uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Adriana Alicia Jordan-Perez

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Mathematics)

GRADE / DEGREE

Department of Mathematics and Statistics

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

P-P Plots and Precedence Tests for Planar Point Processes

TITRE DE LA THÈSE / TITLE OF THESIS

Gail Ivanoff

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Murray D. Burke

Miklos Csorgo

Raluca Balan

David Macdonald

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

P-P PLOTS AND PRECEDENCE TESTS FOR PLANAR POINT PROCESSES

Adriana Jordan

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Mathematics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. Program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Your file Votre référence
ISBN: 978-0-494-49365-6
Our file Notre référence
ISBN: 978-0-494-49365-6

Abstract

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a continuous distribution F and $Y_1, Y_2, \ldots, Y_m$ be a random sample of size m from a continuous distribution G. One of the ways to test the hypothesis of equality of F and G against the alternative that $F < G$ when both distributions are univariate is to perform a precedence test -a test that not only requires only a portion of the samples, but which is distribution-free under the null hypothesis. The initial purpose of this thesis was to extend the notion of a precedence test to higher dimensions. In doing so, we found two different tests that are appropriate for both partial and complete data sets. These tests are based on two different extensions of the usual definition of a procentile-procentile plot -which is closely related to the precedence test statistic on the line- to the plane. The first of the above mentioned extensions involves the contours formed by the distribution function F; the second of our tests uses the marginal quantiles of F. For both extensions of the empirical $p-p$ plot, we have proven a Glivenko-Cantelli type of result. Also, we have developed their asymptotic convergence to Gaussian limits. The choice between tests based on these two plots depends on the kind of information that the data of our experiment generates. All the results presented here, although mostly presented for $\Re^2$, are valid for $\Re^d$-valued data.

Acknowledgements

I would like to thank, first and foremost, my supervisor, Dr. Gail Ivanoff, to whom I am deeply indebted. Her constant encouragement, support, and invaluable guidance made the completion of this thesis successful. I am truly honoured to be counted among her students.

I would also like to thank Justin Francis for his help in carrying out several numerical calculations included in this work.

Finally, I would also like to thank Consejo Nacional de Ciencia y Tecnología (CONACYT), México. This work would not have been possible without the support that they have given to my research for the past five years.

Dedication

I would like to dedicate this work to my parents and to Alberto. Without their love and support I would not have been able to complete -or even attempt to start- an undertaking of this magnitude. Thanks for being my strength during the process of achieving this goal.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the most widely used statistical tools in the area of clinical trials is the test of equality of two distributions. In general, we have two separate samples and we want to find out if they come from the same distribution. Although there exist several ways to approach this problem, in the present text we will only discuss the precedence test since the original motivation for this thesis came from this very simple idea.

Basically, a precedence test is a statistical test which counts the number of observations from one sample that occur prior to a certain observation from the second sample and then decides if the observations involved come from the same distribution or if one of the distributions is stochastically larger than the other.

In addition to its very simple structure, a precedence test has two attractive features: first, the statistic is distribution-free under the null hypothesis; second, it is not necessary to have the entire samples in order to construct the test statistic, resulting in savings in both time and money.

We can explain a precedence test in a more formal way: Given n observations from a distribution function $F(x)$ and m observations from a distribution function $G(y)$ we want to test $H_0 : F(x) = G(x)\ \forall\ x$ versus one of the following alternatives: $H_1 : F(x) \leq G(x)\ \forall\ x$, $H_2 : F(x) \geq G(x)\ \forall\ x$ or $H_3 : F(x) \neq G(x)$ for some x. H_0 will be rejected in favor of H_1 if we observe too many observations from $G(y)$, say r, before the kth observation from $F(x)$. H_0 will be rejected in favor of H_2 if we observe too many observations from $F(x)$, say k, before the rth observation from $G(y)$. Similarly, we will reject H_0 in favor of H_3 whenever we have either too many or too few observations of one sample before the kth or the rth observation of the second sample. In this thesis, we will primarily focus on the alternatives H_1 and H_2.

The first time the phrase "precedence test" appeared in the literature was in [38]. Here, we can find tables for equal sample sizes up to 20 that tell us when to reject the null hypothesis, for both kinds of alternatives: one-sided (for levels of significance $\alpha = 0.05, 0.025$) and two-sided (for levels of significance $\alpha = 0.10, 0.05$).

As we mentioned earlier, precedence tests are distribution-free, so the underlying distributions only play a role when we are calculating the power of the test. Up until this point, the assumption when dealing with the power was that the underlying distributions were normal. But then, questions about how good the test was when normality could not be assumed arose. These questions were addressed in [18]: in light of the application to life-testing, it was assumed that the underlying distributions for the samples were exponential with unknown parameters θ_x and θ_y, respectively. The hypotheses then became $H_0 : \theta_x = \theta_y$ versus $H_1 : \theta_x < \theta_y$, $H_2 : \theta_x > \theta_y$ or $H_3 : \theta_x \neq \theta_y$. Eilbott and Nadler showed how to compute α, and that this calculation did not depend on the underlying distributions being exponential.

Less than two years later [42] was published; this paper extended [18] in the following sense: the power calculations made for underlying exponential distributions were also true for a larger class of distributions. It was shown that since the power of a precedence test was unaffected by applying a strictly increasing continuous function to the data, the power calculated for the exponential distribution could be used for any distribution that can be transformed into an exponential by means of any such function.

After [42], literature focused more on particular applications to life-testing and reliability theory. A notable exception was [26], where Gastwirth proposed the first-median test: a precedence test that counted the number of observations of one sample preceding the (first) median of the other sample. This test is actually a two-sided test, since our test statistic changes depending on which median observation occurred first. It also presented a heuristic proof for the asymptotic distribution of the test.

The research on precedence tests blossomed in several directions: some researchers decided to focus on developing a test for the combined sample (see [45], [33], [34] and [48]), others wanted to develop the theory for the censored case (see [49], [44] and [11]), and yet another group decided to continue the exploration of the power and comparisons among several well-known tests.

It was in [30] that a precedence test with maximized power -the best precedence test- was discussed in depth. This publication obtained powers for the precedence test using Lehmann alternatives (see [42]) with small sample sizes and then chose the test with the greatest power among them. It also presented some power comparisons between the precedence test and other tests, such as the Mann-Whitney-Wilcoxon (MWW) and the Savage test, in order to be able to understand how much power we must lose in favor of an early termination of the experiment.

Interest in precedence tests was revived -as it happened with many other applied results- with the advance of technology. In 1993, Nelson published [39], a short paper explaining once again the basics of the precedence test. The new feature was a Basic computer program that analyzed the result of a precedence test and gave significance levels for some combinations of sample sizes. It was with the help of computers that the research of power calculations of this sort really took off.

Lastly, we will mention [12] and [7]. Despite the fact that the innovative topic in [12] was confidence bounds, it contained a vast review of precedence tests, which is quite complete and very interesting by itself. On the other hand, [7] contains not only a review of what has been done in the past, but further extends the discussion of weighted, censored, three or more samples, maximal and other types of precedence tests and related topics.

A natural question that arises is whether the concept of a test for equality of distributions can be extended to data sets in higher dimensions in such a way that the distribution-free property can be maintained -at least to some degree- and that it is not necessary to have the entire samples in order to construct the test statistic. Our goal with the present work was to answer this question. We will go through the methodology behind the one-dimensional precedence test in order to pave the way for our multidimensional version of the test. We will also add a few notes and observations at some junctures of the test that may appear obvious in one dimension, but will be crucial in our generalization to two or more dimensions.

Suppose that we have two sets of data and we want to know if they come from the same distribution. This situation could arise, for example, if we want to know which of two brands of bulbs would last longer, which gender is more likely to develop heart disease earlier in life or if two kinds of trees are equally distributed in a park.

Formally what we have are n observations $X_1, X_2, \ldots, X_n$ from a distribution function F and m observations $Y_1, Y_2, \ldots, Y_m$ from a distribution function G and we want to test $H_0 : F(x) = G(x)\ \forall\ x$ against any of the alternatives $H_1 : F(x) \leq G(x)$ $\forall\ x$, $H_2 : F(x) \geq G(x)\ \forall\ x$, or $H_3 : F(x) \neq G(x)$ for some x; in other words, we want to determine if the two sets come from the same distribution or in the case of H_1 and H_2, if the distributions are "stochastically ordered", such as, for example, a shift in location of the same distribution. It is obvious that in the alternatives H_1, H_2, we require the inequalities to be strict for some values of x.

To be able to perform the test we need the order statistics from F, namely $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, and $Y_{(1)}, Y_{(2)}, \ldots, Y_{(m)}$, the order statistics from G. Then, the only thing we need to know in order to make a decision is whether $X_{(k)} < Y_{(r)}$ or $Y_{(r)} < X_{(k)}$, for some predetermined values k and r. To show the way that precedence tests work, we will focus our attention on the test of the form $H_0 : F(x) = G(x)\ \forall\ x$ vs $H_1 : F(x) \leq G(x)\ \forall\ x$ (equivalently, $H_0 : X =_{\mathcal{D}} Y$ vs $H_1 : Y \leq X$). We will reject H_0 in favour of H_1 if we find too many Y's before $X_{(k)}$, i.e. if $Y_{(r)} < X_{(k)}$. Equivalently, we will reject H_0 if $G_m(Y_{(r)}) < G_m(X_{(k)})$, where $G_m(y) = \frac{1}{m}\sum_{i=1}^{m} I\{Y_i \leq y\}$ denotes the empirical distribution function of G at the point y. It is at this point where the connection between precedence tests and $p-p$ plots becomes clear.

Let $F^{-}(p) = \inf\{x : F(x) \geq p\}$ denote the left continuous inverse of F, where $F^{-}(0) = \lim_{p\to 0} F^{-}(p) = F^{-}(0+)$. Then, the procentile-procentile, or $p-p$ plot of G against F, is defined as $G(F^{-}(p))$ for $0 \leq p \leq 1$. Similarly, the empirical $p-p$ plot of G against F is defined as $G_m(F_n^{-}(p))$ for $0 \leq p \leq 1$. For an extensive review and bibliography of $p-p$ plots, $q-q$ plots and other graphical methods in nonparametric statistics, see [23]. Going back to our test, an order statistic can be rewritten as $X_{(\lceil np \rceil)} = F_n^{-}(p)$, where $\lceil np \rceil$ denotes the smallest integer greater than or equal to np. Therefore, it is straightforward that the statistic used in precedence tests can be regarded as a snapshot of the empirical $p-p$ plot of G against F at a single point $p = \frac{k}{n}$. Not only do we have the advantage that we require only a portion of the data points, but this statistic will be distribution-free under $H_0 : F = G$. Following our previous method, we will reject H_0 if $G_m(X_{(k)})$ takes an extreme value.

Moreover, if we had all the observations and not only part of them, we could use the entire plot to develop a distribution-free test for the equality of two distributions. For this purpose, we can use the empirical $p-p$ plot process defined by $W^{n,m}(y) = \sqrt{m}[G_m(F_n^{-}(y)) - G(F^{-}(y))]$ for $0 \leq y \leq 1$. Under general regularity conditions, the weak convergence of $W^{n,m}$ to a mean zero Gaussian process is developed in [4].

To be able to extend our test to higher dimensions, we need to analyze the problem from a slightly different perspective.

Let $\tau : \Omega \to [0,1]$ be a random variable. We call τ a stopping time with respect to a filtration $\mathcal{F}$ if $\{\tau \leq t\} \in \mathcal{F}_t \ \forall \ t \in [0,1]$. Now, recall that for the alternative $H_1 : F(x) \leq G(x)$, $H_0 : F(x) = G(x)$ will be rejected in favor of H_1 whenever $Y_{(r)} < X_{(k)}$, i.e. the experiment is terminated at the (random) time of the kth failure in the F-population, or the rth failure in the G-population. Only the data collected up to this time is needed to calculate the test statistic. Actually, the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ can be viewed as n stopping times $\tau_1, \tau_2, \ldots, \tau_n$ with respect to the minimal filtration $\mathcal{F}$ generated by the point process $N(t) = \sum_{i=1}^{n} I\{X_i \leq t\}$; that is, $\mathcal{F}_t = \sigma\{N(s) : 0 \leq s \leq t\}$. Then, our test statistic is reduced to the number of observations from $Y_1, Y_2, \ldots, Y_m$ that lie in the random set $[0, \tau_k]$. Thus, it becomes obvious that before we can extend our test to $\Re^d_+$, we need to somehow extend the concepts of stochastic orders, stopping times, $p-p$ plots and filtrations.

For data taking values in $\Re^d_+$, the setup is similar to that in the one-dimensional case. Given two continuous distributions F and G on $\Re^d_+$, we want to test $H_0 : F = G$ against the alternative that the distributions are stochastically ordered. It is in trying to set up the alternative where we need the extension of stochastic orders. This is due to the fact that $\Re^d_+$ is not a totally ordered space and thus, when we work with d-dimensional data we have more than one stochastic order. There are many such orders (for a full exploration of them see [41]), but because of the application of precedence tests to clinical trials and life testing, we will be focusing first on the lower and upper orthant order. Also, although all the concepts and results that follow can be easily extended to $\Re^d_+$, we chose to restrict our attention to the positive quadrant of the plane to simplify our presentation.

We say that the random variable X is smaller than the random variable Y in the lower orthant order ($X \leq_{lo} Y$) if, for every $t \in \Re^2_+$, the probability that X is less than or equal to t is greater than the probability that Y is less than or equal to t (in the usual partial order on $\Re^2_+$). Intuitively, this means that X is more likely to take values on $[0, t]$ than Y.

We say that the random variable X is smaller than the random variable Y in the upper orthant order ($X \leq_{uo} Y$) if for every $t \in \Re^2_+$, the probability that X is greater than t is smaller than the probability that Y is greater than t. Intuitively, the definition is saying that Y is more likely to take values on $[t, \infty]$ than X. Note that the equivalence $X \leq_{lo} Y \Leftrightarrow X \leq_{uo} Y$, although true on the line, doesn't hold anymore in the bivariate setting. Both of these stochastic orders are usually linked with experiments dealing with clinical data.

Since tests of equality of distributions could also be applied to geographical data, we will also work with another type of stochastic order, similar in spirit to both the Kendall stochastic order and the H-larger order. We say that X is less than Y in the Kendall stochastic order ($X \prec_K Y$) whenever $K_1(t) \geq K_2(t) \ \forall \ t \in \Re$, where K_1 is the distribution function of $F(X)$ and K_2 is the distribution function of $G(Y)$ (see [37]). We say that Y is H-larger than X whenever $\tilde{K}_2(t) \leq \tilde{K}_1(t) \ \forall \ t \in \Re$, where $\tilde{K}_1$ is the distribution function of $F(W)$, $\tilde{K}_2$ is the distribution function of $G(W)$ and H is the distribution function of W (see [36]). The order we want to introduce will be denoted by $\prec^{K^F}$: we say that $X \prec^{K^F} Y$ whenever $\tilde{K}(t) \geq K(t) \ \forall \ t \in \Re$, where $\tilde{K}$ is the distribution function of $F(X)$ and K is the distribution function of $F(Y)$.

Going back to the test of equality of distributions, the alternative hypotheses can be rearranged to fit one of the one-dimensional scenarios: $H_1 : F(x) \prec G(x) \ \forall \ x$, $H_2 : G(x) \prec F(x) \ \forall \ x$, or $H_3 : G(x) \neq F(x)$ for some x, for some specified stochastic order $\prec$. As mentioned earlier, we will focus on H_1 and H_2.

Another problem arises when, following the one-dimensional methodology, we want to define our new test statistic as the number of $G-$observations that we encounter before $X_{(k)}$. Because we can no longer order the observations, we need to find a replacement for the concept of order statistics or, more generally, stopping times. For this purpose we rely on stopping sets, which are the multidimensional generalization of stopping times.

A set $D \subseteq \Re^2_+$ is said to be a lower layer if for every $t \in \Re^2_+$, the event $\{t \in D\}$ implies that $[0,t] \subseteq D$, where $[0,t] = \{x \in \Re^2_+ : 0 \leq x \leq t\}$. A map ξ with the lower layers as range is said to be a stopping set with respect to the filtration $\mathcal{F}$ if $\{t \in \xi\} \in \mathcal{F}_t \ \forall \ t \in \Re^2_+$ (see [29]). Thus, it can be clearly seen that the stopping set we choose will not only depend on all or some of the $F-$observations, but also on the kind of data structure -which is given by the filtration ($\mathcal{F} = \{\mathcal{F}_{(s,t)} : (s,t) \in \Re^2_+\}$)- that the experiment generates.

Again, the fact that precedence tests are widely used in life testing and clinical trials leads us to consider two types of filtrations: the minimal and the product. The minimal filtration is the $\sigma-$field generated by the bivariate empirical distribution function F_n, which is in turn generated by $X_1, X_2, \ldots, X_n$(cf. chapter 2). The simple structure of this filtration gives rise to stopping sets that are defined by the contours of the empirical distribution F_n. The product filtration is the $\sigma-$field generated by the indicator functions of each component of our data (cf. chapter 2). This filtration gives rise to several interesting stopping sets -in addition to those defined by the minimal filtration-, such as the set $[0,(X^1_{(i)}, X^2_{(j)})]$, where $X^s_{(l)}$ denotes the lth order statistic of the sth coordinate of X.

Once we have agreed on the type of stochastic order we are using and a stopping set is chosen -or we have identified the kind of stopping set we can use given the filtration used to obtain the data-, we can go back to our original problem and define a test statistic. If $X_1, X_2, \ldots, X_n$ are i.i.d. F and $Y_1, Y_2, \ldots, Y_m$ are i.i.d. G, our test statistic will be now based on the process

$$G_m(\xi_n(\cdot)) = \frac{1}{m}\sum_{i=1}^{m} I\{Y_i \in \xi_n(\cdot)\},$$

where the sets $\xi_n(\cdot)$ are appropriately chosen random sets depending on $X_1, X_2, \ldots, X_n$ parameterized in some way. The above leads to a sort of $p-p$ plot and to precedence tests on $\Re^2_+$ provided that $\xi_n(\cdot)$ is a stopping set, since we will see that $G_m(\xi_n(\cdot))$ is adapted to the underlying data structure; therefore, we will be able to base our test statistic on partial samples.

In this work we focus on two types of bivariate $p-p$ plots, their asymptotic behaviour, and their respective applications to tests of stochastic orders. We will see that, for example, if we are using the minimal filtration, we consider the statistic $G_m(\xi_p^{F_n})$ that gives rise to the one-dimensional $p-p$ process $\sqrt{m}[G_m(\xi_p^{F_n}) - G(\xi_p^F)]$, where $\xi_p^F = \{x \in \Re^2_+ : F(x-) < p\}$ for $p \in [0,1]$. Under $H_0 : F = G$, this one-dimensional $p-p$ process can be regarded as a two-sample version of Kendall's process (see [9]). If we use the product filtration instead, we consider the statistic $G_m \circ (F_n^{1-}(p), F_n^{2-}(q))$ that gives rise to the two-dimensional $p-p$ process $\sqrt{m}[G_m \circ$

$(F_n^{1-}(p), F_n^{2-}(q)) - G \circ (F_1^{-}(p), F_2^{-}(q))]$ for $(p,q) \in [0,1]^2$. Under $H_0 : F = G$, this two-dimensional $p-p$ process can be viewed as a two-sample version of the empirical copula process (see [22]).

This document is organized in the following way:

In Chapter 2 we will present the definitions of filtrations, stopping sets and stochastic orders on the line, as well as their respective extensions to the plane that will be used to construct our tests for bivariate data. Since the problem of comparing two distributions arises naturally in geographical data and clinical trials, we will be considering the minimal and the product filtrations.

In Chapter 3 we will first formally define the $p-p$ plots on which we will be basing our tests. Once we have defined them, we will develop some Glivenko-Cantelli type of results for those $p-p$ plots. These results, interesting by themselves, will be helpful when we start talking about the possible applications of our tests.

In Chapter 4 we will explore the asymptotic behaviour of $G_m(\xi_p^{F_n})$ -the first of the $p-p$ plots introduced in Chapter 3-, which is the appropriate statistic to study when we deal with data which generates the minimal filtration. We will also talk about specific applications of our test considering two scenarios: when the complete samples are available and when we deal only with partial samples.

In Chapter 5 we will study the asymptotic behaviour of $G_m(F_n^{1-}, F_n^{2-})$ -the second of our $p-p$ plots-, which is the appropriate statistic to explore when the available data generates the product filtration. As in the previous chapter, we will also discuss applications of this test considering the two previously mentioned scenarios.

It will be seen that copulas play an important role in the development of bivariate $p-p$ plots. For this reason, we have included an appendix chapter in which we have gathered some definitions and properties related to copulas. Although by no means a comprehensive chapter, it has everything about copulas that we make use of in the other chapters. In particular, we will be using the Farlie-Gumbel-Morgenstern (FGM) copula to illustrate the theory and techniques developed in this thesis.

Chapter 2

Filtrations, stopping sets and stochastic orders in $\Re^2_+$

In this chapter we will define the concepts of filtrations, stopping sets and stochastic orders on the line and we will make the corresponding extensions to the positive quadrant of the plane.

2.1 Filtrations

In this section we will discuss several types of filtrations associated with point processes on $\Re^2_+$ which will be used later to decide if certain random sets are indeed stopping sets. We start by stating the definition of a filtration on the line.

Definition 2.1.1 *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. $\{\mathcal{F}_t : t \in \Re_+\}$ is a filtration if:*

- *$\mathcal{F}_t \subseteq \mathcal{F}\ \forall\ t \in \Re_+$ and contains all the P-null sets.*
- *$s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t$.*
- *$\{\mathcal{F}_t\}$ is right continuous: $\mathcal{F}_t = \mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$.*

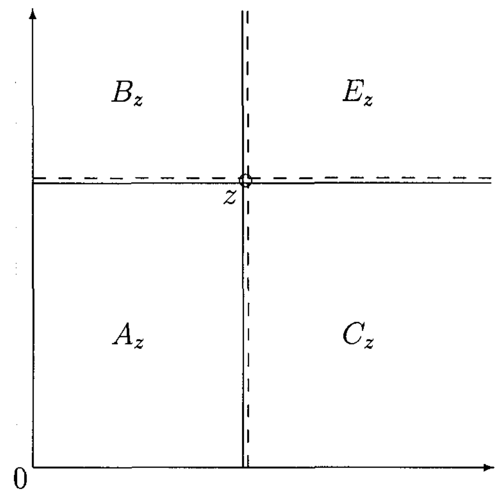

Figure 2.1: $A_z = [0, z_1] \times [0, z_2]$, $B_z = [0, z_1] \times (z_2, \infty)$, $E_z = (z_1, \infty) \times (z_2, \infty)$, $C_z = (z_1, \infty) \times [0, z_2]$.

To state the definition of a two-dimensional filtration, we will make use of the usual partial order on $\Re^2_+$: $z = (z_1, z_2) \leq (z_1^{'}, z_2^{'}) = z^{'}$ if and only if $z_1 \leq z_1^{'}$ and $z_2 \leq z_2^{'}$.

Definition 2.1.2 *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. $\{\mathcal{F}_z : z \in \Re^2_+\}$ is a filtration if:*

- *$\mathcal{F}_z \subseteq \mathcal{F}$ $\forall$ $z \in \Re^2_+$ and contains all the P-null sets.*
- *$z \leq z^{'} \Rightarrow \mathcal{F}_z \subseteq \mathcal{F}_{z^{'}}$.*
- *$\{\mathcal{F}_z\}$ is outer continuous: $\forall$ $z \in \Re^2_+$, $\mathcal{F}_z = \cap_n \mathcal{F}_{z_n}$ for any sequence $z_n \downarrow z$.*

Since we are working on the plane, all of our points will have two components; the notation for random variables will be $X = (X^1, X^2)$, and we will use $t = (t_1, t_2)$ for points. Also, we will be using Figure 2.1 to make the explanation of the information we have available for each filtration at time z a bit clearer.

Before we start to look at filtrations, we need a couple of definitions. Although these definitions can be found for a general space in [47], we will only state them here for $\Re^2_+$.

Definition 2.1.3 *A set* $\mathbf{X} = \{x_1, x_2, \ldots\}$ *of finite and unordered points in* $\Re^2_+$ *is said to be a configuration. A configuration* $\mathbf{X} \subseteq \Re^2_+$ *is said to be locally finite if it places at most a finite number of points in any bounded Borel set* $F \subseteq \Re^2_+$*. The family of all locally finite configurations will be denoted as* $\mathcal{N}^{lf} = \mathcal{N}^{lf}_{\Re^2_+}$.

Definition 2.1.4 *A point process* N *on* $\Re^2_+$ *is a mapping from a probability space* $(\Omega, \mathcal{F}, P)$ *into* $\mathcal{N}^{lf}_{\Re^2_+}$ *such that for all bounded Borel sets* $F \subseteq \Re^2_+$, $N(F)$*, the number of points falling in* F*, is a (finite) random variable.*

For now, we will start looking at the filtrations in the case where we only have a single jump point, X. Note that in the case of a single jump point we have that $N[0, z] = I\{X \in [0, z]\}$.

•Minimal filtration, one jump point

The minimal filtration is the $\sigma-$field generated by the indicator function of our single point:

$$\mathcal{F}^X_z = \sigma\{N[0, z'] : z' \leq z\} = \sigma\{I\{X \in [0, z']\} : z' \leq z\}.$$

Using this filtration we know at time z if our single observation is inside or outside of A_z. If X is in A_z, we know its exact location.

•Product filtration, one jump point

The product filtration is the product of the $\sigma-$fields generated on Ω by the indicator functions of the components of the time X^1 and X^2 of the jump X:

$$\mathcal{F}^p_z = \sigma\{I\{X^1 \leq z_1'\} : z_1' \leq z_1\} \bigotimes \sigma\{I\{X^2 \leq z_2'\} : z_2' \leq z_2\}.$$

Using the product filtration we know in which region X is located. In addition, referring to Figure 2.1, if X is in A_z we know both its components. If X is in B_z we know its first component and if X is in C_z we know its second component.

Now we will do the same kind of analysis in the case where instead of a single point we have a finite number of them, namely, $X_1, X_2, \ldots, X_n$.

•Minimal filtration, n jump points

When working with n random variables, the minimal filtration becomes:

$$\mathcal{F}_z = \sigma\{N[0, z'] : z' \leq z\} = \sigma\{\sum_{i=1}^{n} I\{X_i \in [0, z']\} : z' \leq z\}.$$

When using the minimal filtration we know how many observations lie in A_z and therefore how many lie outside A_z, but not which ones. For the observations in A_z we have their exact locations, but for the ones outside A_z we do not have any other information.

•Identifying filtration, n jump points

The identifying filtration, as suggested by its name, is the $\sigma-$field generated by the indicator functions of $X_1, X_2, \ldots, X_n$:

$$\mathcal{F}_z^I = \bigvee_{i=1}^{n} \mathcal{F}_z^{X_i} = \sigma\{I\{X_i \in [0, z'] : z' \leq z\}, i = 1, \ldots, n\}.$$

Note that if we only have one point, the minimal and the identifying filtrations are the same and this is the reason that we did not mention it before. With this filtration we know how many and which observations lie in A_z and therefore how many and which ones lie outside A_z. In addition, for the observations lying in A_z we also know their exact locations in the region.

•Product filtration, n jump points

For this work we will be using the product filtration associated with the identifying filtration. The product filtration is the product of the $\sigma-$fields generated by the indicator functions of each of the components of the observations $X_1, X_2, \ldots, X_n$:

$$\begin{aligned} \mathcal{F}_z^p &= \sigma\{I\{X_i^1 \leq z_1^{'}\} : z_1^{'} \leq z_1, i = 1, 2, \ldots, n\} \bigotimes \\ &\quad \sigma\{I\{X_j^2 \leq z_2^{'}\} : z_2^{'} \leq z_2, j = 1, 2, \ldots, n\}. \end{aligned}$$

This filtration is much larger than the minimal, since we know exactly which observations lie in the four parts of the plane (A_z, B_z, C_z and E_z). If they lie in A_z we know both their components, if they lie in B_z we know their first component and if they lie in C_z we know their second component. We have no further information on those points lying in E_z.

The choice of the filtration depends on the kind of experiment we are conducting. For example, if we are working with medical trials, where each of the components denote the times that the patients get a certain disease or symptom, the minimal filtration would be of little use. We would have only the following information: how many patients get both diseases or symptoms before a certain time, but we would not be able to identify the sick ones from the rest. If we were using the identifying filtration we would know which patients got both diseases by time t, but we would not be able to say how many or which ones got only one. If we decided to use the product filtration, we would have the same information as with the identifying filtration plus which patients got the first disease by time t_1 or the second disease by time t_2 and at what time.

On the other hand, suppose we are working with geographical data, such as the position of a certain kind of tree in a forest; here, each component of the random variables denotes a position coordinate (latitude and longitude). Then, the minimal filtration would be enough, since we do not need to identify one tree from another; the only thing we are interested in knowing are their positions. This shows that when we are conducting an experiment we always have to keep in mind the data structure behind it. Even if we are not performing the experiment ourselves, we must be able to identify the filtration used to conduct it.

We have almost finished the discussion on filtrations, except for one last thing: in order to show that what we have defined are indeed filtrations, we need to verify that they are outer continuous. Following the method used in [28], we start by proving a lemma that will help us to prove that the filtrations we defined before are indeed outer continuous.

Lemma 2.1.5 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots \subseteq \mathcal{F}$ be sub-σ-fields satisfying $\mathcal{F}_0 \subseteq \mathcal{F}_k$ and $\mathcal{F}_0 \cap H_k = \mathcal{F}_k \cap H_k \ \forall\ k \geq 1$, where $\{H_k\}$ is an increasing sequence in $\mathcal{F}$ such that $\cup_k H_k = \Omega$. Then $\mathcal{F}_0 = \cap_{k=1}^{\infty} \mathcal{F}_k$.*

Proof: i)$\mathcal{F}_0 \subseteq \cap_{k=1}^{\infty} \mathcal{F}_k$:
Since $\mathcal{F}_0 \subseteq \mathcal{F}_k \ \forall\ k \geq 1$, then $\mathcal{F}_0 \subseteq \cap_k \mathcal{F}_k$.

ii)$\cap_{k=1}^{\infty} \mathcal{F}_k \subseteq \mathcal{F}_0$:
Suppose $F \in \cap_{k \geq 1} \mathcal{F}_k$. Then $F \in \mathcal{F}_k$ for each $k \geq 1$.
Therefore, since $\mathcal{F}_0 \cap H_k = \mathcal{F}_k \cap H_k \ \exists\ F_k \in \mathcal{F}_0$ such that $F_k \cap H_k = F \cap H_k \ \forall\ k$. We will show that $F = \liminf F_k$.

ii.a)$F \subseteq \liminf F_k$:

$$\begin{aligned} F &= \cup_{n=1}^{\infty} (F \cap H_n) = \cup_{n=1}^{\infty} (F \cap (\cap_{k=n}^{\infty} H_k)) = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} (F \cap H_k) \\ &= \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} (F_k \cap H_k) = \liminf (F_k \cap H_k) \subseteq \liminf F_k. \end{aligned}$$

We conclude that $F \subseteq \liminf F_k$.

ii.b)$\liminf F_k \subseteq F$:
Take any $\omega \in \liminf F_k$. It follows that $\exists\ n$ such that $\omega \in \cap_{k=n}^{\infty} F_k$. Also, since $H_k \uparrow \Omega$, $\exists\ m$ such that $\omega \in H_k \ \forall\ k \geq m$.
Then we can conclude that whenever $k \geq \max(m, n)$, $\omega \in F_k \cap H_k$. Therefore, $\omega \in F \cap H_k$, and so $\omega \in F$, completing this part of the proof.

Putting together i) and ii) we get that $\liminf F_k = F \in \mathcal{F}_0$. $\diamond$

Now we can prove that the minimal filtration is outer continuous.

