.

i. The parameter θ is known

In the case where the parameter θ is known, this statistic is written in the form

$$A_n^2 = \int_0^1 \frac{B_n^2(u)}{u(1-u)} du. \tag{6.29}$$

The asymptotic result of A_n^2 is given by

$$A_n^2 \xrightarrow{D} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(1+j)}, \tag{6.30}$$

where the $Z_j's$ are defined as in (6.24).

ii. The parameter θ is unknown

When θ is unknown, it is estimated by $\widehat{\theta}_n$, and then the Anderson-Darling statistic is written in the form

$$\widehat{A}_n^2 = \int_0^1 \frac{\widehat{U}_n^2(t)}{t(1-t)} dt, \tag{6.31}$$

and it satisfies

$$\widehat{A}_n^2 \xrightarrow{D} \widehat{A}^2 = \int_0^1 \frac{\widehat{U}^2(t)}{t(1-t)} dt. \tag{6.32}$$

As mentioned in [35], the Cramer-Von Mises and Anderson-Darling tests are quite powerful for alternatives that are close to the uniform distribution and tend to perform better for these alternatives (in terms of power) than the Kolmogorov-Smirnov tests. However, in practice we are typically interested in detecting larger departures from the uniform distribution and, in such cases, all these tests perform well. Other tests related to the Cramer-Von Mises tests as well as the Kolmogorov-Smirnov tests can be found in [16].

e. Likelihood Ratio Test

Let $X_1, \ldots, X_n$ be a random sample from a certain distribution depending on a parameter $\theta \in \Theta \subset \mathbb{R}$. Let $L(\theta)$ be the likelihood function obtained by evaluating the density of $X_1, \ldots, X_n$ at their observed values $x_1, \ldots, x_n$. Assume that Θ is written as $\Theta = \Theta_0 \cup \Theta_1$.

The likelihood ratio (LR) test statistic for testing the null hypothesis $H_0: \ \theta \in \Theta_0$ against the alternative $H_1: \ \theta \in \Theta_1$ is defined by (see [33], page 219)

$$\Lambda = \max_{\theta \in \Theta_0} L(\theta) / \max_{\theta \in \Theta} L(\theta). \tag{6.33}$$

The null hypothesis H_0 is rejected if $\Lambda < c$ for a suitably chosen constant c.

Null Distribution of Λ

In order to use the LR test, we must know, either exactly or approximately, the distribution of the statistic Λ when H_0 is true. As described in [33] (page 220, Result 5.1), when the sample size is large, under the null hypothesis H_0,

$$-2 \ln \Lambda = -2 \ln \left(\max_{\theta \in \Theta_0} L(\theta) / \max_{\theta \in \Theta} L(\theta) \right)$$

is approximately a χ^2 random variable with $\nu - \nu_0$ degrees of freedom, where ν and ν_0 are the dimensions of Θ and Θ_0 respectively.

6.1.2 Univariate Test Statistics for Normality

a. Shapiro-Wilks Test Statistic

Description of the Test

Let $X_1, \ldots, X_n$ be an i.i.d. random sample from a normal distribution, and denote by $X_{1,n}, \ldots, X_{n,n}$, the corresponding order statistics. Let $m = (m_1, \ldots, m_n)$ denote the vector of expected values of standard normal order statistics, and let $V = (v_{ij})$ be the corresponding covariance matrix. The Shapiro-Wilk test statistic for normality is defined by

$$W = \left(\sum_{i=1}^{n} a_i X_{i,n} \right)^2 / \sum_{i=1}^{n} (X_i - \overline{X})^2, \tag{6.34}$$

where

$$(a_1, \ldots, a_n) = \frac{m' V^{-1}}{(m' V^{-1} V^{-1} m)^{1/2}},$$

m' denoting the transpose of m.

Some Properties of W

The W statistic has the following characteristics (see [49]).

1. W is scale and origin invariant.

   This result follows from the fact that for the normal distribution we have $a_i = -a_{n-i+1}$. From the property of invariance, we deduce that the statistic W has a distribution which depends only on the sample size, for samples from a normal distribution. Moreover, W is statistically independent of

   $$S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 \text{ and of } \overline{X},$$

   for samples from a normal distribution.

2. The maximum value of W is 1, and the minimum value is $\frac{na_1^2}{(n-1)}$.

3. The half and the first moments of W are given by

   $$E(W^{1/2}) = \frac{R^2 \Gamma((n-1)/2)}{C \Gamma(n/2) \sqrt{2}} \tag{6.35}$$

and

$$E(W) = \frac{R^2(R^2+1)}{C^2(n-1)}, \tag{6.36}$$

where $R^2 = m'V^{-1}m$ and $C^2 = m'V^{-1}V^{-1}m$. Proofs of the two previous properties can be found in [49].

4. **Approximations Associated With W.**

   The elements of V are generally difficult to obtain for large samples (for example the samples of size larger than 20 as mentioned in [49]). To overcome this problem, Shapiro and Wilk [49] set $a^\star = m'V^{-1}$, and then they suggested the following approximations for $a^\star$.

$$\widehat{a}_i^\star = 2m_i, \text{ for } i = 3, \ldots, n-1; \tag{6.37}$$

   and

$$\widehat{a}_1^\star = \widehat{a}_n^\star = \begin{cases} \frac{\Gamma(n/2)}{\sqrt{2}\Gamma((n+1)/2)}, & n \leq 20 \\ \\ \frac{\Gamma((n+1)/2)}{\sqrt{2}\Gamma(1+n/2)}, & n > 20. \end{cases}$$

b. De Wet-Venter Test Statistics

Consider $X_1, X_2, \ldots, X_n$, an i.i.d. random sample from a distribution function F. Let Φ be the standard normal distribution function. In order to test the simple hypothesis $H_0 : F = \Phi$, De Wet and Venter [10] suggested the statistic

$$L_n^0 = \sum_{i=1}^{n} \left(X_{i,n} - \Phi^{-1}(\frac{i}{n+1}) \right)^2, \tag{6.38}$$

where $X_{i,n}$'s are the order statistics. They also proposed the statistic

$$L_n = \sum_{i=1}^{n} \left(\frac{X_{i,n} - \overline{X}}{S} - \Phi^{-1}\left(\frac{i}{n+1}\right) \right)^2, \tag{6.39}$$

for testing the composite hypothesis $H_0^c : F(x) = \Phi(\frac{x-\mu}{\sigma})$, where μ and σ are unknown,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

In order to give the asymptotic behavior of L_n^0 and L_n, let $H = \Phi^{-1}$ and $H(i/n+1) = H_{in},\ i = 1,\ldots,n$. Let $\nu(x) = x(1-x)$, $h(x) = (H'(x))^2$ and

$$T(x,y) = \begin{cases} x(1-y) \text{ for } x \leq y \\ y(1-x) \text{ for } x > y. \end{cases}$$

The asymptotic result of L_n^0 is described in the following theorem (see [10]).

Theorem 6.1. Under the hypothesis H_0, we have

$$D(L_n^0 - a_n^0) \to D(\sum_{m=1}^{\infty}(Y_m^2 - 1)/m), \tag{6.40}$$

where

$$a_n^0 = \frac{1}{n+1}\sum_{k=1}^{n} h_{kn}\nu_{kn},$$

and $Y_1, Y_2, \ldots$ are i.i.d. $N(0,1)$ random variables.

Now let

$$a_n' = \frac{1}{n(n+1)}\sum_{m=1}^{\infty} H_{in}'H_{jn}'T_{ijn},$$

and

$$a_n'' = \frac{1}{n(n+1)}\sum_{m=1}^{\infty} H_{in}'H_{in}H_{jn}'H_{jn}T_{ijn}.$$

The asymptotic distribution of L_n is described in the theorem below.

Theorem 6.2. Under the hypothesis H_0^c, we have

$$D(L_n - a_n) \to D(\sum_{m=3}^{\infty}(Y_m^2 - 1)/m), \tag{6.41}$$

where $a_n = a_n^0 - a_n' - a_n''$, a_n^0 and Y_m's are i.i.d. $N(0,1)$-distributed.

Complete proofs of the two previous results can be found in [10].

c. A Goodness-of-Fit Test Based on Gini's Index of Spacings

In this section we discuss a goodness-of-fit test statistic based on Gini's index and we give its asymptotic distribution under the null hypothesis.

i. Description of the Test

Consider a random sample $X_1, \ldots, X_n$ from a continuous distribution function F on the real line and let $X_{1,n}, \ldots, X_{n,n}$ denote the order statistics from this sample. Consider testing the hypothesis

$$H_0 : F(x) = F_0(x)$$

against the alternative

$$H_1 : F(x) \neq F_0(x),$$

where F_0 is a completely specified continuous distribution function. Without loss of generality, let us reduce the above problem of goodness-of-fit to testing the hypothesis of uniformity on the unit interval, by means of the probability integral transformation $U = F(X)$. That is, on the basis of the transformed sample,

$$U_i = F_0(X_i),\ i = 1, \ldots, n,$$

the problem becomes one of testing uniformity, i.e.,

$$H_0' : G(u) = u,\ 0 < u < 1$$

against the alternative

$$H_1' : G(u) \neq u,\ 0 < u < 1,$$

where G is the distribution function of U.

Let $U_{0,n} = 0,\ U_{n+1,n} = 1$, and then define the one-step or simple spacings $\{D_i\}_{1 \le i \le n+1}$ by

$$D_i = U_{i,n} - U_{i-1,n}.$$

This can be generalized to the m-step spacings in the following ways (see [30]).

1. The overlapping m-step spacings are defined by

$$D_i^{(m)} = \begin{cases} U_{i+m,n} - U_{i,n}, \text{ for } i = 0, 1, \ldots, n+1-m \\ 1 + U_{i+m-n-1,n} - U_{i,n}, \text{ for } i = n+2-m, \ldots, n. \end{cases}$$

2. The non-overlapping m-step spacings are defined by

$$D_i^{(m)} = U_{(i+1).m} - U_{i.m},\ i = 0, 1, \ldots, [n/m] - 1,$$

where $[n/m]$ is the largest integer less than or equal to n/m.

Greenwood [26] proposed the statistic

$$T = \sum_{i=1}^{n+1} D_i^2$$

to test whether certain events such as the spread of disease occur at random (or follow a Poisson process with fixed rate) on the time axis. The spacings $\{D_i,\ i = 1, \ldots, n+1\}$ under the null hypothesis of uniformity, form an exchangeable set of random variables with an expected value of $\frac{1}{n+1}$. Thus tests of uniformity can be constructed by measuring how different $\{D_i,\ i = 1, \ldots, n+1\}$ are from their average value $\frac{1}{n+1}$. The test based on the squared deviation

$$V_n = \sum_{i=1}^{n+1} (D_i - \frac{1}{n+1})^2$$

corresponds to the Greenwood statistic. An alternate measure of deviation, namely the absolute deviation, leads to the statistic

$$R_n = \sum_{i=1}^{n+1} |D_i - \frac{1}{n+1}|.$$

Gini, in introducing his measure of dispersion, raised the objection to using the variance or the mean absolute deviation, since they measure the deviation of individual observations from the "center", thus interlinking the concept of location with variability. According to him, there are two distinct properties, needing distinct measures which do not depend on each other (see [30]). He then proposed the sum of pairwise distances between the observations, as a measure of deviation. Applying this idea to $\{D_i,\ i = 1, \ldots, n+1\}$, the Gini statistic G_n is defined as (see [30])

$$G_n = \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} |D_i - D_j|, \tag{6.42}$$

and the generalized Gini statistic is defined as

$$G_n(r) = \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} |D_i - D_j|^r,\ r > 0. \tag{6.43}$$

Note that $G_n(1) = G_n$, and $G_n(2)$ corresponds to the Greenwood statistic.

ii. Asymptotic Distribution of the Gini Statistic Under the Null Hypothesis

The Gini measure of dispersion, G_n, can be rewritten in terms of ordered spacing as follows:

$$\begin{aligned} G_n &= \sum_{i=1}^{n+1}\sum_{j=1}^{n+1}|D_{i,n} - D_{j,n}| \\ &= 4L_n - 2(n+2), \end{aligned}$$

where

$$L_n = \sum_{i=1}^{n+1} iD_{i,n}.$$

Defining

$$E_i = (n-i+2)[D_{(i)} - D_{(i-1)}],$$

for $i = 1, 2, \ldots, n+1$, we can write L_n in the form

$$L_n = \frac{n+1}{2} + \frac{1}{2}\sum_{i=1}^{n+1} iE_i.$$

Using the fact that under the null hypothesis $(E_1, E_2, \ldots, E_{n+1})$ are distributionally equivalent to $(D_1, D_2, \ldots, D_{n+1})$, Jammalamadaka and Goria [30] found that, under the null hypothesis,

$$\begin{aligned} L_n &\sim \frac{n+1}{2} + \frac{1}{2}\sum_{i=1}^{n+1} iD_i \\ &= \frac{n+1}{2} + \frac{L_n^\star}{2}, \end{aligned} \tag{6.44}$$

where

$$L_n^\star = \sum_{i=1}^{n+1} iD_i$$

is related to the sum

$$S_n = \sum_{i=1}^{n} U_i$$

of n uniform random variables on the unit interval. They finally obtained

$$G_n \sim 2(n - S_n), \tag{6.45}$$

under the null hypothesis, which yields the asymptotic normality

$$(3/n)^{1/2}(G_n - n) \xrightarrow{D} N(0,1), \text{ as } n \to \infty. \tag{6.46}$$

6.1.3 Other Univariate Test Statistics

Suppose we want to test if $X_1, \dots, X_n$ comes from a normal distribution $N(\mu, \sigma^2)$. The following univariate statistics are useful in order to carry out such a test.

1. The correlation statistic defined by

$$R = 1 - \widehat{\rho}^2, \; 0 < R < 1,$$

   where $\widehat{\rho}$ is the ordinary correlation coefficient between $X_{i,n}$ and $\mu_{i:n}$, with

$$\mu_{i:n} = E\left[\frac{X_{i,n} - \mu}{\sigma}\right], \; 1 \leq i \leq n.$$

2. The Tiku statistic based on the sample spacings:

$$Z = \left[2\sum_{i=1}^{n-1}(n-1-i)G_i\right] \Big/ \left[(n-2)\sum_{i=1}^{n-1}G_i\right], \; 0 < Z < \infty, \tag{6.47}$$

   where

$$G_i = \frac{X_{i+1,n} - X_{i,n}}{\mu_{i+1:n} - \mu_{i:n}}, \; 1 \leq i \leq n-1$$

   are the generalized sample spacings.

3. The Sürücü statistic for testing univariate normality is given by

$$C = 1 - \{[1 + a_1(a_2 - 1)]W + a_1(1 - a_2)(1 - R)\}$$

   where W is the Shapiro-Wilk statistic, and R is the correlation statistic defined previously,

$$a_1 = \exp\{-(b_1/0.6)^5\}, \; a_2 = \exp\{-(b_2/0.6)^5\},$$

   $\sqrt{b_1}$ and b_2 being respectively the sample skewness and kurtosis.

6.2 Bivariate Test Statistics for Normality

In the previous section we described several statistics for univariate tests. Now we consider bivariate test statistics.

6.2.1 Bivariate Kolmogorov-Smirnov Test Statistic

Description of the Test

Consider a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a bivariate cumulative distribution function H, and define the empirical distribution function associated with this sample by

$$H_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x, Y_i \le y\}}, \text{ for } (x,y) \in \mathbb{R}^2.$$

Suppose we want to test the null hypothesis $H_0 : \ H = H_T$ against a general alternative $H_A : \ H \neq H_T$, where H_T is a completely specified continuous bivariate cumulative distribution function.

The two-dimensional Kolmogorov-Smirnov statistic for this test is defined by (see [25])

$$K_n^b = \sqrt{n} \sup_{(x,y)\in\mathbb{R}^2} |H_n(x,y) - H_T(x,y)|. \tag{6.48}$$

Asymptotic Distribution of K_n^b

It can be shown that under H_0, the process

$$B_n^b = \sqrt{n}(H_n(x,y) - H_T(x,y))$$

converges in distribution to a two-parameter H_T-Brownian bridge B^b with covariance function

$$\mathrm{Cov}(B^b(x,y), B^b(x',y')) = H_T(\min(x,x'), \min(y,y')) - H_T(x,y)H_T(x',y').$$

Therefore, since K_n^b can be written in the form

$$K_n^b = \sup_{(x,y)\in\mathbb{R}^2} |B_n^b|,$$

we have the result

$$K_n^b \xrightarrow{D} \sup_{(x,y)\in\mathbb{R}^2} |B^b(x,y)|. \tag{6.49}$$

6.2.2 Test Based on Chi-square Plots

In this section we describe a testing procedure based on Chi-square plots. Consider a bivariate sample $X_1 = (Y_1, Z_1), \ldots, X_n = (Y_n, Z_n)$. A procedure for testing bivariate normality is the following (see [33], page 184):

1. Find the square distances $d_j^2 = (X_j - \overline{X})' S^{-1} (X_j - \overline{X})$, $j = 1, \ldots, n$;
2. Order the squared distances from the smallest to the largest as $d_{(1)}^2 \leq d_{(2)}^2 \leq \ldots \leq d_{(n)}^2$;
3. Graph the pairs $(q_{c,p}((j - \frac{1}{2})/n), d_j^2)$, where $q_{c,p}((j - \frac{1}{2})/n)$ is the $100(j - \frac{1}{2})/n)$ quantile of the Chi-square distribution with p degrees of freedom.

If the sample comes from a bivariate normal distribution, the plot should resemble a straight line through the origin having a slope one. A systematic curved pattern suggests lack of normality. The null hypothesis of bivariate normality is therefore rejected if the Chi-square plot differs considerably from a straight line through the origin having a slope one.

6.2.3 Statistics Obtained by Transforming Bivariate Data Into Univariate Observations

Let (X, Y) be a bivariate normal random vector with density function

$$f(x,y) = (2\pi\sigma_x\sigma_y\sqrt{1-\rho^2})^{-1} \exp\{-\frac{Q}{2(1-\rho^2)}\},$$

where

$$Q = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2.$$

Consider n pairs of observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ of (X, Y). In order to test the null hypothesis that these pairs are independently drawn from a bivariate normal distribution characterized by $f(x,y)$, Versluis [57] transformed the data set into the set of univariate variables

Z_i given by:

$$Z_i = (1-r^2)^{-1}\left\{\left(\frac{X_i - m_x}{S_x}\right)^2 - \frac{2r(X_i - m_x)(Y_i - m_y)}{S_x S_y} + \left(\frac{Y_i - m_y}{S_y}\right)^2\right\},\ i = 1, 2, \ldots, n,$$

with

$$m_x = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$m_y = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - m_x)^2},$$

$$S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - m_y)^2},$$

and

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - m_x}{S_x}\right)\left(\frac{Y_i - m_y}{S_y}\right).$$

The null distribution $F_0(z_{i,n})$ of the order statistic $Z_{i,n}$ is given by (see [57])

$$F_0(z_{i,n}) = 1 - \exp\left\{-\frac{z_{i,n}}{2}\right\}.$$

Let $F_0(z_{i,n}) =: F_i$. The following test statistics are given (see [57]):

1. The Kuiper statistic

$$T_K = \max_{1\le i\le n}\left\{\frac{i}{n} - F_i\right\} + \max_{1\le i\le n}\left\{F_i - \frac{i-1}{n}\right\}; \tag{6.50}$$

2. The Watson statistic

$$T_W = \frac{1}{12n} + \sum_{i=1}^{n}\left[F_i - \frac{2i-1}{2n}\right]^2 - n\left(\sum_{i=1}^{n}\frac{2F_i - 1}{2n}\right)^2; \tag{6.51}$$

3. The Renyi statistic L_1

$$T_{L_1} = \max_{1\le i\le n}\left\{\frac{nF_i}{i-1}\right\}; \tag{6.52}$$

4. The Renyi statistic L_2

$$T_{L_2} = \max_{1\le i\le n}\left\{\frac{i}{nF_i}\right\}; \tag{6.53}$$

5. The Renyi statistic U_1

$$T_{U_1} = \max_{1\le i\le n}\left\{\frac{n(1-F_i)}{n-i}\right\}; \tag{6.54}$$

6. The Renyi statistic U_2

$$T_{U_2} = \max_{1\le i\le n}\left\{\frac{n+1-i}{n(1-F_i)}\right\}; \tag{6.55}$$

6.2.4 Kim-Bickel Statistics for Bivariate Normality Testing

Let $X_1 = (X_{11}, X_{21})^T, \ldots, X_n = (X_{1n}, X_{2n})^T$ be a sample of independent observations on a 2-dimensional (column) vector $(X_1, X_2)^T$, where T denotes "transpose". Consider the hypothesis H_0 : the law of X is $BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ for some $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. Kim and Bickel [34] generalized the previous statistics L_n^o and L_n proposed by De Wet and Venter [10] for testing univariate normality using Roy's union-intersection principle, to the statistics P_n^o and P_n given by

$$P_n^0 = \sup_{c_1,c_2.\ni.c_1^2+c_2^2+2\rho c_2 c_2=1} \sum_{i=1}^{n}\{(c_1X_1 + c_2X_2)_{(i)} - \Phi^{-1}(\frac{i}{n+1})\}^2 \tag{6.56}$$

and

$$P_n = \sup_{c_1,c_2}\sum_{i=1}^{n} A(c_1,c_2)^2, \tag{6.57}$$

where

$$A(c_1,c_2) = \frac{(c_1X_1 + c_2X_2)_{(i)} - (c_1\overline{X}_1 + c_2\overline{X}_2)}{sd(c_1X_1 + c_2X_2)} - \Phi^{-1}(\frac{i}{n+1}),$$

$$\overline{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ki},$$

$$sd^2(c_1X_1 + c_2X_2) = c_1^2\widehat{\sigma}_1^2 + c_2^2\widehat{\sigma}_2^2 + 2c_1c_2\widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2,$$

$$\widehat{\sigma}_k^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{ki} - \widehat{X}_k)^2$$

and

$$\widehat{\rho} = \frac{1}{n}\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2)/(\widehat{\sigma}_1\widehat{\sigma}_2).$$

In order to characterize the asymptotic limit of P_n^0 and P_n, they introduced their truncated versions P_n^{0T} and P_n^T given by

$$P_n^{0T} = \sup_{c_1,c_2 \ni . c_1^2+c_2^2+2\rho c_2 c_2=1} \sum_{i=I_n}^{n-I_n} \{(c_1X_1 + c_2X_2)_{(i)} - \Phi^{-1}(\frac{i}{n+1})\}^2, \tag{6.58}$$

with $I_n = [n^{1-\delta}]$, $0 \leq \delta \leq 1/8$, and

$$P_n^T = \sup_{c_1,c_2} \sum_{i=I_n}^{n-I_n} A(c_1,c_2)^2 = \sup_{\theta\in[o,2\pi)} \sum_{i=I_n}^{n-I_n} A(\cos\theta, \sin\theta)^2, \tag{6.59}$$

where $\cos\theta = \frac{c_1}{\sqrt{c_1^2+c_2^2}}$ and $\sin\theta = \frac{c_2}{\sqrt{c_1^2+c_2^2}}$.

The following results are obtained for P_n^{0T} and P_n^T (see [34]).

Theorem 6.3.

$$P_n^{0T} - a_n^T \xrightarrow{D} \sup_{\theta\in[o,2\pi)} \int_0^1 \frac{B^2(y,\theta) - y(1-y)}{\phi^2(\Phi^{-1}(y))} dy, \tag{6.60}$$

where a_n^T is given by

$$a_n^T = \sum_{i=I_n}^{n-I_n} \frac{i}{n+1}(1 - \frac{i}{n+1})/\phi^2(\Phi^{-1}(\frac{i}{n+1})). \tag{6.61}$$

Theorem 6.4. Under the composite hypothesis H_0, $P_n^T - a_n^T$ converges in distribution to

$$\sup_{\theta\in[o,2\pi)} \left[\int_0^1 \frac{B^2(y,\theta) - y(1-y)}{\phi^2(\Phi^{-1}(y))} dy - \left(\int_0^1 \frac{B(y,\theta)}{\phi(\Phi^{-1}(y))} dy \right)^2 - \left(\int_0^1 \frac{B(y,\theta)}{\phi(\Phi^{-1}(y))} \Phi^{-1}(y) dy \right)^2 \right]. \tag{6.62}$$

Finally the asymptotic results for P_n^0 and P_n are given in the following conjectures (see [34]).

Conjecture 6.5. Under the simple hypothesis H_0^s : the law of X is $BVN(0,0,1,1,\rho)$,

$$P_n^0 - a_n^0 \xrightarrow{D} \sup_{\theta\in[o,2\pi)} \int_0^1 \frac{B^2(y,\theta) - y(1-y)}{\phi^2(\Phi^{-1}(y))} dy, \tag{6.63}$$

where a_n^0 is defined by

$$a_n^0 = \frac{1}{n} \sum_{i=1}^{n} \frac{i}{n+1}(1 - \frac{i}{n+1})/\phi^2(\Phi^{-1}(\frac{i}{n+1})),$$

$\phi(x) = d\Phi(x)/dx$.

Conjecture 6.6. Under the composite hypothesis H_0, the statistic $P_n - a_n^0$ has the limit distribution on the right side of Equation (6.62).

Kim and Bickel [34] used a Monte Carlo experiment to determine approximated upper percentiles of the null distribution of P_n and to study the power of the test based on P_n. The results indicated slow convergence of the quantiles of the test P_n especially in the upper tail. Since P_n is a bivariate version of De Wet-Venter's L_n, they came to the conclusion that the power results for two dimensions are consistent with those for one dimension. See [34] for more details about their analyzes.

6.3 Multivariate Test Statistics

In this section we consider the goodness-of-fit tests for multivariate distributions.

6.3.1 Goodness-of-Fit Test for Sphericity

We follow the discussion in [15] in order to describe a test for sphericity. We start by first giving the following definition.

Definition 6.1. Let $\mathbf{X}$ be a p-dimensional vector and $\mathbf{O}(p)$ be the set of $p \times p$ orthogonal matrices. The vector $\mathbf{X}$ is said to have a spherically symmetric distribution (or a spherical distribution) if, for every $Q \in \mathbf{O}(p)$,

$$Q\mathbf{X} \overset{D}{=} \mathbf{X},$$

where $\overset{D}{=}$ means that the two sides of the equality have the same distribution.

Description of the Test

Given a p-dimensional distribution function $G(\mathbf{x})$, consider testing the null hypothesis

$$H_0 : \; G(\mathbf{x}) \text{ is spherical}$$

against the alternative

$$H_1 : \; G(\mathbf{x}) \text{ is not spherical.}$$

The following lemma is important in order to carry out this test.

Lemma 6.7. Let $\mathbf{X} = (X_1, \ldots, X_p)^T$ be a random column vector. Then $\mathbf{X}$ has a spherical distribution if and only if for each $\mathbf{a} \in S_p$,

$$\mathbf{a}^T\mathbf{X} \stackrel{D}{=} X_1,$$

where S_p is defined by

$$S_p = \{\mathbf{a} : \mathbf{a} \in \mathbb{R}^p, \mathbf{a}^T\mathbf{a} = 1\}.$$

From Lemma 6.7, the hypothesis H_0 can be changed into

$$H_0: \text{ all } \mathbf{a}^T\mathbf{X},\ \mathbf{a} \in S_p, \text{ have the same distribution.}$$

Let $\mathbf{a}_1, \ldots, \mathbf{a}_m \in S_p$. The hypothesis H_0 can be approximated by

$$H_0^\star: \ \mathbf{a}_1^T\mathbf{X}, \ldots, \mathbf{a}_m^T\mathbf{X} \text{ have the same distribution.} \tag{6.64}$$

Now let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an i.i.d. sample from G. For given $0 < k < l \leq m$, consider the two-sample testing problem:

$$H_0: \ \mathbf{a}^T\mathbf{X}_k \stackrel{D}{=} \mathbf{a}^T\mathbf{X}_l.$$

A Wilcoxon-type statistic for this problem is

$$V_n(\mathbf{a}_k, \mathbf{a}_l) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} I_{\{\mathbf{a}_k^T\mathbf{X}_i < \mathbf{a}_l^T\mathbf{X}_j\}}, \tag{6.65}$$

where I is the indicator function. From there, a statistic for testing the hypothesis $H_0^\star$ is given by

$$T_n = \min_{1 \leq k,l \leq m, k \neq l} \{V_n(\mathbf{a}_k, \mathbf{a}_l)\}. \tag{6.66}$$

Limiting Distributions of V_n and T_n

The limiting results of V_n and T_n are described in the following two theorems (see [15]).

Theorem 6.8. Assume that G is spherically symmetric with no atom at the origin, i.e. $P[\mathbf{X} = \mathbf{0}] = 0$. Then the distribution of the row random rector $\{\sqrt{n}(V_n(\mathbf{a}_k, \mathbf{a}_l) - \frac{1}{2}) : \ 1 \leq k, l \leq p\}$ converges weakly to $(p(p-1)/2)$-dimensional normal distribution $N(0, \frac{1}{12}V)$, where the elements of V have the form

$$V(k,l,k_1,l_1) = \begin{cases} 0 \text{ for } k \neq l \neq l_1 \neq k_1; \\ 1 \text{ for } k = k_1,\ l \neq l_1 \text{ or } k \neq k_1,\ l = l_1; \\ 2 \text{ for } k = k_1,\ l = l_1; \\ -1 \text{ for } k = l_1 \text{ or } k_1 = l. \end{cases} \tag{6.67}$$

Theorem 6.9. Under the same assumption as in Theorem 6.8, we have for any real number λ,

$$\begin{aligned} \lim_{n\to\infty} P[\sqrt{n}(T_n - \frac{1}{n}) \leq -\lambda] &= P[T \leq -\lambda] \\ &= 1 - \int_{-\infty}^{\infty} p(\Phi(\sqrt{12}\lambda + z) - \Phi(z))^{p-1}\phi(z)dz, \end{aligned} \tag{6.68}$$

where T is defined as:

$$T = \min_{1\leq k,l\leq p, k\neq l} \{Y_l - Y_k\}, \tag{6.69}$$

$Y_1, \dots, Y_p$ are independent $N(0, \frac{1}{12})$ random variables, and Φ and ϕ are the cumulative distribution and the density functions of $N(0,1)$ distribution, respectively.

The previous theorems are proved in detail in [15].

6.3.2 Multivariate Cramer-Von Mises Statistic

Let $\mathbf{F}$ be the class of continuous distribution functions on d-dimensional Euclidean space $\mathbb{R}^d$ $(d \geq 1)$, and let $\mathbf{F}_0$ be the subclass consisting of every member of $\mathbf{F}$ which is a product of its associated one-dimensional marginal distribution functions. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent random d-vectors with a common distribution function $F \in \mathbf{F}$ and let $F_n(\mathbf{x})$ be the associated empirical distribution function.

The multivariate Cramer-Von Mises statistic is given by (see [7])

$$W^2_{n,d} = \int_{\mathbb{R}^d} \beta_n(\mathbf{x}) \prod_{i=1}^{d} dF_i(x_i), \tag{6.70}$$

where β_n is the empirical process defined by

$$\beta_n(\mathbf{x}) = \sqrt{n}(F_n(\mathbf{x}) - F(\mathbf{x})),\ \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d,\ d \geq 1.$$

For $F \in \mathbf{F}_0$, letting $y_i = F_i(x_i)$, we have

$$\begin{aligned}\beta_n(x) &= \sqrt{n}\left(F_n(\mathbf{x}) - \prod_{i=1}^{d} F_i(x_i)\right)\\ &= \sqrt{n}(F_n(F_1^{-1}(y_1), \ldots, F_d^{-1}(y_d)) - \lambda(y))\\ &= \sqrt{n}(E_n(y) - \lambda(y))\\ &= \alpha_n(y), \end{aligned} \tag{6.71}$$

where

$$E_n(y) = F_n(F_1^{-1}(y_1), \ldots, F_d^{-1}(y_d)) \text{ and } \lambda(y) = \prod_{i=1}^{d} F_i(x_i) = \prod_{i=1}^{d} y_i.$$

Note that α_n is a uniform empirical process.

With this notation, $W_{n,d}^2$ can be written as

$$W_{n,d}^2 = \int_{I^d} \alpha_n^2(y) \prod_{i=1}^{d} dy_i, \ d \geq 1,$$

where I^d is the d-dimensional unit cube.

Limiting Distribution of $W_{n,d}^2$

Before we describe the limiting result for $W_{n,d}^2$, we give the following important theorem (see [7]).

Theorem 6.10. Let $X_1, \ldots, X_n$ $(n = 1, 2, \ldots)$ be independent random d-vectors with a common distribution function $F \in \mathbf{F}$ and let $\alpha_n(.)$ be as in (6.71). Then one can construct a probability space $(\Omega, \mathbf{A}, P)$ with $\{\alpha_n(y); y \in I^d \ (d \geq 1), n = 1, 2, \ldots\}$ and a sequence of Brownian bridges $\{B_n(y); y \in I^d \ (d \geq 1)\}$ on the space so that for any $\mu > 0$, there exists a constant $C > 0$ such that for each n,

$$P\left[\sup_{y \in I^2} |\alpha_n(y) - B_n(y)| > C(\log n)^{3/2} n^{-1/(2(d+1))}\right] \leq n^{-\mu}. \tag{6.72}$$

Further, if $d = 2$ then for all n and x,

$$P\left[\sup_{y \in I^2} |\alpha_n(y) - B_n(y)| > n^{-1/2}(C \log n + x) \log n\right] < Le^{-\lambda x}, \tag{6.73}$$

where C, L, λ are positive absolute constants.

From Equation (6.73), we have for $d \geq 1$,

$$h(\alpha_n(.)) \xrightarrow{D} h(B(.)),$$

for every continuous functional h on the space of real valued functions on I^d endowed with the supremum topology, and so we have

$$W_{n,d}^2 \xrightarrow{D} W_d^2 = \int_{I^d} B^2(y)dy, \tag{6.74}$$

where

$$dy = \prod_{i=1}^{d} dy_i.$$

As suggested in [7], as to higher dimensions, $d \geq 2$, no analytic results appear to be known about the exact distribution of $W_{n,d}^2$, but the characteristic function of W_d is known and given by the equation

$$\phi^{-2}(t) = 2^d \sum_{n=0}^{\infty} L_{n+1} u^n \exp\left(-\sum_{n=1}^{\infty} L_n^d \frac{u^n}{n}\right), \tag{6.75}$$

where $u = 2it$, $|u| < (\pi/2)^d$, and L_n is defined by

$$L_n = \sum_{j=1}^{\infty} \{(j - 1/2)^2\pi^2\}^{-n} = \left(\frac{2}{\pi}\right)^{2n} \sum_{j=0}^{\infty} (1+2j)^{-2n}.$$

Moreover, a principal component decomposition can be used to write W_d in the form

$$W_d^2 = \sum_{k=1}^{\infty} \lambda_k^{-1} Z_k^2, \;\; d \geq 1, \tag{6.76}$$

where the Z_k's are independent standard normal random variables and the λ_k's are the eigenvalues of the integral equation

$$\int_{I^d} E[B(x_1)B(x_2)]f(x_2)dx_2 = \lambda f(x_1),$$

with the eigenfunctions f and the kernel $E[B(x_1)B(x_2)]$.

Cotterill and Csörgö [7] proved that the mean μ_d and the variance σ_d^2 of the random variable W_d^2 are given by

$$\mu_d = 2^{-d} - 3^{-d} \text{ and } \sigma_d^2 = 2.3^{-d}\left\{2^{-d} - 2\left(\frac{5}{2}\right)^{-d} + 3^{-d}\right\}.$$

It is easy to see that μ_d and σ_d^2 go to zero as d goes to infinity, and so the distribution of W_d^2 concentrates as a unit mass at the origin.

6.3.3 De Wet-Venter Statistics for Multivariate Normality Testing

Description of the Test

Let $X_1, \ldots, X_n$ be i.i.d. p-dimensional random vectors with distribution function F, and consider testing the hypothesis $H_0 : F = N_p(\mu, \Sigma)$ for some unspecified μ and Σ, $N_p(\mu, \Sigma)$ being the p-dimensional normal distribution with mean μ and covariance matrix Σ. Consider a univariate test statistic $T(V_1, \ldots, V_n)$ with observations $V_1, \ldots, V_n$, and suppose that T is chosen so that large values are considered significant. Let a' denote the transpose of a.

Based on the fact that X_i is $N_p(\mu, \Sigma)$-distributed if and only if $a'X_i$ is $N_p(a\mu, a'\Sigma a)$-distributed for all p-dimensional real vectors a, De Wet, Venter and van Wyk [11] proposed the test statistic T_n given by

$$T_n = \sup_{a \neq 0} T(a'X_1, \ldots, a'X_n). \tag{6.77}$$

Assuming that T is location and scale invariant, they came up to the conclusion that under the null hypothesis H_0, the distribution of T_n does not depend on μ and it depends on Σ only through its rank.

Test Based on Skewness

In this case, De Wet, Venter and van Wyk [11] considered $T = T_1$, with

$$T_1(V_1, \ldots, V_n) = n^{-1/2} \sum_{j=1}^{n} (V_j - \overline{V}_n)/s^3,$$

where

$$\overline{V}_n = \frac{1}{n} \sum_{j=1}^{n} V_j \text{ and } s^2 = \frac{1}{n} \sum_{j=1}^{n} (V_j - \overline{V}_n)^2.$$

The test statistic for multivariate normality based on skewness is given by

$$T_{1n} = \sup_{\|a\|=1} T_{1n}(a), \tag{6.78}$$

where $T_{1n}(a) = T_1(a'X_1, \ldots, a'X_n)$.

In order to find the asymptotic distribution of T_n, they first remark that for $V_1, \ldots, V_n$, i.i.d. $N(0,1)$-distributed,

$$T_1(V_1, \ldots, V_n) - (6/n)^{1/2} \sum_{j=1}^{n} h_3(V_j) \xrightarrow{P} 0, \text{ as } n \to \infty,$$

where $h_3(V) = (V^3 - 3V)/\sqrt{6}$ is the third normalized Hermitian polynomial. They then defined the statistic

$$T_{1n}^\star = \sup_{\|a\|=1} T_{1n}^\star(a),$$

where

$$T_{1n}^\star(a) = (6/n)^{1/2} \sum_{j=1}^{n} h_3(a' X_j),$$

and showed that

$$T_{1n} - T_{1n}^\star \xrightarrow{P} 0, \text{ as } n \to \infty.$$

They finally proved that the process $T_{1n}^\star(a)$ (and thus $T_{1n}(a)$) converges weakly to the process

$$Y_3(a) = b'W,$$

where

$$b' = (b^{(1)'}, b^{(2)'}, b^{(3)'})$$

with

$$b^{(1)'} = (a_1^3 a_2^3 \ldots a_p^3),$$

$$b^{(2)'} = \sqrt{3}(a_1^2 a_2 a_1 a_2^2 \ldots a_{p-1} a_p^2),$$

$$b^{(3)'} = \sqrt{6}(a_1 a_2 a_3 \ldots a_{p-2} a_{p-1} a_p),$$

and W is $N_k(0, I)$-distributed with $k = \frac{1}{6}p(p+1)(p+2)$.

Tests Based on Kurtosis

In this case, $T = T_2$, with

$$T_2(V_1, \ldots, V_n) = n^{-1/2} \sum_{j=1}^{n} [(V_j - \overline{V}_n)^4 / s^4 - 3].$$

Letting $T_{2n}(a) = T_2(a'X_1, \ldots, a'X_n)$, De Wet, Venter and van Wyk [11] considered the statistics

$$T_{2n} = \sup_{\|a\|=1} T_{2n}(a), \tag{6.79}$$

$$T_{3n} = \sup_{\|a\|=1} (-T_{2n}(a)) \tag{6.80}$$

and

$$T_{4n} = \sup_{\|a\|=1} |T_{2n}(a)|. \tag{6.81}$$

Using the fact that

$$T_2(V_1, \ldots, V_n) - (24/n)^{1/2} \sum_{j=1}^{n} h_4(v_j) \xrightarrow{P} 0,$$

where

$$h_4(v) = \frac{v^4 - 3v^3 + 6}{\sqrt{24}},$$

the following results were obtained.

$$D(T_{2n}) \to D\left(\sup_{\|a\|=1} (b'W)\right), \tag{6.82}$$

$$D(T_{3n}) \to D\left(\sup_{\|a\|=1} (-b'W)\right) \tag{6.83}$$

and

$$D(T_{4n}) \to D\left(\sup_{\|a\|=1} |b'W|\right), \tag{6.84}$$

where b and W are k-dimensional vectors, $k = \frac{1}{24}p(p+1)(p+2)(p+3)$, such that W is $N_k(0, I)$-distributed (i.e. W is a k-variate normal random variable with zero mean vector and covariance matrix I, $k \times k$ identity matrix), and

$$b' = \sqrt{24}\left(b^{(1)'}, b^{(2)'}, b^{(3)'}, b^{(4)'}, b^{(5)'}\right)$$

with

$$b^{(1)'} = (a_1^4 \ldots a_p^4),$$

$$b^{(2)'} = 2(a_1^3 a_2 \ldots a_{p-1} a_p^3),$$

$$b^{(3)'} = \sqrt{6}(a_1^2 a_2^2 \ldots a_{p-1}^2 a_p^2),$$

$$b^{(4)'} = \sqrt{12}(a_1^2 a_2 a_3 \ldots a_{p-2} a_{p-1} a_p^2)$$

and

$$b^{(5)'} = \sqrt{24}(a_1a_2a_3a_4\dots a_{p-3}a_{p-2}a_{p-1}a_p).$$

Test Based on Graphical Method

In this case, T is chosen in the form

$$T(V_1,\dots,V_n) = \sum_{i=1}^{n}\left[\frac{V_{i,n}-\overline{V}_n}{S} - H_{i,n}\right]^2 - a_n, \tag{6.85}$$

where $V_{1,n},\dots,V_{n,n}$ are the order statistics of $V_1,\dots,V_n$, $\overline{V}$ is their average, S is their standard deviation and $H_{in} = \Phi^{-1}(\frac{i}{n+1})$, Φ being the standard normal distribution function. As in [11], consider

$$a_n = \frac{1}{n+1}\sum_{k=1}^{n} c_{kn} - \frac{3}{2},$$

where

$$c_{kn} = \frac{k}{n+1}\left(1-\frac{k}{n+1}\right)(\phi(H_{kn}))^{-2}$$

ϕ denoting the standard normal density function. Letting

$$T_n(a) = T(a'X_1,\dots,a'X_n),\ a\in\mathbb{R}^p,$$

the test statistic for multivariate normality corresponding to the univariate test (6.85) is given by

$$T_{5n} = \sup_{\|a\|=1} T_n(a). \tag{6.86}$$

It can be shown that

$$T(V_1,\dots,V_n) - \frac{1}{n}\sum_{i\neq j}^{n} Q(U_i,U_j) \xrightarrow{P} 0, \tag{6.87}$$

with $U_i = \Phi(V_i)$, so that $U_1,\dots,U_n$ are i.i.d. uniformly distributed on the interval $(0,1)$, and Q is given by

$$Q(u,v) = Q_0(u,v) - h_1(\Phi^{-1}(u))h_1(\Phi^{-1}(v)) - \frac{1}{2}h_2(\Phi^{-1}(u))h_2(\Phi^{-1}(v)),$$

where

$$Q_0(u,v) = \int_0^1 \left[I_{\{u\le t\}}I_{\{v\le t\}} - t\right]\left(\phi(\Phi^{-1}(t))\right)^{-2},$$

I being the indicator function, $h_1(x) = x$ and $h_2(x) = \frac{x^2-1}{\sqrt{2}}$.