Proposition 2.1.6 *Let $\mathcal{F}_z$ denote the minimal filtration generated by a point process N. Then, the family $\{\mathcal{F}_z : z \in \Re^2_+\}$ is outer continuous.*

Proof: Let $\{z_n\} \in \Re^2_+$ be a decreasing sequence such that $[0, z] = \cap_n [0, z_n]$ and $\{\mathcal{F}_z\}$ the minimal filtration. Here, $z_n \downarrow z$, i.e. $z_{1,n} \downarrow z_1$ and $z_{2,n} \downarrow z_2$.

i)$\mathcal{F}_z \subseteq \cap_n \mathcal{F}_{z_n}$:
$[0, z_1] \times [0, z_2] = [0, z] \subseteq [0, z_n]\ \forall\ n$. Then $\mathcal{F}_z \subseteq \mathcal{F}_{z_n}\ \forall\ n$. It is straightforward that $\mathcal{F}_z \subseteq \cap_n \mathcal{F}_{z_n}$.

ii)$\cap_n \mathcal{F}_{z_n} \subseteq \mathcal{F}_z$:
For $z = (z_1, z_2)$ and $z' = (z'_1, z'_2)$, let $\text{dist}(z, z') = \max(|z_1 - z'_1|, |z_2 - z'_2|)$. For a closed subset $C \subseteq \Re^2_+$, and $t \in \Re^2_+$, let $\text{dist}(t, C) = \inf_{x \in C} \text{dist}(t, x)$. Then, define $\delta(\omega) = \min[\text{dist}(t, [0, z]) : t \in [0, z]^c, N_{\{t\}} = 1]$. If no such t exists, we define $\delta(\omega) = \infty$. Note that $\delta(\omega) > 0$ almost surely.

We can choose a subsequence $\{z_{n_k}\}$ such that $\text{dist}(z, z_{n_k}) = \max(z_{n_{k,1}} - z_1, z_{n_{k,2}} - z_2) < 1/k$.

Now let $H_k = \{\omega : 1/k < \delta(\omega)\}$. It is easy to see that if $\omega \in H_k$, then all the jump points that belong to $[0, z_{n_k}]$ also belong to $[0, z]$. Since $H_k = \{\omega : \delta(\omega) > 1/k\}$ and $\Omega = \{\omega : \delta(\omega) > 0\}$, we have that $H_k \uparrow \Omega$. Since $\{z_n\}$ is decreasing and $\{z_{n_k}\}$ is a subsequence of $\{z_n\}$, we also have that $\cap_n \mathcal{F}_{z_n} = \cap_k \mathcal{F}_{z_{n_k}}$.

Now we look at the generators of the minimal filtration. Take any $[0,a] \subseteq [0,z_{n_k}]$. On H_k, $N([0,z_{n_k}]\backslash[0,z]) = N_{z_{n_k}} - N_z = 0$ holds. Then,

$$\begin{aligned} N[0,a] &= N([0,a]\cap[0,z]) + N([0,a]\cap([0,z_{n_k}]\backslash[0,z])) \\ &\leq N([0,a]\cap[0,z]) + N_{z_{n_k}} - N_z = N([0,a]\cap[0,z]) \\ &\leq N[0,a]. \end{aligned}$$

Therefore, $N[0,a] = N([0,a]\cap[0,z])$ on H_k.

Then $\forall\ [0,a] \subseteq [0,z_{n_k}]$, we have that $N_a I_{H_k} = N([0,a]\cap[0,z])I_{H_k}$ and thus, $H_k \cap \mathcal{F}_{z_{n_k}} = H_k \cap \mathcal{F}_z$. It follows from Lemma 2.1.5 that $\cap_n \mathcal{F}_{z_n} \subseteq \mathcal{F}_z$.

We complete the proof by putting together i) and ii). $\diamond$

Now we move on to the identifying filtration. Remember that the identifying filtration $\{\mathcal{F}_z^I\}$ is generated by the indicator functions of the observations inside the rectangles of the form $[0,z']$, $z' \leq z$.

If we have n observations, we can identify this generator with an n-dimensional vector whose entries are either zero or one. The *ith* component of the vector is one if the *ith* observation is inside a rectangle $[0,z]$, and zero otherwise. For example, the vector $(0,1,1,1,0,1,0,0,0,1,0)$ for an arbitrary rectangle $[0,z_a]$ would mean that from our 11 observations, we have 5 inside $[0,z_a]$, namely Y_2,Y_3,Y_4,Y_6, and Y_{10}. Let this vector be denoted by S_a. The same proof used to verify the outer continuity of the minimal filtration can be used to verify that of the identifying filtration. It is enough to substitute the generators N_a by our new generators S_a.

Finally, we take a look at the outer continuity of the product filtration.

Proposition 2.1.7 *The product filtration* $\mathcal{F}_z^p = \sigma\{I\{X_i^1 \leq z_1'\} : z_1' \leq z_1, i = 1,2,\ldots,n\}$
$\bigotimes \sigma\{I\{X_j^2 \leq z_2'\} : z_2' \leq z_2, j = 1,2,\ldots,n\}$ *is outer continuous.*

Proof: Let $\{z_n\} \in \Re^2_+$ be a decreasing sequence such that $[0, z] = \cap_n[0, z_n]$. Again, $z_n \downarrow z$, i.e. $z_{1,n} \downarrow z_1$ and $z_{2,n} \downarrow z_2$.

i)$\mathcal{F}^p_z \subseteq \cap_n \mathcal{F}^p_{z_n}$:
$[0, z_1] \times [0, z_2] = [0, z] \subseteq [0, z_n] \; \forall \, n$. Then $\mathcal{F}^p_z \subseteq \mathcal{F}^p_{z_n} \forall \, n$ and it follows that $\mathcal{F}^p_z \subseteq \cap_n \mathcal{F}^p_{z_n}$.

ii)$\cap_n \mathcal{F}^p_{z_n} \subseteq \mathcal{F}^p_z$:
We will be needing two 1-dimensional processes denoted by N^1 and N^2 defined in the following way:

$$\begin{aligned} N^1(t_1) &= \sum_{i=1}^{n} I\{X_i^1 \leq t_1\} \text{ and} \\ N^2(t_2) &= \sum_{j=1}^{n} I\{X_j^2 \leq t_2\}. \end{aligned}$$

Now we define $\delta(\omega) = \min[\text{dist}(t, z) : t \in [0, z]^c, N^1_{\{t_1\}} = 1 \text{ and } N^2_{\{t_2\}} = 1]$, where the distance we use is that defined in proposition 2.1.6. If there are no such t_1 and t_2 we define $\delta(\omega) = \infty$.

We can then choose a subsequence $\{z_{n_k}\}$ such that $\text{dist}(z, z_{n_k}) = \max(z_{n_{k,1}} - z_1, z_{n_{k,2}} - z_2) < 1/k$.

Let $H_k = \{\omega : 1/k < \delta(\omega)\}$. Then all the jump points that belong to $[0, z_{n_k}]$ also belong to $[0, z]$. Also, as before, $H_k \uparrow \Omega$, and $\cap_n \mathcal{F}^p_{z_n} = \cap_k \mathcal{F}^p_{z_{n_k}}$.

Now we look at the generators of the product filtration. Take any $[0, a] \subseteq [0, z_{n_k}]$. Then $\forall \, i$, $I\{X_i^1 \leq a_1\} = \begin{cases} I\{X_i^1 \leq a_1\} & \text{if } a_1 < z_1 \\ I\{X_i^1 \leq z_1\} & \text{if } a_1 \geq z_1 \end{cases}$ on H_k.

Thus, $I\{X_i^1 \leq a\} I_{H_k} = \begin{cases} I\{X_i^1 \leq a_1\} I_{H_k} & \text{if } a_1 < z_1 \\ I\{X_i^1 \leq z_1\} I_{H_k} & \text{if } a_1 \geq z_1 \end{cases} \quad \forall \;\; a < z_{n_k}$.

Therefore, $\forall \, [0, a] \subseteq [0, z_{n_k}]$, we have that $I\{X_i^1 \leq a_1\} I_{H_k} = I\{X_i^1 \leq \min(a_1, z_1)\} I_{H_k}$. The same argument applies to every second component, i.e. $\forall \; [0, a] \subseteq [0, z_{n_k}]$, $I\{X_j^2 \leq a_2\} I_{H_k} = I\{X_j^2 \leq \min(a_2, z_2)\} I_{H_k}$. Therefore, $H_k \cap \mathcal{F}^p_{z_{n_k}} = H_k \cap \mathcal{F}^p_z$ and it follows from Lemma 2.1.5 that $\cap_n \mathcal{F}^p_{z_n} \subseteq \mathcal{F}^p_z$.

Putting together i) and ii) we complete the proof. ◇

To summarize, in what follows we will be focusing on two filtrations associated with point processes: the minimal and the product filtration. The first data structure arises naturally from geographical observations and the second from clinical trials.

2.2 Stopping times and stopping sets

This section deals with the stopping sets corresponding to each of the filtrations we talked about in the last section. First we need some notation and definitions. We begin with the definition of a stopping time on $\Re_+$.

Definition 2.2.1 *Let $\{\mathcal{F}_t : t \in [0,k]\}$, $k \in \Re_+$ be a filtration. A map $\tau : \Omega \to \Re_+$ is a stopping time if $\{\tau \le t\} \in \mathcal{F}_t \ \forall\ t$.*

Since we required our filtrations to be right continuous, we could have defined τ to be a stopping time if $\{t \le \tau\} \in \mathcal{F}_t$. This issue is addressed in the following lemma.

Lemma 2.2.2 *A random time τ is a stopping time if and only if $\{t \le \tau\} \in \mathcal{F}_t \ \forall\ t$.*

Proof: i)τ is a stopping time $\Rightarrow \{t \le \tau\} \in \mathcal{F}_t \ \forall\ t$:
Let τ be a stopping time, then $\{t \le \tau\} = \{\tau < t\}^C = \{\cup_n \{\tau \le t - \frac{1}{n}\}\}^C \in \mathcal{F}_t$.

ii)$\{t \le \tau\} \in \mathcal{F}_t \ \forall\ t \Rightarrow \tau$ is a stopping time:
For all m we have that

$$\begin{aligned} \{\tau \le t\} &= \cap_n \{\tau - 1/n < t\} \\ &= \cap_n \{\tau < t + 1/n\} \\ &= \cap_{n \ge m} \{\tau < t + 1/n\} \in \mathcal{F}_{t+1/m}. \end{aligned}$$

Thus, $\{\tau \le t\} \in \cap_m \mathcal{F}_{t+1/m} = \mathcal{F}_{t^+}$, and the result follows directly from the right continuity of the filtrations. ◇

Before proceeding to the two-dimensional case, let us make a brief analysis about what happens when we have observations on the line. Suppose we have n observations on $\Re_+$ denoted as $X_1, X_2, \ldots, X_n$. For this scenario we would have several stopping times. For example, since $\Re_+$ is a totally ordered set, we would have no problem working with the smallest of the observations because it is a stopping time with respect to the minimal filtration generated by the point process $N(t) = \sum_{i=1}^{n} I\{X_i \leq t\}$. Likewise, the kth order statistic $X_{(k)} = \inf\{t : N(t) \geq k\}$ is also a stopping time. As we pointed out earlier (see the introductory chapter), there is no natural analogue for order statistics in higher dimensions, but the situation can be handled if instead of stopping times we work with random sets. As an illustrative example, we could have defined a stopping set on $\Re_+$ in the following way: $[0, \tau]$ is a stopping set if the event $\{t \in [0, \tau]\} \in \mathcal{F}_t \; \forall \; t$. We see from Lemma 2.2.2 that $[0, \tau]$ is a stopping set if and only if τ is a stopping time. Therefore, continuing with the example of the smallest observation, it is straightforward that the smallest stopping set will be, in fact, the intersection of all our stopping sets, as defined in the preceding lines.

We will now introduce some terminology that will help to define formally the stopping sets on $\Re_+^2$. Let $z = (z_1, z_2)$ and $z' = (z_1', z_2')$ be two points on the plane. We will write $z \ll z'$ if and only if $z_1 < z_1'$ and $z_2 < z_2'$. Conversely, we will write $z \gg z'$ whenever $z_1 > z_1'$ and $z_2 > z_2'$. Also, let $A_z = \{z' : z' \leq z\}$ be as in Figure 2.1 and $E_z = \{z' : z' \gg z\}$ and $D_z = E_z^c$, as in Figure 2.2.

The definition of stopping sets, as well as all concepts needed to construct this definition, were developed in [29].

Definition 2.2.3 *A set $D \subseteq \Re_+^2$ is called a lower layer if the event $\{z \in D\}$ implies that $[0, z] = A_z \subseteq D \; \forall \; z$. The set of all the lower layers will be denoted as $\mathcal{L}$.*

Definition 2.2.4 *Let $\{\mathcal{F}_t : t \in \Re_+^2\}$ be a filtration. A map $\xi : \Omega \to \mathcal{L}$ is said to be a stopping set with respect to $\mathcal{F}$ if $\{t \in \xi\} \in \mathcal{F}_t \; \forall \; t \in \Re_+^2$.*

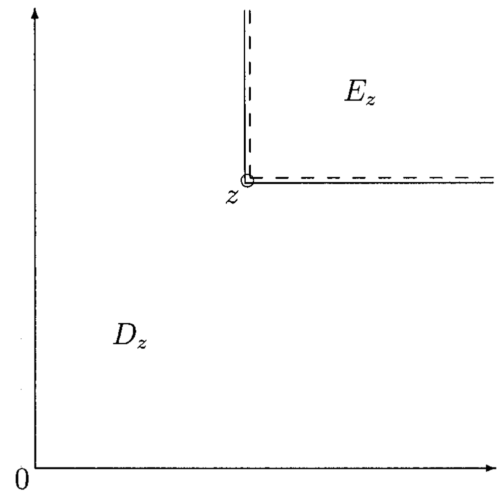

Figure 2.2: $D_z = A_z \cup B_z \cup C_z$, $E_z = D_z^C$.

Some authors have defined and worked with stopping lines instead of stopping sets. These two definitions are equivalent in the sense that stopping lines are the boundaries of stopping sets. For notational convenience, we will use sets instead of lines.

Definition 2.2.5 *Let $\{\mathcal{F}_t : t \in \Re_+^2\}$ be a filtration. A map $(\tau_1, \tau_2) : \Omega \to \Re_+^2$ is a stopping time if $\{\tau \leq t\} \in \mathcal{F}_t \ \forall\ t$.*

Observe that since $\Re_+^2$ is only partially ordered, we do not have the analogue of Lemma 2.2.2 in that if $\{t \leq \tau\} \in \mathcal{F}_t \ \forall\ t$ then τ is necessarily a stopping time, but the converse is no longer true. We address this issue in the following Lemma.

Lemma 2.2.6 *i)If $\{t \leq \tau\} \in \mathcal{F}_t \ \forall\ t \in \Re_+^2$ then τ is a stopping time (i.e. the event $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable $\forall\ t$).*
ii)If $\{\tau \leq t\} \in \mathcal{F}_t \ \forall\ t \in \Re_+^2$ then the event $\{t \leq \tau\}$ is not necessarily $\mathcal{F}_t$-measurable.

Proof: i)Let T' denote the points $t' \in \Re_+^2$ with dyadic coordinates and assume that $\{t \leq \tau\} \in \mathcal{F}_t \ \forall\ t$. We need to show that $\{\tau \in A_t^C\}^C = \{\tau \leq t\} \in \mathcal{F}_t \ \forall\ t$. Since T' is a dense set in $\Re^2_+$, we have that $\forall\ M \in N$:

$$\begin{aligned}\{\tau \in A_t^C\} &= \cup_{s\in A_t^C \cap T'}\{s \leq \tau\} \\ &= \cup_m \cup_{s\in A_t^C\cap T', dist(s,A_t)\leq \frac{1}{m}} \{s \leq \tau\} \\ &= \cup_{m\geq M} \cup_{s\in A_t^C\cap T', dist(s,A_t)\leq \frac{1}{m}} \{s \leq \tau\} \in \mathcal{F}_{t+\frac{1}{M}}.\end{aligned}$$

Then, $\{\tau \leq t\} \in \cap_M \mathcal{F}_{t+\frac{1}{M}} = \mathcal{F}_t$, and τ is a stopping time.

ii)Through $\mathcal{F}_t$ at time t we know whether $\tau \in A_t$ (and therefore its location) or if $\tau \in A_t^C$. However, if $\tau \in A_t^C$ we have no way of knowing whether τ is in E_t or in $B_t \cup C_t$ (cf Figure 2.1). As an example, consider the single jump process with the minimal filtration. Letting $\tau = X$, we see that $\{t \leq \tau\}$ is not $\mathcal{F}_t$-measurable. $\diamond$

At this point the natural question to ask is what kind of sets will be stopping sets for the filtrations we defined in the previous section. A couple of candidates would be A_τ and D_τ, where τ is a stopping time. We have the following two lemmas in this respect.

Lemma 2.2.7 *For a stopping time $\tau = (\tau_1, \tau_2)$, D_τ is always a stopping set.*

Proof: Note that $\{\tau_1 < t_1, \tau_2 < t_2\} = \bigcup_{r_1,r_2\in Q; r_1<t_1, r_2<t_2}\{\tau_1 \leq r_1, \tau_2 \leq r_2\} \in \mathcal{F}_t$. Thus, it is true that $\{t \in D_\tau\} = \{t \in E_\tau\}^c = \{\tau_1 < t_1, \tau_2 < t_2\}^c \in \mathcal{F}_t$. Then, D_τ is a stopping set. $\diamond$

Although the random set D_τ behaves nicely, we have from Lemma 2.2.6 that A_τ may not be a stopping set even if τ is a stopping time. However, in some special cases A_τ will be a stopping set.

Lemma 2.2.8 *Let $\mathcal{F}^p_z$ denote the product filtration generated by the random variables $X_1, X_2, \ldots, X_n$. Also, let $X^1_{(i)}$ be the ith order statistic from $X^1_1, X^1_2, \ldots, X^1_n$ and $X^2_{(j)}$ be the jth order statistic from $X^2_1, X^2_2, \ldots, X^2_n$. If $\tau = (X^1_{(i)}, X^2_{(j)})$, then τ is a stopping time and A_τ is a stopping set with respect to $\mathcal{F}^p_z$.*

Proof: Recall the notation of Lemma 2.1.7: $N^m(t_m) = \sum_{i=1}^n I\{X_i^m \leq t_m\}$, $m = 1, 2$.

First we note that, since each of the indicators inside the sums belong to $\mathcal{F}_t^p$, we have that:

$$\begin{aligned}
&\{N^1[0,t_1) \leq i-1\} \bigcap \{N^2[0,t_2) \leq j-1\} \\
= &\ \{(\sum_{k=1}^n I\{X_k^1 < t_1\}) \leq i-1\} \bigcap \{(\sum_{l=1}^n I\{X_l^2 < t_2\}) \leq j-1\} \\
= &\ \{ \bigcap_{r \in Q, r < t_1} \{\sum_{k=1}^n I\{X_k^1 \leq r\} \leq i-1\}\} \bigcap \{ \bigcap_{s \in Q, s < t_2} \{\sum_{l=1}^n I\{X_l^2 \leq s\} \leq j-1\}\} \\
\in &\ \mathcal{F}_t^p.
\end{aligned}$$

i)τ is a stopping time:

$$\begin{aligned}
\{\tau \leq t\} &= \{X_{(i)}^1 \leq t_1\} \bigcap \{X_{(j)}^2 \leq t_2\} \\
&= \{N^1[0,t_1] \geq i\} \bigcap \{N^2[0,t_2] \geq j\} \\
&= \{N^1[0,t_1] \leq i-1\}^c \bigcap \{N^2[0,t_2] \leq j-1\}^c \in \mathcal{F}_t^p.
\end{aligned}$$

ii)A_τ is a stopping set:

$$\begin{aligned}
\{t \in A_\tau\} &= \{t_1 \leq X_{(i)}^1\} \bigcap \{t_2 \leq X_{(j)}^2\} \\
&= \{N^1[0,t_1) \leq i-1\} \bigcap \{N^2[0,t_2) \leq j-1\} \in \mathcal{F}_t^p.
\end{aligned}$$

◇

When we use the minimal filtration, neither the times $\tau = (X_{(i)}^1, X_{(j)}^2)$ as defined in the previous lemma nor the jump times are, in general, stopping times. The jump times $X_i = (X_i^1, X_i^2)$ are stopping times with respect to the identifying and the product filtrations, but $(X_{(i)}^1, X_{(j)}^2)$ is a stopping time only with respect to the product filtration. As it was mentioned before, A_τ may not be a stopping set with respect to certain filtrations and this is indeed the case for the minimal and the identifying filtration. Nevertheless, there exist other sets that are stopping sets with respect to the minimal filtration and therefore with respect to the larger filtrations as well. These sets are the natural separation sets $L_k = \{z : N(A_{z-}) \leq k-1\} = \{z : N[0,z) \leq k-1\}$, $k = 1, 2, \dots, n$. If we pick any $z \in L_1$ the rectangle $A_{z-} = [0, z)$ will contain no observations, if we pick $z \in L_2$ the rectangle $[0, z)$ will contain at most one observation, and in general, if we pick $z \in L_k$ the rectangle A_{z-} will contain at most $k-1$ observations. It is easy to see that $\{z \in L_k\} \in \mathcal{F}_z$, and therefore, L_k is a stopping set.

Comment 2.2.9 The above mentioned filtrations and stopping sets are defined on $\Re^2_+$. This allows us to define a type of precedence test based on partial data.

As we mentioned before (see the introductory chapter) our test statistic will be based on a process of the form $G_m(\xi_n(\cdot)) = \frac{1}{m}\sum_{i=1}^m I\{Y_i \in \xi_n(\cdot)\}$, where the sets ξ_n are stopping sets depending on $X_1, X_2, \dots, X_n$. If our data generates the product filtration, two appropriate processes to base our test statistic on would be, for example, $G_m(A_\tau)$ or $G_m(D_\tau)$, where $\tau = (X^1_{(i)}, X^2_{(j)})$. If our data generates the minimal filtration, it would be appropriate to base our test statistic on the process $G_m(L_k)$, where the sets L_k, as defined before, are the empirical counterparts of the contours of F.

2.3 Stochastic orders

In this last section we will introduce a new stochastic order which will be appropriate for when we deal with geographical data (and the minimal filtration). We will also present several extensions of the stochastic order $F < G$ to distributions on $\Re^2_+$ which will be used for when we deal with clinical trials (and the product filtration). In the latter case, we will also give a couple of examples to illustrate why we need to have different stopping sets for different kinds of experiments.

We start by reviewing the Kendall stochastic order, which can be found in [37].

Definition 2.3.1 *Let $X = (X^1, X^2)$ and $Y = (Y^1, Y^2)$ be random variables with dis-*

tribution functions F and G, respectively. Also, let $\tilde{H}$ and H denote the distribution functions of $F(X)$ and $G(Y)$. We say that X is less than Y in the Kendall stochastic order ($X \prec_K Y$) if and only if $\tilde{H}(t) \geq H(t)\ \forall\ t \in \Re$.

To clearly see how this order relates to the one we will be defining later on to use with geographical data, we need some notation. Let K be any bivariate distribution function for which we have a random sample of size l. The sets ξ_t^K and its empirical counterpart $\xi_t^{K_l}$ are defined as $\xi_t^K = \{(x,y) : K(x^-, y^-) < t\}$ and $\xi_t^{K_l} = \{(x,y) : K_l(x^-, y^-) < t\}$. Then, $\tilde{H}(t) = P(F(X) \leq t) = P(X \in \xi_t^F) = F(\xi_t^F)$ and $H(t) = P(G(Y) \leq t) = P(Y \in \xi_t^G) = G(\xi_t^G)$. It follows that $X \prec_K Y$ if and only if $F(\xi_t^F) \geq G(\xi_t^G)\ \forall\ t \in \Re$.

We now turn our attention to the H-larger stochastic order, which was developed in [36].

Definition 2.3.2 *Let $X = (X^1, X^2)$, $Y = (Y^1, Y^2)$ and $W = (W^1, W^2)$ be random variables with distribution functions F,G and H respectively. Also, let $\tilde{K}_1$ and $\tilde{K}_2$ denote the distribution functions of $F(W)$ and $G(W)$. We say that Y is H-larger than X if and only if $\tilde{K}_2(t) \leq \tilde{K}_1(t)\ \forall\ t \in \Re$.*

For any bivariate distribution K, let ξ_t^K and $\xi_t^{K_l}$ be defined as before. Then, $\tilde{K}_1(t) = P(F(W) \leq t) = P(W \in \xi_t^F) = H(\xi_t^F)$ and $\tilde{K}_2(t) = P(G(W) \leq t) = P(W \in \xi_t^G) = H(\xi_t^G)$. It follows that Y is H-larger than X if and only if $H(\xi_t^G) \leq H(\xi_t^F)\ \forall\ t \in \Re$.

What we want to do is define a stochastic order similar to the ones defined above. To do it we will apply an F-transformation on the random variables X and Y, namely, $\hat{X} = F(X)$ and $\hat{Y} = F(Y)$. Then, if $\tilde{K}$ and K denote the distribution functions of $\hat{X}$ and $\hat{Y}$, we have that $\tilde{K}(t) = P(\hat{X} \leq t) = P(F(X) \leq t) = P(X \in \xi_t^F) = F(\xi_t^F)$ and $K(t) = P(\hat{Y} \leq t) = P(F(Y) \leq t) = P(Y \in \xi_t^F) = G(\xi_t^F)$. Now that the similarities with the two orders described previously are obvious, we can define the stochastic order we will use.

Definition 2.3.3 *Let X and Y be two-dimensional random variables with distribution functions F and G, and let $\tilde{K}$ and K denote the distribution functions of $\hat{X} = F(X)$ and $\hat{Y} = F(Y)$, respectively. We say that X is less than Y in the F-transformation stochastic order ($X \prec^{K^F} Y$) if and only if $\tilde{K}(t) \geq K(t)\ \forall\ t \in \Re$.*

Having defined the stochastic order we will use when the data we use is filtered by the minimal filtration, we can see that our test statistic will be based on the empirical $p-p$ plot $G_m(\xi_t^{F_n})$; its asymptotic behaviour will be studied in chapter 4.

Now we turn our attention to the type of stochastic order that is appropiate when the product is the underlying filtration. Recall that $\forall\ t = (t_1, t_2)$, we defined the sets $A_t = [0, t_1] \times [0, t_2]$, $E_t = (t_1, 1] \times (t_2, 1]$ and $D_t = E_t^c = \{[0, t_1] \times [0, 1]\} \cup \{[0, 1] \times [0, t_2]\}$ (see Figure 2.1 and Figure 2.2). We will be using this notation as we describe how the multivariate stochastic orders follow naturally from the usual univariate stochastic order. This extension can be found in [41].

Definition 2.3.4 *A set $U \subseteq \Re^d$ is called an upper set if $s \in U$ whenever $s \geq t$ and $t \in U$. In $\Re$, U is an upper set if and only if it is of the form (c, ∞) or $[c, \infty)$, $c \in \Re$.*

Definition 2.3.5 *Let X and Y be continuous random variables with univariate distribution functions F and G respectively. We say that X is smaller than Y in the usual stochastic order (denoted $X \leq_{st} Y$) whenever the following equivalent statements hold:*

(1) $P(X > u) \leq P(Y > u) \forall\ u \in (-\infty, \infty)$.

(2) $P(X \leq u) \geq P(Y \leq u) \forall\ u \in (-\infty, \infty)$.

(3) $P(X \geq u) \leq P(Y \geq u) \forall\ u \in (-\infty, \infty)$.

(4) $P(X \in U) \leq P(Y \in U) \forall$ upper set $U \subseteq (-\infty, \infty)$.

(5) $E(I_U(X)) \leq E(I_U(Y)) \forall$ upper set $U \subseteq (-\infty, \infty)$.

Proof:(that the statements are equivalent)

$1 \Rightarrow 2$)Statement 1 is saying that the random variable X is less likely to take large values than the random variable Y, i.e. that X is more likely to take small values than Y, which is exactly what statement 2 says.

$2 \Rightarrow 1$)We have that $\forall\ u \in (-\infty, \infty)$,

$$\begin{aligned} P(X \leq u) \geq P(Y \leq u) &\Rightarrow 1 - P(X > u) \geq 1 - P(Y > u) \\ &\Rightarrow -P(X > u) \geq -P(Y > u) \\ &\Rightarrow P(X > u) \leq P(Y > u). \end{aligned}$$

$1 \Leftrightarrow 3$)This is straightforward because we required F and G to be continuous.

$4 \Rightarrow 1$)Take any upper set U, it will be of the form (u, ∞) or $[u, \infty)$.

$$\begin{aligned} i)P(X \in U) \leq P(Y \in U) &\Rightarrow P(X \in (u, \infty)) \leq P(Y \in (u, \infty)) \\ &\Rightarrow P(X > u) \leq P(Y > u). \end{aligned}$$

$$\begin{aligned} ii)P(X \in U) \leq P(Y \in U) &\Rightarrow P(X \in [u, \infty)) \leq P(Y \in [u, \infty)) \\ &\Rightarrow P(X \geq u) \leq P(Y \geq u) \\ &\Rightarrow P(X > u) \leq P(Y > u). \end{aligned}$$

$3 \Rightarrow 4$) i)For every upper set of the form $[u, \infty)$, we have that

$$\begin{aligned} P(X \geq u) \leq P(Y \geq u) &\Rightarrow P(X \in [u, \infty)) \leq P(Y \in [u, \infty)) \\ &\Rightarrow P(X \in U) \leq P(Y \in U). \end{aligned}$$

ii)For every upper set of the form (u, ∞), we have that

$$\begin{aligned} P(X \geq u) \leq P(Y \geq u) &\Rightarrow P(X \in [u, \infty)) \leq P(Y \in [u, \infty)) \\ &\Rightarrow P(X \in (u, \infty)) \leq P(Y \in (u, \infty)) \\ &\Rightarrow P(X \in U) \leq P(Y \in U). \end{aligned}$$

$4 \Leftrightarrow 5$)For any upper set U we have that

$$P(X \in U) \leq P(Y \in U) \Leftrightarrow E(I\{X \in U\}) \leq E(I\{Y \in U\})$$
$$\Leftrightarrow E(I_U(X)) \leq E(I_U(Y)).$$

◇

It is now our aim to extend the above definition from $\Re$ to $\Re^2$, and in particular to $\Re^2_+$. In what follows X and Y will be random variables with bivariate distribution functions F and G respectively. From the very definition of an upper set, we have that the natural extensions of (4) and (5) are:

$$(4') \quad P(X \in U) \leq P(Y \in U) \forall \text{ upper set } U \subseteq \Re^2, \text{ and}$$
$$(5') \quad E(I_U(X)) \leq E(I_U(Y)) \forall \text{ upper set } U \subseteq \Re^2.$$

If one of these conditions hold, we say that X is smaller than Y in the usual (multivariate) stochastic order (denoted $X \leq_{st} Y$). Or if we prefer to define it in terms of the distribution functions F and G, the extension would be $F(U) \leq G(U)$ $\forall$ upper sets $U \subseteq \Re^2$.

If we extend (2), we say that X is smaller than Y in the lower orthant order (denoted $X \leq_{lo} Y$) if

$$(2') \quad P(X \leq t) \geq P(Y \leq t) \forall\, t \in \Re^2, \text{or}$$
$$(2') \quad P(X \in A_t) \geq P(Y \in A_t) \forall\, t \in \Re^2_+.$$

Note that the second inequality applies to $\Re^2_+$-valued random variables only. In terms of F and G, $X \leq_{lo} Y$ if $F(t) \geq G(t)$ $\forall\, t \in \Re^2$. It is easy to see that, when F and G have support $\Re^2_+$, the appropriate region to work with the lower orthant order is A_t.

If we extend (1) (or (3)), we say that X is smaller than Y in the upper orthant order (denoted $X \leq_{uo} Y$) if

$$(1') \quad P(X > t) \leq P(Y > t) \forall\, t \in \Re^2, \text{or}$$
$$(1') \quad P(X \in E_t) \leq P(Y \in E_t) \forall\, t \in \Re^2, \text{or}$$
$$(1') \quad P(X \in D_t) \geq P(Y \in D_t) \forall\, t \in \Re^2_+.$$

Note that the last inequality applies to $\Re^2_+$-valued random variables only. In terms of distributions F and G with support $\Re^2_+$, $X \leq_{uo} Y$ if $F(D_t) \geq G(D_t)$ (or $F(E_t) \leq G(E_t)$) holds $\forall\ t \in \Re^2_+$. Since $E_t^c = D_t$, the choice between these two sets when we use the upper orthant order will depend on the specific kind of data we have.

Comment 2.3.6 Note that, since E_t is an upper set, $X \leq_{st} Y$ implies that $X \leq_{uo} Y$. However, the rest of our stochastic orders are not related.

Focusing on the upper and lower orthant orders, we now describe particular situations in which our test could be useful, along with the stopping sets and the stochastic order to be used.

•Hypertension is a health condition in which the force of blood against the artery walls is too strong; symptoms include headaches, visual problems and nausea. Suppose that we are asked to prove that a new drug is effective against hypertension. We design the experiment such that all the participants involved will have high blood pressure and related symptoms. Half of them will be taking the new drug and they will act as our F sample, the rest will be taking placebo and they will act as our G sample (the so-called control group). For both groups we record the observations as follows: the first component will be the time when their blood pressure reaches or drops below an already fixed level, and the second component will be the time when all related symptoms have disappeared for at least a fixed amount of time. We will follow the experiment until we can record both times for 20 people in the F group. If indeed the drug is working, we expect to have more observations from the sample F than from G before the end of the experiment, and we will reject H_0 to conclude that $F \geq G$ ($X \leq_{lo} Y$). In this particular case it makes sense to use the A region and the lower orthant order since the components of our observations will be recordings of time and we will follow the experiment from time zero.

•Leukemia is a disease that makes the body produce abnormal white cells which destroy all the other healthy cells (white, red and plaquettes); symptoms usually include fever, headaches, weakness and night sweats. Suppose now that we are asked to prove that a new drug is better than an existing one (both combined with chemotherapy). The experiment is designed so that all our patients will have leukemia and related symptoms. A third of them will be taking the new drug (they will act as our F sample) and the rest will be taking the existing drug (they will be our G sample). For both groups we record the observations as follows: the first component will be the time when their healthy cell count reaches or surpasses a fixed level, and the second component will be the time when all related symptoms have disappeared for at least a fixed amount of time. Suppose the patients go through a series of chemotherapy sessions followed by a period of rest, all while taking the drug. The chemotherapy may cause headaches and weakness, so the second component of the observations cannot be measured with accuracy during the sessions and shortly after. To accommodate this condition we will only monitor the patients during the last part of the experiment. The patients who have an acceptable cell count and have no symptoms prior to the first check-up will be assumed to have the illness in remission during the non-check-up time. If indeed the new drug is better than the existing one we expect to see fewer patients taking the former during our next check-ups, and therefore to reject H_0. In this case it makes sense to use the E region and the upper orthant order since we cannot have any observations before the first check-up and their components are measuring times of occurrence.

•Suppose a new vaccine for hepatitis is found. The ingredients of the vaccine lead us to think that patients might exhibit fatigue and/or nausea in the month after receiving the vaccine. Also, if the patient exhibits one of them, it is highly likely that he or she will exhibit the other eventually. Assume that we are trying to prove that this new vaccine will not produce any side effects, i.e. that there is not enough evidence to say that either nausea or fatigue are caused by the shot. We will have two groups of patients: the first group will take the vaccine (they will be the G sample) and the second group will take a vitamin shot (they will be the F sample). The components of the observations will be once again the time when the patient exhibits fatigue and the time when he or she exhibits nausea. We will monitor the patients until we can record 10 patients which took the vitamin shot with at least one symptom; after that, we will only monitor those patients who showed one of the side effects until they show the other one. If we are correct, we expect to find only a few observations from both groups during the first part of our experiment and therefore we will not be able to reject H_0. This test suggest that we use the D region and the upper orthant order since the observations are markers for time and the kind of follow-up required.

Before finishing, it is worth mentioning that we outlined the test in an ideal way. In practice we may be given the data and we may have to choose a different test from that mentioned here for the purpose of extracting as much information as possible from the observations we have.

Looking at the upper and lower orthant orders, it is clear that our test statistic should be based on the empirical $p-p$ plot $G_m \circ (F_n^{1-}, F_n^{2-})$; its asymptotic behaviour will be discussed in chapter 5.

The discussion in this chapter on filtrations, stopping and stochastic orders in higher dimensions has focused on $\Re_+^2$-valued random variables, and motivates the two definitions of $p-p$ plots to be studied in subsequent chapters. It will be seen that these definitions may be applied to arbitrary distributions on $\Re^d$, and so henceforth we do not restrict our attention to the positive quadrant.

Chapter 3

$P - p$ plots and Glivenko-Cantelli theorems for multivariate distributions

In this chapter we will introduce two types of $p-p$ plots for multivariate distributions, as well as their empirical counterparts. We will also prove a Glivenko-Cantelli type of result for each of them that will be used in subsequent applications.

3.1 Multivariate $p - p$ plots

Before defining the $p - p$ plots we will be using, we have to go through some of the notation and assumptions that will be used throughout the entire document.

We will consider n independent identically distributed $\Re^d$-valued random variables $X_1 = (X_1^1, X_1^2, \ldots, X_1^d), X_2 = (X_2^1, X_2^2, \ldots, X_2^d), \ldots, X_n = (X_n^1, X_n^2, \ldots, X_n^d)$ with distribution function F and m independent identically distributed $\Re^d$-valued random variables $Y_1 = (Y_1^1, Y_1^2, \ldots, Y_1^d), Y_2 = (Y_2^1, Y_2^2, \ldots, Y_2^d), \ldots, Y_m = (Y_m^1, Y_m^2, \ldots, Y_m^d)$ with distribution function G . We begin with some standard definitions.

Definition 3.1.1 *The random distribution function which assigns mass $\frac{1}{n}$ to each X_i*

is called the empirical distribution function of $X_1, X_2, \ldots, X_n$*:*

$$F_n(x_1, x_2, \ldots, x_d) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i^1 \le x_1, X_i^2 \le x_2, \ldots, X_i^d \le x_d\}.$$

Similarly for $Y_1, Y_2, \ldots, Y_m$*,*

$$G_m(y_1, y_2, \ldots, y_d) = \frac{1}{m}\sum_{i=1}^{m} I\{Y_i^1 \le y_1, Y_i^2 \le y_2, \ldots, Y_i^d \le y_d\}.$$

The next theorem we will mention is the classic Glivenko-Cantelli theorem for empirical distributions on $\Re^d$.

Theorem 3.1.2 *With probability one, the empirical distribution function* $F_n(x)$ *converges uniformly to the distribution function* $F(x)$*, i.e. as* $n \to \infty$*,*

$$\sup_{x \in \Re^d} |F_n(x) - F(x)| \to_{a.s.} 0.$$

Definition 3.1.3 *The multivariate empirical process based on* $X_1, X_2, \ldots, X_n$ *will be denoted as* U_n^F *and is defined as* $U_n^F(t) = \sqrt{n}[F_n(t) - F(t)]$*, for* $t \in \Re^d$*.*

Definition 3.1.4 *A stochastic process* $\{Z(t), t \in \Re^d\}$ *is called a Gaussian process if* $Z(t_1), Z(t_2), \ldots, Z(t_n)$ *has a multivariate normal distribution* $\forall\ t_1, t_2, \ldots, t_n$*.*

Definition 3.1.5 *A stochastic process* $\{U(t), 0 \le t \le 1\}$ *is called a standard Brownian Bridge if it is a Gaussian process such that* $E(U(t)) = 0$ *and* $Cov(U(s), U(t)) = (s \wedge t) - st\ \forall\ 0 \le s, t \le 1$*.*

Definition 3.1.6 *A stochastic process* $\{U^H(t), t \in [0,1]^d\}$ *is called a Brownian Bridge (that depends on a distribution* H *on* $[0,1]^d$*) if it is a Gaussian process such that* $E(U^H(t)) = 0$ *and* $Cov(U^H(s), U^H(t)) = H(s \wedge t) - H(s)H(t)\ \forall\ s, t \in [0,1]^d$*.*

Before we continue with the definitions we need, we will define the spaces that will be used throughout the entire document. Given an arbitrary set $T \subseteq [0, \infty)^d$, the Banach space $\ell^\infty(T)$ is the set of all functions $f : T \to \Re$ that are uniformly bounded equipped with the sup norm $\|f\| = \sup_x |f(x)|$. As usual, $C(T)$ is the space of all continuous functions $f : T \to \Re$ equipped with the sup norm. Furthermore, $D([0,\infty]^d)$ denotes the Skorokhod space of functions that are continuous from above with limits from all quadrants, also equipped with the sup norm.

The following theorem is a well-known result, it can be found in [19].

Theorem 3.1.7 *The empirical process $U_n^F(t) = \sqrt{n}[F_n(t) - F(t)]$ based on the random variables $X_1, X_2, \ldots, X_n$ with distribution function F with support on $[0,1]^d$, converges weakly in ℓ^∞ to a Brownian Bridge $U^F(t)$.*

Now that we have defined the notation we will use, we can focus our attention on $p - p$ plots. First, let us state the usual (one-dimensional) definition of a $p - p$ plot.

Definition 3.1.8 *Let F and G be the distribution functions of two one-dimensional random variables. The procentile-procentile, or $p - p$ plot of G against F is defined as $G(F^-(p))$ for $0 \leq p \leq 1$, where $F^-(p)$ is the left continuous inverse of F and $F^-(0) = \lim_{p\to 0} F^-(p) = F^-(0+)$.*

Our goal is to extend this definition to the d-dimensional setting. We present two natural approaches. The first involves the contours of a distribution function, the second is based on marginal order statistics.

Comment 3.1.9 In what follows, we will denote by F and G the measures generated by the distribution functions F and G, respectively.

Definition 3.1.10 *Let $\xi_p^F = \{(x_1, x_2, \ldots, x_d) : F(x_1^-, x_2^-, \ldots, x_d^-) < p)\}$ for $p > 0$ and $\xi_0^F = \cap_{p>0}\xi_p^F$. The pth contour of F is defined as the upper boundary of ξ_p^F. The one-dimensional $p-p$ plot of G against F is defined by $G(\xi_p^F)$. The empirical $p-p$ plot of G against F is $G_m(\xi_p^{F_n})$, where $\xi_p^{F_n} = \{(x_1, x_2, \ldots, x_d) : F_n(x_1^-, x_2^-, \ldots, x_d^-) < p\}$ for $p > 0$ and $\xi_0^{F_n} = \cap_{p>0}\xi_p^{F_n}$. When we use this definition of a $p - p$ plot, we have a process defined on the unit interval $[0,1]$.*

Definition 3.1.11 *We define the $d-$dimensional $p-p$ plot of G against F -using the marginal order statistics- as $G(F_1^-, F_2^-, \ldots, F_d^-)$. Using this definition, the empirical $p-p$ plot of G against F will be $G_m(F_n^{1-}, F_n^{2-}, \ldots, F_n^{d-})$. In this case, we have a process defined on $[0,1]^d$.*

3.2 Glivenko-Cantelli results for bivariate $p-p$ plots

In this section we will present some Glivenko-Cantelli theorems for the $p-p$ plots we defined in the previous section, as well as for some other closely related statistics. For notational convenience, the theorems are presented for distributions on $\Re^2$, but all results are valid for $\Re^d$.

We start by looking at the one-dimensional $p-p$ plot: $G(\xi_p^F)$.

In what follows, we assume that $\sup_{x,y}|F_n(x,y) - F(x,y)| \to 0 \;\forall\; \omega \in \Omega'$, where $P(\Omega') = 1$. To prove the Glivenko-Cantelli result we want for $G_m(\xi_p^{F_n})$ we will be using the following lemma.

Lemma 3.2.1 *Fix ω and assume that for $n = n(\omega)$ large enough, we have that $\sup_{x,y}|F_n(x,y) - F(x,y)| < \epsilon$, where F is a continuous and strictly increasing distribution function on $\Re^2$. Then, $\forall\; 0 \leq p \leq 1$ the following inclusions hold: $\xi_{p-\epsilon}^F \subseteq \xi_p^{F_n} \subseteq \xi_{p+\epsilon}^F$.*

Proof: Since $|F_n(x,y) - F(x,y)| \leq \sup_{x,y}|F_n(x,y) - F(x,y)| < \epsilon$, we have that $\forall$ x, y:

$$\begin{aligned} & |F_n(x,y) - F(x,y)| < \epsilon \\ \Leftrightarrow\; & -\epsilon < F_n(x,y) - F(x,y) < \epsilon \\ \Leftrightarrow\; & -\epsilon + F(x,y) < F_n(x,y) < \epsilon + F(x,y). \end{aligned}$$

i)$\xi_{p-\epsilon}^F \subseteq \xi_p^{F_n}$:

To prove the first inclusion we take $(x,y) \in \xi_{p-\epsilon}^F$, and we show that $(x,y) \in \xi_p^{F_n}$.

$$\begin{aligned}
(x,y) &\in \xi_{p-\epsilon}^{F} \\
&\Rightarrow F(x^-,y^-) < p-\epsilon \\
&\Rightarrow F_n(x^-,y^-) \leq \epsilon + F(x^-,y^-) < p \\
&\Rightarrow F_n(x^-,y^-) < p \\
&\Rightarrow (x,y) \in \xi_p^{F_n}.
\end{aligned}$$

ii)$\xi_p^{F_n} \subseteq \xi_{p+\epsilon}^{F}$:

To prove the second inclusion we take $(x,y) \in \xi_p^{F_n}$, and prove that $(x,y) \in \xi_{p+\epsilon}^{F}$.

$$\begin{aligned}
(x,y) &\in \xi_p^{F_n} \\
&\Rightarrow F_n(x^-,y^-) < p \\
&\Rightarrow -\epsilon + F(x^-,y^-) \leq F_n(x^-,y^-) < p \\
&\Rightarrow F(x^-,y^-) < p+\epsilon \\
&\Rightarrow (x,y) \in \xi_{p+\epsilon}^{F}.
\end{aligned}$$

◇

Before we continue, we would like to talk a little about Kendall distribution functions (see [37]).

- If X^1 and X^2 are two continuous random variables with joint distribution function F, then the Kendall distribution function of (X^1, X^2) is the distribution function of the random variable $F(X^1, X^2)$, i.e. $K^F(t) = P[F(X^1,X^2) \leq t] = P[(X^1,X^2) \in \xi_t^F] = F(\xi_t^F)$.

- Given a sample $(X_1^1, X_1^2), (X_2^1, X_2^2), \ldots, (X_n^1, X_n^2)$ from a distribution F, the empirical Kendall distribution of F (i.e. the Kendall distribution of the empirical distribution F_n) is defined as $F_n(\xi_t^{F_n})$.

- Thus, the $p-p$ plot $G(\xi_p^F)$ is, under $H_0 : F = G$, a two-sample version of the Kendall distribution function.

The following theorem will become useful later when we talk about the possible applications of our tests, depending on the information we have available. It is easily seen from the discussion above, that it can be regarded as a Glivenko-Cantelli theorem for the empirical Kendall distribution function of F .

Theorem 3.2.2 *Let F be a continuous and strictly increasing distribution function on $\Re^2$. For any $\delta > 0$ and $\forall\ \omega \in \Omega'$, where $P(\Omega') = 1$, there exists $n(\omega)$ such that $\forall$ $n \geq n(\omega)$*

$$\sup_{0 \leq p \leq 1} |F_n(\xi_p^{F_n}) - F(\xi_p^F)| < \delta.$$

Proof: Lemma 3.2.1 gives us that for fixed $\omega \in \Omega'$ and $n = n(\omega)$ sufficiently large, $\xi_{p-\epsilon}^F \subseteq \xi_p^{F_n} \subseteq \xi_{p+\epsilon}^F$ holds $\forall\ p$, $0 \leq p \leq 1$. Since the empirical distribution function F_n is increasing, it follows that for large enough n,

$$F_n(\xi_{p-\epsilon}^F) \leq F_n(\xi_p^{F_n}) \leq F_n(\xi_{p+\epsilon}^F) \text{ and therefore,}$$
$$F_n(\xi_{p-\epsilon}^F) - F(\xi_p^F) \leq F_n(\xi_p^{F_n}) - F(\xi_p^F) \leq F_n(\xi_{p+\epsilon}^F) - F(\xi_p^F). \qquad (3.2.1)$$

On the other hand, $F_n(\xi_p^F) = \frac{1}{n}\sum_{i=1}^n I\{X_i \in \xi_p^F\} = \frac{1}{n}\sum_{i=1}^n I\{F(X_i) \leq p\}$, so $F_n(\xi_p^F) = \tilde{K}_n(p)$, where $\tilde{K}_n$ denotes the (univariate) empirical distribution function corresponding to $\tilde{K}$, the distribution function of $F(X)$ (i.e. $\tilde{K}(p) = F(\xi_p^F)$). Therefore, from Theorem 3.1.2 we have that $\forall\ n = n(\omega)$ large enough

$$\sup_{0 \leq p \leq 1} |F_n(\xi_p^F) - F(\xi_p^F)| < \epsilon. \qquad (3.2.2)$$

Equation 3.2.2 implies that $F - \epsilon < F_n < F + \epsilon$ and thus combining Equation 3.2.1 and Equation 3.2.2 we get that $\forall\ n = n(\omega)$ sufficiently large

$$F(\xi_{p-\epsilon}^F) - \epsilon - F(\xi_p^F) \leq F_n(\xi_p^{F_n}) - F(\xi_p^F) \leq F(\xi_{p+\epsilon}^F) + \epsilon - F(\xi_p^F),$$

and therefore that

$$|F_n(\xi_p^{F_n}) - F(\xi_p^F)| \leq |F(\xi_{p+\epsilon}^F) - F(\xi_{p-\epsilon}^F)| + 2\epsilon. \qquad (3.2.3)$$

But, since $\tilde{K}(p) = P(F(X) \leq p) = P(X \in \xi_p^F) = F(\xi_p^F)$ is a continuous distribution function over $[0,1]$, it is uniformly continuous in p and so for any $\delta > 0$, an $\epsilon > 0$ can be chosen such that

$$\sup_{0\leq p\leq 1} |F(\xi_{p+\epsilon}^F) - F(\xi_{p-\epsilon}^F)| \leq \epsilon, \text{where } \epsilon < \frac{\delta}{3}. \quad (3.2.4)$$

Finally, combining Equation 3.2.3 and Equation 3.2.4 we get that $\forall\ n = n(\omega)$ large enough,

$$\sup_{0\leq p\leq 1} |F_n(\xi_p^{F_n}) - F(\xi_p^F)| \leq 3\epsilon < \delta.$$

$\diamond$

The following theorem is a two-sample version of the previous one.

Theorem 3.2.3 *Let F and G be continuous distribution functions on $\Re^2$ and assume that F is strictly increasing. For any $\delta > 0$ and $\forall\ \omega \in \Omega'$, where $P(\Omega') = 1$, there exist $n(\omega)$ and $m(\omega)$ such that $\forall\ n \geq n(\omega)$ and $m \geq m(\omega)$*

$$\sup_{0\leq p\leq 1} |G_m(\xi_p^{F_n}) - G(\xi_p^F)| < \delta.$$

Proof: Lemma 3.2.1 gives us that for fixed $\omega \in \Omega'$ and $n = n(\omega)$ sufficiently large, $\xi_{p-\epsilon}^F \subseteq \xi_p^{F_n} \subseteq \xi_{p+\epsilon}^F$ holds $\forall\ p$, $0 \leq p \leq 1$. Since the empirical distribution function G_m is increasing, it follows that for n large enough,

$$\begin{aligned} &G_m(\xi_{p-\epsilon}^F) \leq G_m(\xi_p^{F_n}) \leq G_m(\xi_{p+\epsilon}^F) \text{ and therefore} \\ &G_m(\xi_{p-\epsilon}^F) - G(\xi_p^F) \leq G_m(\xi_p^{F_n}) - G(\xi_p^F) \leq G_m(\xi_{p+\epsilon}^F) - G(\xi_p^F). \quad (3.2.5) \end{aligned}$$

Let $G_m(\xi_p^F) = \frac{1}{m}\sum_{i=1}^m I\{Y_i \in \xi_p^F\} = \frac{1}{m}\sum_{i=1}^m I\{F(Y_i) \leq p\}$ be as in the preceding proof, so $G_m(\xi_p^F) = K_m(p)$, where K_m is the (univariate) empirical distribution function of $F(Y)$ (i.e. $K(p) = G(\xi_p^F)$). Therefore, from Theorem 3.1.2 we have that $\forall\ m = m(\omega)$ large enough,

$$\sup_{0\leq p\leq 1} |G_m(\xi_p^F) - G(\xi_p^F)| < \epsilon. \quad (3.2.6)$$

Equation 3.2.6 implies that $G-\epsilon < G_m < G+\epsilon$ and thus combining Equation 3.2.5 and Equation 3.2.6 we get that $\forall\ n = n(\omega)$ and $m = m(\omega)$ sufficiently large,

$$G(\xi_{p-\epsilon}^F) - \epsilon - G(\xi_p^F) \le G_m(\xi_p^{F_n}) - G(\xi_p^F) \le G(\xi_{p+\epsilon}^F) + \epsilon - G(\xi_p^F),$$

and therefore that

$$|G_m(\xi_p^{F_n}) - G(\xi_p^F)| \le |G(\xi_{p+\epsilon}^F) - G(\xi_{p-\epsilon}^F)| + 2\epsilon. \tag{3.2.7}$$

But, since $K(p) = P(F(Y) \le p) = P(Y \in \xi_p^F) = G(\xi_p^F)$ is a continuous distribution over $[0,1]$, it is uniformly continuous in p and so for any $\delta > 0$, an $\epsilon > 0$ can be chosen so that

$$\sup_{0\le p\le 1} |G(\xi_{p+\epsilon}^F) - G(\xi_{p-\epsilon}^F)| \le \epsilon, \text{where } \epsilon < \frac{\delta}{3}. \tag{3.2.8}$$

Finally, putting together Equation 3.2.7 and Equation 3.2.8 we get that $\forall\ n = n(\omega)$ and $m = m(\omega)$ large enough,

$$\sup_{0\le p\le 1} |G_m(\xi_p^{F_n}) - G(\xi_p^F)| \le 3\epsilon < \delta. \quad \diamond$$

We will now move on to the bivariate $p-p$ plot: $G(F_1^-, F_2^-)$.

Although we have included an appendix that has all the definitions and properties related to copulas that we make use of, we will mention a couple of copula-related concepts here.

- A copula C can be regarded as a distribution on $[0,1]^2$ with uniform marginals.
- $C^F(u,v) = F(F_1^-(u), F_2^-(v))$ is a copula for any continuous distribution F with marginals F_1 and F_2.
- Given a sample $X_1, X_2, \ldots, X_n$ from a distribution F, the empirical copula is defined by $C_n^F(u,v) = F_n(F_n^{1-}(u), F_n^{2-}(v))$.

Thus, it is easily seen that the empirical $p-p$ plot can be viewed, under H_0 : $F=G$, as a two-sample empirical copula. For this reason, we will start by proving a Glivenko-Cantelli theorem for empirical copulas. We start with the following lemma which is a well-known result (see [35]), but we are including it here for completeness.

Lemma 3.2.4 *The copula C satisfies a Lipschitz condition of order 1:*

$$|C(p,q)-C(u,v)| < \sqrt{2}||(p,q)-(u,v)||,$$

where $||\cdot||$ denotes the Euclidian norm in $\Re^2$, i.e. $||s,t|| = ||(s_1,s_2),(t_1,t_2)|| = \sqrt{(s_1-t_1)^2+(s_2-t_2)^2}$.

Proof: First we note that $\forall\, x,y \in \Re$ we have that $2(x^2+y^2) \geq (x+y)^2$, from which it follows that $\sqrt{2}\sqrt{x^2+y^2} \geq |x+y|$.

In what follows, we will also make use of the fact that the copula is a distribution function with uniform marginals. We have that:

$$\begin{aligned}
|C(p,q)-C(u,v)| &= |C(p,q)-C(u,q)+C(u,q)-C(u,v)| \\
&\leq |C(p,q)-C(u,q)|+|C(u,q)-C(u,v)| \\
&\leq |C(p,1)-C(u,1)|+|C(1,q)-C(1,v)| \\
&= |p-u|+|q-v| \\
&\leq \sqrt{2}\sqrt{(p-u)^2+(q-v)^2} \\
&= \sqrt{2}||(p-u,q-v)|| \\
&= \sqrt{2}||(p,q)-(u,v)||.
\end{aligned}$$

◇

The following theorem is most likely known; however, we have been unable to find a proof in the literature.

Theorem 3.2.5 *Suppose that $(X_1^1,X_1^2),(X_2^1,X_2^2),\ldots$ are independent and identically distributed with continuous distribution F and copula C^F. If C_n^F is the empirical copula, then $sup_{p,q}|C_n^F(p,q)-C^F(p,q)| \rightarrow_{a.s.} 0$*

Proof: We start by noting that:

$$
\begin{aligned}
\sup_{p,q}|C_n^F(p,q) - C^F(p,q)| &= \sup_{p,q}|F_n(F_n^{1-}(p), F_n^{2-}(q)) - F(F_1^-(p), F_2^-(q))| \\
&= \sup_{p,q}|F_n(F_n^{1-}(p), F_n^{2-}(q)) - F(F_n^{1-}(p), F_n^{2-}(q)) \\
&+ F(F_n^{1-}(p), F_n^{2-}(q)) - F(F_1^-(p), F_2^-(q))| \\
&\leq \sup_{p,q}|F_n(F_n^{1-}(p), F_n^{2-}(q)) - F(F_n^{1-}(p), F_n^{2-}(q))| \\
&+ \sup_{p,q}|F(F_n^{1-}(p), F_n^{2-}(q)) - F(F_1^-(p), F_2^-(q))| \\
&= \underbrace{\sup_{p,q}|(F_n - F)(F_n^{1-}(p), F_n^{2-}(q))|}_{(1)} \\
&+ \underbrace{\sup_{p,q}\left|C^F(F_1(F_n^{1-}(p)), F_2(F_n^{2-}(q))) - C^F(p,q)\right|}_{(2)}.
\end{aligned}
$$

We know from Theorem 3.1.2 that (1) $\rightarrow_{a.s.} 0$.

Since we proved in the previous lemma that the copula satisfies the Lipschitz condition, we can further bound (2) as follows:

$$
\begin{aligned}
(2) &\leq \sup_{p,q}\sqrt{2}||(F_1(F_n^{1-}(p)), F_2(F_n^{2-}(q))) - (p,q)|| \\
&\leq \underbrace{\sup_p \sqrt{2}|F_1(F_n^{1-}(p)) - p|}_{(3)} + \underbrace{\sup_q \sqrt{2}\left|F_2(F_n^{2-}(q)) - q\right|}_{(4)}.
\end{aligned}
$$

We will only show the way to handle (3), since (4) is managed similarly.

$$
\begin{aligned}
(3) &= \sup_p \sqrt{2}|F_1(F_n^{1-}(p)) - F_n^1(F_n^{1-}(p)) + F_n^1(F_n^{1-}(p)) - p| \\
&\leq \underbrace{\sup_p \sqrt{2}|F_1(F_n^{1-}(p)) - F_n^1(F_n^{1-}(p))|}_{(5)} + \underbrace{\sup_p \sqrt{2}\left|F_n^1(F_n^{1-}(p)) - p\right|}_{(6)}.
\end{aligned}
$$

Again, Theorem 3.1.2 gives us that (5) $\rightarrow_{a.s.} 0$.

On the other hand, since $F_n^1(x) = \frac{[np]}{n}$ if $X^1_{([np])} \leq x < X^1_{([np]+1)}$ and $F_n^{1-}(p) = X^1_{([np])}$, we have that:

$$
\begin{aligned}
(6) &= \sup_p \sqrt{2}|\frac{[np]}{n} - p| \\
&= \sup_p \sqrt{2}|\frac{[np] - np}{n}| \rightarrow 0.
\end{aligned}
$$

The convergence of (5) and (6) imply that (3) $\rightarrow_{a.s.} 0$ (and therefore (4) does too), so it is also true that (2) $\rightarrow_{a.s.} 0$. This completes the result. $\diamond$

Before we can proceed with the Glivenko-Cantelli theorem for $G(F_1^-, F_2^-)$, we will prove a lemma that we will need to prove the desired result.

Lemma 3.2.6 *Let $X_1, X_2, \ldots$ be (univariate) independent random variables with a common distribution function F that is continuous and strictly increasing on its open support, and let G be any continuous distribution function. Then, for any $\delta > 0$ and $\forall\ \omega \in \Omega'$, where $P(\Omega') = 1$, there exists $n = n(\omega)$ such that $\forall\ n \geq n(\omega)$,*

$$\sup_{0 \leq p \leq 1} |G(F_n^-(p)) - G(F^-(p))| < \delta.$$

Proof: Fix $\omega \in \Omega'$ and as in the proof of Lemma 3.2.1, for $n = n(\omega)$ large enough, $\sup_x |F_n(x) - F(x)| < \epsilon$ and $-\epsilon + F(x) < F_n(x) < \epsilon + F(x)$, $\forall\ x$. This last expression, combined with the definitions of F_n^- and F^- gives us that $\forall\ 0 \leq p \leq 1$, $F^-(p-\epsilon) \leq F_n^-(p) \leq F^-(p+\epsilon)$ holds.

Thus, for $n = n(\omega)$ sufficiently large we have that $\forall\ 0 \leq p \leq 1:$

$$\begin{aligned} & F^-(p-\epsilon) \leq F_n^-(p) \leq F^-(p+\epsilon) \\ \Rightarrow\ & G(F^-(p-\epsilon)) \leq G(F_n^-(p)) \leq G(F^-(p+\epsilon)) \\ \Rightarrow\ & |G(F_n^-(p)) - G(F^-(p))| \leq G(F^-(p+\epsilon)) - G(F^-(p-\epsilon)). \end{aligned}$$

The result follows from the fact that GF^- is continuous. $\diamond$

We are finally ready to present the Glivenko-Cantelli theorem for the $p-p$ plot $G_m(F_n^{1-}(\cdot), F_n^{2-}(\cdot))$.

Theorem 3.2.7 *Let $(X_1^1, X_1^2), (X_2^1, X_2^2), \ldots$ be independent with common distribution function F and let $(Y_1^1, Y_1^2), (Y_2^1, Y_2^2), \ldots$ be also independent with ditribution G. If F_1, F_2 and G_1, G_2 are the marginals of F and G respectively, then for any $\delta > 0$ and $\forall\ \omega \in \Omega'$, where $P(\Omega') = 1$, there exist $n(\omega)$ and $m(\omega)$ such that $\forall\ n \geq n(\omega)$ and*

$m \geq m(\omega)$,

$$\sup_{0\leq p,q\leq 1} |G_m(F_n^{1-}(p), F_n^{2-}(q)) - G(F_1^-(p), F_2^-(q))| < \delta.$$

Proof: We have that:

$$\begin{aligned}
&\sup_{0\leq p,q\leq 1} |G_m(F_n^{1-}(p), F_n^{2-}(q)) - G(F_1^-(p), F_2^-(q))| \\
= &\sup_{0\leq p,q\leq 1} |G_m(F_n^{1-}(p), F_n^{2-}(q)) - G(F_n^{1-}(p), F_n^{2-}(q)) \\
+ &\ G(F_n^{1-}(p), F_n^{2-}(q)) - G(F_1^-(p), F_2^-(q))| \\
\leq &\sup_{0\leq p,q\leq 1} |G_m(F_n^{1-}(p), F_n^{2-}(q)) - G(F_n^{1-}(p), F_n^{2-}(q))| \\
+ &\sup_{0\leq p,q\leq 1} |G(F_n^{1-}(p), F_n^{2-}(q)) - G(F_1^-(p), F_2^-(q))| \\
= &\underbrace{\sup_{0\leq p,q\leq 1} |(G_m - G)(F_n^{1-}(p), F_n^{2-}(q))|}_{(1)} \\
+ &\underbrace{\sup_{0\leq p,q\leq 1} |C^G(G_1(F_n^{1-}(p)), G_2(F_n^{2-}(q))) - C^G(G_1(F_1^-(p)), G_2(F_2^-(q)))|}_{(2)}.
\end{aligned}$$

From Theorem 3.1.2 we get that for $\epsilon = \frac{\delta}{3}$ we can find $m = m(\omega)$ large enough so that $(1) < \epsilon$.

On the other hand, since the copula satisfies the Lipschitz condition, we have that:

$$\begin{aligned}
(2) \leq &\sup_{0\leq p,q\leq 1} \sqrt{2}\|(G_1(F_n^{1-}(p)), G_2(F_n^{2-}(q))) - (G_1(F_1^-(p)), G_2(F_2^-(q)))\| \\
\leq &\underbrace{\sup_{0\leq p,q\leq 1} \sqrt{2}|G_1(F_n^{1-}(p)) - G_1(F_1^-(p))|}_{(3)} \\
+ &\underbrace{\sup_{0\leq p,q\leq 1} \sqrt{2}|G_2(F_n^{2-}(q)) - G_2(F_2^-(q))|}_{(4)}.
\end{aligned}$$

In the previous lemma we showed that for $\epsilon = \frac{\delta}{3}$ we can find $n = n(\omega)$ such that both $(3), (4) < \epsilon$. Therefore $(2) < \frac{2}{3}\delta$, and that completes the proof. $\diamond$

Chapter 4

Asymptotic behaviour of $G_m(\xi_p^{F_n})$ and applications

In this chapter we will develop the asymptotic behaviour of the $p-p$ plot $G_m(\xi_p^{F_n})$, as well as some applications of this result. This is the appropriate process to study when the minimal filtration is the one generated by the available data. In order to obtain the desired limiting distribution, we will be closely following [9], where Barbe, Genest, Ghoudi and Rémillard studied the weak convergence of Kendall's process.