Now let a_1 and a_2 be two-dimensional vectors with $\|a_1\| = \|a_2\| = 1$ and let $U_{1i} = \Phi(a_1' X_i)$ and $U_{2i} = \Phi(a_2' X_i)$. From Equation (6.87), we have

$$\begin{pmatrix} T_n(a_1) \\ \\ T_n(a_2) \end{pmatrix} - \begin{pmatrix} \frac{1}{n}\sum_{i\neq j} Q(U_{1i}, U_{1j}) \\ \\ \frac{1}{n}\sum_{i\neq j} Q(U_{2i}, U_{2j}) \end{pmatrix} \xrightarrow{P} \begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}. \tag{6.88}$$

Through the bilinear expansion

$$Q(u,v) = \sum_m \gamma_m h_m(\Phi^{-1}(u)) h_m(\Phi^{-1}(v)),$$

De Wet, Venter and van Wyk [11] showed that

$$D \begin{pmatrix} \frac{1}{n}\sum_{i\neq j} Q(U_{1i}, U_{1j}) \\ \\ \frac{1}{n}\sum_{i\neq j} Q(U_{2i}, U_{2j}) \end{pmatrix} \to D \begin{pmatrix} \sum_{m=3}^{\infty} (Y_{1m}^2 - 1)/m \\ \\ \sum_{m=3}^{\infty} (Y_{2m}^2 - 1)/m \end{pmatrix}, \tag{6.89}$$

$$\begin{pmatrix} Y_{1m} \\ \\ Y_{2m} \end{pmatrix},\ m = 3, 4, \ldots$$

being i.i.d. bivariate normal random vectors with mean vector

$$\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}$$

and covariance matrix

$$\begin{pmatrix} 1 & \rho_m \\ \\ \rho_m & 1 \end{pmatrix},$$

where

$$\rho_m = E\left[h_m(a_1' X_1) h_m(a_2' X_1)\right].$$

Using the fact that $a_1' X_1$ and $a_2' X_1$ are jointly bivariate normal with mean vector

$$\begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}$$

and covariance matrix

$$\begin{pmatrix} 1 & a_1'a_2 \\ & \\ a_1'a_2 & 1 \end{pmatrix},$$

ρ_m can be written as $\rho_m = (a_1'a_2)^m$.

Finally consider a sequence of independent Gaussian processes $\{Y_m(a)\},\ a \in \mathbb{R}^p$ such that $\|a\| = 1$, with $E[Y_m(a)] = 0$ and $E[Y_m(a_1)Y_m(a_2)] = (a_1'a_2)^m$. Then the sequence of stochastic processes $T_n(a)$ converges weakly to the process

$$Y(a) = \sum_{m=3}^{\infty}(Y_m^2(a) - a)/m,$$

and thus we have

$$D(T_{5n}) \to D\left(\sup_{\|a\|=1} Y(a)\right). \tag{6.90}$$

6.3.4 The Average Projection Type Weighted Cramer-Von Mises Statistics

Cui [9] constructed Cramer-Von Mises test statistics for testing some multivariate distributions, especially (I) the uniform distributions on p-dimensional unit sphere, (II) p-dimensional standard normal distributions and (III) p-dimensional normal distributions with unknown mean vector and covariance matrix.

Define the unit sphere surface by

$$S^{p-1} = \{a : a \in \mathbb{R}, \| a \|= 1\}.$$

Let

$$F^a(x) = P\{a^T X_1 < x\}$$

and

$$F_n^a(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{a^T X_i < x\}},\ a \in S^{p-1},\ x \in \mathbb{R}^1.$$

Test for the Uniform Distributions on p-Dimensional Unit Sphere

Suppose that $\{X_1, \ldots, X_n\}$ is a random sample from F_0, the uniform distribution function on S^{p-1} and consider testing the hypothesis $H_0 : F = F_0$ versus $H_1 : F \neq F_0$. The AP type weighted Cramer-Von Mises test statistic for testing the uniform distributions on p-dimensional unit sphere is defined by (see [9])

$$\begin{aligned} W_n &= n \int_{S^{p-1}} \int_{\mathbb{R}^1} [F_n^a(t) - F_0^a(t)]^2 w(F_0^a(t)) dF_0^a(t) d\mu(a) \\ &=: \int_{S^{p-1}} W_n(a) d\mu(a), \end{aligned} \tag{6.91}$$

where

$$W_n(a) =: n \int_{\mathbb{R}^1} [F_n^a(t) - F_0^a(t)]^2 w(F_0^a(t)) dF_0^a(t).$$

$$F_0^a(t) = P\{a^T U < t\} = A_0(p) \int_{-1}^{t} (1 - u^2)^{(p-3)/2} du =: G(t)$$

is the distribution function of $a^T U$,

$$f_0(u) = A_0(p)(1 - u^2)^{(p-3)/2} I_{\{|u| \leq 1\}}$$

is the density function of $a^T U$,

$$w(F_0^a(t)) = w(G(t)) = (1 - t^2)^{(p-3)/2} I_{\{|t| \leq 1\}}, \ a \in S^{p-1},$$

$$A_0(p) = \Gamma(p/2)/[\Gamma(1/2)\Gamma((p-1)/2)],$$

and $\mu(.)$ is the uniform measure on S^{p-1}.

Test for p-Dimensional Standard Normal Distributions

Let $X_1, \ldots, X_n$ be i.i.d. with the common p-dimensional distribution function $F(x)$, let $F_0(x)$ be the p-dimensional standard normal distribution function. Consider the testing hypothesis $H_0 : \ F = F_0$ versus $H_1 : \ F \neq F_0$. The AP type weighted Cramer-Von Mises test statistic for testing the p-dimensional standard normal distributions is defined by (see [9])

$$\begin{aligned}
T_n =: & \int_{S^{p-1}} \left[n \int_{\mathbb{R}^1} \left(F_n^a(t) - \Phi(t)\right)^2 \phi^{-2}(t) d\Phi(t) - \frac{1}{n} \sum_{i=1}^{n} Q(a^T X_i, a^T X_i) \right] d\mu(a) \\
= & \frac{1}{n} \sum_{i \neq j} \int_{S^{p-1}} Q(a^T X_i, a^T X_j) d\mu(a) \\
=: & \int_{S^{p-1}} T_n(a) d\mu(a),
\end{aligned} \tag{6.92}$$

where

$$T_n(a) = \frac{1}{n} \sum_{i \neq j} Q(a^T X_i, a^T X_j),$$

$$Q(x, y) = \int_{\mathbb{R}^1} \left[I_{\{x<t\}} - \Phi(t) \right] \left[I_{\{y<t\}} - \Phi(t) \right] \phi^{-1}(t),$$

and $\phi(t)$, $\Phi(t)$ are respectively the density and the distribution functions of the standard normal distribution.

Test for p-Dimensional Standard Normal Distributions with Unknown Mean Vector and Covariance Matrix

Let $X_1, \ldots, X_n$ be i.i.d. with the common p-variate distribution function $F(x)$, let $F_0(x, \mu, \Sigma)$ be a p-variate normal distribution function with unknown parameters μ and $\Sigma > 0$. Consider the testing hypothesis $H_0: \ F = F_0(\mu, \Sigma)$ versus $H_1: \ F \neq F_0(\mu, \Sigma)$.

The AP type weighted Cramer-Von Mises test statistic for testing the p-dimensional normal distributions with unknown mean vector and covariance matrix is defined by (see [9])

$$T_n^{\star} = \int_{S^{p-1}} T_n^{\star}(a) d\mu(a), \tag{6.93}$$

where

$$\begin{aligned}
T_n^{\star}(a) = & \frac{1}{n} \sum_{i \neq j} \int_{S^{p-1}} \int_{\mathbb{R}^1} \left[1_{\{a^T X_i < x\}} - \Phi(x; \widehat{\theta}_1(a), \widehat{\theta}_2(a)) \right] \times \\
& \left[1_{\{a^T X_j < x\}} - \Phi(x; \widehat{\theta}_1(a), \widehat{\theta}_2(a)) \right] w(\Phi(x; \widehat{\theta}_1(a), \widehat{\theta}_2(a))) d\Phi(x; \widehat{\theta}_1(a), \widehat{\theta}_2(a))
\end{aligned} \tag{6.94}$$

and

$$\Phi(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} \int_{-\infty}^{x} \exp\{-(t - \theta_1)^2 / (2\theta_2)\} dt,$$

$w(.)$ is a weight function.

In his work, Cui [9] has provided the asymptotic results for $W_n,\ T_n$ and $T_n^{\star}$. These results are given in the following three theorems.

Theorem 6.11. Under the null hypothesis H_0 given above and for $p \geq 2$, we have

$$W_n \rightsquigarrow \sum_{j=1}^{\infty} \lambda_j \frac{\chi_j^2(h_j^2(1))}{h_j^2(1)} =: W \text{ (as } n \to \infty), \tag{6.95}$$

where

$$\lambda_j = \left(\frac{\Gamma(p/2)}{\Gamma(1/2)\Gamma((p-1)/2)} \right)^2 \frac{1}{j(j+p-2)},\ h_j^2 = \binom{p-2+j}{p-2} + \binom{p-3+j}{p-2}\ (j \geq 1),$$

and $\chi_1^2(h_1^2(1)), \chi_2^2(h_2^2(1)), \ldots$ are independent, chi-square distributions with $h_j^2(1)$ degree of freedom for $j = 1, 2, \ldots$.

We also have the following corollary.

Corollary 6.12. If $p = 2$ and the null hypothesis H_0 is true, then $h_j^2(1) = 2,\ \lambda_j = \frac{1}{\pi^2 j^2}$, and

$$W_n \rightsquigarrow \sum_{j=1}^{\infty} \frac{\chi_j^2(2)}{2\pi^2 j^2} =: W^{\star}$$

with $E(W^{\star}) = 1/6,\ \mathrm{Var}(W^{\star}) = 1/90$.

Theorem 6.13. Under the null hypothesis H_0 given above, we have

$$T_n \rightsquigarrow \sum_{k=1}^{\infty} \frac{1}{k} \left[\int_{S^{p-1}} Z_k^2(a) d\mu(a) - 1 \right] =: T, \tag{6.96}$$

where $Z_1(a), Z_2(a) \ldots$ are independent Gaussian processes with mean zero and

$$E[Z_j(a) Z_j(b)] = (a^T b)^j,\ j = 1, 2, \ldots.$$

Theorem 6.14. Under the null hypothesis H_0 in section 6.3.4, we have

$$T_n^{\star} \rightsquigarrow \sum_{k=3}^{\infty} \frac{\int_{S^{p-1}} Z_k^2(a) d\mu(a) - 1}{k} - \frac{3}{2} =: T^{\star}\ (n \to \infty), \tag{6.97}$$

and

$$E(T^{\star}) = -3/2,\ \mathrm{Var}(T^{\star}) = \left(2 \sum_{k=3}^{\infty} (\frac{1}{2}) k \right) / \left(k^2 (\frac{p}{2}) k \right),$$

where $Z_3(a), Z_4(a), \ldots$ are independent Gaussian processes with mean zero and

$$E[Z_k(a)Z_k(b)] = (a^T b)^k$$

for any $a, b \in S^{p-1},\ k = 3, 4, \ldots$. Moreover, the distributions of $\int_{S^{p-1}} Z_k^2(a)d\mu(a),\ k = 3, 4, \ldots$ are the same as in Theorem 5 in [9]. If $k = 3$ or 4, the distributions of $\int_{S^{p-1}} Z_3^2(a)d\mu(a)$, $\int_{S^{p-1}} Z_4^2(a)d\mu(a)$ are the same as in Theorem 4 in [9].

6.3.5 Other Statistics for Testing Multivariate Normal Distributions

The p-variate statistics Z_p, C_p, R_p, W_p we will describe in this section are generalizations of the univariate statistics Z, C, and R given in Section 6.1.3, and W given in Section 6.1.2.

Suppose that $(X_1, \ldots, X_p)$ has a p-variate normal distribution. Then (see [54]) $Y_1 = X_1, Y_2 = X_2 - \beta_{2.1}X_1, \ldots, Y_p = X_p - \beta_{p1.q1}X_1 - \ldots - \beta_{p(p-1).q_{p-1}}X_{p-1}$ are independently distributed as univariate normal, β_j's being the partial correlation coefficients.

Given a sample $(X_{1i}, \ldots, X_{pi}),\ (1 \le i \le n)$ of size n, consider the random observation

$$Y_{1i} = X_{1i}, Y_{2i} = X_{2i} - \widehat{\beta}_{2.1}X_{1i}, \ldots, Y_{pi} = X_{pi} - \widehat{\beta}_{p1.q1}X_{1i} - \ldots - \widehat{\beta}_{p(p-1).q_{p-1}}X_{(p-1)i},$$

where $\widehat{\beta}_j$'s are the least squares estimators of β_j's.

This transformation leads to the p-variate statistics Z_p, C_p, R_p and W_p given as follow.

1. $$Z_p = \sum_{j=1}^{p} (\frac{Z_j - 1}{\sqrt{V}})^2,\ 0 < Z_p < \infty, \tag{6.98}$$

   where Z_j's, $1 \le j \le p$, are the values of Z calculated from the samples $Y_{1i}, \ldots, Y_{pi},\ 1 \le i \le n$ and $V = var(Z_j)$ for large n.

   The asymptotic distribution of Z_p is a χ^2 distribution with p degrees of freedom.

2. $$C_p = \sum_{j=1}^{p} C_j,\ 0 < C_p < \infty \tag{6.99}$$

   where C_j's, $1 \le j \le p$, are the values of C calculated from the samples $Y_{1i}, \ldots, Y_{pi},\ 1 \le i \le n$.

   The null distribution of C_p is well approximated by a three-moment chi-square with ν

degrees of freedom, and this is given by the relation

$$\chi^2_\nu = \frac{C_p + a}{b},$$

where ν, a and b are obtained by equating the first three moments on both sides.

3.
$$R_p = \sum_{j=1}^{p} R_j,\ 0 < R_p < \infty, \tag{6.100}$$

where R_j's, $1 \le j \le p$, are the values of R calculated from the samples $Y_{1i}, \ldots, Y_{pi},\ 1 \le i \le n$.

4. The p-variate version of the Shapiro-Wilk statistic is given by

$$W_p = \left(\sum_{i=1}^{n} a_{i:n} u_{i,n} \right) / (X_m - \overline{X}) S^{-1} (X_m - \overline{X})^T,\ 0 \le W_p \le \infty, \tag{6.101}$$

where x_m satisfies

$$(x_m - \overline{x}) S^{-1} (x_m - \overline{x})^T = \max_{1 \le i \le n} u_i,$$

$$u_i = (x_i - \overline{x}) S^{-1} (x_i - \overline{x})^T,\ 1 \le i \le n,$$

and the coefficients $a_{i:n}$ are the same as in the case of W.

Once these statistics are constructed, the power properties of the aforementioned tests have been studied by generating $p-$variate normal and a wide variety of nonnormal distributions. This study leads to the conclusion that C_p is overall the most powerful and is effective against skew, long-tailed as well as short-tailed symmetric alternatives. Moreover it has been shown that Z_p and R_p extend to nonnormal distributions. See [54] for the details.

6.4 Parametric Bootstrap Procedure for Goodness-of-Fit Testing

When testing if a given distribution P belongs to a parameterized family **P**, one often compares a nonparametric estimate A_n of some functional A of P with an element A_{θ_n} corresponding to an estimate θ_n of the parameter θ. In many cases, the asymptotic distribution of goodness-of-fit test statistics derived from the process $\sqrt{n}(A_n - A_{\theta_n})$ depends on the unknown distribution P.

In this section we consider the parametric bootstrap methods proposed by Stute et al [52], and we give the validity of the methods when testing the goodness-of-fit of families of multivariate distributions and copulas. We also discuss the validity of a two-level bootstrap in cases that the parametric estimate can not be computed easily.

6.4.1 The Parametric Bootstrap Proposed by Stute et al.

Let $X_1, \ldots, X_n$ be independent copies of a random variable X with cumulative distribution function $F : \mathbb{R}^d \to \mathbb{R}$, and suppose we want to test the hypothesis

$$H_0 : F \in \mathbf{F} = \{F_\theta : \theta \in \mathbf{O}\},$$

i.e. the hypothesis that F comes from a parametric family of distributions whose members are indexed by a parameter $\theta \in \mathbf{O} \subset \mathbb{R}$. Let F_{θ_n} be a parametric estimate of F derived under H_0 from some consistent estimate $\theta_n = T(X_1, \ldots, X_n)$ of the true parameter value θ_0.

Cramer-Von Mises, Kolmogorov-Smirnov and many other standard goodness-of-fit procedures are based on statistics expressed as continuous functionals $S_n = \phi(\mathbb{G}_n^F)$ of the empirical process

$$\mathbb{G}_n^F = n^{1/2}(F_n - F_{\theta_n}),$$

where F_n is the empirical distribution of X. Formal tests, however, require knowledge of the asymptotic null distribution of S_n, which often depends on θ. To solve this problem, Stute et al [52] suggested the following bootstrap procedure.

For some large integer N and every $k \in \{1, \ldots, N\}$, repeat the steps below:

1. Given $\theta_n = T(X_1, \ldots, X_n)$, generate n independent observations $X^\star_{1,k}, \ldots, X^\star_{n,k}$ from the distribution F_{θ_n};

2. Compute

$$\theta^\star_{n,k} = T_n(X^\star_{1,k}, \ldots, X^\star_{n,k})$$

and let

$$F^\star_{n,k}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X^\star_{i,k} \le x\}};$$

3. Compute

$$S^\star_{n,k} = \phi(\mathbb{G}^\star_{n,k}),$$

where

$$\mathbb{G}^{\star}_{n,k} = n^{1/2}(F^{\star}_{n,k} - F_{\theta^{\star}_{n,k}}).$$

Using the convention that large values of S_n lead to the rejection of H_0, Stute et al [52] showed that under appropriate regularity conditions, an approximate p-value for the test is given by

$$\frac{1}{N}\sum_{k=1}^{N} 1_{\{S^{\star}_{n,k} > S_n\}}.$$

6.4.2 Validity of the Parametric Bootstrap Procedure

Motivation

Consider testing the appropriateness of various dependence structures on the basis of a random sample $X_1 = (X_{11}, \ldots, X_{1d}), \ldots, X_n = (X_{n1}, \ldots, X_{nd})$ from a continuous random vector X with cumulative distribution function F. Denote by $F_1, \ldots, F_d$ the univariate marginal distributions of X and let $C : \mathbb{R}^d \to \mathbb{R}$ be the copula associated with F. C is the cumulative distribution function of $U = \xi(X)$, where $\xi : \mathbb{R}^d \to \mathbb{R}^d$ is defined for all $x_1, \ldots, x_d \in \mathbb{R}$ by

$$\xi(x_1, \ldots, x_d) = (F_1(x_1), \ldots, F_d(x_d)).$$

The vectors $U_1 = \xi(X_1), \ldots, U_n = \xi(X_n)$ are only observed if the marginals $F_1, \ldots, F_d$ are known. However, F_j can be estimated by

$$F_{jn} = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\{X_{ij} \leq t\}},$$

for all $t \in \mathbb{R}$ and $j \in \{1, \ldots, d\}$.

Letting $\xi_n(x_1, \ldots, x_d) = (F_{1n}(x_1), \ldots, F_{dn}(x_d))^T$, for all $x_1, \ldots, x_d \in \mathbb{R}$, we can base a test of the hypothesis

$$H_0 : C \in \mathbf{C} = \{C_\theta : \theta \in \mathbf{O}\}$$

on the pseudo-observations

$$\widehat{U_1} = \xi_n(X_1), \ldots, \widehat{U_n} = \xi_n(X_n).$$

There are many possible ways of carrying out the test. One way is to base the test on the empirical copula. In this case, one can use e.g. the Cramer-Von Mises or Kolmogorov-Smirnov statistic

$$S_n = \phi(\mathbb{G}_n^C) \tag{6.102}$$

with

$$\mathbb{G}_n^C = n^{1/2}(C_n - C_{\theta_n}),$$

where C_{θ_n} is a parametric estimate of C_θ derived from the estimation $\theta_n = T(X_1, \ldots, X_n)$ of θ under H_0 while C_n is the empirical copula defined for all $u \in [0,1]^d$ by

$$C_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{\widehat{U}_i \le u\}}.$$

Another way is to base the test on Kendall's distribution, i.e. the distribution function K of the probability integral transformation $W = F(X)$. Since we can write W in the form $W = C(U)$, a consistent estimator of K is given by the empirical distribution K_n of the pseudo-observations $\widehat{W}_1 = C_n(\widehat{U}_1), \ldots, \widehat{W}_n = C_n(\widehat{U}_n)$, defined by

$$K_n(w) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{\widehat{W}_i \le w\}}.$$

Therefore, if K_θ denotes the distribution function of W when $C = C_\theta \in \mathbf{O}$, and if K_{θ_n} is a parametric estimate of K_θ derived from $\theta_n = T(X_1, \ldots, X_n)$ under the subsidiary hypothesis

$$H_0^K : K \in \mathbf{K} = \{K_\theta : \theta \in \mathbf{O}\},$$

a goodness-of-fit test can be based on a continuous functional

$$S_n = \phi(\mathbb{G}_n^K) \tag{6.103}$$

of

$$\mathbb{G}_n^K = n^{1/2}(K_n - K_{\theta_n}).$$

Whether H_0^C is tested using $\mathbb{G}_n^C$ or H_0^K is tested using $\mathbb{G}_n^K$, the limiting distribution of the test S_n depends not only on the unknown parameter θ but also possibly on the nuisance parameters $F_1, \ldots, F_d$. Although a parametric bootstrap may help to find valid P-values, this cannot be done on the basis of the results of Stute et al [52] (see [23]), because of the presence of dependence among the set of pseudo-observations $\widehat{U}_1, \ldots, \widehat{U}_n$ and $\widehat{W}_1, \ldots, \widehat{W}_n$. It then becomes necessary to establish the validity of the parametric bootstrap in situations where the hypothesis to be tested concerns the distribution P of an unobservable s-variate random vector U, viz.

$$H_0 : P \in \mathbf{P} = \{P_\theta : \theta \in \mathbf{O}\},$$

where $\mathbf{O}$ is an open subset of $\mathbb{R}^p$, and $U = \xi(X)$ for some function $\xi : \mathbb{R}^d \to \mathbb{R}^s$ of an observable d-variate random vector X.