4.1 Limiting distribution

Although some of the notation we will be using was described in chapter 3, we will go through it again here for the sake of convenience. $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ will be random samples from continuous bivariate distribution functions F and G with marginals F_1, F_2 and G_1, G_2, respectively. The sets $\xi_t^{F_n}$ and ξ_t^F are defined as $\xi_t^{F_n} = \{(x,y) : F_n(x^-, y^-) < t\}$ and $\xi_t^F = \{(x,y) : F(x^-, y^-) < t\}$. Also, let $\hat{X} = F(X)$ and $\hat{Y} = F(Y)$ be random variables with distribution functions $\tilde{K}$ and K, i.e. $\tilde{K}(t) = P(F(X) \leq t) = P(X \in \xi_t^F) = F(\xi_t^F)$ and $K(t) = P(F(Y) \leq t) = P(Y \in \xi_t^F) = G(\xi_t^F)$, and densities $\tilde{k}$ and k, respectively. We are interested in the behaviour of the process $\alpha_{n,m}(t) = \sqrt{m}[K_{n,m}(t) - K(t)]$, where $K_{n,m}(t)$ is defined as

$K_{n,m}(t) = \frac{1}{m}\sum_{i=1}^{m} I\{F_n(Y_i) \leq t\} = \frac{1}{m}\sum_{i=1}^{m} I\{Y_i \in \xi_t^{F_n}\} = G_m(\xi_t^{F_n})$.

Although the results developed in this section can be extended to $\Re^d$ as in [9] (which we will be closely following), for the sake of clarity we will restrict ourselves to bivariate distributions.

The following assumptions will be made on the distribution functions F and G. They are analogues of Hypotheses I and II in [9].

Assumption 1)The distribution functions $K(t) = G(\xi_t^F)$ of $\hat{Y}$ and $\tilde{K}(t) = F(\xi_t^F)$ of $\hat{X}$ admit continuous densities $k(t)$ and $\tilde{k}(t)$ on $(0,1]$ respectively, that verify $k(t) = o\{t^{-\frac{1}{2}} \ln^{-\frac{1}{2}-\epsilon}(\frac{1}{t})\}$ and $\tilde{k}(t) = o\{t^{-\frac{1}{2}} \ln^{-\frac{1}{2}-\epsilon}(\frac{1}{t})\}$ for some $\epsilon > 0$ as $t \to 0$.

Comment 4.1.1 Note that if $\tilde{K}(t)$ admits a continuous density $\tilde{k}(t)$ on $(0,1]$ verifying $\tilde{k}(t) = o\{t^{-\frac{1}{2}} \ln^{-\frac{1}{2}-\epsilon}(\frac{1}{t})\}$ for some $\epsilon > 0$ as $t \to 0$, and the distribution functions F and G have densities f and g respectively, such that $sup_{x,y}\frac{g(x,y)}{f(x,y)} \leq c$ for some $c < \infty$, then $K(t) = \int_{\xi_t^F} g(z)dz \leq \int_{\xi_t^F} cf(z)dz = c\tilde{K}(t)$, and we would not need an additional restriction for $K(t)$, as we have in assumption 1.

Assumption 2)There exists a version of the conditional distribution of the vector $Y^* \equiv (F_1(Y^1), F_2(Y^2))$ given $F(Y) = t$ and a countable family $\mathcal{P}$ of partitions $\mathcal{C}$ of $[0,1]^2$ into a finite number of Borel sets satisfying $\inf_{\mathcal{C}\in\mathcal{P}} \max_{C\in\mathcal{C}} \operatorname{diam}(C) = 0$ such that $\forall\ C \in \mathcal{C}$ the mapping

$$t \to \mu_t^{F,G}(C) = k(t)P\{Y^* \in C | F(Y) = t\}$$

is continuous on $(0,1]$ with $\mu_1^{F,G}(C) = k(1)I\{(1,1) \in C\}$.

Comment 4.1.2 Defining the pseudo-variables X^*, Y^* as $X^* = (F_1(X^1), F_2(X^2))$ and $Y^* = (F_1(Y^1), F_2(Y^2))$, it is easily seen that $K_{n,m}(t) = K^*_{n,m}(t)$, where $K^*_{n,m}$ is defined as above using the pseudo-variables $X_1^*, X_2^*, \ldots, X_n^*$ and $Y_1^*, Y_2^*, \ldots, Y_m^*$ instead. Since X^* has distribution C^F and Y^* is a random variable on $[0,1]^2$, we may assume without loss of generality that F is a copula and that G is a distribution on $[0,1]^2$. The assumption that F is a copula implies that $Y^* = Y = (Y^1, Y^2)$ in what follows.

The main result of this chapter is the following theorem, which defines the asymptotic behaviour of $\alpha_{n,m}$.

Theorem 4.1.3 *Under assumptions* 1 *and* 2, *and if* $\frac{m}{n} \to \lambda$ *as* $m, n \to \infty$, *the empirical process* $\alpha_{n,m}(t) = \sqrt{m}[K_{n,m}(t) - K(t)]$ *converges in distribution to a continuous, centered Gaussian process* α *with zero mean and covariance function*

$$\begin{aligned}\Gamma(s,t) &= G(\xi_t^F \cap \xi_s^F) - G(\xi_t^F)G(\xi_s^F) \\ &+ \lambda \int\int (F(z \wedge z') - F(z)F(z'))\mu_t^{F,G}(dz)\mu_s^{F,G}(dz').\end{aligned}$$

Moreover, for $t \in [0,1]$, *the limiting process has the following representation in terms of the weak limits* U^F *of* $\sqrt{n}(F_n - F)$ *and* U^G *of* $\sqrt{m}(G_m - G)$:

$$\alpha(t) = U^G(\xi_t^F) - \sqrt{\lambda}\int_{[0,1]^2} U^F(z)\mu_t^{F,G}(dz). \tag{4.1.1}$$

Before proceeding with its proof, we will prove a series of results that will be needed to show the required convergence.

Lemma 4.1.4 *i)The process* $\alpha_{n,m}(t) = \sqrt{m}[G_m(\xi_t^{F_n}) - G(\xi_t^F)]$ *can be expressed as* $\alpha_{n,m}(t) = \beta_m(t) + \gamma_{n,m}(t)$, *where*

$$\begin{aligned}\beta_m(t) &= \sqrt{m}[\frac{1}{m}\sum_{i=1}^{m} I\{\hat{Y}_i \le t\} - G(\xi_t^F)] \text{ and} \\ \gamma_{n,m}(t) &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}[I\{F_n(Y_i) \le t\} - I\{\hat{Y}_i \le t\}].\end{aligned}$$

ii)If $(F_n - F)^+$ *and* $(F_n - F)^-$ *denote the positive and negative parts of* $(F_n - F)$, *then we have that* $\gamma_{n,m}(t) = \delta_{n,m}(t) - \epsilon_{n,m}(t)$*, where*

$$\begin{aligned}\delta_{n,m}(t) &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m} I\{t < F(Y_i) \leq t + (F_n - F)^{-}(Y_i)\} \text{ and} \\ \epsilon_{n,m}(t) &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m} I\{t - (F_n - F)^{+}(Y_i) < F(Y_i) \leq t\}.\end{aligned}$$

Proof: i)$\alpha_{n,m}(t) = \beta_m(t) + \gamma_{n,m}(t)$:

The result follows when we express $\beta_m(t)$ and $\gamma_{n,m}(t)$ in terms of G_m:

$$\begin{aligned}\beta_m(t) &= \sqrt{m}[\frac{1}{m}\sum_{i=1}^{m} I\{\hat{Y}_i \leq t\} - G(\xi_t^F)] \\ &= \sqrt{m}[G_m(\xi_t^F) - G(\xi_t^F)]. \\ \gamma_{n,m}(t) &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}[I\{F_n(Y_i) \leq t\} - I\{\hat{Y}_i \leq t\}] \\ &= \sqrt{m}[G_m(\xi_t^{F_n}) - G_m(\xi_t^F)].\end{aligned}$$

ii)$\gamma_{n,m}(t) = \delta_{n,m}(t) - \epsilon_{n,m}(t)$:

We have that:

$$\begin{aligned}&(\delta_{n,m} - \epsilon_{n,m})(t) \\ =\ & \frac{1}{\sqrt{m}}[\sum_{i=1}^{m} I\{t < F(Y_i) \leq t + (F_n - F)^{-}(Y_i)\} \\ -\ & \sum_{i=1}^{m} I\{t - (F_n - F)^{+}(Y_i) < F(Y_i) \leq t\}] \\ =\ & \frac{1}{\sqrt{m}}[\sum_{i=1}^{m} I\{t - (F_n - F)^{-}(Y_i) < F(Y_i) - (F_n - F)^{-}(Y_i) \leq t\} \\ -\ & \sum_{i=1}^{m} I\{t < F(Y_i) + (F_n - F)^{+}(Y_i) \leq t + (F_n - F)^{+}(Y_i)\}] \\ =\ & \frac{1}{\sqrt{m}}[\sum_{i=1}^{m} I\{t - (F_n - F)^{-}(Y_i) + (F_n - F)^{+}(Y_i) < F(Y_i) \\ -\ & (F_n - F)^{-}(Y_i) + (F_n - F)^{+}(Y_i) \leq t + (F_n - F)^{+}(Y_i)\}] \\ -\ & \frac{1}{\sqrt{m}}[\sum_{i=1}^{m} I\{t - (F_n - F)^{-}(Y_i) < F(Y_i) + (F_n - F)^{+}(Y_i) \\ -\ & (F_n - F)^{-}(Y_i) \leq t + (F_n - F)^{+}(Y_i) - (F_n - F)^{-}(Y_i)\}]\end{aligned}$$

	$I\gamma_{n,m}(t)$	$I\delta_{n,m}(t)$	$I\epsilon_{n,m}(t)$	$I(\delta_{n,m} - \epsilon_{n,m})(t)$
$F_n(Y_i) \leq t$ and $F(Y_i) \leq t$	0	0	0	0
$F_n(Y_i) \leq t$ and $F(Y_i) > t$	1	1	0	1
$F_n(Y_i) > t$ and $F(Y_i) \leq t$	-1	0	1	-1
$F_n(Y_i) > t$ and $F(Y_i) > t$	0	0	0	0

Table 4.1: $\gamma_{n,m}(t) = \delta_{n,m}(t) - \epsilon_{n,m}(t)$

$$
\begin{aligned}
&= \frac{1}{\sqrt{m}}[\sum_{i=1}^{m} I\{t + (F_n - F)(Y_i) < F(Y_i) + (F_n - F)(Y_i) \leq \\
&\quad t + (F_n - F)^+(Y_i)\} - \sum_{i=1}^{m} I\{t - (F_n - F)^-(Y_i) < \\
&\quad F(Y_i) + (F_n - F)(Y_i) \leq t + (F_n - F)(Y_i)\}] \\
&= \frac{1}{\sqrt{m}}[\sum_{i=1}^{m} I\{t + (F_n - F)(Y_i) < F_n(Y_i) \leq t + (F_n - F)^+(Y_i)\} \\
&- \sum_{i=1}^{m} I\{t - (F_n - F)^-(Y_i) < F_n(Y_i) \leq t + (F_n - F)(Y_i)\}].
\end{aligned}
$$

Let $I\omega_{n,m}(t)$ denote the indicator of the process $\omega_{n,m}(t)$. For each Y_i, $i = 1, 2, \ldots, m$ consider the cases shown in Table 4.1.

It becomes clear that $\gamma_{n,m}(t)$ can be rewritten as $(\delta_{n,m} - \epsilon_{n,m})(t)$. $\diamond$

Comment 4.1.5 We know by Theorem 3.1.7 that $\beta_m(\cdot)$ converges weakly to a Gaussian bridge $U^K(\cdot)$ with covariance function $K(s \wedge t) - K(s)K(t) = G(\xi_t^F \cap \xi_s^F) - G(\xi_t^F)G(\xi_s^F)$.

Since the asymptotic behaviour of β_m is known, we will focus our attention on the process $\gamma_{n,m}$ (and therefore on $\delta_{n,m}$ and $\epsilon_{n,m}$). We will consider the behavior of

$\gamma_{n,m}$ in two separate cases: when t is bounded away from the origin and when t is in the neighborhood of the origin. Following [9], we start with the former.

Lemma 4.1.6 *The following quantities converge in probability to zero for any* $0 < t_0 \leq 1$*:*
i)$\sup_{t_0 \leq t \leq 1} |\delta_{n,m}(t) - \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^-(z)\mu_t^{F,G}(dz)|$.
ii)$\sup_{t_0 \leq t \leq 1} |\epsilon_{n,m}(t) - \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^+(z)\mu_t^{F,G}(dz)|$.
iii)$\sup_{t_0 \leq t \leq 1} |\gamma_{n,m}(t) + \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)(z)\mu_t^{F,G}(dz)|$.

Proof: iii)If i and ii are true, then:

$$\begin{aligned}
& \sup_{t_0 \leq t \leq 1} |\gamma_{n,m}(t) + \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)(z)\mu_t^{F,G}(dz)| \\
= & \sup_{t_0 \leq t \leq 1} |\delta_{n,m}(t) - \epsilon_{n,m}(t) - \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^-(z)\mu_t^{F,G}(dz) \\
+ & \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^+(z)\mu_t^{F,G}(dz)| \\
\leq & \sup_{t_0 \leq t \leq 1} |\delta_{n,m}(t) - \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^-(z)\mu_t^{F,G}(dz)| \\
+ & \sup_{t_0 \leq t \leq 1} |\epsilon_{n,m}(t) - \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^+(z)\mu_t^{F,G}(dz)| \to 0.
\end{aligned}$$

ii)The proof of ii is analogous to that of i (below) and therefore omitted.

i)For any element C of a partition $\mathcal{C} = (C_j)_{j=1}^r \in \mathcal{P}$, define

$$\begin{aligned}
I_{n,j} &= \inf_{y \in C_j} \sqrt{n}(F_n - F)^-(y), \\
S_{n,j} &= \sup_{y \in C_j} \sqrt{n}(F_n - F)^-(y), \\
\varepsilon_{m,C}(t) &= \sqrt{m}[\frac{1}{m} \sum_{i=1}^m I\{F(Y_i) \leq t, Y_i \in C\} - P(F(Y) \leq t, Y \in C)], \text{ and} \\
\delta_{n,m,j}(t) &= \frac{1}{\sqrt{m}} \sum_{i=1}^m I\{t < F(Y_i) \leq t + (F_n - F)^-(Y_i)\} I\{Y_i \in C_j\}.
\end{aligned}$$

Then,

$$\begin{aligned}\sum_{j=1}^{r}\delta_{n,m,j}(t) &= \sum_{j=1}^{r}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}I\{t<F(Y_i)\le t+(F_n-F)^{-}(Y_i)\}I\{Y_i\in C_j\}\\ &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}I\{t<F(Y_i)\le t+(F_n-F)^{-}(Y_i)\}\sum_{j=1}^{r}I\{Y_i\in C_j\}\\ &= \delta_{n,m}(t).\end{aligned}$$

For $1\le j\le r$:

$$\begin{aligned}\delta_{n,m,j}(t) &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}I\{t<F(Y_i)\le t+(F_n-F)^{-}(Y_i)\}I\{Y_i\in C_j\}\\ &\le \frac{1}{\sqrt{m}}\sum_{i=1}^{m}I\{t<F(Y_i)\le t+\frac{S_{n,j}}{\sqrt{n}}\}I\{Y_i\in C_j\}\\ &= \varepsilon_{m,C_j}(t+\frac{S_{n,j}}{\sqrt{n}})-\varepsilon_{m,C_j}(t)+\sqrt{m}\int_t^{t+\frac{S_{n,j}}{\sqrt{n}}}\mu_s^{F,G}(C_j)ds\\ &= \varepsilon_{m,C_j}(t+\frac{S_{n,j}}{\sqrt{n}})-\varepsilon_{m,C_j}(t)+\sqrt{m}\int_t^{t+\frac{S_{n,j}}{\sqrt{n}}}(\mu_s^{F,G}(C_j)-\mu_t^{F,G}(C_j))ds\\ &+ \mu_t^{F,G}(C_j)S_{n,j}\frac{\sqrt{m}}{\sqrt{n}}\\ &= [\varepsilon_{m,C_j}(t+\frac{S_{n,j}}{\sqrt{n}})-\varepsilon_{m,C_j}(t)]+\sqrt{m}\int_t^{t+\frac{S_{n,j}}{\sqrt{n}}}(\mu_s^{F,G}(C_j)-\mu_t^{F,G}(C_j))ds\\ &+ [\mu_t^{F,G}(C_j)S_{n,j}\frac{\sqrt{m}}{\sqrt{n}}-\int_{C_j}\sqrt{m}(F_n-F)^{-}(z)\mu_t^{F,G}(dz)]\\ &+ \int_{C_j}\sqrt{m}(F_n-F)^{-}(z)\mu_t^{F,G}(dz)\\ &\le [\varepsilon_{m,C_j}(t+\frac{S_{n,j}}{\sqrt{n}})-\varepsilon_{m,C_j}(t)]+\sqrt{m}\int_t^{t+\frac{S_{n,j}}{\sqrt{n}}}(\mu_s^{F,G}(C_j)-\mu_t^{F,G}(C_j))ds\\ &+ [\mu_t^{F,G}(C_j)S_{n,j}\frac{\sqrt{m}}{\sqrt{n}}-\frac{\sqrt{m}}{\sqrt{n}}I_{n,j}\mu_t^{F,G}(C_j)]\\ &+ \int_{C_j}\sqrt{m}(F_n-F)^{-}(z)\mu_t^{F,G}(dz).\end{aligned}$$

An analogous argument gives us that

$$\delta_{n,m,j}(t) \geq [\varepsilon_{m,C_j}(t+\frac{I_{n,j}}{\sqrt{n}}) - \varepsilon_{m,C_j}(t)] + \sqrt{m}\int_t^{t+\frac{I_{n,j}}{\sqrt{n}}} (\mu_s^{F,G}(C_j) - \mu_t^{F,G}(C_j))ds$$

$$-[\mu_t^{F,G}(C_j)S_{n,j}\frac{\sqrt{m}}{\sqrt{n}} - \frac{\sqrt{m}}{\sqrt{n}}I_{n,j}\mu_t^{F,G}(C_j)] + \int_{C_j}\sqrt{m}(F_n - F)^-(z)\mu_t^{F,G}(dz).$$

For arbitrary $C \in \mathcal{C}$, the finite-dimensional distributions of the pseudo-empirical process $\varepsilon_{m,C}$ converge in law to those of a Gaussian process on $[0,1]$ with mean zero and covariance function $P(Y \in C, F(Y) \leq t \wedge s) - P(Y \in C, F(Y) \leq t)P(Y \in C, F(Y) \leq s)$. Moreover, since it can be shown that $\varepsilon_{m,C}$ is tight, it follows that $\varepsilon_{m,C}$ converges in distribution, as $m \to \infty$ to a continuous Gaussian process (for details, see [9]).

Furthermore, $\sqrt{n}(F_n - F)$ converges in $\ell^\infty[0,1]^2$ to a continuous Gaussian process $U^F = (U^F)^+ - (U^F)^-$. Therefore, $I_{n,j} \to_{\mathcal{D}} \inf_{z\in C_j}(U^F)^-(z)$, $S_{n,j} \to_{\mathcal{D}} \sup_{z\in C_j}(U^F)^-(z)$, $\frac{I_{n,j}}{\sqrt{n}} \to_P 0$ and $\frac{S_{n,j}}{\sqrt{n}} \to_P 0$.

But then, since the ε_{m,C_j} are tight, and r is fixed, we have that the processes

$$\begin{aligned}
\wp^1_{n,m} &= \sum_{j=1}^r \sup_{0\leq t\leq 1} |\varepsilon_{m,C_j}(t+\frac{I_{n,j}}{\sqrt{n}}) - \varepsilon_{m,C_j}(t)|, \text{ and} \\
\wp^2_{n,m} &= \sum_{j=1}^r \sup_{0\leq t\leq 1} |\varepsilon_{m,C_j}(t+\frac{S_{n,j}}{\sqrt{n}}) - \varepsilon_{m,C_j}(t)|
\end{aligned}$$

both converge to zero in probability.

Also, since $\mu_s^{F,G}(C_j)$ is continuous for $s \in [t_0, 1]$, we have that for $0 < t_0 \leq 1$, the processes

$$\begin{aligned}
\wp^3_{n,m} &= \sqrt{m}\sum_{j=1}^r \sup_{t_0\leq t\leq 1} |\int_t^{t+\frac{S_{n,j}}{\sqrt{n}}} (\mu_s^{F,G}(C_j) - \mu_t^{F,G}(C_j))ds| \text{ and} \\
\wp^4_{n,m} &= \sqrt{m}\sum_{j=1}^r \sup_{t_0\leq t\leq 1} |\int_t^{t+\frac{I_{n,j}}{\sqrt{n}}} (\mu_s^{F,G}(C_j) - \mu_t^{F,G}(C_j))ds|
\end{aligned}$$

converge in probability to zero.

Finally, consider $\wp^5_{n,m}$ as defined below:

$$
\begin{aligned}
\wp^5_{n,m} &= \frac{\sqrt{m}}{\sqrt{n}} \sum_{j=1}^{r} \sup_{t_0 \le t \le 1} |\mu_t^{F,G}(C_j) S_{n,j} - \mu_t^{F,G}(C_j) I_{n,j}| \\
&= \frac{\sqrt{m}}{\sqrt{n}} \sum_{j=1}^{r} \sup_{t_0 \le t \le 1} \mu_t^{F,G}(C_j) | \sup_{z_1 \in C_j} \sqrt{n}(F_n - F)^-(z_1) - \inf_{z_2 \in C_j} \sqrt{n}(F_n - F)^-(z_2)| \\
&\le \frac{\sqrt{m}}{\sqrt{n}} \sup_{t_0 \le t \le 1} \sum_{j=1}^{r} \sup_{j} \mu_t^{F,G}(C_j) \sup_{z_1, z_2 \in C_j} \sqrt{n} |(F_n - F)^-(z_1) - (F_n - F)^-(z_2)| \\
&\le \frac{\sqrt{m}}{\sqrt{n}} \sup_{t_0 \le t \le 1} k(t) \max_{1 \le j \le r} \sup_{z_1, z_2 \in C_j} \sqrt{n} |(F_n - F)^-(z_1) - (F_n - F)^-(z_2)| \\
&\le \frac{\sqrt{m}}{\sqrt{n}} \sup_{t_0 \le t \le 1} k(t) \omega\{\sqrt{n}(F_n - F)^-, \max_{1 \le j \le r} \operatorname{diam}(C_j)\},
\end{aligned}
$$

where $\omega\{f, s\} = \sup_{z_1, z_2 \in [0,1]^2, ||z_1, z_2|| \le s} |f(z_1) - f(z_2)|$.

Since $\frac{m}{n} \to \lambda$ as $n, m \to \infty$, it is possible to make $\wp^5_{n,m}$ arbitrarily small with probability arbitrarily close to 1 when n and m are large by choosing a partition with an appropriate mesh.

Then,

$$
\begin{aligned}
&\sup_{t_0 \le t \le 1} |\delta_{n,m}(t) - \frac{\sqrt{m}}{\sqrt{n}} \int_{[0,1]^2} \sqrt{n}(F_n - F)^-(z) \mu_t^{F,G}(dz)| \\
&= \sup_{t_0 \le t \le 1} | \sum_{j=1}^{r} \delta_{n,m,j}(t) - \sum_{j=1}^{r} \int_{C_j} \sqrt{m}(F_n - F)^-(z) \mu_t^{F,G}(dz)| \\
&\le \max(\wp^1_{n,m}, \wp^2_{n,m}) + \max(\wp^3_{n,m}, \wp^4_{n,m}) + \wp^5_{n,m} \\
&\to_P 0.
\end{aligned}
$$

◇

Lemma 4.1.7 *The restriction of the process $\gamma_{n,m}(t)$ to the interval $[t_0, 1]$, $0 < t_0 \le 1$, converges in law to a centered, continuous Gaussian process having the representation $-\sqrt{\lambda} \int_{[0,1]^2} U^F(z) \mu_t^{F,G}(dz)$.*

Proof: There exists a continuous version $\dot{F}_n$ of F_n such that $\sup_{z \in [0,1]^2} |\dot{F}_n(z) - F_n(z)| \le \frac{1}{n}$ and $\sqrt{n}(\dot{F}_n - F) \to_{\mathcal{D}} U^F$ in $C[0,1]^2$.

Also note that:

$$\begin{aligned}&\sup_{t_0\le t\le 1}|\int_{[0,1]^2}\sqrt{n}(\dot{F}_n-F)(z)\mu_t^{F,G}(dz)-\int_{[0,1]^2}\sqrt{n}(F_n-F)(z)\mu_t^{F,G}(dz)|\\ \le&\sup_{t_0\le t\le 1}\frac{k(t)}{\sqrt{n}}.\end{aligned}$$

But in view of Lemma 4.1.6, it is enough to show that for any $f\in C[0,1]^2$ the function $t\mapsto\int_{[0,1]^2}f(z)\mu_t^{F,G}(dz)\in C[t_0,1]$. If that is true, then $f\mapsto\int_{[0,1]^2}f(z)\mu_t^{F,G}(dz)$ will be a continuous functional from $C[0,1]^2$ to $C[t_0,1]$.

Given a partition $\mathcal{C}=(C_j)_{j=1}^r\in\mathcal{P}$, we have by hypothesis that the function $t\mapsto\mu_t^{F,G}(C_j)$ is continuous on $[t_0,1]$ for any $1\le j\le r$. Then, for any sequence (t_l) in $[t_0,1]$ converging to t:

$$\begin{aligned}\overline{L}&=\limsup_{l\to\infty}\int f(z)\mu_{t_l}^{F,G}(dz)\le\sum_{j=1}^r\mu_t^{F,G}(C_j)\sup_{z\in C_j}f(z)<\infty\text{ and}\\ \underline{L}&=\liminf_{l\to\infty}\int f(z)\mu_{t_l}^{F,G}(dz)\ge\sum_{j=1}^r\mu_t^{F,G}(C_j)\inf_{z\in C_j}f(z)>-\infty.\end{aligned}$$

Therefore,

$$\begin{aligned}0&\le\overline{L}-\underline{L}\le\sum_{j=1}^r\mu_t^{F,G}(C_j)[\sup_{z\in C_j}f(z)-\inf_{z\in C_j}f(z)]\\ &\le\sum_{j=1}^r\mu_t^{F,G}(C_j)\sup_{z_1,z_2\in C_j}|f(z_1)-f(z_2)|\\ &\le k(t)\omega\{f,\max_j\operatorname{diam}(C_j)\}.\end{aligned}$$

Since the above does not depend on the partition chosen, $\overline{L}=\underline{L}$ and the result follows from the fact that $\frac{m}{n}\to\lambda$ as $n,m\to\infty$. $\diamond$

To prove our next lemma, we will make use of the following result, the proof of which can be found in [9] (corollary, page 215) and therefore is presented without proof.

Lemma 4.1.8 *Let* $q(t)=\sqrt{t}\ln^p(\frac{1}{t})$ *and for arbitrary* $M>0$, $1<2p<r$ *and an integer* $n\ge 1$, *let* $H_{n,M}=\{\sup_{z:F(z)>t_n}\sqrt{n}\frac{|F_n(z)-F(z)|}{q\{F(z)\}}\le M\}$, *where* $t_n=\frac{\ln^r(n)}{n}$. *If*

$\tilde{k}(t) = o\{t^{-1/2}\ln^{-1/2-\epsilon}(\frac{1}{t})\}$ *as* $t \to 0$, *then*

$$\lim_{M\to\infty}\liminf_{n\to\infty} P(H_{n,M}) = 1$$

and as $n \to \infty$,

$$\sup_{z:F(z)>t_n} \frac{|F_n(z) - F(z)|}{F(z)} \leq M \ln^{p-r/2}(n) \to 0$$

when $H_{n,M}$ *is realized.*

In the next lemma we consider the behaviour of $\gamma_{n,m}$ in the neighborhood of the origin.

Lemma 4.1.9 *For arbitrary* $\rho > 0$, *one has:*
i)$\lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} \delta_{n,m}(t) \geq \rho) = 0$.
ii)$\lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} \epsilon_{n,m}(t) \geq \rho) = 0$.
iii)$\lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |\gamma_{n,m}(t)| \geq \rho) = 0$.

Proof: iii)If i and ii hold, then:

$$\begin{aligned}
& \lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |\gamma_{n,m}(t)| \geq \rho) \\
= & \lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |\delta_{n,m}(t) - \epsilon_{n,m}(t)| \geq \rho) \\
\leq & \lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |\delta_{n,m}(t)| \geq \rho) + \lim_{t_0\to 0}\limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |\epsilon_{n,m}(t)| \geq \rho) \\
= & 0.
\end{aligned}$$

ii)The proof of ii is close to that of i (below) and therefore omitted.

i)Choose p and r such that $1 < 2p < r < 1 + 2\epsilon$. Since $P(A) \leq P(A \cap B) + P(B^C)$ for any events A and B, for given $\rho > 0$:

$$P(\sup_{0\leq t\leq t_0} \delta_{n,m}(t) \geq \rho) \leq P(\sup_{0\leq t\leq t_0} \delta_{n,m}(t) \geq \rho, H_{n,M}) + (1 - P(H_{n,M})).$$

Since $1 - P(H_{n,M})$ can be made arbitrarily small by choosing M large enough, it is enough to show that

$$\limsup_{t_0 \to 0} \limsup_{n,m \to \infty} P(\sup_{0 \le t \le t_0} \delta_{n,m}(t) \ge \rho, H_{n,M}) = 0.$$

Proceeding as in [9], first note that $\{F_n(Y_i) \le t < F(Y_i)\} = \{t < F(Y_i) \le t + (F_n - F)^-(Y_i)\}$. Then, observe that if $H_{n,M}$ is realized, we have that $(F_n - F)^-(z) \le |F_n - F|(z) \le \frac{Mq\{F(z)\}}{\sqrt{n}}$. Therefore, the event $H_{n,M} \cap \{t < F(Y_i) \le t + (F_n - F)^-(Y_i)\} \cap \{F(Y_i) \ge t_n\}$ is equal to the event $H_{n,M} \cap \{t < F(Y_i) \le t + \frac{M}{\sqrt{n}} q[F(Y_i)]\} \cap \{F_n(Y_i) \le t < F(Y_i)\} \cap \{F(Y_i) \ge t_n\}$.

Now take n sufficiently large such that $M \ln^{p-r/2}(n) \le \frac{1}{2}$. It follows from Lemma 4.1.8 that $\sup_{z:F(z)>t_n} |\frac{F_n(z)}{F(z)} - 1| = \sup_{z:F(z)>t_n} \frac{|F_n(z) - F(z)|}{F(z)} \le \frac{1}{2}$, implying that $\frac{F_n(Y_i)}{F(Y)} \ge \frac{1}{2}$, i.e. $H_{n,M} \cap \{F(Y_i) \ge t_n\} \subseteq \{\frac{F_n(Y_i)}{F(Y_i)} \ge \frac{1}{2}\}$. It then follows that

$$\begin{aligned} & H_{n,M} \cap \{F_n(Y_i) \le t < F(Y_i)\} \cap \{F(Y_i) \ge t_n\} \\ \subseteq \; & \{F_n(Y_i) \le t < F(Y_i)\} \cap \{\frac{F_n(Y_i)}{F(Y_i)} \ge \frac{1}{2}\} \\ = \; & \{\frac{F(Y_i)}{2} \le t < F(Y_i)\} \subseteq \{F(Y_i) \le 2t\}. \end{aligned}$$

Putting these two facts together, we get that $H_{n,M} \cap \{t < F(Y_i) \le t + (F_n - F)^-(Y_i)\} \cap \{F(Y_i) \ge t_n\} \subset \{t < F(Y_i) \le t + \frac{M}{\sqrt{n}} q(2t)\}$, an inclusion that will be used in the following set of inequalities.

The following inequalities hold when $H_{n,M}$ is realized:

$$\begin{aligned} \delta_{n,m}(t) &= \frac{1}{\sqrt{m}} \sum_{i=1}^{m} I\{t < F(Y_i) \le t + (F_n - F)^-(Y_i)\} \\ &\le \frac{1}{\sqrt{m}} \sum_{i=1}^{m} I\{t < F(Y_i) \le t + (F_n - F)^-(Y_i)\} I\{F(Y_i) \ge t_n\} \\ &+ \frac{1}{\sqrt{m}} \sum_{i=1}^{m} I\{F(Y_i) \le t_n\} \\ &\le \frac{1}{\sqrt{m}} \sum_{i=1}^{m} I\{t < F(Y_i) \le t + \frac{M}{\sqrt{n}} q(2t)\} + \frac{1}{\sqrt{m}} \sum_{i=1}^{m} I\{F(Y_i) \le t_n\} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{m}}\sum_{i=1}^{m} I\{F(Y_i) \le t + \frac{M}{\sqrt{n}}q(2t)\} - \frac{1}{\sqrt{m}}\sum_{i=1}^{m} I\{F(Y_i) \le t\} \\
&+ \frac{1}{\sqrt{m}}\sum_{i=1}^{m} I\{F(Y_i) \le t_n\} \\
&= [\beta_m(t + \frac{M}{\sqrt{n}}q(2t)) + \sqrt{m}K(t + \frac{M}{\sqrt{n}}q(2t))] - [\beta_m(t) + \sqrt{m}K(t)] \\
&+ [\beta_m(t_n) + \sqrt{m}K(t_n)] \\
&= [\beta_m(t_n) + \sqrt{m}K(t_n)] + [\beta_m(t + \frac{M}{\sqrt{n}}q(2t)) - \beta_m(t)] \\
&+ \sqrt{m}[K(t + \frac{M}{\sqrt{n}}q(2t)) - K(t)].
\end{aligned}
$$

If we define

$$
\wp^6_{n,m} = \sup_{0\le t\le t_0} |\beta_m(t + \frac{M}{\sqrt{n}}q(2t)) - \beta_m(t)| \text{ and}
$$

$$
\wp^7_{n,m} = \sup_{0\le t\le t_0} \sqrt{m}[K(t + \frac{M}{\sqrt{n}}q(2t)) - K(t)],
$$

then

$$
\begin{aligned}
P(\sup_{0\le t\le t_0} \delta_{n,m}(t) \ge \rho, H_{n,M}) &\le P(|\beta_m(t_n)| \ge \frac{\rho}{4}) + P(\sqrt{m}K(t_n) \ge \frac{\rho}{4}) \\
&+ P(\wp^6_{n,m} \ge \frac{\rho}{4}) + P(\wp^7_{n,m} \ge \frac{\rho}{4})
\end{aligned}
$$

and it remains to show that each term can be made arbitrarily small when n, m are large and t_0 is sufficiently small.

Recall Comment 4.1.5: β_m converges in distribution to a process β which is continuous and vanishes at the origin, so that the first term is taken care of.

We can rewrite the second term as:

$$
\begin{aligned}
\sqrt{m}K(t_n) &= \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}K(t_n) \\
&= \frac{\sqrt{m}}{\sqrt{n}} o\{\frac{\sqrt{nt_n}}{\ln^{\frac{1}{2}+\epsilon}(1/t_n)}\} \\
&= \frac{\sqrt{m}}{\sqrt{n}} o\{\ln^{(r-1)/2-\epsilon}(n)\} \to 0.
\end{aligned}
$$

The second equality above follows from assumption 1 and the convergence follows from the fact that $\frac{\sqrt{m}}{\sqrt{n}} \to \sqrt{\lambda}$.

Since β_m is tight, as $n, m \to \infty$ $\wp^6_{n,m} \to 0$.

Now let $\kappa_{t_0,M} = \sup_{0<t\leq t_0+Mq(2t_0)} k(t)q(t)$ and rewrite $\wp^7_{n,m}$ as:

$$\begin{aligned}
\wp^7_{n,m} &= \sup_{0\leq t\leq t_0} \sqrt{m}[K(t+\frac{M}{\sqrt{n}}q(2t)) - K(t)] \\
&= \sqrt{m} \sup_{0\leq t\leq t_0} \int_t^{t+\frac{M}{\sqrt{n}}q(2t)} k(s)ds \\
&\leq \sqrt{m}\kappa_{t_0,M} \sup_{0\leq t\leq t_0} \int_t^{t+\frac{M}{\sqrt{n}}q(2t)} \frac{1}{q(s)}ds \\
&\leq \frac{\sqrt{m}}{\sqrt{n}} M\kappa_{t_0,M} \sup_{0\leq t\leq t_0} \frac{q(2t)}{q(t)} \\
&\leq \frac{\sqrt{m}}{\sqrt{n}} M\kappa_{t_0,M}\sqrt{2} \to_{t_0\to 0} 0.
\end{aligned}$$

Therefore, $\lim_{t_0\to 0} \limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} \delta_{n,m}(t) \geq \rho) = 0$. $\diamond$

We are now in a position to prove Theorem 4.1.3, which deals with the asymptotic behaviour of $\alpha_{n,m}$.

Proof of theorem 4.1.3: Since we have established that $\alpha_{n,m}(t) = \beta_m(t) + \gamma_{n,m}(t)$, given $0 < t_0 \leq 1$, it follows from Comment 4.1.5 and Lemma 4.1.7 that $\alpha_{n,m}(t)$ converges weakly to a continuous process $\alpha(t)$ that may be represented as in equation (4.1.1) for $t_0 \leq t \leq 1$. From Lemma 4.1.9 and the fact that the limit β of the sequence (β_m) is continuous with mean zero, we may conclude that

$$\begin{aligned}
&\lim_{t_0\to 0} \limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |\alpha_{n,m}(t)| \geq \rho) \\
&= \lim_{t_0\to 0} \limsup_{n,m\to\infty} P(\sup_{0\leq t\leq t_0} |(\beta_m + \gamma_{n,m})(t)| \geq \rho) = 0 \ \forall\ \rho > 0.
\end{aligned}$$

Therefore, $\alpha_{n,m}(t)$ converges weakly to $\alpha(t)$. It remains to check the covariance.

$$\Gamma(s,t) = E(\alpha(t)\alpha(s)) = E(U^G(\xi^F_t)U^G(\xi^F_s)$$

$$
\begin{aligned}
&- \sqrt{\lambda}\int U^G(\xi_t^F)U^F(z')\mu_s^{F,G}(dz') - \sqrt{\lambda}\int U^G(\xi_s^F)U^F(z)\mu_t^{F,G}(dz)\\
&+ \lambda\int\int U^F(z)U^F(z')\mu_t^{F,G}(dz)\mu_s^{F,G}(dz'))\\
&= E(U^G(\xi_t^F)U^G(\xi_s^F)) + \lambda\int\int E(U^F(z)U^F(z'))\mu_t^{F,G}(dz)\mu_s^{F,G}(dz')\\
&= G(\xi_t^F \cap \xi_s^F) - G(\xi_t^F)G(\xi_s^F)\\
&+ \lambda\int\int (F(z\wedge z') - F(z)F(z'))\mu_t^{F,G}(dz)\mu_s^{F,G}(dz').
\end{aligned}
$$

◇

The following theorem is Theorem 5 in [9]. We will state it without proof and with minor changes to be consistent with the notation we have been using.

Theorem 4.1.10 *Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from a bivariate distribution function F satisfying Assumptions 1 and 2 (with $F = G$). If $\tilde{K}$ denotes the distribution function of $F(X)$ and $\tilde{K}_n$ its empirical counterpart, then the process $\alpha_n(t) = \sqrt{n}[\tilde{K}_n(t) - \tilde{K}(t)]$ converges weakly to a continuous centered Gaussian process α. This limiting process has the following representation:*

$$\alpha(t) = U^F(\xi_t^F) - \int_{[0,1]^2} U^F(z)\mu_t^{F,F}(dz), t \in [0,1].$$

We can use Theorem 4.1.3 and Theorem 4.1.10 to show the convergence of the process $\psi_{n,m}(t) = \sqrt{m}[G_m(\xi_t^{F_n}) - F_n(\xi_t^{F_n})]$ under $H_0 : F = G$.

Theorem 4.1.11 *Under the null hypothesis $H_0 : F = G$, and assumptions 1 and 2, the process $\psi_{n,m}(t) = \sqrt{m}[G_m(\xi_t^{F_n}) - F_n(\xi_t^{F_n})]$ converges weakly to a centered, continuous Gaussian process which is equal in distribution to $\psi(t) = \sqrt{1+\lambda}U^F(\xi_t^F)$.*

Proof: We can rewrite $\psi_{n,m}(t)$ as:

$$
\begin{aligned}
\psi_{n,m}(t) &= \sqrt{m}[G_m(\xi_t^{F_n}) - F_n(\xi_t^{F_n})]\\
&= \sqrt{m}[G_m(\xi_t^{F_n}) - F_n(\xi_t^{F_n}) - G(\xi_t^{F_n}) + G(\xi_t^{F_n}) - F(\xi_t^{F_n}) + F(\xi_t^{F_n})]
\end{aligned}
$$

$$= \underbrace{\sqrt{m}[G_m(\xi_t^{F_n}) - G(\xi_t^F)]}_{(1)} - \underbrace{\frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}[F_n(\xi_t^{F_n}) - F(\xi_t^F)]}_{(2)}$$
$$+ \underbrace{\sqrt{m}[G(\xi_t^F) - F(\xi_t^F)]}_{(3)}.$$

Using Theorem 4.1.3, (1) $\rightarrow U^G(\xi_t^F) - \sqrt{\lambda}\int_{[0,1]^2} U^F(z)\mu_t^{F,G}(dz)$. Using Theorem 4.1.10 and the fact that $\frac{m}{n} \rightarrow \lambda$, (2) $\rightarrow \sqrt{\lambda}U^F(\xi_t^F) - \sqrt{\lambda}\int_{[0,1]^2} U^F(z)\mu_t^{F,F}(dz)$.

To simplify the result when $H_0 : F = G$ is true, $\overline{U}^F$ will denote the bridge associated with $\sqrt{m}[G_m(\xi_t^F) - G(\xi_t^F)]$. Note that although $\overline{U}^F$ and U^F will have the same distribution, they will remain independent. Also note that under H_0, (3) = 0.

Then, under $H_0 : F = G$, we have that:

$$\begin{aligned}
\psi_{n,m}(t) &\rightarrow \overline{U}^F(\xi_t^F) - \sqrt{\lambda}\int_{[0,1]^2} U^F(z)\mu_t^{F,F}(dz) - \sqrt{\lambda}U^F(\xi_t^F) \\
&+ \sqrt{\lambda}\int_{[0,1]^2} U^F(z)\mu_t^{F,F}(dz) \\
&= \overline{U}^F(\xi_t^F) - \sqrt{\lambda}U^F(\xi_t^F) \\
&=_{\mathcal{D}} \sqrt{1+\lambda}U^F(\xi_t^F).
\end{aligned}$$

◇

Applications of Theorem 4.1.11 will be discussed in the next section.

What we want to do now, following Theorem 2 in [9], is to show that Assumption 2 is not as restrictive as it might seem at first glance. We have the following theorem.

Theorem 4.1.12 *Suppose that F is a copula with a continuous and strictly positive density f, and that G is a distribution on $[0,1]^2$ with continuous density g. Then there exists a version of the conditional distribution of $Y^* = Y$ given $F(Y) = t$ such that for any rectangle C in $[0,1]^2$, the mapping $t \rightarrow \mu_t^{F,G}(C) = k(t)P\{Y \in C|F(Y) = t\}$ is continuous on $(0,1]$ and $\mu_1^{F,G}$ is the Dirac measure with mass $k(1)$ at point $(1,1)$. Moreover, for any Borel set C in $[0,1]^2$ and for any $0 < t < 1$, $\mu_t^{F,G}(C)$ admits the* *representation*

$$\mu_t^{F,G}(C) = \int_{(0,1)} I\{(x, R_x(t)) \in C\} g_t^F(x) dx,$$

where

$$R_x(t) = F_x^-(t) = \{y : F(x,y) = t\}$$

and

$$g_t^F(x) = [\frac{\partial}{\partial t} R_x(t)] g(x, R_x(t)) I\{x \in (0,1) : F(x,1) > t\}.$$

In particular, $k(t) = \int_{(0,1)} g_t^F(x) dx$.

Comment 4.1.13 Note that the mapping $t \longrightarrow \mu_t^{F,G}(C) = k(t)P\{Y \in C | F(Y) = t\}$ is only required to be continuous on $(0, 1]$. The reason that we do not ask for continuity at 0 is that the proof will make use of the transformation $(x, y) \longrightarrow (x, F(x, y))$ and its inverse. The inverse is unique whenever $F(x, y) \neq 0$, but if we take $F(x, y) = 0$ and $x = 0$, then $(0, y) \longrightarrow (0, 0)\ \forall\ y$ and the inverse does not exist.

Proof: Following [9], the proof will consist of two parts: the first is to show that a version of the conditional distribution of Y given $F(Y) = t$ can be found such that for any continuous j on $[0, 1]^2$, the mapping $t \longrightarrow m_t(j) = \int_{(0,1)^2} j(z) \mu_t^{F,G}(dz)$ is continuous on $(0, 1]$. The second part will be to show that for any non-empty rectangle $C = \{z \in [0, 1]^2 : z_1 < z \leq z_2\}$, the function $t \longrightarrow \mu_t^{F,G}(C)$ is continuous on $(0, 1]$.

i)Take a continuous function j on $[0, 1]^2$ and for $t \in [0, 1]$ define

$$M_t(j) = E(j(Y) I\{F(Y) \leq t\}).$$

Then,

$$M_t(j) = \int_{(0,1)} \int_{(0,1)} j(x, y) g(x, y) I\{F(x, y) \leq t\} dy dx.$$

If we apply the change of variable $s = F(x, y) = F_x(y)$ for fixed x such that $F(x, 1) > s$, we have that $y = R_x(s)$ and $\frac{\partial y}{\partial s} = \frac{\partial}{\partial s} R_x(s) = R_x'(s)$. Note that since F_x is strictly increasing, its inverse $R_x(s)$ is unique and that R_x' is continuous since f is. If we let

$g_s^F(x) = [\frac{\partial}{\partial s} R_x(s)] g(x, R_x(s)) I\{x \in (0,1) : F(x,1) > s\}$, we have that:

$$\begin{aligned} M_t(j) &= \int_{(0,1)} (\int_{(0,1)} j(x, F_x^-(s)) g(x, F_x^-(s)) I\{s \leq t\} R_x^{'}(s) \\ &\qquad I\{x : F(x,1) > s\} ds) dx \\ &= \int_0^t (\int_{(0,1)} j(x, R_x(s)) g(x, R_x(s)) R_x^{'}(s) I\{x : F(x,1) > s\} dx) ds \\ &= \int_0^t (\int_{(0,1)} j(x, R_x(s)) g_s^F(x) dx) ds. \end{aligned}$$

On the other hand, applying the definition of $\mu_t^{F,G}$, we have that for $t \in [0,1]$:

$$\begin{aligned} M_t(j) &= E(j(Y) I\{F(Y) \leq t\}) \\ &= \int_{(0,1)^2} j(z) I\{F(z) \leq t\} dG(z) \\ &= \int_0^1 \int_{(0,1)^2} j(z) k(s) P((Y^1, Y^2) \in dz | F(Y) = s) I\{s \leq t\} dz ds \\ &= \int_0^t \int_{(0,1)^2} j(z) \mu_s^{F,G}(dz) ds. \end{aligned}$$

From the above we deduce that $\frac{\partial M_t}{\partial t}(j) = m_t(j) = \int_{(0,1)^2} j(z) \mu_t^{F,G}(dz)$. Also, note that

$$\int_{(0,1)^2} j(z) \mu_t^{F,G}(dz) = \int_{(0,1)} j(x, R_x(t)) g_t^F(x) dx \tag{4.1.2}$$

must be true for any bounded measurable function j.

Therefore, $\mu_t^{F,G}$ is a version of $k(t)$ times the conditional distribution of Y given $F(Y) = t$ for which the mapping

$$t \to m_t(j) = \int_{(0,1)^2} j(z) \mu_t^{F,G}(dz) = \int_{(0,1)} j(x, R_x(t)) g_t^F(x) dx$$

is continuous on $(0,1)$ for all continuous functions j on $(0,1)^2$, because j and g are both continuous and the derivative of R_x is continuous on $(0,1)$. Taking j identically equal to one we obtain that for $t \in (0,1)$,

$$M_t(1) = E(I\{F(Y) \leq t\}) = P(F(Y) \leq t) = G(\xi_t^F)$$

and

$$m_t(1) = \int_{(0,1)} g_t^F(x)dx = k(t).$$

Next, observe that $F(y) = 1 \Leftrightarrow y = (1,1)$. Therefore $E(j(Y)|F(Y) = 1) = j(1,1)$ and

$$\begin{aligned} m_1(j) &= \int_{(0,1)^2} j(z)\mu_1^{F,G}(dz) \\ &= \int_{(0,1)^2} j(z)k(1)P((Y^1,Y^2) \in dz|F(Y) = 1) \\ &= j(1,1)k(1)\int_{(0,1]^2} P((Y^1,Y^2) \in dz|F(Y) = 1) \\ &= j(1,1)k(1). \end{aligned}$$

To complete the first step of the proof, it remains to show that $m_t(j)$ converges to $m_1(j)$ as $t \to 1$.

$$\begin{aligned} &|m_1(j) - m_t(j)| \\ = \quad & |j(1,1)k(1) - \int_{(0,1)} j(x,R_x(t))g_t^F(x)dx| \\ = \quad & |j(1,1)k(1) - j(1,1)k(t) - \int_{(0,1)} j(x,R_x(t))g_t^F(x)dx \\ + \quad & j(1,1)k(t)| \\ = \quad & |j(1,1)(k(1) - k(t)) - \int_{(0,1)} j(x,R_x(t))g_t^F(x)dx \\ + \quad & \int_{(0,1)} j(1,1)g_t^F(x)dx| \\ \leq \quad & |j(1,1)||k(1) - k(t)| + \int_{(0,1)} |j(x,R_x(t)) - j(1,1)|g_t^F(x)dx \\ \leq \quad & |j(1,1)||k(1) - k(t)| + \sup_{x:F(x,1)>t} |j(x,R_x(t)) - j(1,1)|k(t) \\ \leq \quad & |j(1,1)||k(1) - k(t)| + k(t) \sup_{y\in[t,1]^2} |j(y) - j(1,1)| \\ \to_{t\to 1} \quad & 0. \end{aligned}$$

The above is true since k is continuous on $(0,1]$, j is continuous on $[0,1]^2$ and the last inequality holds from the argument below.

We need to prove that $\sup_{x:F(x,1)>t}|j(x,R_x(t))-j(1,1)| \leq \sup_{y\in[t,1]^2}|j(y)-j(1,1)|$. Note that if $s<t$, then $F(x,s)\leq s<t$ because F is a copula. Then, if $F(x,1)>t$ we have that $R_x(t)>s$ and $R_x(t)\geq t$. Moreover, the fact that $F(x,1)>t$, implies that $x=F(x,1)>t$.

ii)The second part of the proof consists of showing that for fixed $t\in(0,1]$, the mapping $t\rightarrow\mu_t^{F,G}(C)$ is continuous at t for any non-empty rectangle C of the form $\{z\in[0,1]^2: z_1<z\leq z_2\}$.

Following the method presented in [9], we will first deal with the case $t<1$. Consider a rectangle C of the form $\{z\in[0,1]^2: z_1<z\leq z_2\}$. Since the boundary of C is included in a finite union of sets of the form $\{z: z^{(i)}=c\}$, it will be enough to show that $\mu_t^{F,G}\{z: z^{(i)}=c\}=0$ for arbitrary $c\in[0,1]$, $i=1,2$. Then, it will follow that $\mu_t^{F,G}(\partial C)=0$ and so $\mu_t^{F,G}([0,z])$ is continuous in z for every t and $\lim_{s\rightarrow t}\mu_s(C)=\mu_t(C)$.

Since

$$\begin{aligned}\mu_t^{F,G}\{[t,1)^2\} &= k(t)P((Y^1,Y^2)\in[t,1)^2|F(Y)=t)\\ &= k(t),\end{aligned}$$

it is sufficient to consider $c\in(0,1)$. Let π_i be the projection function that omits the ith coordinate: $\pi_1(z)=z_2$ and $\pi_2(z)=z_1$. Since the representation

$$\int_{(0,1)^2} j(z)\mu_t^{F,G}(dz_1\times dz_2)=\int_{(0,1)} j(x,R_x(t))g_t^F(x)dx,$$

is valid for any continuous function j on $[0,1]$, the measure $\mu_t^{F,G}\circ\pi_2^{-1}$ has density $g_t^F(x)$ with respect to Lebesgue measure. Hence, $\mu_t^{F,G}\{z: z^{(1)}=c\}=0$ for arbitrary $c\in(0,1)$.

To show that $\mu_t^{F,G}\{z: z^{(2)}=c\}=0$, let $\partial_2F(x,c)$ denote the partial derivative of $F(z_1,z_2)$ with respect to $z^{(2)}$ evaluated at (x,c). Equation 4.1.2 is valid for any bounded, measurable function j, yielding

$$
\begin{aligned}
&\mu_t^{F,G}\{z \in [0,1]^2 : z^{(2)} = c\} \\
&= \int_{(0,1)^2} I\{z \in [0,1]^2 : z^{(2)} = c\}\mu_t^{F,G}(dz) \\
&= \int_{(0,1)} I\{(x, R_x(t)) \in [0,1]^2 : R_x(t) = c\} g_t^F(x)dx \\
&= \int_{(0,1)} I\{(x, R_x(t)) \in [0,1]^2 : R_x(t) = c\}\frac{\partial}{\partial t}R_x(t) \\
&\qquad g(x, R_x(t))I\{x \in (0,1) : F(x,1) > t\}dx \\
&= \int_{(0,1)} I\{x \in (0,1) : F(x,c) = t\}g(x,c)\frac{1}{F_x'(R_x(t))}dx \\
&= \int_{(0,1)} g(x,c)\frac{1}{\partial_2 F(x,c)}I\{x \in (0,1) : F(x,c) = t\}dx.
\end{aligned}
$$

But $P(I\{x \in (0,1) : F(x,c) = t\} = 1) = 0 \ \forall \ x \in [0,1]$ because F has a strictly positive density, and it follows that $\mu_t^{F,G}\{z \in [0,1]^2 : z^{(2)} = c\} = 0$.

Finally, we consider the case $t = 1$. Take $C = \{z \in [0,1]^2 : z_1 < z \leq z_2\}$ and let $z_2 \neq (1,1)$. Then $(1,1) \notin \partial C$, and it follows from the continuity of $\mu_1^{F,G}$ that $\mu_1^{F,G}(\partial C) = k(1)I\{(1,1) \in \partial C\} = 0$. Hence $\mu_s^{F,G}(C) \to \mu_1^{F,G}(C)$ as $s \to 1$.

Now let $z_2 = (1,1)$. We have that whenever $s > \max_{i=1,2} z_1^{(i)}$,

$$\mu_1(C) = k(1) \text{ and } \mu_s(C) = k(s)P((Y^1, Y^2) \in C | F(Y) = s) = k(s),$$

because for such an s, $F(x,1) > s$ implies that $(x, R_x(s)) \subset [s,1]^2 \subset C$. Since k is continuous we have that $k(s) \to k(1)$ as $s \to 1$ and the proof is complete. $\diamond$

Comment 4.1.14 We are requiring an extra assumption that is not needed in [9]: that f be a strictly positive density. We need this assumption to be able to prove that $\mu_t^{F,G}\{z : z^{(2)} = c\} = 0$; in the one-sample case, studied in [9], it is not necessary because when we express $\mu_t^{F,F}\{z : z^{(2)} = c\}$ as an integral, it is immediate that the indicator function inside the integral will be greater than zero only when the density f is equal to zero, yielding that part of the proof .

4.2 Applications

In this section it is our aim to discuss potential applications of Theorem 4.1.3, and in particular to focus on the different scenarios that are created based on the information that is available to the experimenter.

Recall that Theorem 4.1.11 in the previous section showed that the process $\psi_{n,m}(t) = \sqrt{m}[G_m(\xi_t^{F_n}) - F_n(\xi_t^{F_n})]$ converges, under H_0, to a continuous Gaussian process which is equal in distribution to $\psi(t) = \sqrt{1+\lambda}U^F(\xi_t^F)$. To point out that this process will lead to consistent test statistics, we make the following observation.

Comment 4.2.1 *If $H_0 : F = G$ is not true, then*

$$\sup_p \sqrt{m}|G(\xi_p^F) - F(\xi_p^F)| \to \infty,$$

and therefore

$$\sup_p \psi_{n,m}(p) \to_P \infty.$$

Before talking about the possible test statistics to use, let us show with an example, as in [9], that in order to verify that our distributions comply with assumption 1, it is not necessary to calculate $k(t)$ and $\tilde{k}(t)$ explicitly.

Example 4.2.2 The Farlie-Gumbel-Morgenstern (FGM) family of copulas is given by $C(u,v) = uv + \theta uv(1-u)(1-v)$, where $-1 \leq \theta \leq 1$. Take F to be the FGM copula with parameter θ and G to be the FGM copula with parameter ε, so that:

$$\begin{aligned}
F(x,y) &= xy + \theta xy(1-x)(1-y) \\
f(x,y) &= 1 + \theta(1-2x)(1-2y) \\
G(x,y) &= xy + \varepsilon xy(1-x)(1-y) \\
g(x,y) &= 1 + \varepsilon(1-2x)(1-2y).
\end{aligned}$$

As defined before,

$$R_x(t) = F_x^-(t) = \{y : F(x,y) = t\}$$

and

$$g_t^F(x) = [\frac{\partial}{\partial t} R_x(t)] g(x, R_x(t)) I\{x \in (0,1) : F(x,1) > t\}.$$

Since $F(x,y)$ can be rewritten as $F(x,y) = y^2(\theta x^2 - \theta x) + y(\theta x - \theta x^2 + x)$, it is easy to see that $R_x(t) = \{y : y^2(\theta x - \theta x^2) + y(\theta x^2 - \theta x - x) + t = 0\}$. Therefore, following [9] by letting $c_x = \theta(1-x)$ and $r_x(t) = \sqrt{(1+c_x)^2 - 4c_x(\frac{t}{x})}$, we have that:

$$\begin{aligned}
R_x(t) &= \frac{\theta x + x - \theta x^2 - \sqrt{(\theta x^2 - \theta x - x)^2 - 4(\theta x - \theta x^2)(t)}}{2(\theta x - \theta x^2)} \\
&= \frac{x(\theta + 1 - \theta x) - \sqrt{[x(\theta x - \theta - 1)]^2 - 4t\theta x^2(\frac{1}{x} - 1)}}{2\theta x(1-x)} \\
&= \frac{\theta(1-x) + 1 - \sqrt{[\theta(x-1) - 1]^2 - 4t\theta(\frac{1-x}{x})}}{2\theta(1-x)} \\
&= \frac{c_x + 1 - \sqrt{[-c_x - 1]^2 - 4tc_x(\frac{1}{x})}}{2c_x} \\
&= \frac{1 + c_x - \sqrt{(1+c_x)^2 - 4c_x(\frac{t}{x})}}{2c_x} \\
&= \frac{1 + c_x - r_x}{2c_x}.
\end{aligned}$$

Since $\frac{\partial R_x(t)}{\partial t} = \frac{1}{2c_x}\frac{1}{2}[(1+c_x)^2 - 4c_x(\frac{t}{x})]^{-\frac{1}{2}}\frac{4c_x}{x} = \frac{1}{x\sqrt{(1+c_x)^2 - 4c_x(\frac{t}{x})}}$, we have that:

$$\begin{aligned}
g_t^F(x) &= \frac{1 + \varepsilon(1-2x)(1 - \frac{1+c_x-r_x}{c_x})}{x\sqrt{(1+c_x)^2 - 4c_x(\frac{t}{x})}} I\{x \in (0,1) : F(x,1) > t\} \\
&= \frac{(1-x) + \frac{\varepsilon}{\theta}\theta(1-x)(1-2x)(1 - \frac{1+c_x-r_x}{c_x})}{xr_x(1-x)} I\{x \in (0,1) : F(x,1) > t\} \\
&= \frac{(1-x) + \frac{\varepsilon}{\theta}c_x(1-2x)(\frac{c_x-1-c_x+r_x}{c_x})}{xr_x(1-x)} I\{x \in (0,1) : F(x,1) > t\} \\
&\leq \frac{(1-x) + \delta(1-2x)(r_x - 1)}{xr_x(1-x)} I\{x \in (0,1) : F(x,1) > t\}, \text{where } \delta = (\frac{\varepsilon}{\theta} \vee 1) \\
&\leq \delta[\frac{(1-x) + (1-2x)(r_x - 1)}{xr_x(1-x)}] I\{x \in (0,1) : F(x,1) > t\} \\
&= \delta[\frac{(1-x) + (r_x - 1 - 2xr_x + 2x)}{xr_x(1-x)}] I\{x \in (0,1) : F(x,1) > t\} \\
&= \delta[\frac{r_x - 2xr_x + x}{x(1-x)r_x}] I\{x \in (0,1) : F(x,1) > t\}
\end{aligned}$$

$$\begin{aligned}
&= \delta[\frac{x(1-r_x)+(1-x)r_x}{x(1-x)r_x}]I\{x \in (0,1) : F(x,1) > t\} \\
&= \delta[\frac{1-r_x}{(1-x)r_x} + \frac{1}{x}]I\{x \in (0,1) : F(x,1) > t\} \\
&= \delta[\frac{1-r_x^2}{(1-x)r_x(1+r_x)} + \frac{1}{x}]I\{x \in (0,1) : F(x,1) > t\} \\
&\leq \delta[\frac{1+2|\theta|}{x}]I\{x \in (0,1) : F(x,1) > t\}.
\end{aligned}$$

The first two inequalities follow from the fact that, depending on the (one-sided) alternative we are using, δ will be either $\frac{\epsilon}{\theta}$ or 1 and the last inequality ($1 - r_x^2 \leq 2|\theta|(1-x)$) follows from [9], page 203.

Thus we can see that

$$\begin{aligned}
k(t) &= \int_0^1 g_t^F(x)dx \leq \delta(1+2|\theta|)\int_t^1 \frac{1}{x}dx \\
&= \delta(1+2|\theta|)(-\ln t) \\
&= \delta(1+2|\theta|)\ln\frac{1}{t}.
\end{aligned}$$

This is enough because we have that $k(t) = o\{t^{-1/2}\ln^{-1/2-\epsilon}(1/t)\}\ \forall -1 \leq \theta \leq 1$ and arbitrary ϵ. The procedure to check assumption 1 for $\tilde{k}(t)$ is similar and developed fully in [9]. Thus, although explicit forms for $k(t)$ and $\tilde{k}(t)$ cannot be found, the FGM family of copulas verifies assumption 1.

Now we can consider applications of Theorem 4.1.3. We will be discussing two scenarios: complete and partial samples.

•Complete samples known.

We have completed our experiment and recorded the values of the observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$. We are working with the process $\psi_{n,m}(p) = \sqrt{m}[G_m(\xi_p^{F_n}) - F_n(\xi_p^{F_n})] = \sqrt{m}[K_{n,m}(p) - \tilde{K}_n(p)]$. In the previous section, Theorem 4.1.11 proved that under $H_0 : F = G$ this process converges weakly to a continuous Gaussian process with mean zero and covariance $(1+\lambda)F(\xi_s^F \cap \xi_t^F) - F(\xi_s^F)F(\xi_t^F)$, i.e. its variance is given by $(1+\lambda)F(\xi_p^F)(1-F(\xi_p^F)) = (1+\lambda)\tilde{K}(p)(1-\tilde{K}(p))$. Thus, it is straightforward that the process $\psi_{n,m}$ converges under H_0 to a process which is equal in distribution to $\sqrt{1+\lambda}U^{\tilde{K}}(p)$. Since $\tilde{K}$ is a continuous increasing function, it is also true that $\sup_p U^{\tilde{K}}(p) =_{\mathcal{D}} \sup_p U(\tilde{K}(p)) =_{\mathcal{D}} \sup_p U(p)$, where U is the usual Brownian bridge on $[0,1]$.