In order to encompass procedures based on $\mathbb{G}_n^C$ and $\mathbb{G}_n^K$ as special cases, suppose that a test of H_0 is to be derived from a continuous functional

$$S_n = \phi(\mathbb{G}_n^A) \tag{6.104}$$

of an abstract empirical process of the form

$$\mathbb{G}_n^A = n^{1/2}(A_n - A_{\theta_n}),$$

where A_{θ_n} and A_n are respectively parametric and nonparametric estimate of an abstract quantity A that depends on P. More generally, A is a function mapping a closed rectangle $\mathcal{T} \subset [-\infty, \infty]$ into $\mathbb{R}^s$, and A_θ denotes the form taken by A when $P = P_\theta$ for some $\theta \in \mathbf{O}$. Therefore, $\mathcal{T} = [0,1]^d$, $s = d$ and $A_\theta = C_\theta$ for the test based on $\mathbb{G}_n^C$; $\mathcal{T} = [0,1]$, $s = 1$ and $A_\theta = K_\theta$ for a test based on $\mathbb{G}_n^K$.

In order to show that the parametric bootstrap yields a valid approximation to the null distribution of the empirical process $\mathbb{G}_n^A$ under appropriate conditions, the processes

$$\Theta_n = n^{1/2}(\theta_n - \theta)$$

and

$$\mathbb{A}_n = n^{1/2}(A_n - A)$$

need to converge weakly, as $n \to \infty$, respectively to a centered random variable Θ and a centered process $\mathbb{A}$ in the space $D(\mathcal{T}, \mathbb{R}^s)$ of càdlàg processes from $\mathcal{T}$ to $\mathbb{R}^s$.

Symbolically, we write

$$\Theta_n = n^{1/2}(\theta_n - \theta) \rightsquigarrow \Theta \tag{6.105}$$

and

$$\mathbb{A}_n = n^{1/2}(A_n - A) \rightsquigarrow \mathbb{A}. \tag{6.106}$$

Validity of the One-Level Parametric Bootstrap

Let $U_1, \ldots, U_n$ be a random sample from some distribution P, and assume that we want to test the hypothesis

$$H_0 : P \in \mathbf{P} = \{P_\theta : \theta \in \mathbf{O}\},$$

where $\mathbf{P}$ is a family of probability measures on $\mathbb{R}^d$ indexed by the parameter θ living in an open set $\mathbf{O} \subset \mathbb{R}^p$. Assume that $\mathbf{P}$ is identifiable, i.e.,

$$\theta \neq \theta' \Rightarrow P_\theta \neq P_{\theta'}.$$

Let $\mathcal{T} \subset [-\infty, \infty]$ be a closed rectangle and suppose that the test of H_0 is to be based on an abstract mapping $A : \mathbf{A} \to \mathbb{R}^s$. Suppose that $A = A_\theta$ when $P = P_\theta$, and let $\mathbf{A} = \{A_\theta : \theta \in \mathbf{O}\}$. Then the identifiability is ensured if for each $\epsilon > 0$,

$$\inf \left\{ \sup_{t \in \mathcal{T}} \|A_\theta(t) - A_{\theta_0}(t)\| : \theta \in \mathbf{O} \text{ and } |\theta - \theta_0| > \epsilon \right\} > 0.$$

Furthermore, assume that the mapping $\theta \to A_\theta$ is Frechet differentiable with derivative $\dot{A}_\theta$, i.e., for all $\theta_0 \in \mathbf{O}$,

$$\lim_{\|h\| \to 0} \sup_{t \in \mathcal{T}} \frac{\|A_{\theta_0 + h}(t) - A_{\theta_0}(t) - \dot{A}_\theta(t) h\|}{\|h\|} = 0. \tag{6.107}$$

Finally, let $\theta_n = T_n(U_1, \ldots, U_n)$ be a consistent estimate of θ and assume that the $D(\mathcal{T}, \mathbb{R}^s)$-valued process $A_n = \Upsilon_n(U_1, \ldots, U_n)$ estimates A consistently. Suppose specifically that the process $\Theta_n = n^{1/2}(\theta_n - \theta)$ and $\mathbb{A}_n = n^{1/2}(A_n - A)$ have centered Gaussian limits when $n \to \infty$ as in 6.105 and 6.106.

Before we give the conditions under which the weak limits of the processes $\mathbb{G}_n^A = n^{1/2}(A_n - A)$ and $\mathbb{G}_n^{A^\star} = n^{1/2}(A_n^\star - A^\star)$ are independent and identically distributed and then guarantee that a parametric bootstrap based on the process $\mathbb{A}_n$ is valid, let us give the following definitions (see [23]).

Definition 6.2. A family $\mathbf{P} = \{P_\theta : \theta \in \mathbf{O}\}$ is said to belong the class $S(\lambda)$ for a given measure λ (independent of θ) if

1. The measure P_θ is absolutely continuous with respect to λ for all $\theta \in \mathbf{O}$.
2. The density $p_\theta = \frac{dP_\theta}{d\lambda}$ admits first and second order derivatives with respect to all components of $\theta \in \mathbf{O}$. The gradient (row) vector with respect to θ is denoted $\dot{p}_\theta$, and the Hessian matrix is represented by $\ddot{p}_\theta$.

3. For arbitrary $u \in \mathbb{R}^d$ and every $\theta \in \mathbf{O}$, the mappings $\theta \to \frac{\dot{p}_\theta(u)}{p_\theta(u)}$ and $\theta \to \frac{\ddot{p}_\theta(u)}{p_\theta(u)}$ are continuous at θ_0, P_{θ_0} almost surely.

4. For every $\theta_0 \in \mathbf{O}$, there exists a neighborhood N of θ_0 and a λ-integrable function $h : \mathbb{R}^d \to \mathbb{R}$ such that for all $u \in \mathbb{R}^d$,

$$\sup_{\theta \in N} \|\dot{p}_\theta(u)\| \leq h(u).$$

5. For every $\theta_0 \in \mathbf{O}$, there exists a neighborhood N of θ_0 and P_{θ_0}-integrable functions $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ such that for every $u \in \mathbb{R}^d$,

$$\sup_{\theta \in N} \|\frac{\dot{p}_\theta(u)}{p_\theta(u)}\|^2 \leq h_1(u)$$

and

$$\sup_{\theta \in N} \|\frac{\ddot{p}_\theta(u)}{p_\theta(u)}\| \leq h_2(u).$$

Definition 6.3. Let $U_1, \ldots, U_n$ be a random sample from $P = P_\theta$ and let

$$\mathbb{W}_{P,n} = n^{-1/2} \sum_{i=1}^{n} \frac{\dot{p}^T(U_i)}{p(U_i)},$$

where $\dot{p}^T$ denotes the transpose of $\dot{p}$.

A sequence A_n is said to be P_{θ_0}-regular for $A = A_{\theta_0}$ if, as $n \to \infty$, the process $(\mathbb{A}_n, \mathbb{W}_n)$ converges weakly in $D(T, \mathbb{R}^s) \times \mathbb{R}^p$ to a centered Gaussian pair $(\mathbb{A}, \mathbb{W})$ and the Fréchet derivative $\dot{A}$ of A defined by Equation (6.107) satisfies

$$\dot{A}(t) = E[\mathbb{A}(t)\mathbb{W}_P(t)]$$

for every $t \in T$. The sequence is said to be **P**-regular for **A** if it is P_θ-regular for A_{θ_0} at all $\theta_0 \in \mathbf{O}$.

Now let $U_1^\star, \ldots, U_n^\star$ be a bootstrap sample from P_{θ_n}, and set

$$\theta_n^\star = T_n^\star(U_1^\star, \ldots, U_n^\star),$$

$$\Theta_n^\star = n^{1/2}(\theta_n^\star - \theta),$$

$$A_n^\star = \Upsilon_n^\star(U_1^\star, \ldots, U_n^\star),$$

and

$$\mathbb{A}_n^\star = n^{1/2}(A_n^\star - A).$$

The conditions under which the weak limits of the processes $\mathbb{G}_n^A = n^{1/2}(A_n - A_{\theta_n})$ and $\mathbb{G}_n^{A^\star} = n^{1/2}(A_n^\star - A_{\theta_n^\star})$ are independent and identically distributed are given by the following theorem, which then guarantees that a parametric bootstrap based on the process $\mathbb{A}_n$ is valid.

Theorem 6.15. Assume that $\mathbf{P} \in S(\lambda)$ and that as $n \to \infty$,

$$(\mathbb{A}_n, \Theta_n, \mathbb{W}_{P,n}) \rightsquigarrow (\mathbb{A}, \Theta, \mathbb{W}_P) \text{ in } D(T, \mathbb{R}^s) \times \mathbb{R}^{p\otimes 2},$$

where the limit is a centered Gaussian process.

Let $\Gamma = E[\Theta \mathbb{W}_P^T]$ and set $a(t) = E[\mathbb{A}(t)\mathbb{W}_P]$ for every $t \in T$. Then, as $n \to \infty$,

$$(\mathbb{A}_n, \mathbb{A}_n^\star, \Theta_n, \Theta_n^\star) \rightsquigarrow (\mathbb{A}, \mathbb{A}^\star, \Theta, \Theta^\star) \text{ in } D(T, \mathbb{R}^s)^{\otimes 2} \times \mathbb{R}^{p\otimes 2}.$$

In the limit, $\mathbb{A}^\star = \mathbb{A}^\perp + a\Theta$ and $\Theta^\star = \Theta^\perp + \Gamma\Theta$ are defined in terms of an independent copy $(\mathbb{A}^\perp, \Theta^\perp)$ of $(\mathbb{A}, \Theta)$. If in addition (A_n, θ_n) is $\mathbf{P}$-regular for $\mathbf{A} \times \mathbf{O}$, then

$$(\mathbb{G}_n^A, \mathbb{G}_n^{A^\star}) \rightsquigarrow (\mathbb{A} - \dot{A}\Theta, \mathbb{A}^\perp - \dot{A}\Theta^\perp) \text{ in } D(T, \mathbb{R}^s)^{\otimes 2},$$

as $n \to \infty$.

A Two-Level Parametric Bootstrap

When performing a goodness-of-fit test based on a continuous functional $S_n = \phi(\mathbb{G}_n^A)$ of a process $\mathbb{G}_n^A = n^{1/2}(A_n - A_{\theta_n})$, we have to compute A_{θ_n} at various points, but this is not always easy. For tests based on the empirical copula, we have $A_{\theta_n} = C_{\theta_n}$, and many copula families are not algebraically closed. In this case, a simple way to solve the problem is to generate a random sample $V_1^\star, \ldots, V_m^\star$ from the probability measure Q_{θ_n} with distribution function C_{θ_n} and for $u \in [0,1]^d$, to approximate $C_{\theta_n}(u)$ by

$$\check{C}_n^\star(u) = \frac{1}{m}\sum_{j=1}^m 1_{\{V_j^\star \leq u\}}.$$

In other words, we replace A_{θ_n} by an approximation $\check{A}_n^\star = \Psi_m(V_1^\star, \ldots, V_m^\star)$ built from a random sample $V_1^\star, \ldots, V_m^\star$ from

$$Q_{\theta_n} \in \mathbf{Q} = \{Q_\theta : \theta \in \mathbf{O}\}.$$

For the approach to make sense, we assume that if $A = A_{\theta_0}$ and $\breve{A}_n = \Psi_m(V_1, \ldots, V_m)$ for a random sample $V_1, \ldots, V_m$ from $Q = Q_{\theta_0}$, then

$$\breve{\mathbb{A}}_n = n^{1/2}(\breve{A}_n - A) \rightsquigarrow \breve{\mathbb{A}} \ \in D(T, \mathbb{R}^s), \tag{6.108}$$

as $n \to \infty$ (and hence $m \to \infty$).

Given that such a process exists, the following method can be used to circumvent the lack of a closed form for A_{θ_n} in the computation of the test S_n.

1. Compute $\theta_n = T_n(U_1, \ldots, U_n)$ and let $A_n = \Upsilon_n(U_1, \ldots, U_n)$.
2. Given $U_1, \ldots, U_n$, generate a random sample $V_1^\star, \ldots, V_m^\star$ from Q_{θ_n}.
3. Let $\breve{A}_n^\star = \Psi_m(V_1^\star, \ldots, V_m^\star)$ and compute

$$S_n = \phi(\mathbb{G}_n^{\breve{A}^\star}), \tag{6.109}$$

   where

$$\mathbb{G}_n^{\breve{A}^\star} = n^{1/2}(A_n - \breve{A}_n^\star). \tag{6.110}$$

A second parametric bootstrap procedure is necessary to approximate the distribution of S_n. To this end (see [23]), take N large and repeat the following steps for every $k \in \{1, \ldots, N\}$:

1. Given $U_1, \ldots, U_n, V_1, \ldots, V_n$, generate a random sample $U_{1,k}^\star, \ldots, U_{n,k}^\star$ from P_{θ_n}.
2. Compute $\theta_{n,k}^\star = T_n(U_{1,k}^\star, \ldots, U_{n,k}^\star)$ and let $A_{n,k}^\star = \Upsilon(U_{1,k}^\star, \ldots, U_{n,k}^\star)$.
3. Given $U_1, \ldots, U_n, V_1^\star, \ldots, V_n^\star$ and $U_{1,k}^\star, \ldots, U_{n,k}^\star$, generate a random sample $V_{1,k}^{\star\star}, \ldots, V_{n,k}^{\star\star}$ from $Q_{\theta_{n,k}^\star}$.
4. Let

$$\breve{A}_{n,k}^{\star\star} = \Psi_m(V_{1,k}^{\star\star}, \ldots, V_{n,k}^{\star\star})$$

   and compute

$$S_{n,k}^\star = \phi(\mathbb{G}_{n,k}^{\breve{A}^\star}), \tag{6.111}$$

   where

$$\mathbb{G}_{n,k}^{\breve{A}^\star} = n^{1/2}(A_{n,k}^\star - \breve{A}_{n,k}^{\star\star}). \tag{6.112}$$

Under the convention that large values of S_n lead to the rejection of H_0, and under regularity conditions, a valid P-value for the test based on $S_n = \phi(\mathbb{G}_n^{\breve{A}^\star})$ is given by

$$\frac{1}{N}\sum_{k=1}^{N} 1_{\{S^\star_{n,k} > S_n\}}.$$

In order to establish the validity of the conditions of the previous two-level parametric bootstrap, Genest and Rémillard [23] first introduced the following notation. Let $U_1, \ldots, U_n$ and $V_1, \ldots, V_m$ be two mutually independent random samples from $P = P_{\theta_0} \in \mathbf{P}$ and $Q = Q_{\theta_0} \in \mathbf{Q}$, respectively. Let $\mathbb{W}_{P,n}$ and $\mathbb{W}_{Q,n}$ be defined by

$$\mathbb{W}_{P,n} = n^{-1/2}\sum_{i=1}^{n}\frac{\dot{p}^T(U_i)}{p(U_i)}$$

and

$$\mathbb{W}_{Q,n} = n^{-1/2}\sum_{i=1}^{m}\frac{\dot{p}^T(V_i)}{p(V_i)},$$

where $q = q_{\theta_0}$ and $\dot{q} = \dot{q}_{\theta_0}$. Assume the following.

1. Given $\theta_n = T_n(U_1, \ldots, U_n)$, the random vectors $U_1^\star, \ldots, U_n^\star$ and $V_1^\star, \ldots, V_m^\star$ are mutually independent random samples from P_{θ_n} and Q_{θ_n}, respectively.

2. Given $U_1^\star, \ldots, U_n^\star, V_1^\star, \ldots, V_m^\star$ and $\theta_n^\star = T_n(U_1^\star, \ldots, U_n^\star)$, the random vectors $V_1^{\star\star}, \ldots, V_n^{\star\star}$ constitute a random sample from $Q_{\theta_n^\star}$.

Finally, let

$$\breve{A}_n = \Psi_m(V_1, \ldots, V_m),$$

$$\breve{A}_n^\star = \Psi_m(V_1^\star, \ldots, V_m^\star),$$

$$\breve{A}_n^{\star\star} = \Psi_m(V_1^{\star\star}, \ldots, V_m^{\star\star}),$$

$$\breve{\mathbb{A}}_n = n^{1/2}(\breve{A}_n - A),$$

$$\breve{\mathbb{A}}_n^\star = n^{1/2}(\breve{A}_n^\star - A),$$

$$\breve{\mathbb{A}}_n^{\star\star} = n^{1/2}(\breve{A}_n^{\star\star} - A).$$

The following result (see [23]) gives the conditions under which the weak limits of the processes

$$\mathbb{G}_n^{\breve{A}^\star} = n^{1/2}(A_n - \breve{A}_n^\star)$$

and

$$\mathbb{G}_n^{\breve{A}^{\star\star}} = n^{1/2}(A_n^\star - \breve{A}_n^{\star\star})$$

are independent and identically distributed, and then proves the validity of a two-level parametric bootstrap.

Theorem 6.16. Assume that $\mathbf{P} \in S(\lambda)$, $\mathbf{Q} \in S(\nu)$ and that as $n \to \infty$,

$$(\mathbb{A}_n, \breve{\mathbb{A}}_n, \Theta_n, \mathbb{W}_{P,n}, \mathbb{W}_{Q,n}) \rightsquigarrow (\mathbb{A}, \breve{\mathbb{A}}, \Theta, \mathbb{W}_P, \mathbb{W}_Q)$$

and that the limit is a centered Gaussian process in $D(T, \mathbb{R}^s)^{m\otimes 2} \times \mathbb{R}^{p\otimes 3}$. Let $\Gamma = E[\Theta \mathbb{W}_p^T]$ and set $a(t) = E[\mathbb{A}(t)\mathbb{W}_p^T]$ and $\breve{a}(t) = \breve{\mathbb{A}}(t)\mathbb{W}_Q^T$ for every $t \in T$. Then as $n \to \infty$,

$$(\mathbb{A}_n, \mathbb{A}_n^\star, \breve{\mathbb{A}}_n, \breve{\mathbb{A}}_n^\star, \breve{\mathbb{A}}_n^{\star\star}, \Theta_n, \Theta_n^\star) \rightsquigarrow (\mathbb{A}, \mathbb{A}^\star, \breve{\mathbb{A}}, \breve{\mathbb{A}}^\star, \breve{\mathbb{A}}^{\star\star}, \Theta, \Theta^\star)$$

in $D(T, \mathbb{R}^s)^{\otimes 5} \times \mathbb{R}^{p\otimes 2}$. In the limit, $\mathbb{A}^\star = \mathbb{A}^\perp + a\Theta$, $\Theta^\star = \Theta^\perp + \Gamma\Theta$, $\breve{\mathbb{A}}^\star = \breve{\mathbb{A}}^\perp + \breve{a}\Theta$, $\breve{\mathbb{A}}^{\star\star} = \breve{\mathbb{A}}^{\perp\perp} + \breve{a}\Theta^\star$ where $(\mathbb{A}^\perp, \Theta^\perp)$ is an independent copy of $(\mathbb{A}, \Theta)$. In addition, the processes $\breve{\mathbb{A}}$, $\breve{\mathbb{A}}^\perp$ and $\breve{\mathbb{A}}^{\perp\perp}$ are mutually independent and identically distributed, as well as independent of $\mathbb{A}$, $\mathbb{A}^\perp$, Θ and $\Theta^\perp$. Moreover, if (A_n, θ_n) is $\mathbf{P}$-regular for $\mathbf{A} \times \mathbf{O}$ and $\breve{A}_n$ is $\mathbf{Q}$-regular for $\mathbf{A}$, then

$$(\mathbb{G}_n^{\breve{A}^\star}, \mathbb{G}_n^{\breve{A}^{\star\star}}) \rightsquigarrow (\mathbb{A} - \breve{\mathbb{A}}^\perp - \dot{\mathbb{A}}\Theta, \mathbb{A}^\perp - \breve{\mathbb{A}}^{\perp\perp} - \dot{\mathbb{A}}\Theta^\perp)$$

in $D(T, \mathbb{R}^s)^{\otimes 2}$ as $n \to \infty$.

6.5 Power Study of Goodness-of-Fit Tests

As we have seen previously, there are several test statistics for testing goodness-of-fit. This may cause confusion about which test to use when one has a testing problem to handle. A well-reflected choice of tests requires some knowledge about preferences concerning alternatives which may come from the practical experiment. The idea of power helps in choosing a relatively better test. In this section we discuss the power of goodness-of-fit tests.