If we are testing $H_0 : F = G$ against $H_1 : F \prec^{K^F} G$ (i.e. $H_0 : X =_{\mathcal{D}} Y$ vs $H_1 : Y \prec^{K^F} X$), an appropriate test statistic would be $V_{n,m}^+ \equiv \sup_p \sqrt{m}[G_m(\xi_p^{F_n}) - F_n(\xi_p^{F_n})]$ which, from the argument above, converges under $H_0 : F = G$ to $\sqrt{1+\lambda}\sup_p U(p)$. Similarly, for the alternative $H_2 : X \prec^{K^F} Y$ we use the test statistic $V_{n,m}^- \equiv \sup_p \sqrt{m}[F_n(\xi_p^{F_n}) - G_m(\xi_p^{F_n})] \to_{\mathcal{D},H_0} \sqrt{1+\lambda}\sup_p U(p)$. Finally, if our alternative is two-sided, we can use the statistic $|V_{n,m}(p)| \equiv \sup_p \max(V_{n,m}^+(p), V_{n,m}^-(p))$ that will converge under $H_0 : F = G$ to $\sqrt{1+\lambda}\sup_p |U(p)|$.

Now that we have identified our test statistics, it remains to calculate the corresponding critical values c_α (for different levels α) to know when to reject H_0. Let us start with the one-sided alternative $H_1 : Y \prec^{K^F} X$; the alternative $H_2 : X \prec^{K^F} Y$ is treated similarly.

We want to find a value c_α such that $P(\sqrt{1+\lambda}\sup_p U(p) \geq c_\alpha) \leq \alpha$. Since it is known that $P(\sup_p U(p) > x) = \exp\{-2x^2\}$ $\forall x \geq 0$ (see [43], page 142), we want to find c_α such that

$$\begin{aligned} & P(\sqrt{1+\lambda}\sup_p U(p) \geq c_\alpha) = \alpha \\ \Leftrightarrow \quad & P(\sup_p U(p) \geq \frac{c_\alpha}{\sqrt{1+\lambda}}) = \alpha \\ \Leftrightarrow \quad & \exp\{-2(\frac{c_\alpha}{\sqrt{1+\lambda}})^2\} = \alpha \\ \Leftrightarrow \quad & -2(\frac{c_\alpha}{\sqrt{1+\lambda}})^2 = \ln\alpha \\ \Leftrightarrow \quad & c_\alpha = \sqrt{-\frac{(1+\lambda)\ln\alpha}{2}}. \end{aligned}$$

For testing $H_0 : F = G$ against $H_3 : F \neq G$, we need to find the critical values

c_α such that $P(\sqrt{1+\lambda}\sup_p |U(p)| \geq c_\alpha)$, or $P(\sup_p |U(p)| \geq \frac{c_\alpha}{\sqrt{1+\lambda}})$. It is known that $P(\sup_p |U(p)| > x) \equiv 1 - L(x) = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp\{-2k^2x^2\}\ \forall\ x \geq 0$, so we are looking for c_α such that $2\sum_{k=1}^{\infty}(-1)^{k+1}\exp\{-2k^2(\frac{c_\alpha}{\sqrt{1+\lambda}})^2\} = \alpha$. Alternatively, tables for $L(x)$, for several values of x (ranging from .28 to 3.00) can be found in [43].

Comment 4.2.3 Although this document is intended to focus on one-sided alternatives, we are also presenting a statistic to use when we are testing $H_0 : F = G$ versus $H_3 : F \neq G$. However, for the two-sided alternative, the Cramer-Von Mises statistic $\int_0^1 \psi_{n,m}^2(p)dp$ might be a much better choice in terms of power than the statistic that we are suggesting. This is a topic for further study.

Comment 4.2.4 An important property of these tests that should be highlighted, is that they are all distribution free, i.e. they do not depend on the underlying distributions of our samples.

•Partial samples known.

In this scenario we have already completed our experiment but we only know the values of some of our observations, in particular, the process $\psi_{n,m}(p) = \sqrt{m}[G_m(\xi_p^{F_n}) - F_n(\xi_p^{F_n})] = \sqrt{m}[K_{n,m}(p) - \tilde{K}(p)]$ can only be observed over the region $0 \leq p \leq p_0$.

We will start to describe the procedure when the alternative is $H_1 : Y \prec^{K^F} X$ (resp. $H_2 : X \prec^{K^F} Y$). Since we don't have complete samples, we can no longer use the statistic $V_{n,m}^{+}$ ($V_{n,m}^{-}$ resp.), but we can define a similar test statistic: $V_{n,m}^{*,+}(p_0) \equiv \sup_{p\leq p_0}\sqrt{m}[G_m(\xi_p^{F_n}) - F_n(\xi_p^{F_n})]$ ($V_{n,m}^{*,-}(p_0) \equiv \sup_{p\leq p_0}\sqrt{m}[F_n(\xi_p^{F_n}) - G_m(\xi_p^{F_n})]$ resp.). Under $H_0 : F = G$, $V_{n,m}^{*,+}(p_0) \to_{\mathcal{D}} \sqrt{1+\lambda}\sup_{p\leq p_0}(U^{\tilde{K}})(p)$ (resp. $V_{n,m}^{*,-}(p_0) \to_{\mathcal{D}} \sqrt{1+\lambda}\sup_{p\leq p_0}(U^{\tilde{K}})(p)$). In a similar way, if we are testing $H_0 : F = G$ against $H_3 : F \neq G$, we can define our test statistic as $|V_{n,m}^{*}(p_0)| \equiv \sup_{p\leq p_0}\max(V_{n,m}^{*,+}(p), V_{n,m}^{*,-}(p))$; it will converge under H_0 to $\sqrt{1+\lambda}\sup_{p\leq p_0}|U^{\tilde{K}}(p)| =_{\mathcal{D}} \sqrt{1+\lambda}\sup_{p\leq p_0}U(\tilde{K}(p)) =_{\mathcal{D}} \sqrt{1+\lambda}\sup_{p\leq \tilde{K}(p_0)}U(p)$.

If we know what $\tilde{K}$ is explicitly, the critical values c_α can be found without any further problems, see [14]. Here, we can use Corollary I.3.1 (page 9 of [14]), which gives us a formula that we can use to calculate $\sup_{p\leq\tilde{K}(p_0)} U(p)$, used to find c_α when we have a one-sided alternative. We can also find tables ([14], pages 22 through 34) that calculate the critical values c_α for the two-sided alternative, for different values of $\tilde{K}(p_0)$ $(.1, .2, \ldots, 1)$ and different values of α $(.01, .025, .05, .1, .2, \ldots, .9)$.

However, in practice, the usual scenario is that $\tilde{K}$ is unknown. One way to solve this problem is to observe that the $*-$test statistics defined above will always be at most equal to their complete samples counterpart. For example, since it is true that $\sup_{p\leq\tilde{K}(p_0)} U(p) \leq \sup_p U(p)$, it is also true that $P(\sqrt{1+\lambda}\sup_{p\leq\tilde{K}(p_0)} U(p) \geq c_\alpha) \leq P(\sqrt{1+\lambda}\sup_p U(p) \geq c_\alpha) \leq \alpha$. Thus, we can use the critical values found in the complete sample scenario, though it must be said that they will be a conservative choice.

Another way to solve the problem that $\tilde{K}(p) = F(\xi_p^F)$ is unknown, is to use Theorem 3.2.2, where it was proven that $\sup_{0\leq p\leq 1} |F_n(\xi_p^{F_n}) - F(\xi_p^F)| \rightarrow_{a.s.} 0$. Since U is a.s. continuous, we can just replace $\tilde{K}(p) = F(\xi_p^F)$ with its empirical counterpart $\tilde{K}_n(p) = F_n(\xi_p^{F_n})$ for $0 \leq p \leq p_0$, and proceed as if $\tilde{K}(p)$ was known to approximate the critical values c_α.

To end the chapter we will describe a couple of scenarios in which the alternative $H_1 : Y \prec^{K^F} X$ would be appropriate. As a first example, take two distributions F and G with exactly the same dependence structure, i.e. $C^F = C^G$, but with, possibly, different marginals. Our test then becomes a test on the marginals: $H_0 : F_1 = G_1$ and $F_2 = G_2$ versus $H_1 : F_1 < G_1$ or $F_2 < G_2$ or both. Second, suppose that we are working with two distributions F and G with equal marginals and whose dependence structure is given by FGM copulas with unknown parameters θ_F and θ_G. In other words, F and G are the same except for, possibly, the parameters θ_F and θ_G. Under these assumptions, our test will actually become $H_0 : \theta_F = \theta_G$ versus $H_1 : \theta_F < \theta_G$. While it is not true that $\theta_F < \theta_G$ implies that $Y \prec^{K^F} X$, it can be shown by numerical calculation that $F(\xi_p^F) > G(\xi_p^F)$ for all θ_F, θ_G and $p > p_0$, where $.1 \leq p_0 \leq .2$ (see the procedure following this paragraph). Therefore, an appropriate test statistic would be $V^{*;-}_{n,m}(.2) \equiv \sup_{p\geq .2} \sqrt{m}[F_n(\xi_p^{F_n}) - G_m(\xi_p^{F_n})]$. The discussion on critical values preceding this paragraph is easily adapted to this test.

To perform the above mentioned numerical calculations, note that if $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, F(x, y) \leq p\}$, then:

$$\begin{aligned} G(\xi_p^F) - F(\xi_p^F) &= \int_D\int g - f dxdy = (\theta_G - \theta_F)\int_D\int (1-2x)(1-2y)dxdy \\ &= (\theta_G - \theta_F)\int_p^1\int_0^{p(x)} (1-2x)(1-2y)dydx, \end{aligned}$$

where $p(x)$ is given by:

$$p(x) = \begin{cases} \frac{1}{2} - \frac{1}{2\theta_F(x-1)} + \frac{\sqrt{x^2(1+\theta_F - x\theta_F)^2 + 4p\theta_F x(x-1)}}{2x\theta_F(x-1)} & \text{if } p < x < 1, \theta_F \neq 0 \\ p & \text{if } x = 1 \\ \frac{p}{x} & \text{if } \theta_F = 0, x \neq 0 \\ 1 & \text{if } x = p \\ \text{not defined} & \text{otherwise.} \end{cases}$$

The definition of $p(x)$ when $p < x < 1, \theta_F \neq 0$ comes from solving the equation $xy + xy\theta_F(1-x)(1-y) = p$ for y. In Table 4.2 we can find different values of $\int_p^1\int_0^{p(x)}(1-2x)(1-2y)dydx$ for different values of θ_F and p (Thanks to Justin Francis for these calculations).

	p	p	p	p	p	p	p	p	p	p
θ	0 or 1	.1	.2	.3	.4	.5	.6	.7	.8	.9
-1	0	.0005	-.0362	-.0547	-.0592	-.0541	-.0429	-.0289	-.0150	-.0043
-.9	0	.0013	-.0356	-.0545	-.0593	-.0544	-.0432	-.0291	-.0151	-.0043
-.8	0	.0020	-.0349	-.0543	-.0595	-.0546	-.0435	-.0293	-.0152	-.0043
-.7	0	.0027	-.0343	-.0541	-.0596	-.0549	-.0437	-.0294	-.0153	-.0044
-.6	0	.0034	-.0336	-.0538	-.0597	-.0552	-.0440	-.0296	-.0153	-.0044
-.5	0	.0040	-.0329	-.0536	-.0598	-.0554	-.0443	-.0298	-.0154	-.0044
-.4	0	.0045	-.0322	-.0533	-.0599	-.0557	-.0445	-.0300	-.0155	-.0044
-.3	0	.0050	-.0315	-.0530	-.0600	-.0560	-.0448	-.0302	-.0156	-.0044
-.2	0	.0055	-.0307	-.0527	-.0601	-.0563	-.0451	-.0303	-.0156	-.0044
-.1	0	.0059	-.0300	-.0524	-.0602	-.0565	-.0454	-.0305	-.0157	-.0044
0	0	.0063	-.0293	-.0520	-.0602	-.0568	-.0457	-.0307	-.0158	-.0044
.1	0	.0066	-.0286	-.0517	-.0603	-.0571	-.0460	-.0309	-.0159	-.0045
.2	0	.0068	-.0279	-.0513	-.0603	-.0573	-.0463	-.0311	-.0160	-.0045
.3	0	.0070	-.0272	-.0509	-.0603	-.0576	-.0466	-.0313	-.0161	-.0045
.4	0	.0072	-.0265	-.0505	-.0603	-.0579	-.0469	-.0316	-.0161	-.0045
.5	0	.0074	-.0258	-.0500	-.0603	-.0582	-.0472	-.0318	-.0162	-.0045
.6	0	.0075	-.0251	-.0496	-.0603	-.0584	-.0475	-.0320	-.0163	-.0045
.7	0	.0075	-.0244	-.0491	-.0603	-.0587	-.0478	-.0322	-.0164	-.0045
.8	0	.0076	-.0238	-.0486	-.0602	-.0589	-.0481	-.0324	-.0165	-.0046
.9	0	.0076	-.0231	-.0481	-.0602	-.0592	-.0485	-.0327	-.0166	-.0046
1	0	.0076	-.0225	-.0476	-.0601	-.0595	-.0488	-.0329	-.0167	-.0046

Table 4.2: Numerical calculations of $\int_p^1 \int_0^{p(x)} (1-2x)(1-2y)dydx$.

Chapter 5

Asymptotic behaviour of $G_m(F_n^{1-}, F_n^{2-})$ and applications

In this chapter we will study the asymptotic behavior of the $p-p$ plot $G_m(F_n^{1-}, F_n^{2-})$, as well as some applications of this result. This is the appropriate process to explore when the information associated with the data we obtain yields the product filtration. In order to obtain the desired limiting distribution, we will be following [22] closely, where Fermanian, Radulovic and Wegkamp studied the convergence in $\ell^\infty([0,1]^2)$ of an empirical copula process; it is worth mentioning that Gaenssler and Stute (see [25]) had previously proved the weak convergence of the said process in $D([0,1]^2)$.

5.1 Limiting distribution

We will be making use of some basic properties of copulas (for more information see the copula appendix). Let F and G have continuous marginal distribution functions F_1, F_2 and G_1, G_2, respectively. The associated copulas will be denoted as C^F and C^G, i.e.

$$C^F(F_1(x_1), F_2(x_2)) = F(x_1, x_2) \text{ and}$$

$$C^G(G_1(y_1), G_2(y_2)) = G(y_1, y_2).$$

Since the marginals are continuous, we may write

$$C^F(x_1, x_2) = F(F_1^-(x_1), F_2^-(x_2)) \text{ and}$$

$$C^G(y_1, y_2) = G(G_1^-(y_1), G_2^-(y_2)),$$

where $F_i^-(p) = \inf\{x : F_i(x) \geq p\}$ and $G_i^-(q) = \inf\{y : G_i(y) \geq q\}$ are the left continuous inverses of F_i, G_i, $i = 1, 2$ ($F_i^-(0) = F_i^-(0+)$ and $G_i^-(0) = G_i^-(0+)$).

It follows from the definition of the empirical distribution functions that the marginal empirical distribution functions will be:

$$\begin{aligned} F_n^1(x_1) &= F_n(x_1, +\infty), F_n^2(x_2) = F_n(+\infty, x_2) \text{ and} \\ G_m^1(y_1) &= G_m(y_1, +\infty), G_m^2(y_2) = G_m(+\infty, y_2). \end{aligned}$$

Therefore, we can define the empirical copula functions as:

$$\begin{aligned} C_n^F(x_1, x_2) &= F_n(F_n^{1-}(x_1), F_n^{2-}(x_2)) \text{ and} \\ C_m^G(y_1, y_2) &= G_m(G_m^{1-}(y_1), G_m^{2-}(y_2)), \end{aligned}$$

where $F_n^{i-}(p) = \inf\{x : F_n^i(x) \geq p\}$ and $G_m^{i-}(q) = \inf\{y : G_m^i(y) \geq q\}$ are the left continuous inverses of F_n^i and G_m^i, $i = 1, 2$ ($F_n^{i-}(0) = F_n^{i-}(0+)$ and $G_m^{i-}(0) = G_m^{i-}(0+)$).

As in Chapter 4, we start by defining, following the method of [22], the pseudo-variables $X^* = (X_1^*, X_2^*) = (F_1(X_1), F_2(X_2))$ and $Y^* = (Y_1^*, Y_2^*) = (F_1(Y_1), F_2(Y_2))$, with distribution functions F^* and G^* given by

$$\begin{aligned} F^*(x_1, x_2) &= P(X_1^* \leq x_1, X_2^* \leq x_2) = P(X_1 \leq F_1^-(x_1), X_2 \leq F_2^-(x_2)) \\ &= F(F_1^-(x_1), F_2^-(x_2)) = C^F(x_1, x_2), \\ G^*(y_1, y_2) &= P(Y_1^* \leq y_1, Y_2^* \leq y_2) = P(Y_1 \leq F_1^-(y_1), Y_2 \leq F_2^-(y_2)) \\ &= G(F_1^-(y_1), F_2^-(y_2)), \end{aligned}$$

and marginals F_1^*, F_2^* and G_1^*, G_2^* respectively. Although $F_1^*(x_1)$ and $F_2^*(x_2)$ are both uniform distributions as noted in [22], $G_1^*(y_1)$ and $G_2^*(y_2)$ no longer have this property under our transformation.

We will also use the copulas C^{F^*}, C^{G^*} associated with X^* and Y^*, as well as the empirical distributions $F_n^*, F_n^{1*}, F_n^{2*}$ and $G_m^*, G_m^{1*}, G_m^{2*}$ associated with $(X_{i,1}^*, X_{i,2}^*)$, $i = 1, \ldots, n$ and $(Y_{j,1}^*, Y_{j,2}^*)$, $j = 1, \ldots, m$.

The following Lemma allows us to reduce our problem to the distributions F^* and G^* defined on $[0,1]^2$.

Lemma 5.1.1 *Let F_1, F_2, G_1, G_2 be continuous distribution functions and assume that F_1 and F_2 are strictly increasing. We have the following equalities:*

$$
\begin{aligned}
&i) \quad C^F(x_1, x_2) = C^{F^*}(x_1, x_2) = F^*(x_1, x_2).\\
&ii) \quad C^G(y_1, y_2) = C^{G^*}(y_1, y_2).\\
&iii) \quad G^*(F_1^{*-}(x_1), F_2^{*-}(x_2)) = G(F_1^-(x_1), F_2^-(x_2)).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&iv) \quad C_n^F(\frac{i}{n}, \frac{j}{n}) = C_n^{F^*}(\frac{i}{n}, \frac{j}{n}).\\
&v) \quad G_m^*(F_n^{1*-}(\frac{i}{n}), F_n^{2*-}(\frac{j}{n})) = G_m(F_n^{1-}(\frac{i}{n}), F_n^{2-}(\frac{j}{n})).
\end{aligned}
$$

Since the proofs of equalities $i)$ and $iv)$ can be found in [22], they are omitted. It is worth mentioning that with our definition of (Y_1^*, Y_2^*), $C_n^G(\frac{i}{n}, \frac{j}{n}) \neq C_n^{G^*}(\frac{i}{n}, \frac{j}{n})$.

Proof: ii)$C^G(y_1, y_2) = C^{G^*}(y_1, y_2)$:

$$
\begin{aligned}
C^G(y_1, y_2) &= G(G_1^-(y_1), G_2^-(y_2))\\
&= P(Y_1 \leq G_1^-(y_1), Y_2 \leq G_2^-(y_2))\\
&= P(F_1(Y_1) \leq F_1G_1^-(y_1), F_2(Y_2) \leq F_2G_2^-(y_2))\\
&= P(Y_1^* \leq F_1G_1^-(y_1), Y_2^* \leq F_2G_2^-(y_2))\\
&= G^*(F_1G_1^-(y_1), F_2G_2^-(y_2))\\
&= C^{G^*}(G_1^*F_1G_1^-(y_1), G_2^*F_2G_2^-(y_2))\\
&= C^{G^*}(G_1F_1^-F_1G_1^-(y_1), G_2F_2^-F_2G_2^-(y_2))\\
&= C^{G^*}(y_1, y_2).
\end{aligned}
$$

iii)$G^*(F_1^{*-}(x_1), F_2^{*-}(x_2)) = G(F_1^-(x_1), F_2^-(x_2))$:

$$\begin{aligned} G^*(F_1^{*-}(x_1), F_2^{*-}(x_2)) &= P(Y_1^* \le F_1^{*-}(x_1), Y_2^* \le F_2^{*-}(x_2)) \\ &= P(F_1^*(Y_1^*) \le x_1, F_2^*(Y_2^*) \le x_2) \\ &= P(F_1^* F_1(Y_1) \le x_1, F_2^* F_2(Y_2) \le x_2) \\ &= P(F_1 F_1^- F_1(Y_1) \le x_1, F_2 F_2^- F_2(Y_2) \le x_2) \\ &= P(F_1(Y_1) \le x_1, F_2(Y_2) \le x_2) \\ &= P(Y_1 \le F_1^-(x_1), Y_2 \le F_2^-(x_2)) \\ &= G(F_1^-(x_1), F_2^-(x_2)). \end{aligned}$$

v)$G_m^*(F_n^{1*-}(\frac{i}{n}), F_n^{2*-}(\frac{j}{n})) = G_m(F_n^{1-}(\frac{i}{n}), F_n^{2-}(\frac{j}{n}))$:
Let $i_n = \frac{i}{n}$ and $j_n = \frac{j}{n}$, then:

$$\begin{aligned} G_m^*(F_n^{1*-}(\frac{i}{n}), F_n^{2*-}(\frac{j}{n})) &= \frac{1}{m}\sum_{k=1}^{m} I\{Y_k^{1*} \le F_n^{1*-}(i_n), Y_k^{2*} \le F_n^{2*-}(j_n)\} \\ &= \frac{1}{m}\sum_{k=1}^{m} I\{Y_k^1 \le F_1^- F_n^{1*-}(i_n), Y_k^2 \le F_2^- F_n^{2*-}(j_n)\} \\ &= G_m(F_1^- F_n^{1*-}(i_n), F_2^- F_n^{2*-}(j_n)). \end{aligned}$$

It remains only to show that $F_n^{1-}(i_n) = F_1^- F_n^{1*-}(i_n)$ (the result for j_n is similar and therefore omitted). First note that $F_n^1(X_{(i)}^1) = \frac{1}{n}\sum_{k=1}^n I\{X_k^1 \le X_{(i)}^1\} = i_n$, so $F_n^{1-}(i_n) = X_{(i)}^1$. We also have that

$$\begin{aligned} F_n^{1*}(F_1(X_{(i)}^1)) &= \frac{1}{n}\sum_{k=1}^{n} I\{X_k^{1*} \le F_1(X_{(i)}^1)\} \\ &= \frac{1}{n}\sum_{k=1}^{n} I\{X_k^1 \le F_1^- F_1(X_{(i)}^1)\} \\ &= \frac{1}{n}\sum_{k=1}^{n} I\{X_k^1 \le X_{(i)}^1\} = i_n, \end{aligned}$$

hence $F_n^{1*-}(i_n) = F_1(X_{(i)}^1)$.

Therefore, $F_1^- F_n^{1*-}(i_n) = F_1^- F_1(X_{(i)}^1) = X_{(i)}^1 = F_n^{1-}(i_n)$ and our proof is complete. ◇

To prove the main theorem of the section we will be making use of the functional delta method, which in turn uses the concept of Hadamard differentiability. We start by stating the definition of the latter, which is taken from [46].

Definition 5.1.2 *Let D and E be Banach spaces. A map $\Phi : D_\Phi \subset D \to E$ is called Hadamard differentiable at $\theta \in D_\Phi$ tangentially to a set $D_0 \subset D$ if there is a continuous linear map $\Phi'_\theta : D \to E$ such that $\frac{\Phi(\theta+t_n h_n)-\Phi(\theta)}{t_n} \to \Phi'_\theta(h)$ for all converging sequences $t_n \to 0$ and $h_n \to h \in D_0$ such that $\theta + t_n h_n \in D_\Phi$ for every n.*

Now we can move on to prove that the bivariate $p-p$ plot $G \circ (F_1^-, F_2^-)$ is Hadamard differentiable.

Lemma 5.1.3 *Let $F(x_1, x_2)$ and $G(y_1, y_2)$ have compact support $[0,1]^2$ and marginal distributions $F_1(x_1), F_2(x_2), G_1(y_1)$ and $G_2(y_2)$ that are continuously differentiable on their support with strictly positive densities $f_1(x_1), f_2(x_2)$ and $g_1(y_1), g_2(y_2)$ respectively. Then, the map $\Phi : (D[0,1]^2)^2 \to \ell^\infty([0,1]^2)$ defined by $\Phi(F,G) = G \circ (F_1^-, F_2^-)$ is Hadamard differentiable tangentially to $(C[0,1]^2)^2$, with derivative given by*

$$\begin{aligned}\Phi'(F,G)(\alpha,\beta) &= \beta(F_1^-, F_2^-) - \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial G_1 F_1^-}\frac{g_1(F_1^-)}{f_1(F_1^-)}\alpha^1(F_1^-) \\ &- \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial G_2 F_2^-}\frac{g_2(F_2^-)}{f_2(F_2^-)}\alpha^2(F_2^-).\end{aligned}$$

Proof: We can decompose our map Φ into three simpler maps as follows:

$$(F,G) \xrightarrow{\varphi_1} (F_1, F_2, G) \xrightarrow{\varphi_2} (F_1^-, F_2^-, G) \xrightarrow{\varphi_3} G \circ (F_1^-, F_2^-).$$

Then we have that $\Phi = \varphi_3 \circ \varphi_2 \circ \varphi_1$, and using the chain rule (see [46], Theorem 3.9.3) we get that $\Phi'_\theta = \varphi'_3\,(\varphi_2 \circ \varphi_1(\theta)) \circ \varphi'_2\,(\varphi_1(\theta)) \circ \varphi'_1\,(\theta)$.

The map φ_1 is linear and continuous, hence Hadamard differentiable. Its derivative is given by

$$\varphi'_1(F,G)(\alpha,\beta) = (\alpha^1, \alpha^2, \beta),$$

where $\alpha^1(\cdot) = \alpha(\cdot, 1)$ and $\alpha^2(\cdot) = \alpha(1, \cdot)$.

The map φ_2 is Hadamard differentiable tangentially to $C([0,1]^2)$ by Lemma 3.9.23 in [46] and its derivative is

$$\varphi_2'(F_1, F_2, G)(\gamma^1, \gamma^2, \zeta) = (-\frac{\gamma^1}{f_1} \circ F_1^-, -\frac{\gamma^2}{f_2} \circ F_2^-, \zeta).$$

Finally, the map φ_3 is Hadamard differentiable by Lemma 3.9.27 in [46], with derivative

$$\begin{aligned} \varphi_3'(F_1^-, F_2^-, G)(\mu^1, \mu^2, \nu) &= \nu(F_1^-, F_2^-) + (\frac{\partial G(F_1^-, F_2^-)}{\partial F_1^-}, \frac{\partial G(F_1^-, F_2^-)}{\partial F_2^-}) \begin{pmatrix} \mu^1 \\ \mu^2 \end{pmatrix} \\ &= \nu(F_1^-, F_2^-) + \frac{\partial G(F_1^-, F_2^-)}{\partial F_1^-}\mu^1 + \frac{\partial G(F_1^-, F_2^-)}{\partial F_2^-}\mu^2. \end{aligned}$$

Combining these three results we obtain that our map Φ is Hadamard differentiable as a composition of Hadamard differentiable functions. Its derivative is given by:

$$\begin{aligned} \Phi' \quad & (F,G)(\alpha,\beta) = \varphi_3'(\varphi_2 \circ \varphi_1(F,G)) \circ \varphi_2'(\varphi_1(F,G)) \circ \varphi_1'(F,G)(\alpha,\beta) \\ = \quad & \varphi_3' \circ (-\frac{\alpha^1}{f_1} \circ F_1^-, -\frac{\alpha^2}{f_2} \circ F_2^-, \beta) \\ = \quad & \beta(F_1^-, F_2^-) + \frac{\partial G(F_1^-, F_2^-)}{\partial F_1^-}(\frac{-\alpha^1(F_1^-)}{f_1(F_1^-)}) + \frac{\partial G(F_1^-, F_2^-)}{\partial F_2^-}(\frac{-\alpha^2(F_2^-)}{f_2(F_2^-)}) \\ = \quad & \beta(F_1^-, F_2^-) - \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial F_1^-}\frac{\alpha^1(F_1^-)}{f_1(F_1^-)} - \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial F_2^-}\frac{\alpha^2(F_2^-)}{f_2(F_2^-)} \\ = \quad & \beta(F_1^-, F_2^-) - \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial G_1F_1^-}\frac{\partial G_1F_1^-}{\partial F_1^-}\frac{\alpha^1(F_1^-)}{f_1(F_1^-)} \\ - \quad & \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial F_2^-}\frac{\partial G_2F_2^-}{\partial F_2^-}\frac{\alpha^2(F_2^-)}{f_2(F_2^-)} \\ = \quad & \beta(F_1^-, F_2^-) - \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial G_1F_1^-}\frac{g_1(F_1^-)}{f_1(F_1^-)}\alpha^1(F_1^-) \\ - \quad & \frac{\partial C^G(G_1F_1^-, G_2F_2^-)}{\partial G_2F_2^-}\frac{g_2(F_2^-)}{f_2(F_2^-)}\alpha^2(F_2^-). \end{aligned} \tag{5.1.1}$$

◇

The following theorem (the functional delta method) can also be found in [46].

Theorem 5.1.4 *Let D and E be metrizable topological vector spaces. Let $\Phi : D_\Phi \subset D \to E$ be Hadamard differentiable at θ tangentially to D_0. Let $X_n : \Omega_n \to D_\Phi$ be maps such that $r_n(X_n - \theta) \to_{\mathcal{D}} X$ for some sequence of constants $r_n \to \infty$, where X is separable and takes its values in D_0. Then $r_n(\Phi(X_n) - \Phi(\theta)) \to_{\mathcal{D}} \Phi'_\theta(X)$.*

Finally, in what follows, recall Definition 3.1.6: if H is a distribution function on $[0,1]^2$, we will be denoting by U^H the Gaussian process with covariance $E(U^H(s_1,s_2)U^H(t_1,t_2)) = H(\min\{(s_1,s_2),(t_1,t_2)\}) - H(s_1,s_2)H(t_1,t_2)$. We are ready to state and prove the main theorem of the section.

Theorem 5.1.5 *Suppose F and G are continuous and differentiable distribution functions with marginal distributions F_1, F_2, G_1, G_2 that have positive derivatives on their open support. Let $v_p^{F_1} = F_1^-(p)$ and $v_q^{F_2} = F_2^-(q)$. If $\frac{m}{n} \to \lambda$ as $n, m \to \infty$, then $\sqrt{m}[G_m \circ (F_n^{1-}, F_n^{2-}) - G \circ (F_1^-, F_2^-)]$ converges weakly to a Gaussian process in $\ell^\infty([0,1]^2)$ which is equal in distribution to*

$$\begin{aligned} W(p,q) &= U^{C^G}(G_1(v_p^{F_1}), G_2(v_q^{F_2})) \\ &- \sqrt{\lambda}\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})}\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}U^{C^F}(p,1) \\ &- \sqrt{\lambda}\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})}\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}U^{C^F}(1,q), \end{aligned}$$

where U^{C^F} and U^{C^G} are independent bridges.

Proof: First recall that for any distribution function H on $[0,1]^2$ and its empirical distribution H_n we have that $\sqrt{n}[H_n - H] \to_{\mathcal{D}} U^H$ (Theorem 3.1.7). Thus, it is straightforward that $\sqrt{m}[(F_n^*, G_m^*) - (F^*, G^*)] = \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}[(F_n^*, G_m^*) - (F^*, G^*)] \to_{\mathcal{D}} (\sqrt{\lambda}U^{F^*}, U^{G^*})$, where U^{F^*} and U^{G^*} are independent Brownian bridges.

Then, notice that $\forall\, x, y \in [0,1]$ there exists i_n, j_n such that $G_m(F_n^{1-}(x), F_n^{2-}(y)) = G_m(F_n^{1-}(i_n), F_n^{2-}(j_n))$. This and Lemma 5.1.1 give us that

$$\begin{aligned} &\sqrt{m}[G_m \circ (F_n^{1-}, F_n^{2-}) - G \circ (F_1^-, F_2^-)](x,y) \\ &= \sqrt{m}[G_m^* \circ (F_n^{*1-}, F_n^{*2-}) - G^* \circ (F_1^{*-}, F_2^{*-})](x,y)\ \forall\ (x,y). \end{aligned}$$

We now use Theorem 5.1.4 applied to $\sqrt{m}[(F_n^*, G_m^*) - (F^*, G^*)]$ with Φ as in Lemma 5.1.3, and since $G^*(F_1^{*-}(x_1), F_2^{*-}(x_2)) = G(F_1^-(x_1), F_2^-(x_2))$ (from Lemma 5.1.1), we get that

$$\begin{aligned}\sqrt{m}[\Phi(F_n, G_m) - \Phi(F,G)] &= \sqrt{m}[\Phi(F_n^*, G_m^*) - \Phi(F^*, G^*)] \\ &\to_{\mathcal{D}} \Phi'(\sqrt{\lambda}U^{F^*}, U^{G^*}) \text{ in } \ell^\infty([0,1]^2).\end{aligned}$$

Substituting (α, β) by $(\sqrt{\lambda}U^{F^*}, U^{G^*})$ in Lemma 5.1.3 and recalling that F_1^* and F_2^* are uniform, we find that:

$$\begin{aligned}\Phi'(F^*,G^*)(\sqrt{\lambda}U^{F^*}, U^{G^*})(p,q) &=_{\mathcal{D}} U^{G*}(p,q) - \frac{\partial C^{G^*}(G_1^*(p), G_2^*(q))}{\partial G_1^*(p)} g_1^*(p)\sqrt{\lambda}U^{F^*}(p,1) \\ &- \frac{\partial C^{G^*}(G_1^*(p), G_2^*(q))}{\partial G_2^*(q)} g_2^*(q)\sqrt{\lambda}U^{F^*}(1,q) \\ &=_{\mathcal{D}} U^G(F_1^-(p), F_2^-(q)) - \frac{\partial C^G(G_1F_1^-(p), G_2F_2^-(q))}{\partial G_1F_1^-(p)} \frac{g_1(F_1^-(p))}{f_1(F_1^-(p))}\sqrt{\lambda}U^F(F_1^-(p), \infty) \\ &- \frac{\partial C^G(G_1F_1^-(q), G_2F_2^-(q))}{\partial G_2F_2^-(q)} \frac{g_2(F_2^-(q))}{f_2(F_2^-(q))}\sqrt{\lambda}U^F(\infty, F_2^-(q)) \\ &=_{\mathcal{D}} U^G(v_p^{F_1}, v_q^{F_2}) - \frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})} \frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}\sqrt{\lambda}U^F(v_p^{F_1}, \infty) \\ &- \frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})} \frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}\sqrt{\lambda}U^F(\infty, v_q^{F_2}) \\ &=_{\mathcal{D}} U^{C^G}(G_1(v_p^{F_1}), G_2(v_q^{F_2})) - \sqrt{\lambda}\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})} \frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}U^{C^F}(p,1) \\ &- \sqrt{\lambda}\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})} \frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}U^{C^F}(1,q).\end{aligned}$$

◇

Corollary 5.1.6 *The limiting Gaussian variable $W(p,q)$ described in Theorem 5.1.5 has variance*

$$\begin{aligned}Var(W(p,q)) &= C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))(1 - C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))) \\ &+ \lambda(\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})})^2(\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})})^2 p(1-p)\end{aligned}$$

$$
\begin{aligned}
&+ \quad \lambda(\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})})^2(\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})})^2 q(1-q) \\
&+ \quad 2\lambda(\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})})(\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})})\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})}\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})} \\
&\qquad \times [C^F(p,q) - C^F(p,1)C^F(1,q)].
\end{aligned}
$$

Proof: From the usual covariance form of a Brownian bridge and noting that U^{C^F} and U^{C^G} are independent we have

$$
\begin{aligned}
&Var(W(p,q)) = C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))(1 - C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))) \\
&+ \quad \lambda(\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})})^2(\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})})^2 p(1-p) \\
&+ \quad \lambda(\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})})^2(\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})})^2 q(1-q) \\
&+ \quad 2\lambda(\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})})(\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})})\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})}\frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})} \\
&\times \quad [C^F(p,q) - C^F(p,1) * C^F(1,q)].
\end{aligned}
$$

◇

Corollary 5.1.7 *Under the null hypothesis $H_0 : F = G$, the process $W(p,q)$ simplifies to $\overline{U}^{C^F}(p,q) - \sqrt{\lambda}\frac{\partial C^F(p,q)}{\partial p}U^{C^F}(p,1) - \sqrt{\lambda}\frac{\partial C^F(p,q)}{\partial q}U^{C^F}(1,q)$, where U^{C^F} and $\overline{U}^{C^F}$ are independent, identically distributed bridges.*

Comment 5.1.8 It is important to remark that although U^{C^F} and $\overline{U}^{C^F}$ have the same distribution, they are the limits of independent empirical processes, and thus they remain independent. This fact makes the covariance structure simpler than in the single sample case.

Proof: If $F = G$ the limiting distribution becomes:

$$
\begin{aligned}
W(p,q) &= \overline{U}^F(v_p^{F_1}, v_q^{F_2}) - \frac{\partial C^F(F_1(v_p^{F_1}), F_2(v_q^{F_2}))}{\partial F_1(v_p^{F_1})}\frac{f_1(v_p^{F_1})}{f_1(v_p^{F_1})}\sqrt{\lambda}U^{F_1}(v_p^{F_1}) \\
&- \frac{\partial C^F(F_1(v_p^{F_1}), F_2(v_q^{F_2}))}{\partial F_2(v_q^{F_2})}\frac{f_2(v_q^{F_2})}{f_2(v_q^{F_2})}\sqrt{\lambda}U^{F_2}(v_q^{F_2})
\end{aligned}
$$

$$= \overline{U}^{C^F}(p,q) - \sqrt{\lambda}\frac{\partial C^F(p,q)}{\partial p}U^{C^F}(p,1) - \sqrt{\lambda}\frac{\partial C^F(p,q)}{\partial q}U^{C^F}(1,q).$$

$\diamond$

What we want to do now is to bring into the mix the fact that, depending on the kind of experiment that we are dealing with, we work with different regions of the plane (see discussion of these regions in chapter 2).

Recall that to prove Theorem 5.1.5 we were closely following [22]. As we mentioned, this paper developed the weak convergence of the copula process; note that the process $\sqrt{m}[G_m \circ (F_n^{1-}, F_n^{2-}) - G \circ (F_1^-, F_2^-)]$ can be regarded as a two-sample copula process when $F = G$. For this reason we will state the following theorem (Theorem 3 in [22]) with minor changes to fit the notation we are using.

Theorem 5.1.9 *Suppose that F has continuous marginal distribution functions and that the copula function C^F has continuous partial derivatives. Then the empirical copula process $\sqrt{n}[C_n^F(p,q) - C^F(p,q)]$ converges weakly to the Gaussian process $U^{C^F}(p,q) - \frac{\partial C^F(p,q)}{\partial p}U^{C^F}(p,1) - \frac{\partial C^F(p,q)}{\partial q}U^{C^F}(1,q)$ in $\ell^\infty([0,1]^2)$ $\forall\ 0 \le p,q \le 1$.*

Now, using the result of Theorem 5.1.5 and Theorem 5.1.9 we can find the limiting distribution of the process

$$\begin{aligned} &\sqrt{m}[\{G_m(V(F_n^{1-}(p), F_n^{2-}(q))) - F_n(V(F_n^{1-}(p), F_n^{2-}(q)))\} \\ &- \{G(V(F_1^-(p), F_n^{2-}(q))) - F(V(F_1^-(p), F_2^-(q)))\}], \end{aligned}$$

where V stands for the different regions we have introduced in Figure 2.1 and Figure 2.2: A_z and D_z (or equivalently E_z).

In everything that follows, we will indicate convergence in distribution in $\ell^\infty([0,1]^2)$ by "$\rightarrow_{\mathcal{D}}$".

•*A* region

When working with the A region, our process becomes

$$\sqrt{m}[\{G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))\}$$

$$- \quad \{G(F_1^-(p), F_2^-(q)) - F(F_1^-(p), F_2^-(q))\}].$$

Combining Theorem 5.1.5 and Theorem 5.1.9, we get that:

$$\begin{aligned}
&\sqrt{m}[\{G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))\} - \{G(F_1^-(p), F_2^-(q)) \\
- \quad & F(F_1^-(p), F_2^-(q))\}] \to_{\mathcal{D}} U^{C^G}(G_1(v_p^{F_1}), G_2(v_q^{F_2})) - \sqrt{\lambda} U^{C^F}(p,q) \\
+ \quad & \sqrt{\lambda} U^{C^F}(p,1)[\frac{\partial C^F(p,q)}{\partial p} - \frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})} \frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}] \\
+ \quad & \sqrt{\lambda} U^{C^F}(1,q)[\frac{\partial C^F(p,q)}{\partial q} - \frac{\partial C^G(G_1(v_p^{F_1}), G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})} \frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}].
\end{aligned}$$

Under $H_0 : F = G$ the last expression becomes:

$$\overline{U}^{C^F}(p,q) - \sqrt{\lambda} U^{C^F}(p,q),$$

where $\overline{U}^{C^F}(p,q)$ is, as stated before, the bridge depending on the G-sample. This is equal in distribution to $\sqrt{1+\lambda} U^{C^F}(p,q)$.

•D region

When working with the D region, our process becomes:

$$\begin{aligned}
&\sqrt{m}[\{G_m^1(F_n^{1-}(p)) - G_1(F_1^-(p))\} - \{F_n^1(F_n^{1-}(p)) - F_1(F_1^-(p))\} \\
+ \quad & \{G_m^2(F_n^{2-}(q)) - G_2(F_2^-(q))\} - \{F_n^2(F_n^{2-}(q)) - F_2(F_2^-(q))\} \\
- \quad & \{G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))\} \\
+ \quad & \{G(F_1^-(p), F_2^-(q)) - F(F_1^-(p), F_2^-(q))\}].
\end{aligned}$$

Therefore, it is clear that to determine the limiting distribution of this process, we need the limiting distribution of each of its 1-dimensional analogues. Although the following is a weaker version of a well known result by Aly, Csörgo and Horváth (see [4]), setting $p = 1$ or $q = 1$ in Theorem 5.1.5 will suffice to give us that:

$$\begin{aligned}
&\sqrt{m}[\{G_m^1(F_n^{1-}(p)) - G_1(F_1^-(p))\} - \{F_n^1(F_n^{1-}(p)) - F_1(F_1^-(p))\}] \\
&\rightarrow_{\mathcal{D}} \; U^{C^G}(G_1(v_p^{F_1}),1) - \sqrt{\lambda}\frac{g_1(F_1^-(p))}{f_1(F_1^-(p))}U^{C^F}(p,1) \text{ and} \\
&\sqrt{m}[\{G_m^2(F_n^{2-}(q)) - G_2(F_2^-(q))\} - \{F_n^2(F_n^{2-}(q)) - F_2(F_2^-(q))\}] \\
&\rightarrow_{\mathcal{D}} \; U^{C^G}(1,G_2(v_q^{F_2})) - \sqrt{\lambda}\frac{g_2(F_2^-(q))}{f_2(F_2^-(q))}U^{C^F}(1,q).
\end{aligned}$$

Next, since the Hadamard derivative of a linear function is the function itself, we may add the result on the A section to get the limiting distribution of the process associated with the D region:

$$\begin{aligned}
&\sqrt{m}[\{G_m^1(F_n^{1-}(p)) - G_1(F_1^-(p))\} - \{F_n^1(F_n^{1-}(p)) - F_1(F_1^-(p))\} \\
+&\quad \{G_m^2(F_n^{2-}(q)) - G_2(F_2^-(q))\} - \{F_n^2(F_n^{2-}(q)) - F_2(F_2^-(q))\} \\
-&\quad \{G_m(F_n^{1-}(p),F_n^{2-}(q)) - F_n(F_n^{1-}(p),F_n^{2-}(q))\} \\
+&\quad \{G(F_1^-(p),F_2^-(q)) - F(F_1^-(p),F_2^-(q))\}] \\
\rightarrow_{\mathcal{D}}&\quad U^{C^G}(G_1(v_p^{F_1}),1) - \sqrt{\lambda}\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}U^{C^F}(p,1) + U^{C^G}(1,G_2(v_q^{F_2})) \\
-&\quad \sqrt{\lambda}\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}U^{C^F}(1,q) - \{U^{C^G}(G_1(v_p^{F_1}),G_2(v_q^{F_2})) - \sqrt{\lambda}U^{C^F}(p,q) \\
+&\quad \sqrt{\lambda}U^{C^F}(p,1)[\frac{\partial C^F(p,q)}{\partial p} - \frac{\partial C^G(G_1(v_p^{F_1}),G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})}\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}] \\
+&\quad \sqrt{\lambda}U^{C^F}(1,q)[\frac{\partial C^F(p,q)}{\partial q} - \frac{\partial C^G(G_1(v_p^{F_1}),G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})}\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}]\} \\
=&\quad U^{C^G}(G_1(v_p^{F_1}),1) + U^{C^G}(1,G_2(v_q^{F_2})) - U^{C^G}(G_1(v_p^{F_1}),G_2(v_q^{F_2})) \\
+&\quad \sqrt{\lambda}U^{C^F}(p,q) - \sqrt{\lambda}U^{C^F}(p,1)[\frac{\partial C^F(p,q)}{\partial p} - \frac{\partial C^G(G_1(v_p^{F_1}),G_2(v_q^{F_2}))}{\partial G_1(v_p^{F_1})}\frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})} \\
+&\quad \frac{g_1(v_p^{F_1})}{f_1(v_p^{F_1})}] - \sqrt{\lambda}U^{C^F}(1,q)[\frac{\partial C^F(p,q)}{\partial q} - \frac{\partial C^G(G_1(v_p^{F_1}),G_2(v_q^{F_2}))}{\partial G_2(v_q^{F_2})}\frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})} \\
+&\quad \frac{g_2(v_q^{F_2})}{f_2(v_q^{F_2})}].
\end{aligned}$$

Under $H_0 : F = G$ the last expression becomes $\overline{U}^{C^F}(p,1) + \overline{U}^{C^F}(1,q) - \overline{U}^{C^F}(p,q) +$

$\sqrt{\lambda}U^{C^F}(p,q) - \sqrt{\lambda}U^{C^F}(p,1) - \sqrt{\lambda}U^{C^F}(1,q)$. This is equal in distribution to

$$\sqrt{1+\lambda}[U^{C^F}(p,1) + U^{C^F}(1,q) - U^{C^F}(p,q)].$$

As we will see in more detail in the applications section of this chapter, when we work with the D region we have the option to work with the (equivalent) E region. In this case, instead of the usual copula structure, we will use the survival copula structure instead.

To end the section, we will mention that although the applications of Theorem 5.1.5 that we will talk about next can be easily modified to include the results for the D_z (or E_z) region, we will focus on the case of the A_z region for the sake of clarity.

5.2 Applications

In this section it is our aim to discuss applications of Theorem 5.1.5 in practice, and in particular to focus on the different scenarios that are created based on the information that is available to the experimenter. Again, it is divided in two parts: the first part deals with the case where we have complete data, the second with precedence tests based on partial data.

•Complete samples known

We have already completed our experiment and know the values of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$. It remains to make the decision whether to reject $H_0 : F = G$ with a specific level of significance or not.

We define the following notation (as in [2]): $|F - G|^+ = \sup(F(x) - G(x))$, $|F - G|^- = \sup(G(x) - F(x))$, and $|F - G| = \max(|F - G|^+, |F - G|^-)$ for $x \in \Re^2$

and any two continuous distribution functions F and G.

Now, assume that $\sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))]$ is the process we are working with (which is the one corresponding to the A_z region) and that our one-sided alternative is $H_1 : Y \leq_{lo} X$, i.e. $H_1 : F \leq G$. In the previous section we showed that under $H_0 : F = G$, this process converges in distribution to $\sqrt{1+\lambda}U^{C^F}$, so an appropriate test statistic would be

$$\begin{aligned} W_{n,m}^+ &\equiv \sup_{p,q} \sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))] \\ &\rightarrow_{\mathcal{D}} \sqrt{1+\lambda} \sup_{p,q} U^{C^F}(p,q). \end{aligned}$$

Let $|W_{n,m}| = \max(W_{n,m}^+, W_{n,m}^-)$, where $W_{n,m}^- \equiv \sup_{p,q} \sqrt{m}[F_n(F_n^{1-}(p), F_n^{2-}(q)) - G_m(F_n^{1-}(p), F_n^{2-}(q))]$; then for a two-sided test with $H_3 : F \neq G$ we use the fact that $|W_{n,m}| \rightarrow_{\mathcal{D}} \sqrt{1+\lambda} \sup_{p,q} |U^{C^F}(p,q)|$.

Comment 5.2.1 As mentioned in Section 4.2, a Cramer Von-Mises statistic is probably a better way to go if our alternative is two-sided. Although it is not our aim to focus on a two-sided alternative, we are giving a test statistic that agrees with the method we are following for the sake of completeness.

To be able to reach a conclusion we need to determine the critical values c_α for each of our different alternatives. If the copula C^F is known, the asymptotic distribution of $|W_{n,m}|$ is completely determined and the values c_α can easily be found or simulated. On the other hand if the F copula is unknown, as in most cases, we need to develop an alternative method to find those critical values. We suggest the following two approaches.

a)Conservative approach

Conservative asymptotic values c_α can be calculated by using the next two results, the first of which was proved in detail in [1]; only a sketch of the proof is presented here.

Theorem 5.2.2 *Let J be the degenerate distribution $J(x) = \max(x_1 + x_2 - 1, 0)$ for $x \in [0,1]^2$. Then for any copula C^F and any $\delta > 0$ we have that $P(\sup_{x\in[0,1]^2} U^{C^F}(x) > \delta) \leq P(\sup_{x\in[0,1]^2} U^J(x) > \delta)$.*

Proof (sketch): A unique map m from $[0,1]^2$ onto $\{(x,y) \in [0,1]^2 : x + y \geq 1\}$ can be defined such that $J(m(x)) = J(m_1(x), m_2(x)) = F(x)$ and $m_2(x) - m_1(x) = x_2 - x_1$ hold $\forall\, x \in [0,1]^2$. It can be shown $\forall\, x, y \in [0,1]^2$, by taking separately the cases $x < y$ (or $y < x$) and $x_1 > y_1, x_2 < y_2$ (or $x_1 < y_1, x_2 > y_2$), that $F(x \wedge y) \geq J(m(x) \wedge m(y))$, from which it follows that $Cov(U^{C^F}(x), U^{C^F}(y)) \geq Cov(U^J(m(x)), U^J(m(y)))$. This together with the fact that $Var(U^{C^F}(x)) = Var(U^J(m(x)))$ gives us, via Slepian's inequality, that $P(\sup_{x\in[0,1]^2} U^{C^F}(x) > \delta) \leq P(\sup_{x\in[0,1]^2} U^J(m(x)) > \delta)$. The result follows since $\sup_x U^J(m(x)) = \sup_x U^J(x)$. $\diamond$

More precisely, we have the following result (Proposition 5.3 and Remark 5.4) in [10]:

Proposition 5.2.3 *Let J be the degenerate distribution $J(x) = \max(x_1 + x_2 - 1, 0)$ for $x \in [0,1]^2$. Then for any distribution F on $\Re^2$ and any $c : [0,1] \to [0,1]$, we have that $P(J_n(y) \geq c(J(y)) \forall\, y \in \Re^2) \leq P(F_n(x) \geq c(F(x)) \forall\, x \in \Re^2)$. In particular, it is true that $P(\sup_{t\in\Re^2}\{\sqrt{n}(F_n(t) - F(t))\} > c_n) \leq P(\sup_{t\in\Re^2}\{\sqrt{n}(J_n(t) - J(t))\} > c_n)$*

In view of this two results, Adler, Brown and Lu [2] found, using simulation techniques, several critical values which we will make use of next.

Going back to our original problem, our null hypothesis was that F and G have the same distribution. We consider first the case when we have $H_1 : Y \leq_{lo} X$ ($H_1 : F \leq G$) as our alternative; the case $H_2 : X \leq_{lo} Y$ ($H_2 : G \leq F$) is managed exactly the same way using $W^-_{n,m}$ instead of $W^+_{n,m}$.

Consider

$$\begin{aligned} W^+_{n,m} &= \sup \sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))] \\ &= \sup[\sqrt{m}\{G_m(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))\} \end{aligned}$$

$$\begin{aligned}
&- \sqrt{m}\{F_n(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))\}] \\
&= \sup[\sqrt{m}\{G_m(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))\} \\
&+ \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}\{H(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))\}] \\
&\leq \sup\sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))] \\
&+ \frac{\sqrt{m}}{\sqrt{n}}\sup\sqrt{n}[H(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))] \\
&\leq \sqrt{m}|G_m - H|^+ + \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}|F_n - H|^-.
\end{aligned}$$

Under $H_0 : F = G = H$ the last expression reduces to $W_{n,m}^+ \leq \sqrt{m}|\overline{F}_m - F|^+ + \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}|F_n - F|^-$, where $\overline{F}_m$ denotes the empirical distribution of the second sample. We are looking for values c_α such that $P(W_{n,m}^+ \geq c_\alpha) \leq \alpha$. This inequality certainly holds if we choose $c_\alpha = 2\max(d_{\frac{\alpha}{2}}^+, d_{\frac{\alpha}{2}}^-)$, where the values $d_{\frac{\alpha}{2}}$ are chosen such that

$$P(\sqrt{m}|\overline{F}_m - F|^+ \geq d_{\frac{\alpha}{2}}^+) \leq \frac{\alpha}{2} \text{ and } P(\sqrt{n}|F_n - F|^- \geq \frac{d_{\frac{\alpha}{2}}^-}{\sqrt{\lambda}}) \leq \frac{\alpha}{2}.$$

We can find in [2] simulated values of $d_{\frac{\alpha}{2}}^+$ and $\frac{d_{\frac{\alpha}{2}}^-}{\sqrt{\lambda}}$ for various values of n and m; the values for our test for different values of n and m are shown in Table 5.1 and Table 5.2. Values for the alternative $H_2 : X \leq_{lo} Y$ are shown in Table 5.3 and Table 5.4, where we use instead the test statistic $W_{n,m}^-$.

Now consider the case where our alternative is two-sided: $H_3 : F \neq G$. This time we will reject $H_0 : F = G = H$ when our test statistic $|W_{n,m}| = \max(W_{n,m}^+, W_{n,m}^-)$ gets too big. Notice that:

$$\begin{aligned}
|W_{n,m}| &= \sup|\sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))]| \\
&= \sup|\sqrt{m}\{G_m(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))\} \\
&- \sqrt{m}\{F_n(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))\}| \\
&\leq \sup|\sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - H(F_n^{1-}(p), F_n^{2-}(q))]| \\
&+ \frac{\sqrt{m}}{\sqrt{n}}\sup|\sqrt{n}[H(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))]|
\end{aligned}$$

$$\leq \sqrt{m}|G_m - H| + \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}|F_n - H|.$$

Under $H_0 : F = G = H$ we have that $|W_{n,m}| \leq \sqrt{m}|\overline{F}_m - F| + \frac{\sqrt{m}}{\sqrt{n}}\sqrt{n}|F_n - F|$. We need to find the critical values c_α such that $P(|W_{n,m}| \geq c_\alpha) \leq \alpha$. This holds if we let $c_\alpha = 2\max(d^1_{\frac{\alpha}{2}}, d^2_{\frac{\alpha}{2}})$, where the values $d^1_{\frac{\alpha}{2}}$ and $d^2_{\frac{\alpha}{2}}$ are such that

$$P(\sqrt{m}|\overline{F}_m - F| \geq d^1_{\frac{\alpha}{2}}) \leq \frac{\alpha}{2} \text{ and } P(\sqrt{n}|F_n - F| \geq \frac{d^2_{\frac{\alpha}{2}}}{\sqrt{\lambda}}) \leq \frac{\alpha}{2}.$$

Again, we can use [2] to find $d^1_{\frac{\alpha}{2}}$ and $\frac{d^2_{\frac{\alpha}{2}}}{\sqrt{\lambda}}$; these values for our test for several levels of significance and combinations of n and m are shown in Table 5.5 and Table 5.6.

The method to calculate critical values that we have just described will give us more conservative values as we increase our sample sizes. Thus, even if we are willing to accept somewhat conservative critical values, this may not be the best approach if we have large sample sizes.

To handle the large samples scenario, for the alternative $H_1 : Y \leq_{lo} X$, we can use Theorem 5.2.2 to find that

$$\begin{aligned} P(W^+_{n,m} \geq d^+_\alpha) &\approx P(\sqrt{1+\lambda} \sup_{x\in[0,1]^2} U^{C^F}(x) \geq d^+_\alpha) \\ &= P(\sup_{x\in[0,1]^2} U^{C^F}(x) \geq \frac{d^+_\alpha}{\sqrt{1+\lambda}}) \\ &\leq P(\sup_{x\in[0,1]^2} U^J(x) \geq \frac{d^+_\alpha}{\sqrt{1+\lambda}}). \end{aligned}$$

Again, simulated values for $\frac{d^+_\alpha}{\sqrt{1+\lambda}}$ can be found in [2] and the resulting critical values are in table 5.7. For the alternative $H_1 : X \leq_{lo} Y$, the argument is a similar one, but replacing $W^+_{n,m}$ and d^+_α with $W^-_{n,m}$ and d^-_α; the corresponding critical values are also shown in table 5.7.

α	λ	$m = 10$	$n = 10$	$m = 20$	$n = 20$	$m = 30$	$n = 30$
		$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$
.20	$\frac{1}{3}$	1.453	0.788	1.474	0.813	1.485	0.825
	$\frac{1}{2}$	1.453	0.965	1.474	0.996	1.485	1.011
	1	1.453	1.365	1.474	1.409	1.485	1.430
	2	1.453	1.930	1.474	1.992	1.485	2.022
	3	1.453	2.364	1.474	2.440	1.485	2.476
.10	$\frac{1}{3}$	1.576	0.859	1.603	0.887	1.615	0.899
	$\frac{1}{2}$	1.576	1.052	1.603	1.087	1.615	1.101
	1	1.576	1.488	1.603	1.538	1.615	1.558
	2	1.576	2.104	1.603	2.175	1.615	2.203
	3	1.576	2.577	1.603	2.663	1.615	2.698
.05	$\frac{1}{3}$	1.685	0.923	1.720	0.956	1.731	0.968
	$\frac{1}{2}$	1.685	1.130	1.720	1.170	1.731	1.185
	1	1.685	1.599	1.720	1.656	1.731	1.677
	2	1.685	2.261	1.720	2.341	1.731	2.371
	3	1.685	2.769	1.720	2.868	1.731	2.904
.02	$\frac{1}{3}$	1.815	0.993	1.855	1.034	1.875	1.046
	$\frac{1}{2}$	1.815	1.216	1.855	1.266	1.875	1.281
	1	1.815	1.721	1.885	1.791	1.875	1.813
	2	1.815	2.433	1.885	2.532	1.875	2.563
	3	1.815	2.980	1.885	3.102	1.875	3.140
.01	$\frac{1}{3}$	1.903	1.044	1.950	1.088	1.976	1.101
	$\frac{1}{2}$	1.903	1.279	1.950	1.332	1.976	1.348
	1	1.903	1.809	1.950	1.885	1.976	1.907
	2	1.903	2.558	1.950	2.665	1.976	2.696
	3	1.903	3.133	1.950	3.264	1.976	3.303

Table 5.1: Values of $d^+_{\frac{\alpha}{2}}$ and $d^-_{\frac{\alpha}{2}}$ for the alternative $H_1 : Y \leq_{lo} X$.

α	λ	$m = 50$	$n = 50$	$m = 100$	$n = 100$	$m = 500$	$n = 500$
		$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$
.20	$\frac{1}{3}$	1.494	0.840	1.501	0.849	1.507	0.866
	$\frac{1}{2}$	1.494	1.028	1.501	1.040	1.507	1.060
	1	1.494	1.455	1.501	1.472	1.507	1.500
	2	1.494	2.057	1.501	2.081	1.507	2.121
	3	1.494	2.520	1.501	2.549	1.507	2.598
.10	$\frac{1}{3}$	1.625	0.915	1.634	0.926	1.640	0.943
	$\frac{1}{2}$	1.625	1.121	1.634	1.134	1.640	1.156
	1	1.625	1.586	1.634	1.604	1.640	1.635
	2	1.625	2.242	1.634	2.268	1.640	2.312
	3	1.625	2.747	1.634	2.778	1.640	2.831
.05	$\frac{1}{3}$	1.744	0.985	1.755	0.994	1.752	1.012
	$\frac{1}{2}$	1.744	1.207	1.755	1.218	1.752	1.239
	1	1.744	1.707	1.755	1.723	1.752	1.753
	2	1.744	2.414	1.755	2.436	1.752	2.479
	3	1.744	2.956	1.755	2.984	1.752	3.036
.02	$\frac{1}{3}$	1.884	1.065	1.898	1.076	1.904	1.103
	$\frac{1}{2}$	1.884	1.305	1.898	1.318	1.904	1.351
	1	1.884	1.846	1.898	1.865	1.904	1.911
	2	1.884	2.610	1.898	2.637	1.904	2.702
	3	1.884	3.197	1.898	3.230	1.904	3.309
.01	$\frac{1}{3}$	1.978	1.121	2.001	1.137	2.004	1.167
	$\frac{1}{2}$	1.978	1.373	2.001	1.393	2.004	1.430
	1	1.978	1.943	2.001	1.970	2.004	2.023
	2	1.978	2.747	2.001	2.786	2.004	2.860
	3	1.978	3.365	2.001	3.412	2.004	3.503

Table 5.2: Values of $d^+_{\frac{\alpha}{2}}$ and $d^-_{\frac{\alpha}{2}}$ for the alternative $H_1 : Y \leq_{lo} X$.

α	λ	$n = 10$	$m = 10$	$n = 20$	$m = 20$	$n = 30$	$m = 30$
		$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$
.20	$\frac{1}{3}$	0.838	1.365	0.851	1.409	0.857	1.430
	$\frac{1}{2}$	1.027	1.365	1.042	1.409	1.050	1.430
	1	1.453	1.365	1.474	1.409	1.485	1.430
	2	2.054	1.365	2.084	1.409	2.100	1.430
	3	2.516	1.365	2.553	1.409	2.572	1.430
.10	$\frac{1}{3}$	0.909	1.488	0.925	1.538	0.932	1.558
	$\frac{1}{2}$	1.114	1.488	1.133	1.538	1.141	1.558
	1	1.576	1.488	1.603	1.538	1.615	1.558
	2	2.228	1.488	2.266	1.538	2.283	1.558
	3	2.729	1.488	2.776	1.538	2.797	1.558
.05	$\frac{1}{3}$	0.972	1.599	0.993	1.656	0.999	1.677
	$\frac{1}{2}$	1.191	1.599	1.216	1.656	1.224	1.677
	1	1.685	1.599	1.720	1.656	1.731	1.677
	2	2.382	1.599	2.432	1.656	2.448	1.677
	3	2.918	1.599	2.979	1.656	2.998	1.677
.02	$\frac{1}{3}$	1.047	1.721	1.070	1.791	1.082	1.813
	$\frac{1}{2}$	1.283	1.721	1.311	1.791	1.325	1.813
	1	1.815	1.721	1.885	1.791	1.875	1.813
	2	2.566	1.721	2.623	1.791	2.651	1.813
	3	3.143	1.721	3.212	1.791	3.247	1.813
.01	$\frac{1}{3}$	1.098	1.809	1.125	1.885	1.140	1.907
	$\frac{1}{2}$	1.345	1.809	1.378	1.885	1.397	1.907
	1	1.903	1.809	1.950	1.885	1.976	1.907
	2	2.691	1.809	2.757	1.885	2.794	1.907
	3	3.296	1.809	3.377	1.885	3.422	1.907

Table 5.3: Values of $d^+_{\frac{\alpha}{2}}$ and $d^-_{\frac{\alpha}{2}}$ for the alternative $H_2 : X \leq_{lo} Y$.