6.5.1 Power Function

Consider the random vector $X = (X_1, X_2, \ldots, X_n)$ with joint distribution function $F(x, \theta)$ for some $\theta \in \Theta$. Let $\Theta = \Theta_0 \cup \Theta_1$ for two disjoint sets Θ_0 and Θ_1, where Θ_0 is often taken to be a lower dimensional sub-space of the parameter space Θ. The determination to decide whether the parameter θ lies in Θ_0 or in Θ_1 is done through a decision function called a test function, ϕ, which can be assumed to take the values 0 and 1; if $\phi(x) = 0$ then we decide that $\theta \in \Theta_0$ while if $\phi(x) = 1$ then we decide that $\theta \in \Theta_1$. The classical approach to this testing is to specify the test function $\phi(X)$ such that for some specified $\alpha > 0$,

$$E_\theta[\phi(X)] \leq \alpha,$$

for all $\theta \in \Theta_0$. (See [35] for more details.)

Definition 6.4. For a given test function ϕ, the power of the test at θ is defined by

$$\pi(\theta) = E_\theta[\phi(X)]. \tag{6.113}$$

For a specified level α, we require $\pi(\theta) \leq \alpha$ for all $\theta \in \Theta_0$ and so we are most interested in $\pi(\theta)$ for $\theta \in \Theta_1$.

Suppose that we want to test the null hypothesis H_0 versus the alternative H_1 at level α, where α is small, and suppose that given data $X = x$, $\phi(x) = 1$. If H_0 is true then this event occurs with probability at most α and so this gives the evidence to believe that H_0 is false (and hence H_1 is true). This assumes that the test is taken such that $P[\phi(X) = 1]$ is larger when H_1 is true. Conversely, if $\phi(x) = 0$, then the test is very much inconclusive; this may tell us that H_0 is true or, alternatively, that H_1 is true but that the test used does not have sufficient power to detect this.

Since the model under the null hypothesis H_0 is generally simpler, the previous approach of testing protects against choosing unnecessarily complicated models but, depending on the power of the test, may prevent us from identifying more complicated models when such models are appropriate. Some methods of finding "good" test functions can be found in [35], pages 354-371.

6.5.2 Linear Interpolated Power for Categorical Goodness-of-Fit Test Statistics

The power of a test is the probability of rejecting the null hypothesis when it is not true, i.e. when the alternative is true. Since it is possible to obtain the same critical value for a number of different significance levels, the required significance level may not have a unique test statistic. That is, a comparison of the relative power of categorical goodness-of-fit test statistics is questionable if different significance levels are used for the null and the alternative distributions of the test statistic. In order to overcome this problem, many authors used linear interpolation of the power of the test statistic using the power for a significance level less than (denoted by α_1) and greater than (denoted by α_2) the desired significance level (denoted by α). Linear interpolation gives (see [51]) a weighting to the power based on how close α_1 and α_2 are to α and this weighting is given by

$$Power = \frac{(\alpha - \alpha_1)P[T \geq X_2(\alpha)|H_1] + (\alpha_2 - \alpha)P[T \geq X_1(\alpha)|H_1]}{\alpha_2 - \alpha_1}, \tag{6.114}$$

where $X_1(\alpha)$ and $X_2(\alpha)$ are the critical values immediately below and above the significance level α, i.e.

$$\alpha_1 = P[T \geq X_1(\alpha)|H_0]$$

and

$$\alpha_2 = P[T \geq X_2(\alpha)|H_0]$$

are the significance levels for $X_1(\alpha)$ and $X_2(\alpha)$ respectively.

6.5.3 Power Function for the Brownian Bridge Shift Experiment

Consider the model

$$B_0(t) + \int_0^t h(u)du, \ 0 \leq t \leq 1, \tag{6.115}$$

on $C[0,1]$ for the Brownian bridge $B_0(.)$ with parameter space

$$\begin{aligned} H &= L_2^0(\lambda_{|(0,1)}) \\ &:= \left\{ h \in L_2(\lambda_{|(0,1)}) : \int h d\lambda_{|(0,1)} = 0 \right\} \end{aligned}$$

which is endowed with the natural inner product

$$< h_1, h_2 >= \int h_1 h_2 d\lambda_{|(0,1)}$$

where $\lambda_{|(0,1)}$ is the uniform distribution on the unit interval. This model is a standard Gaussian shift

$$G = (\Omega, \mathcal{A}, \{P_h : h \in H\})$$

with P_h the distribution of the process in (6.115). The Neyman-Pearson envelope power function at level α and sample size n for testing the null hypothesis $\{P_0^n\}$ against the alternative $\{P_h^n\}$ is given by (see [31])

$$\beta_1(h,n) := \phi(n^{\frac{1}{2}}\|h\| - u_{1-\alpha}) \tag{6.116}$$

and its two-sided counterpart for unbiased testing is

$$\beta_2(h,n) := \Phi(n^{\frac{1}{2}}\|h\| - u_{1-\alpha/2}) + \Phi(-n^{\frac{1}{2}}\|h\| - u_{1-\alpha/2}), \tag{6.117}$$

where

$$u_{1-\alpha} = \Phi(1-\alpha)$$

denotes the $(1-\alpha)$-quantile of the standard normal distribution function Φ.

6.5.4 A Discussion on Global Power Functions

The results for the linear model (6.115) have some important consequences for the goodness-of-fit testing problem on the real line for fixed sample size n. This asymptotic Brownian bridge model is much the same as the suitably normalized finite sample nonparametric test statistic. Let ϕ be a goodness-of-fit test on $C[0,1]$. As suggested in [31], we have the following observations:

1. If ϕ is an integral test of Cramer-Von Mises or Anderson-Darling type, then often a global principal component decomposition of the test statistic and its power function is available.

2. Two-sided goodness-of-fit tests ϕ with centrally symmetric and convex acceptance regions have more general structure than integral tests. Since there is no principal component decomposition of their test statistics, Milbrodt and Strasser [40] proposed a principal decomposition of the curvature of the power function at $h = 0$ in H. A Taylor expansion of the power function along the present exponential family is given by

   $$E_{P_{th}}(\phi) = \alpha + < h, Th > t^2/2 + o(t^2), \; t \in \mathbb{R}, \; h \in H,$$

   at $t = 0$ where $T : H \to H$ is a Hilbert-Schmidt operator with

   $$T(g) = \sum_{i=1}^{\infty} \lambda_i < h_i, g > h_i.$$

   In one of his papers, Strasser obtained global extrapolations for power functions

   $$t \mapsto E_{P_{th}}(\phi) \tag{6.118}$$

   of tests ϕ with centrally symmetric acceptance regions which are based on the curvature

   $$< h, Th >, \; \|h\| = 1.$$

This yields sharp upper bounds for (6.118), given the curvature $< h, Th >$, which are attained in the class of tests with centrally symmetric acceptance regions. Using this bounds, it is easy to see that within this class of goodness-of-fit tests the global power function becomes flat if the curvature is small.

3. Since every test has a preference for some finite dimensional subspace, one may be interested in the construction of tests which have good performance on a given finite dimensional linear subspace of alternatives.

Example 6.1. Two-sided Kolmogorov-Smirnov goodness-of-fit test.

This example is given in [40]. Let

$$\phi_{KS}^{(n)} = I\left(\sup_{0\le t\le 1} |n^{\frac{1}{2}}(\widehat{F}_n(t) - t)| > C_\alpha\right)$$

denotes the two-sided Kolmogorov-Smirnov goodness-of-fit test of asymptotic level α, at sample size n for $\lambda_{|(0,1)}$ versus unspecified continuous alternatives, where $\widehat{F}_n$ is the empirical distribution function. Under the continuous local alternatives of order $n^{-\frac{1}{2}}$, given by alternatives with tangent $h \in L_2^0(\lambda_{|(0,1)})$, the asymptotic power function of $\phi_{KS}^{(n)}$ is given by

$$P\left[\sup_{0\le t\le 1} |B_0(t) + \int_0^t h(u)du| > C_\alpha\right]. \tag{6.119}$$

The curvature $< h, Th >$ of the power function is greater than zero and the power is strictly larger than the level

$$\alpha = P\left[\sup_{0\le t\le 1} |B_0(t)| > C_\alpha\right]$$

for all nontrivial directions $h \neq 0$. This fact is labeled as $\sqrt{n}$-consistency of the sequence of tests, and (6.119) is close to one for tangents sh with $s \in \mathbb{R}$ large enough.

The relation

$$\lim_{n\to\infty} \sup \frac{l_p(\phi_{KS}^{(n)}, \beta, h)}{l_{p_2}(\beta, n)} < \infty$$

holds for the level points for each $\beta < 1$ and each direction $h \neq 0$, where $l_p(\phi, \beta, h)$ is the level point of the test ϕ in direction h, and $l_{p_2}(\beta, n)$ is the level point of the envelope power function defined in (6.117).

Remark 6.2. Level points.
The level point (l_p) of a test ϕ in direction h is given by (see [31])

$$l_p(\phi, \beta, h) = \inf\{|s| : E_{sh}(\phi) \geq \beta\}.$$

Chapter 7

Goodness-of-Fit Tests for Copulas

As one of the main ways of modeling dependence, the copula proved to be a handy instrument in the analysis of multivariate time series. It allows to capture the full dependence within multivariate time series without specifying a shape of marginal distributions. In finance, it is used to characterize the dependency structure between assets. However, to check whether the dependency structure of a data set is approximately modelled by a chosen copula, there is no agreed upon recommended method. Information criteria are not able to provide any understanding about the power of the decision rule employed. Goodness-of-fit approaches on the other hand, are able to reject or fail to reject a parametric copula, and are thus preferred. The goodness-of-fit approach for copulas is a special case of the more general problem of testing multivariate models, but is complicated due to the unspecified marginal distributions. In this chapter, we discuss several methods for carrying out the goodness-of-fit tests for copulas.

7.1 Goodness-of-Fit Procedures for Copulas Based on the Probability Integral Transform

Definition 7.1. Probability Integral Transform (PIT) (see [3]).
Let $\mathbf{X} = (X_1, \ldots, X_d)$ denote a random vector with marginals $F_i(x_i) = P[X_i \leq x_i]$ and conditional distributions

$$F_{i|1,\ldots,i-1}(x_i|x_1, \ldots, x_{i-1})$$

for $i = 1, \ldots, d$. The PIT of $\mathbf{X}$ is defined as

$$T(\mathbf{X}) = (T_1(X_1), \ldots, T_d(X_d))$$

where $T_i(X_i)$ is defined as follows:

$$T_1(x_1) = P[X_1 \leq x_1] = F_1(x_1)$$

$$T_2(x_2) = P[X_2 \leq x_2 | X_1 = x_1] = F_{2|1}(x_2|x_1)$$

$$\vdots$$

$$T_d(x_d) = P[X_d \leq x_d | X_1 = x_1, \dots, X_{d-1} = x_{d-1}] = F_{d|1,\dots,d-1}(x_d|x_1, \dots, x_{d-1}).$$

The random variables $Z_i = T_i(X_i),\ i = 1, \dots, d$ are uniformly and independently distributed on $[0,1]^d$.

7.1.1 Description of the Test Statistics

Let $(X_1, Y_n), \dots, (X_n, Y_n)$ be independent copies of a bivariate random vector (X, Y) from some continuous bivariate copula model $\mathbf{C} = (C_\theta)$ with unknown continuous marginal distributions $F,\ G$. In other words, suppose that the cumulative distribution function H of X is of the form

$$H(x, y) = C(F(x), G(y)),$$

for some copula $C \equiv C_\theta \in \mathbf{C}$, whose parameter θ takes its values in some open set $O \in \mathbb{R}^m$, $x,\ y \in \mathbb{R}$. Let K be the distribution function of the probability integral transformation $V = H(X, Y)$.

a. Case of Archimedean Copula

Assume C belongs to a parametric class $C = (C_{\phi_\theta})$ of Archimedean copulas, where ϕ_θ represents the generators. In this case, K is in the form

$$K(\theta, t) = t - \frac{\phi_\theta(t)}{\phi'_\theta(t)},\ t \in (0, 1].$$

Consider the empirical version $K_n(.)$ of $K(\theta, .)$ defined by

$$K_n(t) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{V_{jn} \leq t\}},\ t \in [0, 1], \tag{7.1}$$

where

$$V_{jn} = \frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k \le X_j, Y_k \le Y_j\}},$$

and the Kendall's process $\sqrt{n}\{K_n(.) - K(\theta, .)\}$. Consider the pseudo observations

$$\nu_i = \frac{1}{n-1}\sharp\{j : X_j < X_i, Y_j < Y_i\}.$$

Let the value θ_n of θ be such that, under the null hypothesis H_0, the population value

$$\tau(\theta) = 4E(V) - 1$$

of Kendall's tau matches its standard empirical version given by

$$\tau_n = 4\overline{\nu} - 1,$$

where

$$\overline{\nu} = \frac{1}{n}\sum_{i=1}^{n} \nu_i.$$

Wang and Wells [58] then proposed the goodness-of-fit test statistic

$$S_{\xi_n} = \int_{\xi}^{1} \{\mathbb{K}_n(t)\}^2 dt, \tag{7.2}$$

which is a continuous functional of the process

$$\mathbb{K}_n(t) = \sqrt{n}\{K_n(t) - K(\theta_n, t)\}.$$

The constant $\xi > 0$ is taken in order to avoid technical difficulties related to unboundedness of the density of $K(., \theta)$ at the origin, which is common in practice.

b. General Case

Define K by

$$K(\theta, t) = P[H(X, Y) \le t],$$

and its empirical version by Equation 7.1. Suppose that for all $\theta \in \mathbf{O}$, the distribution function $K(\theta, t)$ of $H(X, Y)$ admits a density $k(\theta, t)$ which is continuous on $\mathbf{O} \times (0, 1]$, and let $\mathbb{K}$ be the limit of the process $\mathbb{K}_n$.

Genest, Quessy and Remillard [21] proposed the test statistics S_n and T_n given by

$$S_n = \int_0^1 |\mathbb{K}_n(t)|^2 k(\theta_n, t) dt \tag{7.3}$$

and

$$T_n = \sup_{0 \le t \le 1} |\mathbb{K}_n(t)|, \tag{7.4}$$

as alternatives to S_{ξ_n} when C is not necessarily Archimedean, and they proved that their limits are given respectively by

$$S = \int_0^1 |\mathbb{K}(t)|^2 k(\theta, t) dt$$

and

$$T = \sup_{0 \le t \le 1} |\mathbb{K}(t)|.$$

c. Implementation of the Tests

Straightforward calculations (see [21]) show that

$$\begin{aligned} S_n = \frac{n}{3} + n \sum_{j=1}^{n-1} K_n^2\left(\frac{j}{n}\right)\left\{K\left(\theta_n, \frac{j+1}{n}\right) - K\left(\theta_n, \frac{j}{n}\right)\right\} \\ - n \sum_{j=1}^{n-1} K_n^2\left(\frac{j}{n}\right)\left\{K^2\left(\theta_n, \frac{j+1}{n}\right) - K^2\left(\theta_n, \frac{j}{n}\right)\right\} \end{aligned}$$

and

$$T_n = \sqrt{n} \max_{i=0,1; 0 \le j \le n-1} \left\{ |K_n\left(\frac{j}{n}\right) - K\left(\theta_n, \frac{j+i}{n}\right)| \right\}.$$

7.1.2 Performance of the tests

Formal testing procedures based on these statistics consist of rejecting the null hypothesis H_0 : $C \in \mathbf{C}$ when the observed value of S_n or T_n is greater than the $100(1-\alpha)$th percentile of its distribution under H_0. However, this distribution depends on the unknown parameter θ even in the limit. To overcome this problem, an alternative is to use a parametric bootstrap or Monte Carlo testing approach based on C_{θ_n}. Parametric bootstrap procedures for copulas goodness-of-fit tests will be discussed in next section.

7.2 Parametric Bootstrap Procedures for Copulas Goodness-of-Fit

Let X be a continuous d-variate random vector with distribution function $F(.,\theta)$, marginal distributions $F_1,\dots,F_d$, and unique copula C. Consider testing the null hypothesis

$$H_0 : C \in \mathbf{C} = \{C_\theta : \theta \in \mathbf{O}\},$$

i.e, $C = C_{\theta_0}$ for some $\theta_0 \in \mathbf{O}$. Given a random sample $X_1,\dots,X_n$ from F, a natural way to proceed is to compare C_n to the empirical copula C_{θ_n}, where θ_n is an estimate of the unknown parameter $\theta \in \mathbb{R}^p$. Consider the transformation $U_i = \widehat{F}_n(X_i),\ i = 1,\dots,n$, where $\widehat{F}_n(x) = F(x,\theta_n)$. For arbitrary $u \in [0,1]^d$, let

$$B_n(u) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\{U_i \le u\}},$$

and for every $j \in \{1,\dots,d\}$, and $t \in [0,1]$, define

$$B_{jn}(t) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\{U_{ij} \le t\}}.$$

The empirical copula is then asymptotically equivalent to

$$C_n(u) = B_n\{B_{1n}^{-1}(u_1),\dots,B_{dn}^{-1}(u_d)\}$$

at every $u = (u_1,\dots,u_d) \in [0,1]^d$. Assume that $\theta_n = T_n(U_1,\dots,U_d)$ and suppose that

$$S_n = \phi(\mathbb{G}_n^C)$$

is a continuous functional of the empirical process

$$\mathbb{G}_n^C = n^{1/2}(C_n - C_{\theta_n}).$$

The following algorithm is used to compute the p-values for any test statistic based on the process $\mathbb{G}_n^C$.

Algorithm 7.1. (see [21])

1. Estimate θ by a consistent estimator θ_n.

2. Generate N random samples of size n from C_{θ_n} and, for each of these samples, estimate θ by the same method as before and determine the value of the test statistic.

3. If $S^\star_{1,N}, \ldots, S^\star_{N,N}$ denote the ordered values of the statistics calculated in step 2, an estimate of the critical value of the test at level α based on S_n is given by

$$S^\star_{\lfloor(1-\alpha)N\rfloor},$$

and

$$\frac{1}{N}\sum_{j=1}^{N} I_{\{S^\star_j \geq S_n\}}$$

yields an estimate of the p-value associated with the observed value S_n of the statistic.

In order to establish the validity of the parametric bootstrap for such goodness-of-fit statistics, one can use Theorem 6.15 and Theorem 6.16 with $A_\theta = C_\theta$ and P_θ the unique probability measure associated with C_θ and density c_θ. Assume that $\mathbf{P} = \{P_\theta : \theta \in \mathbf{O}\} \in S(\lambda)$, where $S(\lambda)$ is defined as in Definition 6.2 and λ is Lebesgue's measure. Let $c = c_{\theta_0}$, $\dot{c} = \dot{c}_{\theta_0}$ and $\ddot{c} = \ddot{c}_{\theta_0}$. Let

$$\mathbb{C}_n = n^{1/2}(C_n - C)$$

be the empirical copula process and let

$$W_{C,n} = n^{1/2}\sum_{i=1}^{n}\frac{\dot{c}^T(U_i)}{c(U_i)}.$$

With results from Chapter 5 of [20], it is shown that as $n \to \infty$,

$$(\mathbb{B}_n, \mathbb{C}_n, \mathbb{W}_{C,n}) \rightsquigarrow (\mathbb{B}, \mathbb{C}, \mathbb{W}_C)$$

in $D([0,1]^d, \mathbb{R})^{\otimes 2} \times \mathbb{R}^p$, where $\mathbb{W}_C$ is a centered Gaussian variable with variance

$$I_C = \int \frac{\dot{c}^T(x)\dot{c}(x)}{c(x)}\lambda(dx),$$

and $\mathbb{B}$ is **C**-Brownian bridge. Ganßler and Stute [20] showed that the limit $\mathbb{C}$ admits the representation

$$\mathbb{C} = B(u) - \sum_{j=1}^{d}\beta_j(u_j)\frac{\partial}{\partial u_j}C(u), \tag{7.5}$$

for all $u = (u_1, \ldots, u_d) \in [0,1]^d$, where for each $j \in \{1, \ldots, d\}$, β_j is a classical Brownian bridge related to $\mathbb{B}$ via the equation

$$\beta_j(t) = \mathbb{B}(1_{t,j})$$

in which $1_{t,j} = (e_1, \ldots, e_d) \in \mathbb{R}^d$ with $e_i = t$ if $i = j$ and $e_i = 1$ otherwise.

As observed by Genest and Rémillard [23],

$$E[\mathbb{B}(u)\mathbb{W}_C^T] = \int \dot{c}(v) 1_{\{v \le u\}} \lambda(dv) = \dot{C},$$

for all $u \in [0,1]^d$, and hence for all $t \in [0,1]$ and $j \in \{1, \ldots, d\}$, one has

$$E[\beta_j(t)\mathbb{W}_C^T] = \dot{C}(1_{t,j}) = 0.$$

Therefore, for all $u \in [0,1]^d$, the representation given by Equation 7.5 yields

$$E[\mathbb{C}(u)\mathbb{W}_C^T] = \dot{C}(u).$$

By these observations, the following result is obtained.

Proposition 7.1. Let $X_1, \ldots, X_n$ be a random sample from distribution F with unique copula $C = C_{\theta_0}$ for some $\theta_0 \in \mathbf{O}$. If $\mathbf{P} \in \mathcal{S}(\lambda)$, then the empirical copula C_n is $\mathbf{P}$-regular for C.

Assuming that θ_n is a $\mathbf{P}$-regular sequence for $\mathbf{O}$ such that, as $n \to \infty$,

$$(\mathbb{B}_n, \Theta_n, \mathbb{W}_{C,n}) \rightsquigarrow (\mathbb{B}, \Theta, \mathbb{W}_C)$$

in $\mathcal{D}([0,1]^d, \mathbb{R}) \times \mathbb{R}^{p\otimes 2}$, where the weak limit is Gaussian, it follows that (C_n, θ_n) is $\mathbf{P}$-regular for $\mathbf{C} \times \mathbf{O}$ because $E[\mathbb{C}\mathbb{W}_C^T] = \dot{C} = \dot{C}_{\theta_0}$, and $E[\Theta\mathbb{W}_C^T] = I$ by the regularity hypothesis on θ_n.

Finally, all the conditions of Theorem 6.15 and Theorem 6.16 are satisfied with $\mathbf{A} = \mathbf{C}$, $A_n = C_n$ and $\check{A}_n = \check{B}_n$ defined for all $u \in [0,1]^d$ by

$$\check{B}_n(u) = \frac{1}{m} \sum_{i=1}^{m} 1_{\{Y_i \le u\}}$$

in terms of a random sample $Y_1, \ldots, Y_m$ from P_θ that is independent of $X_1, \ldots, X_n$. Therefore, the one-level and the two-level parametric bootstraps yield valid approximations of the distribution of any continuous functional $S_n = \phi(\mathbb{G}_n^C)$.

The same reasoning can be used for the test based on a continuous functional $S_n = \phi(\mathbb{G}_n^K)$ of the process $\mathbb{G}_n^K = n^{1/2}(K_n - K_{\theta_n})$, where K_{θ_n} is the nonparametric estimator of K. In this case also, the conditions are satisfied for the application of Theorem 6.15 and Theorem 6.16, and so the one-level and two-level parametric bootstraps yield valid approximations of the distribution of any continuous functional $S_n = \phi(\mathbb{G}_n^K)$. More details can be found in [23].

7.3 Dimension Reduction Approaches to Copula Goodness-of-Fit Problem

In this section we consider four different approaches for goodness-of-fit for copulas, based on the probability integral transform. An advantage with the PIT is that the null and the alternative hypotheses are the same, regardless of the distribution before the probability integral transformation. In fact, the PIT is a universally applicable way of creating a set of i.i.d. $U(0,1)$ variables from any data set with known distribution. Given a test for multivariate, independent uniformity, this transformation can be used to test the fit of any assumed model. We introduce three dimension reduction approaches. All these approaches can be found in [3].

7.3.1 Breymann, Dias and Embrechts' Approach

Consider the i.i.d. $U(0,1)^d$ variables $\mathbf{Z} = (Z_1, \ldots, Z_d)$, obtained from applying the PIT (see Definition 7.1) to a multivariate data set $(X_1, \ldots, X_d)$. The dimension reduction is then performed as follows:

$$Y_G = \sum_{i=1}^{d} \Phi^{-1}(Z_i)^2,$$

where $\Phi^{-1}(.)$ is the inverse Gaussian cumulative distribution function. Since the Z_i's are i.i.d. $U(0,1)$ under the null hypothesis H_0 (from [47]), the variables $\Phi^{-1}(Z_i)$ are i.i.d. $N(0,1)$. Hence Y_G is χ_d^2 distributed. Let

$$W_G = F_{\chi_d^2}(Y_G).$$

As we can see, the multivariate problem is now reduced to a univariate one, and the approach G is defined as the cumulative distribution function of W_G:

$$F_G(w) = P[W_G \leq w], \ w \in [0,1].$$

Under the null hypothesis, we have $F_G(w) = w$ and the density function is $f_G(w) = 1$. Given n observations of the d-dimensional vector $\mathbf{Z}$, we get n observations of W_G, say $W_{G_1}, \ldots, W_{G_n}$, and so the empirical version of F_G is given by

$$\widehat{F}_G(w) = \sum_{j=1}^{n} 1_{\{W_{G_j} \leq w\}},\ w = \left(\frac{1}{n+1}, \ldots, \frac{n}{n+1}\right). \tag{7.6}$$

Breymann et al. [4] applied the Anderson-Darling test statistic to F_G and they have noticed that this approach (approach G) is computationally very efficient. However, it is inconsistent. The cause of this inconsistency is caused by the projection from a multivariate problem to a univariate problem through Y_G.

7.3.2 Berg and Bakken's Approach

This approach, by Berg et al. [2], is an extension of the one by Breymann et al. [4] we just discussed. Let $\mathbf{Z} = (Z_1, \ldots, Z_d)$ be the i.i.d. $U(0,1)^d$ variables obtained from applying the PIT (see Definition 7.1) to a multivariate data set $(X_1, \ldots, X_d)$. Define the new variable $\mathbf{Z}^\star = (Z_d^\star, \ldots, Z_d^\star)$, where

$$Z_i^\star = 1 - \left(\frac{1 - \tilde{Z}_i}{1 - r_{i-1}}\right)^{d-(i-1)},$$

for $i = 1, \ldots, d$, where

$$\tilde{\mathbf{Z}} = (\tilde{Z}_1, \ldots, \tilde{Z}_d)$$

is the sorted counterpart of $\mathbf{Z}$ and r_i is the [1]rank variable i from $\mathbf{Z}$, and $r_0 = 0$. The dimension reduction is then performed using $\mathbf{Z}^\star$:

$$Y_B = \sum_{i=1}^{d} \gamma(Z_i; \alpha).\Phi(Z_i^\star)^2,$$

where γ is a weight function used for weighting $\Phi(Z_i^\star)^2$ depending on its corresponding value Z_i, and α is the set of weight parameters. Considering $F_{Y_B}(.)$ as the cumulative distribution function of Y_B, we define

$$W_B = F_{Y_B}(Y_B),$$

which is $U(0,1)$ under H_0. The approach B is then defined as the cumulative distribution function of W_B:

$$F_B(w) = P[W_B \leq w],\ w \in [0,1].$$

[1]Rank variables are the observed variables, ordered ascendingly

Here also under the null hypothesis, we have $F_B(w) = w$ and the density function is $f_B(w) = 1$. Given n observations of the d-dimensional vector **Z**, we get n observations of W_B, say $W_{B_1}, \ldots, W_{B_n}$, and so the empirical version of F_B is given by

$$\widehat{F}_B(w) = \sum_{j=1}^{n} 1_{\{W_{B_j} \leq w\}}, \;\; w = \left(\frac{1}{n+1}, \ldots, \frac{n}{n+1}\right). \tag{7.7}$$

As noticed by Berg et al. [2], the approach B is computationally fast, but slower than the approach G, due to the distribution of the linear combination of squared normal variables. Further, the weighting functionality adds valuable flexibility, and any weight function can be applied to any region of the copula. Note that the transformation $\mathbf{Z}^\star$ enables a consistent dimension reduction, without loosing any information.

7.3.3 Genest, Quessey and Rémillard's Procedure

This approach is proposed for Archimedean copulas based on Kendall's process. Genest et al. [22] utilized the empirical copula cumulative distribution function for dimension reduction. They defined

$$F_K(w) = P[F(\mathbf{X}) \leq w] = P[C(U) \leq w], \;\; w \in [0,1],$$

where C is the H_0 copula, with parameter θ and **X** is the observed multivariate data set. The density of F_K is given by

$$f_K(w) = \frac{\partial F_K(w)}{\partial w},$$

and under the null hypothesis, we have $F_K(w) = F_{K,0}(w)$, where $F_{K,0}(w)$ is copula specific and must be derived for any copula that one chooses to use. This can be seen by rewriting $F_K(w)$ in the form

$$F_K(w) = \int_0^1 \ldots \int_0^1 1_{\{C(u_1,\ldots,u_d) \leq w\}} c(u_1, \ldots, u_d) du_1 \ldots du_d,$$

where $c(u_1, \ldots, u_d)$ is the copula density, which is copula specific. The empirical version of $F_K(w)$ is given by

$$\widehat{F}_K(w) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{\widehat{F}(x_j) \leq w\}} = \frac{1}{n} \sum_{j=1}^{n} 1_{\{\widehat{C}(u_j) \leq w\}},$$

where u_j is the j-th observation of the d-dimensional vector **u**, $\widehat{F}_j$'s are the empirical marginal cumulative distribution functions for n observations $X_{1i}, \ldots, X_{ni}$ of X_i, given by

$$\widehat{F}_i(x) = \frac{1}{n+1} \sum_{j=1}^{n} 1_{\{X_{ji} \leq x\}}, \ i = 1, \ldots, d,$$

and $\widehat{C}(u)$ is the empirical copula cumulative distribution function given by

$$\widehat{C}(u) = \frac{1}{n+1} \sum_{j=1}^{n} 1_{\{U_{j1} \leq u_1, \ldots, U_{jd} \leq u_d\}}.$$

Genest et al. [22] specified some hypotheses based on Kendall's process, and they further derived two univariate test statistics that they applied to $F_K(w)$. Unfortunately, these statistics depend on the null hypothesis copula. This leads them to find an alternative approach, say the approach K, described below.

As before, let **Z** be the uniformly and identically distributed variables on $[0,1]^d$, obtained from applying the probability integral transform to the multivariate data set **X**. The approach K is then defined as the cumulative distribution of the copula:

$$F_K(w) = P[C(\mathbf{Z}) \leq w], \ w \in [0,1],$$

and its density is given by

$$f_K(w) = \frac{\partial F_K(w)}{\partial w}.$$

The empirical version of $F_K(w)$ is given by

$$\widehat{F}_K(w) = \sum_{j=1}^{n} 1_{\{\widehat{C}(z_j) \leq w\}}, \ w = \left(\frac{1}{n+1}, \ldots, \frac{n}{n+1}\right),$$

where z_j is the observation number j of the d-dimensional vector **z**.

In order to counter the issue of

$$\lim_{w \to 0} f_K(w) = \infty,$$

the approach K is replaced by the approach M. In this case, by applying the inverse function $F_K^{-1}(w)$ to the empirical results $\widehat{F}_K$, a uniform distribution is obtained. Let $F_K^{-1}(x)$ be a function that solves the equation $F_K(w) = x$ with respect to w. The approach M can be considered as the generalization of the approach K and is defined by

$$F_M(w) = F_K(F_K^{-1}(w)),\ w \in [0,1].$$

Under the null hypothesis, we have $F_M(w) = w$ and the density function is $f_M(w) = 1$. The empirical version of $F_K(w)$ is given by

$$\widehat{F}_M(w) = \sum_{j=1}^{n} 1_{\{F_K(\widehat{C}(z_j)) \le w\}},\ w = \left(\frac{1}{n+1}, \ldots, \frac{n}{n+1}\right). \tag{7.8}$$

The testing procedure of the previous dimension reduction approaches is given by the following algorithm.

Algorithm 7.2. Breymann et al. [4]

1. Construct the copula of **Z** by applying the PIT to the observed data set **X**, given a H_0 copula;
2. Compute $\widehat{F}_T(w)$ according to 7.6, 7.7 or 7.8 respectively;
3. Compute the univariate test $\widehat{T}$ to be used, using $\widehat{F}_T(w)$;
4. Repeatedly (N times) perform steps 1-3, using a simulated observed data $\mathbf{X}^\star$, simulated from the H_0 distribution. The resulting N values of $\widehat{T}^\star$ form the distribution of T;
5. Compute the p-value,
$$p = \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} 1_{\{\widehat{T}_i^\star \ge \widehat{T}\}}\right).$$

7.4 Chi-square and Likelihood Ratio Tests for Bivariate Copulas

Consider a known parametric family $C(u,v,\theta),\ \theta \in \Theta \subset \mathbb{R}^d$ of bivariate copulas, and let (X,Y) denote a random vector with unknown copula $C(u,v)$. Suppose we want to test the null hypothesis

$$H_0 : C(u,v) = C(u,v,\theta)$$

for a parameter $\theta \in \Theta$, i.e the unknown copula $C(u,v)$ is a member of the previous parametric family.

7.4.1 Description of the Tests

Consider n i.i.d. observations $(X_i, Y_i),\ i = 1, \ldots, n$ of (X, Y). Let $B_{ij}\ i = 1, \ldots, r\ j = 1, \ldots, s$ be a partitioning of $[0, 1]^2$ into $r.s$ rectangles of equal area $\frac{1}{r}.\frac{1}{s}$ and let

$$p_{ij} = P[(U, V) \in B_{ij}|\theta] = \int_{B_{ij}} \int dC(u, v, \theta).$$

In order to develop tests of fit for H_0, Dobrić and Schmid [12] dealt with the following two cases.

Case 1: The marginal distribution functions F_X and F_Y are known.
Let $(U_i, V_i) = (F_X(X_i), F_Y(Y_i)),\ i = 1, \ldots, n$ be a i.id sample from $(U, V) = (F_X(X), F_Y(Y))$, and let N_{ij} be the number of (U_i, V_i) in B_{ij}. A suitable test statistic is the Chi-square statistic

$$\chi^2_{emp}(\widehat{\theta}^{(1)}) = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - np_{ij}(\widehat{\theta}^{(1)}))^2}{np_{ij}(\widehat{\theta}^{(1)})}, \tag{7.9}$$

where $\widehat{\theta}^{(1)}$ is the minimum Chi-square estimator for θ, defined by

$$\chi^2_{emp}(\widehat{\theta}^{(1)}) = \min_{\theta \in \Theta} \chi^2_{emp}(\theta).$$

Another test statistic is the Likelihood Ratio (LR) test statistic given by

$$LR^{(1)}(\widehat{\theta}^{(1)}_{ML}) = \sum_{i=1}^{r} \sum_{j=1}^{s} N_{ij} \ln p_{ij}(\widehat{\theta}^{(1)}_{ML}), \tag{7.10}$$

where $(\widehat{\theta}^{(1)}_{ML})$ is the maximum likelihood estimator for θ, defined by

$$LR^{(1)}(\widehat{\theta}^{(1)}_{ML}) = \max_{\theta \in \Theta} LR^{(1)}(\theta).$$

Under the null hypothesis H_0, both the statistics $\chi^2_{emp}(\widehat{\theta}^{(1)})$ and $LR^{(1)}(\widehat{\theta}^{(1)}_{ML})$ have an asymptotic Chi-square distribution with $rs - 1 - d$ degrees of freedom.

Case 1 is unsuitable for practical applications because most of the times, the marginal distributions F_X and F_Y of X and Y are unknown. The most interesting case for applications is the following case.

Case 2: The empirical distribution functions F_X and F_Y are unknown and treated as nuisance parameters.
Define $\widehat{U}_i = \widehat{F}_{X,n}(X_i)$ and $\widehat{V}_i = \widehat{F}_{Y,n}(Y_i)$, where $\widehat{F}_{X,n}$ and $\widehat{F}_{Y,n}$ are the empirical versions of

U_i and V_i respectively. Note that $(\widehat{U}_i, \widehat{V}_i)$, $i = 1, \ldots, n$ are not independent and their joint distribution is not given by the copula $C(u, v)$. Letting N_{ij} be the number of $(\widehat{U}_l, \widehat{V}_l)$ in B_{ij}, a suitable test statistic is given by (see [12])

$$\chi^2_{emp}(\widehat{\theta}^{(2)}) = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - np_{ij}(\widehat{\theta}^{(2)}))^2}{np_{ij}(\widehat{\theta}^{(2)})}, \tag{7.11}$$

where $\widehat{\theta}^{(2)}$ is the minimum Chi-square estimator for θ based on χ^2_{emp}.

We have

$$\sum_{i=1}^{r} \widehat{N}_{ij} = n.\frac{1}{s},\ j = 1, \ldots, s,$$

and

$$\sum_{j=1}^{s} \widehat{N}_{ij} = n.\frac{1}{r},\ i = 1, \ldots, r.$$

Therefore, the margins of the $r \times s$ table containing the frequencies $\widehat{N}_{ij}$ are fixed, and so the degrees of freedom of the Chi-square distribution of $\chi^2_{emp}(\widehat{\theta}^{(2)})$ are

$$df = r.s - (s + r - 1) - d = (r - 1)(s - 1) - d.$$

The critical value for the modified Chi-square is determined as the $(1 - \alpha)$-quantile of the Chi-square distribution with $df = (r - 1)(s - 1) - d$ degrees of freedom.

The same modification can be made to the Likelihood Ratio test $LR^{(1)}(\widehat{\theta}^{(1)}_{ML})$.

7.4.2 Properties of the Modified Chi-Square Test Under H_0

Dobrić and Schmid [12] used a sample of size $n = 2500$ for several families of copulas and carried out the test in the following way: H_0 is rejected if $\chi^2_{emp}(\widehat{\theta}^{(2)}) > c$, where c is the $(1-\alpha)$-quantile of the Chi-square distribution with $df = (r-1)(s-1)-1$ degrees of freedom. They came to the conclusion that $df = (r-1)(s-1)-1$ is a reliable choice of degrees of freedom in the Chi-square distribution of the test statistic for samples of size $n = 2500$, and that the Chi-square test of fit suggested has the power to reject a wrong null hypothesis. In fact, they noticed that Monte Carlo means and variances are close to their expected values and the simulated rejection probabilities are sufficiently close to the nominal value α if the critical value is taken as the $(1 - \alpha)$-quantile of a χ^2-distribution with $rs - r - s$ degrees of freedom.

7.5 A Goodness-of-Fit Test for Copulas Based on Rosenblatt's Transformation

In this section we consider a goodness-of-fit test investigated by Dobrić and Schmid [13], its usefulness and the problems it encounter.

7.5.1 Sketch of the Rosenblatt Transformation Test (RTT)

Consider a random vector (X, Y) with joint distribution function $F_{X,Y}$ and continuous marginals F_X and F_Y. There exists a unique copula C such that

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)).$$

The copula C is the joint distribution function of the variables $U = F_X(X)$ and $V = F_Y(Y)$, i.e. $C(u, v) = P[U \leq u, V \leq v]$ for $(u, v) \in [0, 1]^2$. Define the conditional distribution function of V given $U = u$ by

$$\begin{aligned} C(v|u) &= P[V \leq v|U = u] && (7.12)\\ &= \lim_{\Delta u \to 0} P[V \leq v|u \leq U \leq u + \Delta u] && (7.13)\\ &= \lim_{\Delta u \to 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} && (7.14)\\ &= \frac{\partial C(u, v)}{\partial u}. && (7.15) \end{aligned}$$

Rosenblatt [47] proved that the random variables

$$Z_1 = U = F_X(X)$$

and

$$Z_2 = C(V|U) = C(F_Y(Y)|F_X(X))$$

are independent and uniformly distributed on $[0, 1]$, and so the random variable

$$S(X, Y) = [\Phi^{-1}(F_X(X))]^2 + [\Phi^{-1}(C(F_Y(Y)|F_X(X)))]^2$$

has a χ^2_2-distribution. The null hypothesis of interest is $H_0 : (X, Y)$ has copula $C(u, v)$. In the case where the marginals F_X and F_Y are known, the preliminaries can be used to perform a test for H_0. From the previous result, if $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a random sample from (X, Y), then $S(X_1, Y_1), \ldots, S(X_n, Y_n)$ is a sample from a χ_2^2-distributed random variable. Therefore, if F_X and F_Y are known, the values of $S(X_1, Y_1), \ldots, S(X_n, Y_n)$ can be computed and be used to test the auxiliary null hypothesis $H_0^* : S(X, Y)$ is χ_2^2-distributed. Since H_0^* implies H_0, we reject H_0 if H_0^* is rejected. To test H_0^*, the Anderson Darling (AD) test has been used. The AD test statistic is given by

$$AD = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln(F_0(S_{j,n})) + \ln(1 - F_0(S_{n-j+1,n}))\right].$$

In the empirical applications of this approach, at least two problems arise: First, the hypothesis of interest is a composite hypothesis in the majority of cases; second, the marginal distribution functions F_X and F_Y are usually unknown in applications and need to be estimated by their empirical versions $\widehat{F}_X$ and $\widehat{F}_Y$.

7.5.2 Performance of the Test and Discussion

By means of a Monte Carlo (MC) simulation, the properties of the RTT for Gaussian copulas and $t_\nu-$copulas have been investigated. In the case of a bivariate Gaussian copula $C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$, where Φ is the standard normal distribution function, Dobrić and Schmid [13] dealt with the following three cases:

Case A: The marginal distributions F_X and F_Y are known and ρ is known.

Case B: The marginal distributions F_X and F_Y are unknown and are replaced by the corresponding empirical distribution functions $\widehat{F}_X$ and $\widehat{F}_Y$, and also ρ is unknown and has to be estimated using $\widehat{F}_X$ and $\widehat{F}_Y$.

Case C: The marginal distributions F_X and F_Y are unknown and are replaced by their empirical versions, but ρ is known.

They came to the conclusion that the goodness-of-fit test based on Rosenblatt's transformation works in Case A, and that in Case B the null distribution of the AD test statistic is very different from that in Case A. The $t_\nu-$copulas have been used as an alternative in order to shed some light on the power of the RTT.

Based on the bivariate Gaussian copula, it has been demonstrated by simulation that using critical values of Case A makes the test useless because the rejection probability under H_0 is zero and there is reduced power for rejecting a wrong null hypothesis. As an alternative, the parametric bootstrap is used for the determination of the critical values. By simulation based on a bivariate Gaussian copula, Dobrić and Schmid [13] showed that the bootstrap version of the RTT works well.

7.6 Simulations

Simulation is a widely used tool where the power of the computer helps to experiment with complex stochastic models. It can be used for example to generate a sequence of data from a multivariate distribution, in order to study the properties of functions of such data. In this section we conduct simulation studies to first compare the power of rejection of the null hypothesis of the Clayton copula by four different test statistics under the alternative of the Gumbel-Hougaard copula, and then to compare the power of rejection of the null hypothesis of the Gumbel-Hougaard copula under the alternative of the Clayton copula. Both these family of copulas were given in Table 2.1. We use the S_n and T_n statistics described in Section 7.1 and two of the dimension reduction approaches proposed by Breymann, Dias and Embrecht (denoted by BDA) on one hand, Berg and Bakken (denoted by BB) on the other, both these tests were described in Section 7.3. In all cases we use the bootstrap method to estimate the p-values.

7.6.1 Visualization of the Two Families of Copulas

Before we carry out the tests, we first compare the Clayton copula and the Gumbel-Hougaard copula by visualizing simulated samples from them in terms of contour plots (Figure 7.1) as well as perspective plots (Figure 7.2 and Figure 7.3).

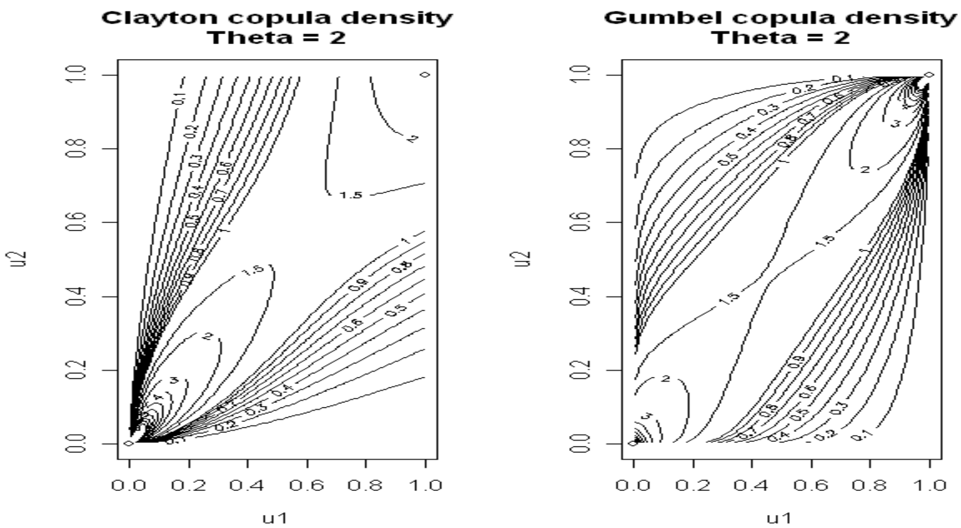

Figure 7.1: Contour plots of the null and the alternative copulas

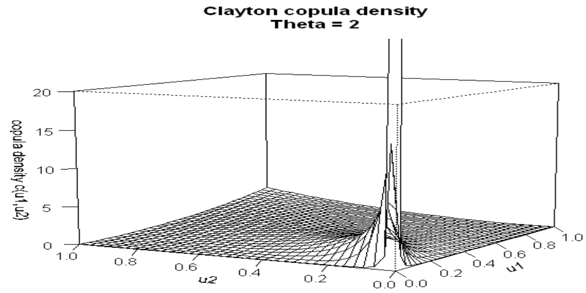

Figure 7.2: Perspective plot of the null copula

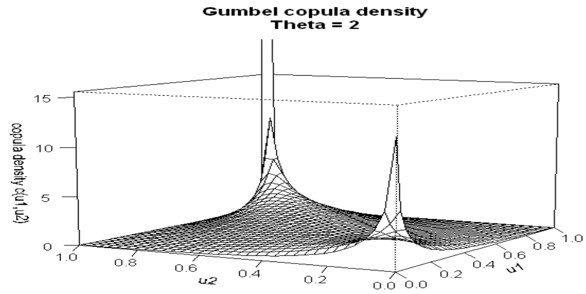

Figure 7.3: Perspective plot of alternative copula

The figures are more or less similar in the lower parts and so when one wants to apply them to a given data set in practice, it is important to distinguish carefully which one to use.

7.6.2 Simulation Results and Interpretations

Using a resampling method, we found the critical values for the test statistics under the null hypothesis. The results are given in Table 7.1.

n	S_n	T_n	BDE	BB
50	0.188	0.967	1.115	9.431
100	0.174	0.895	1.021	9.175
150	0.171	0.913	0.987	8.872
200	0.168	0.856	1.019	8.919

Table 7.1: Critical values of the test statistics under the Clayton copula

Table 7.2 contains the percentages of rejection of the null hypothesis (the Clayton copula) for the four test statistics.

n	S_n	T_n	BDE	BB
50	6	5	4	4
100	5	6	3	3
150	4	5	5	4
200	5	4	5	4

Table 7.2: Percentages of rejection under the null hypothesis

Finally, the power of rejecting the null hypothesis of the Clayton copula, given the alternative of the Gumbel-Hougaard copula, is given in Table 7.3.

n	S_n	T_n	BDE	BB
50	0.78	0.85	0.53	0.81
100	0.87	0.79	0.51	0.89
150	0.89	0.74	0.55	0.84
200	0.83	0.78	0.58	0.79

Table 7.3: Power of rejecting the null hypothesis of the Clayton copula

We see from Table 7.3 that the tests S_n and BB seem to perform better than the two other tests in that they have larger power of rejecting the null hypothesis. The test BDE does the worst of the four tests.

Considering the Gumbel-Hougaard copula as the null hypothesis and testing it against the Clayton copula, we get the power of rejection given in Table 7.4.

n	S_n	T_n	BDE	BB
50	0.67	0.71	0.61	0.74
100	0.74	0.68	0.44	0.76
150	0.81	0.69	0.51	0.82
200	0.80	0.72	0.50	0.81

Table 7.4: Power of rejecting the null hypothesis of the Gumbel-Hougaard copula

From this table, we also see that the test statistics S_n, BB and T_n are powerful for rejecting the null hypothesis of a Gumbel-Hougaard copula. Given a practical data set, suppose we have done prior analyses such as graphical representations and found out that both the Clayton and the Gumbel-Hougaard copulas fit the data. In order to decide which one of them to use, any of the three tests (S_n, BB or T_n) can then be used.

Chapter 8

Application

In this chapter we apply the previous described methods to analyze a practical data set, from the Catholic University of Leuven, Belgium. This data set comes from the university carillon, and consists of the partials of its bells. The carillon consists of two groups of bells, founded in years 1928 and 1983 by respectively Gillett & Johnston, and Eisjbouts. A partial is an identifiable frequency in the sound of a bell, arising from a mode of vibration of the bell. Teugels & Bearda [55] state that the properties of bells, and then specifically carillons, are not fully documented. In the 17th century, bell founders developed techniques that enabled them to tune bells so that the different bells within a carillon could sound harmonically together. Like the sound of all instruments, the sound of a bell is made up of a number of tones or partials, also called overtones. Each partial corresponds to a vibration mode and is expressed in frequencies. In the case of bells, these frequency ratios are $1:2:2.4:3:4:\ldots$, which form the ratios of an ideally tuned bell. The actual tuning can then be expressed by these frequencies of the partials. The classical note denomination uses as reference $a4 = 440$Hz and assuming equal tempered tuning, the octave interval is represented by a frequency ratio of 2, i.e. $a5 = 880$Hz. As there are 12 semitones in an octave, each consecutive note is presented by a ratio increase of $2^{1/12}$. Each semitone interval, for example from 440×2^0Hz to $440 \times 2^{1/12}$Hz, can be divided into 100 cents in the form 440×2^0Hz to $440 \times 2^{100/1200}$Hz. When the bell is struck the following basic frequencies are activated:

1. The hum-note.
2. The fundamental or prime.
3. The minor third.
4. The fifth or quint.

5. The nominal or octave.

The pitch of a bell is indicated by its strike note which is one octave below the nominal and coincides with the fundamental when ideally tuned. A bell can now be tuned either by its partials with the cent correction or by the corresponding frequencies. The cent notation is used more often by musicians. Based on the prime of a bell the theoretical frequencies of the partials are known. In practice however, they differ from the real frequencies obtained. These differences are expressed in terms of cents, and the data set consists of these differences from the theoretical frequencies for the minor third as well as for the quint. Figure 8.1 gives some pictures of typical bells.

Figure 8.1: Typical Bells

Getting correct information on bells and carillons is of interest in campanometry, the quantitative or statistical study of bells, bell-casting and bell-ringing. Our aim here is not to investigate this interesting science, but to analyze a portion of the above mentioned data using statistical techniques described in this thesis. We especially consider the first 33 observations of the third and fifth partials, namely THIRD and QUINT. These form a bivariate data set which will be called Deviations data, and is given below.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
THIRD	-2	-6	-18	-6	-8	-1	-8	-8	-11	-8	-16	-15	-7	0	-15	-19	-23	-13
QUINT	-22	-22	-29	-52	-29	-40	-32	-31	-30	-28	-35	-30	-36	-34	-44	-38	-33	-31

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
THIRD	-21	-10	-14	-20	-23	-11	-14	5	-2	-12	-6	2	-2	-2	-13
QUINT	-36	-47	-35	-38	-30	-51	-47	-45	-50	-38	-48	-51	-50	-60	-45

Table 8.1: Deviations data set

8.1 Graphical Displays

The left panel of Figure 8.2 shows the scatterplot of the data set. In order to explore the variation of the two variables, we use the plot in the right panel of Figure 8.2.

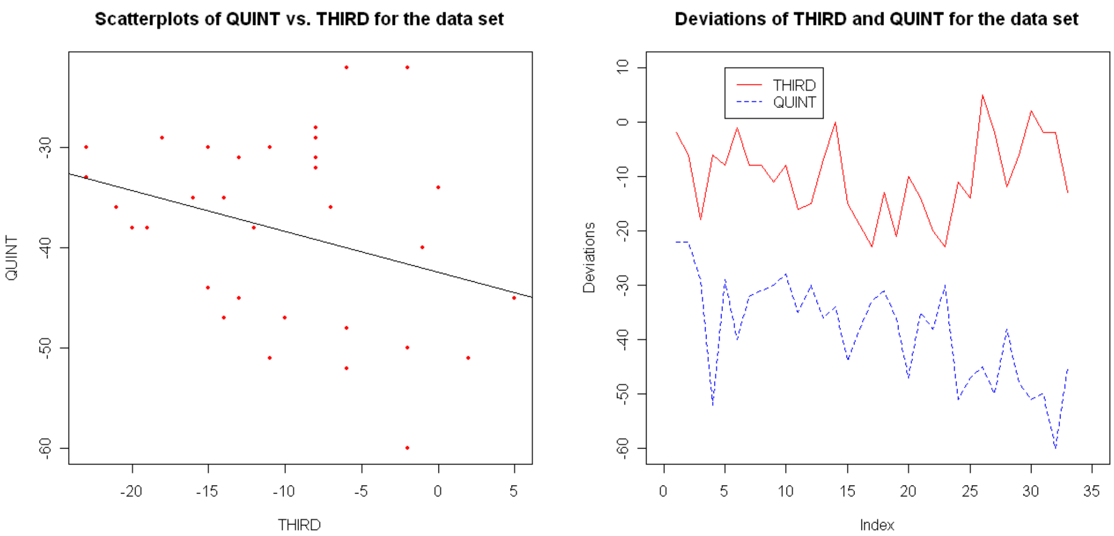

Figure 8.2: Scatterplots (left) and variation plots (right) for the Deviations data set

We see from the plot that the variation of the two variables THIRD and QUINT is quite similar. It is then clear from this plot that there is some form of dependence between the variables. The Pearson correlation coefficient is -0.3155615, i.e. a negative correlation between the variables THIRD and QUINT.

Figure 8.3 gives QQ-plots for the marginal distributions of the data set, and Figure 8.4 gives the Chi-square plot for the bivariate data set, as described in Section 6.2.2.

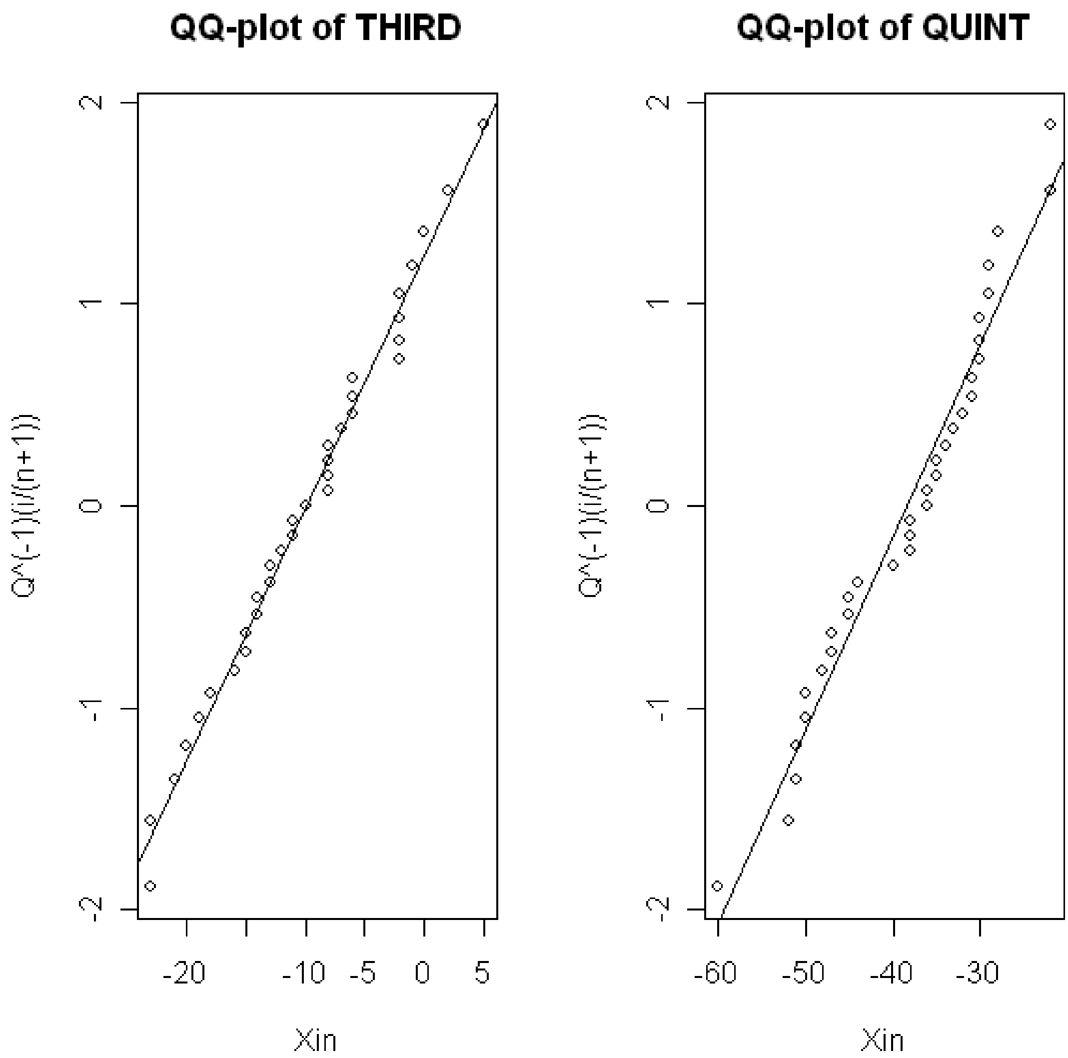

Figure 8.3: QQ-plots for marginal distributions

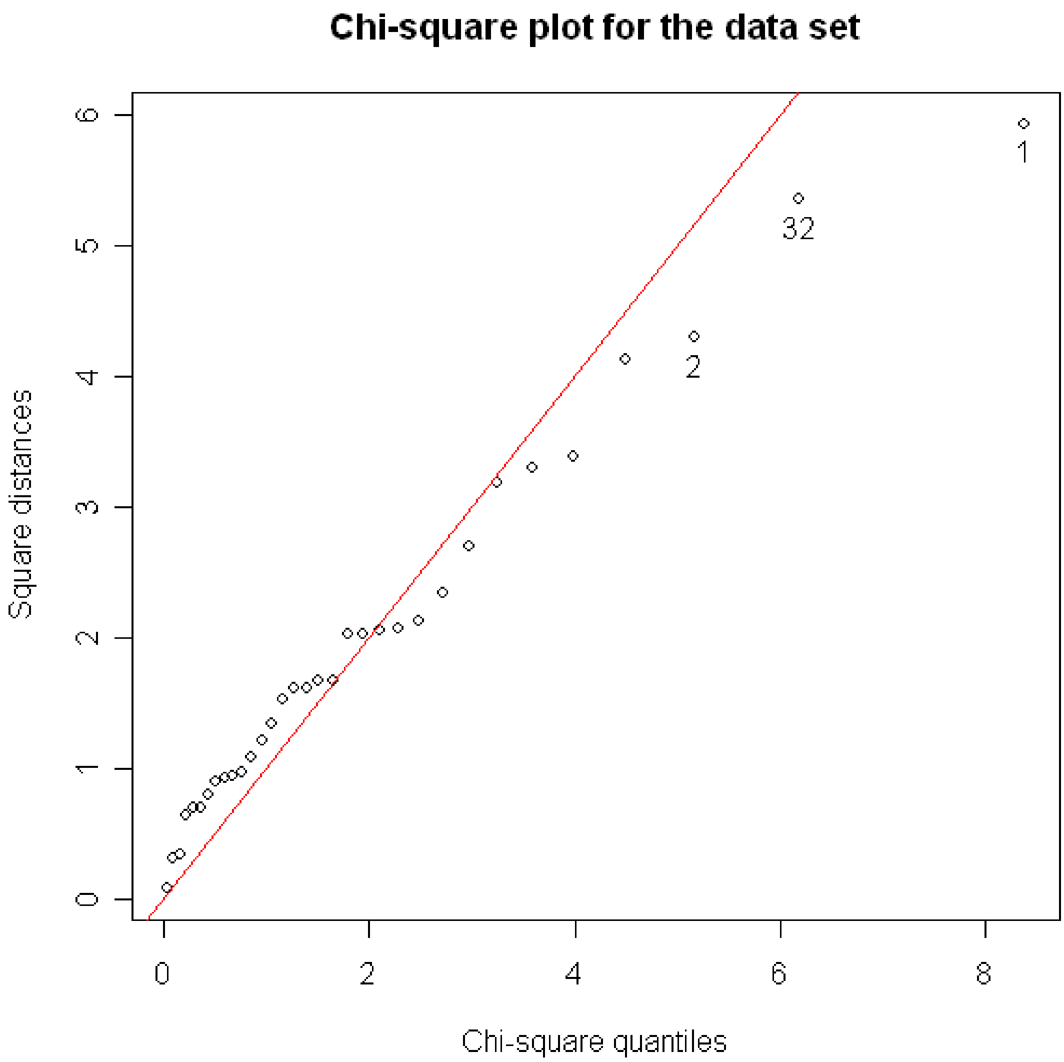

Figure 8.4: Chi-square plot for the Deviations data set

In the case of marginals, we can notice that the points are nearly straight lines, and so we suspect univariate normality to hold. The correlation for the QQ-plot points is 0.9942462 for the variable THIRD and 0.9817476 for the variable QUINT. The means, the standard deviations, the skewness and the kurtosis for these variables are given below.

	THIRD	QUINT
mean	-9.909090909	-38.3939394
standard deviation	7.341507153	9.4668481
skewness	0.002890125	-0.2899351
kurtosis	2.159859999	2.1752236

Moreover, we notice that the Chi-square plot corresponding to the data can be reasonably approximated by a straight line having slope 1 and passing through the origin, with one outlying point (observation 1). Therefore the hypothesis of bivariate normality can be considered for formal testing. The Pearson correlation coefficient is -0.3155615 (negative correlation), denoting the dependence structure between THIRD and QUINT. We also see that the variable THIRD is more symmetric than the variable QUINT. The standard deviations show more or less the same variation.

8.2 Goodness-of-Fit Testing

Before we test the null hypothesis H_0, that the joint distribution is bivariate normal, we first investigate the test for univariate normality for the marginal variables.

8.2.1 Test for Univariate Normality

From the statistics given in the previous section, we see that both these variables THIRD and QUINT have more or less the same kurtosis, not that far from the standard normal kurtosis which is 3. Using the Shapiro-Wilk (described in Section 6.1.2 a.), Anderson-Darling (see Section 6.1.1 d.) and Cramer-Von Mises (see Section 6.1.1 c.) tests for univariate normality, gives the following p-values:

```
Variable THIRD

Number of bootstrap   Shapiro-Wilks   Anderson-Darling   Cramer-Von Mises
           1000.000           0.307              0.817              0.828

Variable QUINT

Number of bootstrap   Shapiro-Wilks   Anderson-Darling   Cramer-Von Mises
           1000.000           0.273              0.795              0.810
```

We see that the hypothesis of univariate normality is not rejected in both cases.

8.2.2 Test for Bivariate Normality

We have seen in Section 8.1 that the Chi-square points (Figure 8.4) are approximately on a straight line having slope 1 and passing through the origin, portraying the possibility of not rejecting the bivariate normal distribution. In this section we consider several tests for bivariate normality in order to formally test this point. Using the Shapiro-Wilk (described in Section 6.3.5), Anderson-Darling (see Section 6.1.1 c.) and Cramer-Von Mises (see Section 6.3.2) test for bivariate normality gives the following results:

```
        Cramer-von Mises normality test

W = 0.0327, p-value = 0.8006


        Anderson-Darling normality test

A = 0.2194, p-value = 0.822


        Shapiro-Wilk normality test

W = 0.8713, p-value = 0.5641


Bootstrap test for bivariate normality

   Test    Cramer-Von Mises    Anderson-Darling       Shapiro-Wilks
p-value               0.437               0.874               0.656
```

We see that in all cases, the null hypothesis of bivariate normality is not rejected, the p-values being much larger than 0.05. Therefore, the Gaussian copula can be used to analyze the dependence structure of the data. To confirm that, we carry out a test for the Gaussian copula and we obtain a p-value of 0.5431776 when using the Kolmogorov-Smirnov distance and 0.7664668 for the Anderson-Darling distance. The QQ-plots associated with the Gaussian copula test are given in Figure 8.5, for two simulations.

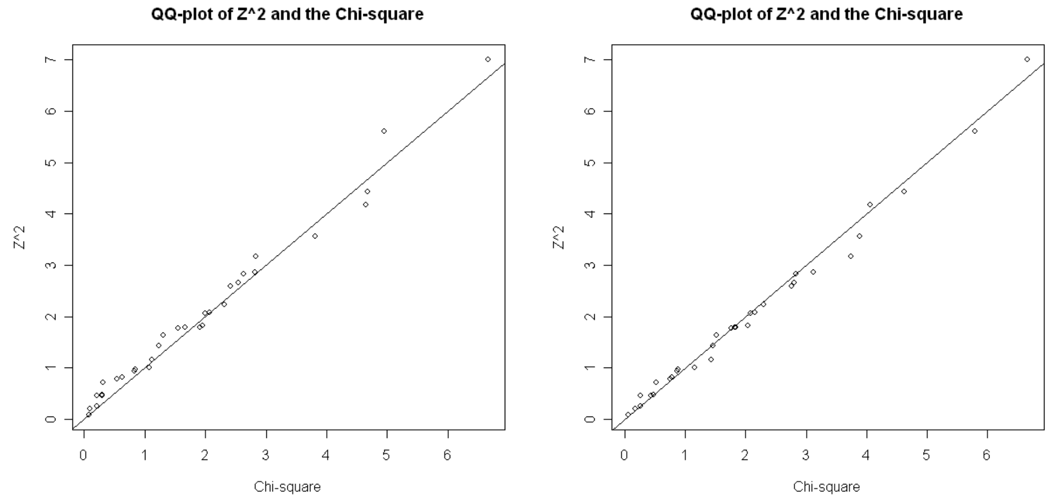

Figure 8.5: QQ-plots associated with the Gaussian copula test

This graphs show how well the Gaussian copula fits the data set.

8.3 Testing Other Copulas

In order to give an application of the goodness-of-fit methods we discussed in Chapter 7, we next carry out a goodness-of-fit test to compare the Clayton family of copulas to the Gumbel-Hougaard family in the case of the Deviations data set. The estimated Kendall tau (see (3.5)) associated with the data is -0.1998065, and the Pearson correlation coefficient is -0.3155615. We use these quantities to generate samples from the Clayton and Gumbel-Hougaard copulas. The estimated parameter for the Clayton copula is -0.3330645, and that for the Gumbel-Hougaard is 0.83346771. We obtain the contour plots of the estimated copulas displayed in Figure 8.6.

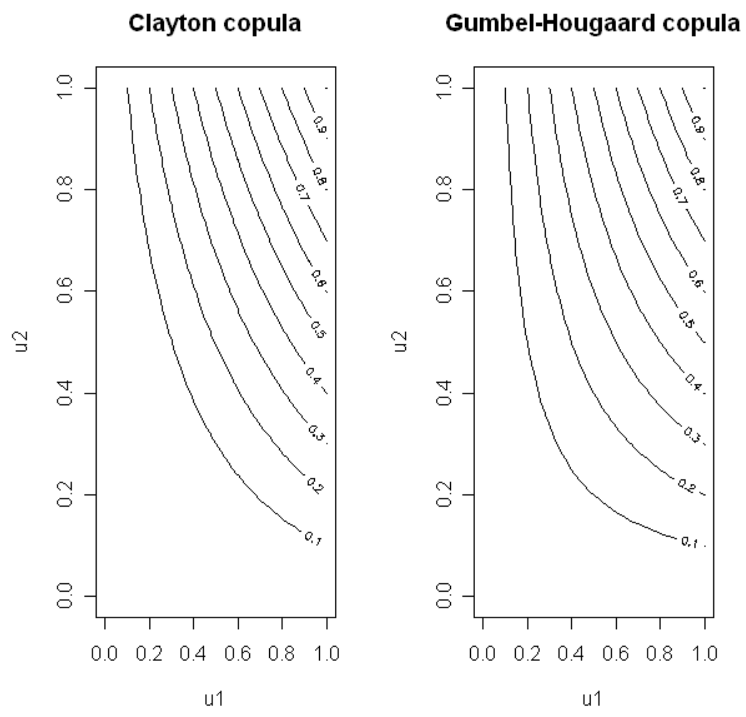

Figure 8.6: Contour plots of the estimated copulas

We see that these plots are very similar and so we need further work to decide which one of the two copulas better fits the data. For this reason, we compare the QQ-plots associated with these copulas to that of the data. We obtain the plots in Figure 8.7, which show that Gumbel-Hougaard copula seems to perform better than the Clayton copula.

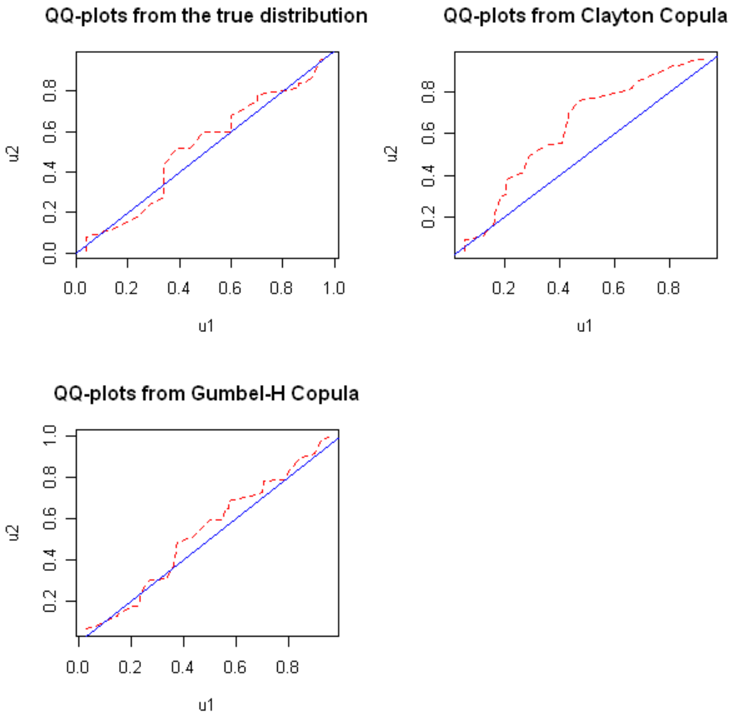

Figure 8.7: QQ-plots for comparison of the estimated copulas to the joint distribution

We further test the null hypothesis H_0: the copula describing the data is from the Clayton family, using the test statistics considered for simulation in Chapter 7. We obtain the results in Table 8.2.

Statistic	S_n	T_n	BDE	BB
p-value	0.02970297	0.04950495	0.1584158	0.0007030
H_0 rejected	YES	YES	NO	YES

Table 8.2: Rejection of the null hypothesis of the Clayton copula for the Deviations data set

From the results in this table, we can see that the null hypothesis is rejected by three tests (S_n, T_n and BB). Therefore the Gumbel-Hougaard copula seems to fit the data better, compared to the Clayton copula. Before we draw any conclusion, we finally consider the null hypothesis to be that of the Gumbel-Hougaard copula and then we test it against the alternative of the Clayton copula. We obtain the results in Table 8.3.

Statistic	S_n	T_n	BDE	BB
p-value	0.9504950	0.6237624	0.10056436	0.01980198
H_0 rejected	NO	NO	NO	YES

Table 8.3: Rejection of the null hypothesis of the Gumbel-Hougaard copula for the Deviations data set

From Table 8.3, we see that the null hypothesis of a Gumbel-Hougaard copula is strongly not rejected by the test S_n, and only rejected by BB. Although it is weakly not rejected by BDE, we can see by putting together the results from the two analyses that inference about the Deviations data set can be done using a copula from the Gumbel-Hougaard family.

Chapter 9

Summary and Further Work

Measures of goodness-of-fit typically summarize the discrepancy between observed values and the values expected under a statistical model. Such measures can be used in statistical hypothesis testing, for example to test whether outcome frequencies follow a specified distribution. Copulas on the other hand proved to be one of the most widely used tools to study multivariate outcomes. In the case of dependent multivariate data, multivariate copulas provide a useful tool to assist in the process of model building.

In this thesis, we studied some aspects of copulas and goodness-of-fit test statistics. A copula is a function that relates a multivariate distribution function to its one-dimensional marginal distribution functions. In Chapter 2, we defined a copula function, and we gave some examples. We reviewed its properties, such as the invariance under strictly monotone transformations in Chapter 3. Sklar's theorem, also given in that chapter, proves the one-to-one relationship that exists between the joint distribution function on one hand, and the copula function combined with all marginal distribution functions on the other hand. The inversion method of constructing bivariate copulas and an algorithm to generate Archimedean copulas were also considered.

Since copulas are parametric families, standard techniques such as the maximum likelihood and inference functions for margins (IFM) methods, are useful for estimating their parameters. These methods are described in Chapter 4. Regression analysis is a statistical technique intensively used to measure the degree of relationship between two or more variables. In Chapter 5, we discussed an alternative way of looking at regression analysis by using copulas. Specifically, linear and non linear copula regression functions were described.

In Chapter 6, we investigated the goodness-of-fit tests for general distributions. There, we described several test statistics, and we looked at the asymptotic behavior of some of them. There are in fact many situations in statistics where we want to test whether a particular distribution fits our observed data. In the univariate case, we described the Chi-square type goodness-of-fit tests, the Kolmogorov-Smirnov, Cramer-Von Mises and Anderson-Darling test statistics. An attractive feature of the chi-square goodness-of-fit test is that it can be applied to any univariate distribution for which one can calculate the cumulative distribution function. Moreover, it is an alternative to the Anderson-Darling and Kolmogorov-Smirnov goodness-of-fit tests. The chi-square goodness-of-fit test can in fact be applied to discrete distributions such as the Binomial and the Poisson distributions while Kolmogorov-Smirnov and Anderson-Darling tests are restricted to continuous distributions. In order to handle the cases where parameters are estimated, we also gave a description of the estimated empirical process. Univariate tests for normality, such as the Shapiro-Wilk and De Wet-Venter tests were discussed.

Several bivariate goodness-of-fit test statistics were described in Section 6.2. Particularly, we described the bivariate Kolmogorov-Smirnov test, as well as the Kim-Bickel test for bivariate normality. We also described in Section 6.3 some multivariate test statistics, namely the test for sphericity, the multivariate Cramer-Von Mises tests and De Wet-Venter tests for multivariate normality, as well as others. When testing if a given distribution P belongs to a parameterized family **P**, one often compares a nonparametric estimate A_n of some functional A of P with an element A_{θ_n} corresponding to an estimate θ_n of the parameter θ. In many cases, the asymptotic distribution of goodness-of-fit test statistics derived from the process $\sqrt{n}(A_n - A_{\theta_n})$ depends on the unknown distribution P. In Section 6.4, we considered the parametric bootstrap methods proposed by Stute et al [52], and gave the validity of the methods when testing for families of multivariate distributions. We also discussed the validity of a two-level bootstrap in cases where the parametric estimate can not be computed easily. We ended Chapter 6 by giving an indication of the power of goodness-of-fit tests.

In Chapter 7, we described some goodness-of-fit test statistics for copulas. One interesting approach discussed is the probability integral transform method. Dimension reduction approaches, the likelihood ratio tests as well as the parametric bootstrap method for copula goodness-of-fit testing were considered.

With simulations studies conducted in Section 7.6, we compared the power of rejection of the null hypothesis of the Clayton copula by four different test statistics under the alternative of a Gumbel-Hougaard copula, and also the power of rejection of the null hypothesis of the Gumbel-Hougaard copula under the alternative of a Clayton copula. The results from Table 7.3 yield the conclusion that the tests S_n and BB perform better than the tests T_n and BDE, in that they have larger power of rejecting the null hypothesis. We also found that the test BDE does the worst of the four tests. Considering the Gumbel-Hougaard copula as the null hypothesis and testing it against the Clayton copula, we obtained the power of rejection given by Table 7.4, from which we concluded that the test statistics S_n, BB and T_n are powerful for rejecting the null hypothesis of a Gumbel-Hougaard copula. Therefore, given a practical data set, suppose we have done prior analyses such as graphical representations and found out that both the Clayton and the Gumbel-Hougaard copulas reasonably fit the data. In order to decide which one of them to use, any of the three tests (S_n, BB or T_n) could be used. Finally in Chapter 8, we applied the previously described methods to analyze a practical data set, from the Catholic University of Leuven, consisting of partials of bells in the carillon in the University library.

The area of Extreme Value Statistics in general offers a wide variety of problems. In modelling the extreme events of a random variable, Extreme Value Theory is the counterpart of the Central Limit Theorem for sums. Extreme Value Theory deals with these extreme events, providing a classification of continuous distributions according to the behavior of the tail region or their extreme realizations. The theory distinguishes three limiting stable distributions for the maximum values of a random variable, called Generalized Extreme Value Distributions and the associated Generalized Pareto Distributions. Given an extreme value problem, the correct choice of distribution is of some importance as the three asymptotic distributions may differ considerably. Moreover, several methods are available for estimating the extreme value parameters. Therefore, having fitted a model to a data set, one should evaluate how well the model describes or explains the available data. In further research, we will investigate goodness-of-fit tests in the field of Extreme Value Theory.

References

[1] Belzunce F., Castano A., Olveira-Cervantes A., and Suarez-Llornz A. Quantile Curves And Dependence Structure For Bivariate Distributions. *Computational Statistics and Data Analysis*, 51:5112–5129, 2007.

[2] Berg D. and Bakken H. A Goodness-of-Fit Test for Copula Based on the Probability Integral Transform. 2005. Norwegian Computing Centre SAMBA/41/05.

[3] Berg D. and Bakken H. Copula Goodness-of-fit Tests: A Comparative Study. Technical report, 2006. Link: http://www.danielberg.no/publications/CopulaGOF.pdf.

[4] Breymann W.A., Dia A., and Embrechts P. Dependence Structures for Multivariate High-frequency Data in Finance. *Quantitative Finance*, 1:1–14, 2003.

[5] Clayton D.G. A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, 65:141–151, 1978.

[6] Cook R.D and Johnson M.E. A Family of Distributions for Modelling Non-Elliptically Symmetric Multivariate Data. *Journal of the Royal Statistical Society*, B43:210–218, 1981.

[7] Cotterill D.S. and Csörgö M. On the Limiting Distribution of and Critical Values for the Multivariate Cramer-Von Mises Statistic. *The Annals of Statistics*, 10(1):233–244, 1982.

[8] Cressie N. and Read T.R.C. Multinomial Goodness of Fit Tests. *Journal of Royal Statistical Society*, 46:440–464, 1984.

[9] Cui H. The Average Projection Type Weighted Cramer-Von Mises Statistics for Testing Some Distributions. *Science in China Series A: Mathematics*, 45:562–577, 2002.

[10] De Wet T. and Venter J.H. Asymptotic Distributions of Certain Test Criteria of Normality. *South African Statististical Journal*, 6:135–149, 1972.

[11] De Wet T., Venter J.H., and Van Wyk J.W.J. The Null Distributions of some Test Criteria of Multivariate Normality. *South African Statististical Journal*, 13:153–176, 1979.

[12] Dobrić J. and Schmid F. Testing Goodness-of-Fit for Parametric Families of Copulas - Application to Financial Data. *Communications in Statistic: Simulation and Computation*, 34, 2005.

[13] Dobrić J. and Schmid F. A Goodness-of-fit Test for Copula Based on Rosenblatt's Transformation. *Computational Statistics and Data Analysis*, 51:4633–4642, 2007.

[14] Embrechts P, Klüppelberg C, and Mikosch T. *Modeling Extremal Events*. Springer, 1997.

[15] Fang K-T., Zhu L-X., and Bentler P.M. A Necessary Test of Goodness-of-fit for Sphericity. *Journal of Multivariate Analysis*, 45:34–55, 1993.

[16] Fligner M.A. Goodness of Fit. Encyclopedia of Statistical Sciences, Vol 3:451–461, 1983. Editors-in-chief: Kotz S., Norman L. J.

[17] Frank M.J. On the Simultaneous Associativity of F(x,y) and x+y-F(x,y). *Acquationes Mathematicae*, 19:194–226, 1979.

[18] Freeman M.F. and Tukey J.W. Transformations Related to the Angular and the Square root. *Annals of Mathematical Statistics*, 91:607–609, 1950.

[19] Frees E. W. and Valdez E. A. Understanding Relationships Using Copulas. *North American Actuarial Journal*, 1998.

[20] Ganßler P. and Stute W. Seminar on Empirical Processes. *Birkhäuser Verlag, Basel*, 1987.

[21] Genest C., Quessy J.-F., and Rémillard B. Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation. 2005. Preprint.

[22] Genest C., Quessy J.-F., and Rémillard B. Goodness-of-Fit Procedures for Copula Models Based on the Probability Integral Transform. *Scandinavian Journal of Statistics*, 33, 2006.

[23] Genest C. and Rémillard B. Validity of the Parametric Bootstrap for Goodness-of-Fit Testing in Semiparametric Models. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 0:1–34, 2008.

[24] Genest C. and Rivest L. Statistical Inference Procedures for Archimedean Copulas. *Journal of the American Statistical Association*, 88:1034–1043, 1993.

[25] Greenberg S.L. Bivariate Goodness-of-fit Test Based on Kolmogorov-Smirnov Type Statistics. Master's thesis, Faculty of Science, University of Johannesburg, South Africa, 2006.

[26] Greenwood M. The Statistical Study of Infection Diseases. *Journal of the Royal Statistical Society*, 109, Series A:85–110, 1946.

[27] Gumbel E.J. Bivariate Exponential Distributions. *Journal of the American Statistical Association*, 55:698–707, 1960.

[28] Gumbel E.J. Distribution des Valeurs Extremes en Plusieurs Dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9:171–173, 1960.

[29] Hougaard P. A Class of Multivariate Failure Time Distributions. *Biometrika*, 73:671–678, 1986.

[30] Jammalamadaka S.R. and Goria M.N. A Test of Goodness-of-Fit Based on Gini's Index of Spacings. *Statistics and Probability Letters*, 68:177–187, 2004.

[31] Janssen A. Global Power Functions of Goodness-of-Fit Tests. *The Annals of Statistics*, 38(1):239–253, 2000.

[32] Joe H. and Xu J.J. The Estimation Method of Inference Functions for Margins for Multivariate Models. Technical report, Department of Statistics, University of British Columbia, 1996.

[33] Johnson R.A. and Wichern D.W. *Applied Multivariate Statistical Analysis*. Pearson Education International, 2007. Sixth Edition.

[34] Kim N. and Bickel P.J. The Limit Distribution of a Test Statistic for Bivariate Normality. *Statistica Sinica*, 13:327–349, 2003.

[35] Knight K. *Mathematical Statistics*. Chapman and Hall/CRC, 2000.

[36] Kullback S. *Information Theory and Statistics*. John Wiley, New York, 1997.

[37] Marshall A.W and Olkin. A Generalized Bivariate Exponential Distribution. *Journal of Applied Probability*, 4:291–302, 1967.

[38] Marshall A.W. and Olkin. A Multivariate Exponential Distribution. *Journal of the American Statistical Association*, 62:30–44, 1988.

[39] Mikosch T. Copulas: Tales and facts. *Springer Science + Business Media, LLC*, 2006.

[40] Milbrodt H. and Strasser H. On the Asymptotic Power of the Two-sided Kolmogorov-Smirnov Test. *Journal of Statistical Planning and Inference*, 26:1–23, 1990.

[41] Nelsen R.B. *An Introduction to Copulas*. Springer, 2006. Second Edition.

[42] Neyman J. Contribution to the Theory of the Chi-Square Test. In *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, volume 58, pages 894–898, 1949.

[43] Oakes D. A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society*, B44:414–422, 1982.

[44] Pearson K. On the Criterion that a Given System of Deviations from the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to Have Arisen from Random Sampling. *Philosophical Magazine*, 50:157–175, 1900.

[45] Pettitt A.N. Generalized Cramer-Von Mises Statistic for Gamma Distribution. *Biometrika*, 65:232–235, 1978.

[46] Quasada-Molina J.J. What are copulas? *Monografias Del Semi. Matem. Garcia de Galdeono. 27*, 2003.

[47] Rosenblatt M. Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23:470–472, 1952.

[48] Schweizer B. and Sklar A. *Probabilistic Metric Spaces*. North-Holland, Amsterdam, 1983.

[49] Shapiro S.S. and Wilk M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52:591, 1965.

[50] Shorack G.R. and Wellner J.A. *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics, 1986.

[51] Steele M.C. *The Power of Categorical Goodness-of-Fit Test Statistics*. PhD thesis, School of Environmental Studies, Faculty of Environmental Sciences, Griffith University, 2002.

[52] Stute W., Gonzalez-Manteiga W., and Presedo-Quindimil M. Bootstrap Based Goodness-of-Fit Tests. *Metrika*, 40:243–256, 1993.

[53] Sungur E.A. Some Observations On Copula Regression Functions. *Communications in Statistics-Theory and Methods*, 34:1967–1978, 2005.

[54] Sürücü B. Goodness-of-Fit Test for Multivariate Distributions. *Communications in Statistics-Theory and Methods*, 35:1319–1331, 2006.

[55] Teugels J. L. and Bearda T. Campanometry. Umpublished Manuscript, 2008.

[56] Veraverbeke N. *Multivariate Data and Copulas*. Hasselt University, Belgium, 31 October 2005. Unpublished Lecture Notes.

[57] Versluis C. Comparison of Tests for Bivariate Normality with Unknown Parameters by Transformation to an Univariate Statistic. *Communications in Statistics: Theory and Methods*, 25:647–665, 1996.

[58] Wang W. and Wells M.T. Model Selection and Semiparametric Inference for Bivariate Failure-time Data. *Journal of the American Statistical Association*, 95:62–72, 2000.

[59] Wilks S.S. The Likelihood Test of Independence in Contingency Tables. *Annals of Mathematical Statistics*, 6:190–196, 1935.