α	λ	$n=50$	$m=50$	$n=100$	$m=100$	$n=500$	$m=500$
		$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$	$d^+_{\alpha/2}$	$d^-_{\alpha/2}$
.20	$\frac{1}{3}$	0.862	1.455	0.866	1.472	0.870	1.500
	$\frac{1}{2}$	1.056	1.455	1.061	1.472	1.065	1.500
	1	1.494	1.455	1.501	1.472	1.507	1.500
	2	2.112	1.455	2.122	1.472	2.131	1.500
	3	2.587	1.455	2.599	1.472	2.610	1.500
.10	$\frac{1}{3}$	0.938	1.586	0.943	1.604	0.946	1.635
	$\frac{1}{2}$	1.149	1.586	1.155	1.604	1.159	1.635
	1	1.625	1.586	1.634	1.604	1.640	1.635
	2	2.298	1.586	2.310	1.604	2.319	1.635
	3	2.814	1.586	2.830	1.604	2.840	1.635
.05	$\frac{1}{3}$	1.006	1.707	1.013	1.723	1.011	1.753
	$\frac{1}{2}$	1.233	1.707	1.240	1.723	1.238	1.753
	1	1.744	1.707	1.755	1.723	1.752	1.753
	2	2.466	1.707	2.481	1.723	2.477	1.753
	3	3.020	1.707	3.039	1.723	3.034	1.753
.02	$\frac{1}{3}$	1.087	1.846	1.095	1.865	1.099	1.911
	$\frac{1}{2}$	1.332	1.846	1.342	1.865	1.346	1.911
	1	1.884	1.846	1.898	1.865	1.904	1.911
	2	2.664	1.846	2.684	1.865	2.692	1.911
	3	3.263	1.846	3.287	1.865	3.297	1.911
.01	$\frac{1}{3}$	1.141	1.943	1.155	1.970	1.157	2.023
	$\frac{1}{2}$	1.398	1.943	2.414	1.970	1.417	2.023
	1	1.978	1.943	2.001	1.970	2.004	2.023
	2	2.797	1.943	2.829	1.970	2.834	2.023
	3	3.425	1.943	3.465	1.970	3.471	2.023

Table 5.4: Values of $d^+_{\frac{\alpha}{2}}$ and $d^-_{\frac{\alpha}{2}}$ for the alternative $H_2 : X \leq_{lo} Y$.

α	λ	$m=10$	$n=10$	$m=20$	$n=20$	$m=30$	$n=30$
		$d^1_{\alpha/2}$	$d^2_{\alpha/2}$	$d^1_{\alpha/2}$	$d^2_{\alpha/2}$	$d^1_{\alpha/2}$	$d^2_{\alpha/2}$
.20	$\frac{1}{3}$	1.536	0.886	1.574	0.908	1.588	0.916
	$\frac{1}{2}$	1.536	1.086	1.574	1.112	1.588	1.122
	1	1.536	1.536	1.574	1.574	1.588	1.588
	2	1.536	2.172	1.574	2.225	1.588	2.245
	3	1.536	2.660	1.574	2.726	1.588	2.750
.10	$\frac{1}{3}$	1.646	0.950	1.690	0.975	1.705	0.984
	$\frac{1}{2}$	1.646	1.163	1.690	1.195	1.705	1.205
	1	1.646	1.646	1.690	1.690	1.705	1.705
	2	1.646	2.327	1.690	2.390	1.705	2.411
	3	1.646	2.850	1.690	2.927	1.705	2.953
.04	$\frac{1}{3}$	1.776	1.025	1.826	1.054	1.845	1.065
	$\frac{1}{2}$	1.776	1.255	1.826	1.291	1.845	1.304
	1	1.776	1.776	1.826	1.826	1.845	1.845
	2	1.776	2.511	1.826	2.582	1.845	2.609
	3	1.776	3.076	1.826	3.162	1.845	3.195
.02	$\frac{1}{3}$	1.864	1.076	1.919	1.107	1.942	1.121
	$\frac{1}{2}$	1.864	1.318	1.919	1.356	1.942	1.373
	1	1.864	1.864	1.919	1.919	1.942	1.942
	2	1.864	2.636	1.919	2.713	1.942	2.746
	3	1.864	3.228	1.919	3.323	1.942	3.363

Table 5.5: Values of $d^1_{\frac{\alpha}{2}}$ and $d^2_{\frac{\alpha}{2}}$ for the alternative $H_3 : F \neq G$.

α	λ	$m = 50$	$n = 50$	$m = 100$	$n = 100$	$m = 500$	$n = 500$
		$d^1_{\alpha/2}$	$d^2_{\alpha/2}$	$d^1_{\alpha/2}$	$d^2_{\alpha/2}$	$d^1_{\alpha/2}$	$d^2_{\alpha/2}$
.20	$\frac{1}{3}$	1.607	0.927	1.620	0.935	1.638	0.945
	$\frac{1}{2}$	1.607	1.136	1.620	1.145	1.638	1.158
	1	1.607	1.607	1.620	1.620	1.638	1.638
	2	1.607	2.272	1.620	2.291	1.638	2.316
	3	1.607	2.783	1.620	2.805	1.638	2.837
.10	$\frac{1}{3}$	1.727	0.997	1.739	1.004	1.752	1.011
	$\frac{1}{2}$	1.727	1.221	1.739	1.229	1.752	1.238
	1	1.727	1.727	1.739	1.739	1.752	1.752
	2	1.727	2.442	1.739	2.459	1.752	2.477
	3	1.727	2.991	1.739	3.012	1.752	3.034
.04	$\frac{1}{3}$	1.865	1.076	1.882	1.086	1.906	1.100
	$\frac{1}{2}$	1.865	1.318	1.882	1.330	1.906	1.347
	1	1.865	1.865	1.882	1.882	1.906	1.906
	2	1.865	2.637	1.882	2.661	1.906	2.695
	3	1.865	3.230	1.882	3.259	1.906	3.301
.02	$\frac{1}{3}$	1.961	1.132	1.984	1.145	2.012	1.161
	$\frac{1}{2}$	1.961	1.386	1.984	1.402	2.012	1.422
	1	1.961	1.961	1.984	1.984	2.012	2.012
	2	1.961	2.773	1.984	2.805	2.012	2.845
	3	1.961	3.396	1.984	3.436	2.012	3.484

Table 5.6: Values of $d^1_{\frac{\alpha}{2}}$ and $d^2_{\frac{\alpha}{2}}$ for the alternative $H_3 : F \neq G$.

α	λ	$H_1 : Y \leq_{lo} X$	$H_2 : X \leq_{lo} Y$
.10	$\frac{1}{3}$	1.740	1.732
	$\frac{1}{2}$	1.845	1.837
	1	2.131	2.121
	2	2.610	2.598
	3	3.014	3.000
.05	$\frac{1}{3}$	1.893	1.887
	$\frac{1}{2}$	2.008	2.002
	1	2.319	2.312
	2	2.840	2.831
	3	3.280	3.270
.025	$\frac{1}{3}$	2.023	2.024
	$\frac{1}{2}$	2.145	2.146
	1	2.477	2.479
	2	3.034	3.036
	3	3.504	3.506
.01	$\frac{1}{3}$	2.198	2.206
	$\frac{1}{2}$	2.331	2.340
	1	2.692	2.702
	2	3.297	3.309
	3	3.808	3.822
.005	$\frac{1}{3}$	2.314	2.335
	$\frac{1}{2}$	2.454	2.477
	1	2.834	2.860
	2	3.471	3.503
	3	4.008	4.046

Table 5.7: Large sample critical values for the alternatives $H_1 : Y \leq_{lo} X$ and $H_2 : X \leq_{lo} Y$.

α	λ	$H_3 : F \neq G$
.10	$\frac{1}{3}$	1.891
	$\frac{1}{2}$	2.006
	1	2.316
	2	2.837
	3	3.276
.05	$\frac{1}{3}$	2.023
	$\frac{1}{2}$	2.145
	1	2.477
	2	3.034
	3	3.504
.02	$\frac{1}{3}$	2.200
	$\frac{1}{2}$	2.334
	1	2.695
	2	3.301
	3	3.812
.01	$\frac{1}{3}$	2.323
	$\frac{1}{2}$	2.464
	1	2.845
	2	3.484
	3	4.024

Table 5.8: Large sample critical values for the alternatives $H_3 : F \neq G$.

Now consider the case where our alternative is $H_1 : F \neq G$ and the statistic $|W_{n,m}| = \max(W^+_{n,m}, W^-_{n,m})$. In the same way as before, we get that $P(|W_{n,m}| \geq d_\alpha) \leq P(\sup_{x\in[0,1]^2} U^J(x) \geq \frac{d_\alpha}{\sqrt{1+\lambda}})$ and we have in [2] simulated values such that $P(\sup_{x\in[0,1]^2} U^J(x) \geq \frac{d_\alpha}{\sqrt{1+\lambda}}) \leq \alpha$. Critical values are presented in table 5.8.

b)Bootstrap approach

As we mentioned before, in practice, we do not usually know what the copula C^F is and need to estimate it from the data we have. One of the ways to approach this situation is to estimate C^F via the bootstrap technique, which was first introduced in [17]. The general idea behind the bootstrap is to obtain a sample from an unknown distribution, construct an empirical distribution applying $1/nth$ of mass to each observation, resample with replacement from that empirical distribution and then use this last sample to estimate the initial unknown distribution or a function of it. As under our null hypothesis the samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are actually two samples from the same distribution, we will present a bootstrap that uses pooled data.

Following the two-sample bootstrap procedure described in section 3.7.2 of [46], we denote the pooled data as

$$(Z_{1N}, Z_{2N}, \ldots, Z_{NN}) = (X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m),$$

where $N = n + m$. To follow this approach we will sample with replacement from $(Z_{1N}, Z_{2N}, \ldots, Z_{NN})$ to obtain $(\overset{B}{Z}_{1N}, \overset{B}{Z}_{2N}, \ldots, \overset{B}{Z}_{NN})$. This can be regarded as a single bootstrapped sample from the distribution $H = \frac{1}{1+\lambda}F + \frac{\lambda}{1+\lambda}G$. We define the pooled empirical distribution function H_N and the two-sample bootstrap empirical distribution functions $\overset{B}{F}_{n,N}$ and $\overset{B}{G}_{m,N}$ as follows:

$$\begin{aligned} H_N(t) &= \frac{1}{N}\sum_{i=1}^{N} I\{Z_{iN} \leq t\} = \frac{n}{N}F_n(t) + \frac{m}{N}G_m(t) \\ \overset{B}{F}_{n,N}(t) &= \frac{1}{n}\sum_{i=1}^{n} I\{\overset{B}{Z}_{iN} \leq t\} \end{aligned}$$

$$\overset{B}{G}_{m,N}(t) = \frac{1}{m}\sum_{i=1}^{m} I\{\overset{B}{Z}_{n+i,N} \leq t\}.$$

First, consider the process

$$\sqrt{m}[(\overset{B}{F}_{n,N}, \overset{B}{G}_{m,N}) - (H_N, H_N)] = (\sqrt{m}(\overset{B}{F}_{n,N} - H_N), \sqrt{m}(\overset{B}{G}_{m,N} - H_N)).$$

This will converge to $(\sqrt{\lambda}U^{C^H}, \overline{U}^{C^H})$ given $(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)$, where as before, the bridges U^{C^H} and $\overline{U}^{C^H}$ are independent. We will again use the fact that the function $\Phi(F, G) = G \circ (F_1^-, F_2^-)$ is Hadamard differentiable with derivative given by Lemma 5.1.3.

We have that:

$$\begin{aligned}
&\sqrt{m}[\overset{B}{G}_{m,N}(\overset{B^{1-}}{F}_{n,N}, \overset{B^{2-}}{F}_{n,N}) - H_N(H_N^{1-}, H_N^{2-})](p,q) \\
&= \sqrt{m}[\Phi(\overset{B}{F}_{n,N}, \overset{B}{G}_{m,N}) - \Phi(H_N, H_N)](p,q) \\
&\to_{\mathcal{D}} \Phi'(H,H)(\sqrt{\lambda}U^{C^H}, \overline{U}^{C^H})(p,q) \\
&= \overline{U}^{C^H}(p,q) - \sqrt{\lambda}\frac{\partial C^H(p,q)}{\partial p}U^{C^H}(p,1) - \sqrt{\lambda}\frac{\partial C^H(p,q)}{\partial q}U^{C^H}(1,q)
\end{aligned}$$

given $(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)$.

Similarly:

$$\begin{aligned}
&\sqrt{m}[\overset{B}{F}_{n,N}(\overset{B^{1-}}{F}_{n,N}, \overset{B^{2-}}{F}_{n,N}) - H_N(H_N^{1-}, H_N^{2-})](p,q) \\
&= \sqrt{m}[\Phi(\overset{B}{F}_{n,N}, \overset{B}{F}_{n,N}) - \Phi(H_N, H_N)](p,q) \\
&\to_{\mathcal{D}} \Phi'(H,H)(\sqrt{\lambda}U^{C^H}, \sqrt{\lambda}U^{C^H})(p,q) \\
&= \sqrt{\lambda}U^{C^H}(p,q) - \sqrt{\lambda}\frac{\partial C^H(p,q)}{\partial p}U^{C^H}(p,1) - \sqrt{\lambda}\frac{\partial C^H(p,q)}{\partial q}U^{C^H}(1,q)
\end{aligned}$$

given $(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)$.

Both asymptotic behaviors above follow from the delta method for bootstrap (see Theorem 3.9.11 in [46]).

Putting the two expressions together, we get that the following holds:

$$
\begin{aligned}
&\sqrt{m}[\overset{B}{G}_{m,N}(\overset{B^{1-}}{F}_{n,N}, \overset{B^{2-}}{F}_{n,N}) - \overset{B}{F}_{n,N}(\overset{B^{1-}}{F}_{n,N}, \overset{B^{2-}}{F}_{n,N})](p,q) \\
\rightarrow_{\mathcal{D}} \quad & \overline{U}^{C^H}(p,q) - \sqrt{\lambda}U^{C^H}(p,q) + \sqrt{\lambda}U^{C^H}(p,1)[\frac{\partial C^H(p,q)}{\partial p} - \frac{\partial C^H(p,q)}{\partial p}] \\
+ \quad & \sqrt{\lambda}U^{C^H}(1,q)[\frac{\partial C^H(p,q)}{\partial q} - \frac{\partial C^H(p,q)}{\partial q}] \\
= \quad & \overline{U}^{C^H}(p,q) - \sqrt{\lambda}U^{C^H}(p,q) \\
=_{\mathcal{D}} \quad & \sqrt{1+\lambda}U^{C^H}(p,q)
\end{aligned}
$$

given $(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)$.

Now we define our new bootstrapped process:

$$W^{B,+}_{n,m}(p,q) = \sup_{p,q} \sqrt{m}[\overset{B}{G}_{m,N}(\overset{B^{1-}}{F}_{n,N}, \overset{B^{2-}}{F}_{n,N}) - \overset{B}{F}_{n,N}(\overset{B^{1-}}{F}_{n,N}, \overset{B^{2-}}{F}_{n,N})].$$

We just proved that $W^{B,+}_{n,m} \rightarrow_{\mathcal{D}} \sqrt{1+\lambda}\sup_{p,q} U^{C^H}$. Therefore, we can take a large number of samples of size N from our pooled empirical distribution H_N and use $W^{B,+}_{n,m}$ to approximate observations of $\sqrt{1+\lambda}\sup_{p,q} U^{C^H}$. The upper $(1-\alpha)100$th percentile of our bootstrapped observations will approximate c_α, the appropriate critical value for testing $H_0 : F = G = H$ against the alternative $H_1 : Y \leq_{lo} X$. The alternatives $H_2 : X \leq_{lo} Y$ and $H_3 : F \neq G$ are handled similarly using, respectively, bootstrapped observations of $W^{B,-}_{n,m}$ and $|W^B_{n,m}|$

•Partial samples known

In this scenario we have already completed our experiment but we only know the values of some of our observations, in particular, the process $\sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))]$ can only be observed over the region $0 \leq p \leq p_0$ and $0 \leq q \leq q_0$.

a)Conservative approach

Assume we are working with the process $\sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))]$ and the alternative $H_1 : Y \leq_{lo} X$. Under the precedence scheme we can no longer use the test statistic $W_{n,m}$, but we will define a new statistic

$$\begin{aligned} W^*_{n,m} &= W^*_{n,m}(p_0, q_0) \\ &\equiv \sup_{p \leq p_0, q \leq q_0} \sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))] \end{aligned}$$

to takes its place. It is easy to see that under $H_0 : F = G$,

$$W^*_{n,m}(p_0, q_0) \to_{\mathcal{D}} \sqrt{1+\lambda} \sup_{p \leq p_0, q \leq q_0} U^{C^F}(p, q).$$

Now we are in a familiar position: we need to find the critical values c_α for each alternative. If the copula C^F is known the critical values are simulated without any further complications.

If the F copula is unknown we can certainly compare this new statistic with that we already used in the case of complete samples. We know that $W^{*,+}_{n,m} \leq W^+_{n,m}$, from which follows that $P(W^{*,+}_{n,m} \geq c_\alpha) \leq P(W^+_{n,m} \geq c_\alpha) \leq \alpha$. Then, if we let $c_\alpha = 2\max(d^+_{\frac{\alpha}{2}}, d^-_{\frac{\alpha}{2}})$, where the values $d_{\frac{\alpha}{2}}$ are chosen as in the case of complete samples, we have found critical values that will work. It is obvious that these values are not the best in the sense that $P(W^{*,+}_{n,m} \geq c_\alpha)$ could be significantly smaller than $P(W^+_{n,m} \geq c_\alpha)$, which in turn is almost always smaller than α. As a consequence of accepting as critical values the values found in Table 5.1, Table 5.2 or Table 5.7 we are definitely setting a conservative critical region and in doing so, choosing a less powerful test. Using the same logic we can conclude that the values found in Table 5.3, Table 5.4 or Table 5.7 can be seen as conservative critical values for a test $H_0 : F = G$ vs $H_2 : X \leq_{lo} Y$.

Similarly, when we have a two-sided alternative $H_3 : F \neq G$, we use that $|W^*_{n,m}(p_0, q_0)| \to_{\mathcal{D}} \sqrt{1+\lambda} \sup_{p \leq p_0, q \leq q_0} |U^{C^F}(p, q)|$. Since $|W^*_{n,m}| \leq |W_{n,m}|$, it follows that $P(|W^*_{n,m}| \geq c_\alpha) \leq P(|W_{n,m}| \geq c_\alpha) \leq \alpha$ and once more we can use the values in Table 5.5, Table 5.6 or Table 5.8 as conservative critical values for our test.

If we have some additional information about the copulas involved, this approach can be refined. To explain the path to follow we need to go back to the sketch of the proof of Theorem 5.2.2. There, we worked with the supremum over $[0,1]^2$ of the two bridges U^{C^F} and U^J. If we now want to restrict x to $[0,p_0] \times [0,q_0]$, the proof will still hold up until the last line: we will have that $P(\sup_{x\in[0,p_0]\times[0,q_0]} U^{C^F}(x) > \delta) \leq P(\sup_{x\in[0,p_0]\times[0,q_0]} U^J(m(x)) > \delta)$, but $\sup_{x\in[0,p_0]\times[0,q_0]} U^J(m(x)) = \sup_{x\in[0,p_0]\times[0,q_0]} U^J(x)$ will not be true anymore.

We have that:

$$\begin{aligned} P(\sup_{x\in[0,p_0]\times[0,q_0]} U^{C^F}(x) > \delta) &\leq P(\sup_{x\in[0,p_0]\times[0,q_0]} U^J(m(x)) > \delta) \\ &\leq P(\sup_{x_1\leq \sup\{m_1(z):z\in[0,p_0]\times[0,q_0]\}, x_2\leq \sup\{m_2(z):z\in[0,p_0]\times[0,q_0]\}} U^J(x) > \delta). \end{aligned}$$

With that in mind, let us pick a specific example to show the way this method works. Suppose we know that both F and G share the FGM copula $C(u,v) = uv + \theta uv(1-u)(1-v)$ with unknown parameter θ, and we want to test $H_0 : F = G$ vs $H_1 : Y \leq_{lo} X$. Under this assumption, we are actually testing $H_0 : \theta_F = \theta_G$ vs $H_1 : \theta_F \leq \theta_G$.

The function $m(x)$ for the FGM copula is $\frac{1}{2}(x_1 - x_2 + 1 + x_1x_2(1+\theta(1-x_1)(1-x_2)), x_2 - x_1 + 1 + x_1x_2(1+\theta(1-x_1)(1-x_2)))$. To continue with the example, fix $p_0 = q_0 = \frac{1}{2}$. Some of the values of $m(x)$ are: $m_1(\frac{1}{2},\frac{1}{2}) = m_2(\frac{1}{2},\frac{1}{2}) = \frac{5}{8}+\frac{\delta}{32}$, $m_1(0,\frac{1}{2}) = m_2(\frac{1}{2},0) = \frac{1}{4}$, and $m_1(\frac{1}{2},0) = m_2(0,\frac{1}{2}) = \frac{3}{4}$. We can check that m_1 is increasing in x_1, but decreasing in x_2 and that m_2 is decreasing in x_1, but increasing on x_2. Therefore, the supremum ($\frac{3}{4}$ for both functions) is attained at $m_1(\frac{1}{2},0)$ and at $m_2(0,\frac{1}{2})$. Then we can say that $P(\sup_{x\in[0,\frac{1}{2}]\times[0,\frac{1}{2}]} U^{C^F}(x) > \lambda) \leq P(\sup_{x\in[0,\frac{3}{4}]^2} U^J(x) > \lambda)$.

In general, if we can find the function $m(x)$ as defined in Theorem 5.2.2, which in turn means we have some information about the copula structure, we can use this alternative approach to obtain more accurate critical values.

b)Standarization

Up to this point, we have assumed that the process

$$T_{n,m}(p,q) \equiv \sqrt{m}[G_m(F_n^{1-}(p), F_n^{2-}(q)) - F_n(F_n^{1-}(p), F_n^{2-}(q))]$$

can be completely observed on some region $[p_0, q_0]$. However, there may be situations in which only aggregate data is available, and we observe only $T_{n,m}(p_0, q_0)$ at one fixed value of (p, q). It is our intention to prove that under H_0,

$$\frac{T_{n,m}(p_0, q_0)}{\sqrt{1+\frac{m}{n}}\sqrt{C_n^F(p_0, q_0)[1 - C_n^F(p_0, q_0)]}} \to_{\mathcal{D}} N(0,1), \tag{5.2.1}$$

where $C_n^F = F_n(F_n^{1-}, F_n^{2-})$ is the empirical copula of the F distribution. In this case, the appropriate critical values for each of the alternatives are easily determined.

We will rewrite $\frac{T_{n,m}(p_0,q_0)}{\sqrt{(1+\frac{m}{n})C_n^F(p_0,q_0)[1-C_n^F(p_0,q_0)])}}$ as:

$$\underbrace{\frac{T_{n,m}(p_0, q_0)}{\sqrt{(1+\lambda)C^F(p_0, q_0)[1 - C^F(p_0, q_0)]}}}_{(a)} \underbrace{\frac{\sqrt{1+\lambda}}{\sqrt{1+\frac{m}{n}}}}_{(b)} \underbrace{\frac{\sqrt{C^F(p_0, q_0)[1 - C^F(p_0, q_0)]}}{\sqrt{C_n^F(p_0, q_0)[1 - C_n^F(p_0, q_0)]}}}_{(c)}.$$

Since under the null hypothesis $T_{n,m}(p_0, q_0) \to_{\mathcal{D}} \sqrt{1+\lambda}U^{C^F}(p_0, q_0)$, which is normal with mean 0 and variance $(1 + \lambda)C^F(p_0, q_0)[1 - C^F(p_0, q_0)]$, it immediately follows that $(a) \to_{\mathcal{D}} N(0, 1)$. From the fact that $\frac{m}{n} \to \lambda$, it is clear that $(b) \to 1$. Finally, to deal with (c) we first recall that the empirical copula C_n^F converges to C^F a.s. (Theorem 3.2.5). Then we observe that the functions $f(x) = x(1 - x)$ and $g(x) = \sqrt{x}$ (and therefore their composition) are continuous. This implies that $C_n^F(p_0, q_0)[1 - C_n^F(p_0, q_0)] \to_{a.s} C^F(p_0, q_0)[1 - C^F(p_0, q_0)]$, and we can deduce that $(c) \to_{a.s.} 1$. Equation (5.2.1) follows from the convergence of $(a), (b), (c)$ and Slutsky's theorem.

Comment 5.2.4 It is important to remark that the particular bootstrap technique that we used in the complete samples case cannot be used when we only have partial data, since the full samples are needed for resampling.

Although all the results in this section are for the A region, it is worth mentioning that to handle information for the D region we can take two approaches. The first one is to work with the limiting distribution that was developed in the previous section. The second, and more convenient, one is to work with the equivalent region E instead (see figure 2.2). We have not developed the limiting distribution of this region because it is possible to use the asymptotics found for the A region for this purpose. What we need to do is to invert all our data so that the points $(0, 0)$ and $(1, 1)$ become $(1, 1)$ and $(0, 0)$, respectively. By "flipping all our data over" we will end up with observations in the A region, and we have already presented different alternatives to work with this type of data. It should be noted that, since we are inverting all our data, the resulting copula structure under $H_0 : F = G$ will correspond to the survival copula of F (see Definition 7.1.11). All the results developed for the A region hold for the E region under the above mentioned transformation and the corresponding survival copula structure.

Chapter 6

Conclusion

The goal of this thesis has been to suitably extend the concepts of a $p-p$ plot and a precedence test to higher dimensions. Although the focus has been on the underlying theory, potential applications have been proposed.

For multivariate data that generates the minimal filtration -like geographical data- and data that generates the product filtration -like clinical experiments-, we have defined two different extensions of a $p-p$ plot to $\Re^d$. We developed a Glivenko-Cantelli type of result for each of these $p-p$ plots, as well as described their respective asymptotic Gaussian behaviour. We have used these results to construct tests of stochastic order between two distributions. These tests have the advantage that they do not necessarily need the complete samples to be effective; instead, partial data may be enough.

At this point, we would like to mention that, although we have managed to achieve all the goals we set for ourselves at the beginning of this work -and that our theoretical results are useful as they are-, there is still plenty of research to be done on this topic.

The first step in this direction would be to do some simulations in order to calculate the power of our tests. Once we have their power we would be able to compare them both to other tests (for example, to two one-dimensional tests) and to other test statistics (for example, a Cramer-Von Mises test statistic for a two sided alternative). We could, as well, assess the adequacy of tests based on partial data.

Further research is also needed in the case that our samples are small, instead of large. We did not go into the details of the small sample scenario because our focus was on the asymptotic behaviour of our two test statistics. Regarding the large sample case, we have developed a weak approximation for both our processes; it would be interesting to obtain a strong approximation. We would also suggest, as future research, to look for weak and strong approximations for weighted processes.

Also, in essence, our tests can set a cut-off point, where only the data found before such point will be considered. This can be regarded as a particular form of censoring. Therefore, extending our schemes and results to include censored data might turn out to be a productive idea.

Chapter 7

Appendix A

7.1 Copulas

For this appendix chapter we have gathered some definitions, theorems and properties concerning copulas. Although the literature on copulas is very extensive, we will only be covering the small part that is needed for the present work. All the following results, and their corresponding proofs when applicable, can be found in [35].

Definition 7.1.1 *Let S_1 and S_2 be nonempty subsets of $\overline{\Re} = [-\infty, \infty]$, and let H be a function with domain $S_1 \times S_2$. Let $B = [x_1, x_2] \times [y_1, y_2]$ be a rectangle whose vertices are in $S_1 \times S_2$. The $H-$volume of B is denoted as $V_H(B)$ and it is defined as $V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$.*

Definition 7.1.2 *A real function H with domain $S_1 \times S_2 \subseteq \overline{\Re}^2$ and range in the real numbers is 2-increasing if $V_H(B) \geq 0$ for all rectangles B whose vertices lie in $S_1 \times S_2$.*

Definition 7.1.3 *Suppose that S_1 has a least element a_1 and that S_2 has a least element a_2. A function H from $S_1 \times S_2$ to $\Re$ is said to be grounded if $H(x, a_2) = 0 = H(a_1, y)$ for all $(x, y) \in S_1 \times S_2$.*

Definition 7.1.4 *A copula is a function C with the following properties:*

- *The domain of C is $[0,1]^2$.*

- *C is grounded and 2-increasing.*

- *For every u in S_1 and every v in S_2, $C(u,1) = u$ and $C(1,v) = v$.*

Equivalently, a copula is a function C from $[0,1]^2$ to $[0,1]$ with the following properties:

- *For every $u, v \in [0,1]$, $C(u,0) = 0 = C(0,v)$, $C(u,1) = u$ and $C(1,v) = v$.*

- *For every $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$.*

Theorem 7.1.5 *If C is a copula, then for every $(u,v) \in [0,1]^2$, $\max(u+v-1,0) \leq C(u,v) \leq \min(u,v)$.*

Theorem 7.1.6 *If C is a copula, it is uniformly continuous on $[0,1]^2$: for every $u_1, u_2, v_1, v_2 \in [0,1]$, $|C(u_2,v_2) - C(u_1,v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$.*

Theorem 7.1.7 *Let C be a copula. For any $v \in [0,1]$, the partial derivative $\frac{\partial C}{\partial u}$ exists for almost all u, and for such v and u, $0 \leq \frac{\partial C(u,v)}{\partial u} \leq 1$. Similarly, for any $u \in [0,1]$, the partial derivative $\frac{\partial C}{\partial v}$ exists for almost all v, and for such u and v, $0 \leq \frac{\partial C(u,v)}{\partial v} \leq 1$.*

Theorem 7.1.8 *Let H be a joint distribution function with marginals F and G. Then there exists a copula C such that for all $x, y \in \overline{\Re}$,*

$$H(x,y) = C(F(x), G(y)). \tag{7.1.1}$$

If F and G are continuous, then C is unique. Conversely, if C is a copula and F and G are distribution functions, then the function defined by equation 7.1.1 is a joint distribution function with marginals F and G.

Corollary 7.1.9 *Let H be a joint distribution function with continuous marginals F and G, and let F^- and G^- be the left continuous inverses of F and G, respectively. Then for any $(u,v) \in [0,1]^2$, $C(u,v) = H(F^-(u), G^-(v))$.*

We can restate the last theorem in terms of random variables and their distribution functions.

Theorem 7.1.10 *Let X and Y be random variables with distribution functions F and G, respectively, and joint distribution function H. Then there exists a copula C such that equation 7.1.1 holds. If F and G are continuous, C is unique.*

Definition 7.1.11 *For a pair (X, Y) of random variables with distribution function H, the joint survival function is given by $\overline{H}(x, y) = P(X > x, Y > y)$. If the marginals of $\overline{H}$ are denoted by $\overline{F}$ and $\overline{G}$, then the survival copula is defined as $\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$. As with the usual copula structure, we have that $\overline{H}(x, y) = \hat{C}(\overline{F}(x), \overline{G}(y))$.*

Theorem 7.1.12 *Let X and Y be continuous random variables. Then X and Y are independent if and only if $C(u, v) = \prod(u, v) = uv$. $\prod$ is called the product copula.*

Example 7.1.13 The Farlie-Gumbel-Morgenstern (FGM) family of copulas is given by $C(u, v) = uv + \theta uv(1 - u)(1 - v)$, where $-1 \leq \theta \leq 1$. In view of their simple form, FGM copulas have been used, traditionally, in tests of association and non-parametric studies.

Bibliography

[1] Adler, R.J. and Brown L.D. (1986). Tail behaviour for suprema of empirical processes. *The Annals of Probability* **14**, pages 1-30.

[2] Adler, R.J., Brown L.D. and Lu K.L. (1990). Tests and confidence bands for bivariate cumulative distribution functions. *Communications in Statistics: Simulation and Computation* **19**, pages 25-36.

[3] Al-Hussaini, A. and Elliott, R.J. (1981). Martingales, potentials and exponentials associated with a two-parameter jump process. *Stochastics* **6**, pages 23-42.

[4] Aly, E.A.A., Csörgo, M. and Horváth, L. (1987). P-P plots, rank processes, and Chernoff-Savage theorems. *New Perspectives in Theoretical and Applied Statistics* John Wiley & Sons, Inc., pages 135-156.

[5] Arenas, C. (1989). On point processes in the plane. *Statistics & Probability* **8**, pages 325-328.

[6] Babu, G.J. and Rao, C.R. (1988). Joint asymptotic distribution of marginal quantiles and quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis* **27**, pages 15-23.

[7] Balakrishnan, N. and Tony Ng, H.K. (2006). *Precedence-Type Tests and Applications.* First Edition, John Wiley & Sons, U.S.A.

[8] Barbe, P., Genest, C., Ghoudi, K. and Rémillard, B. (1989). On point processes in the plane. *Statistics & Probability* **8**, pages 325-328.

[9] Barbe, P., Genest, C., Ghoudi, K. and Rémillard, B. (1996). On Kendall's Process. *Journal of Multivariate Analysis* **58**, pages 197-229.

[10] Brown, L.D. and Rinott, Y. (1988). Inequalities for multivariate infinitely divisible processes. *The Annals of Probability* **16**, pages 642-657.

[11] Chakraborti, S. and Van der Laan, P. (1997). An overview of precedence-type tests for censored data. *Biometrika* **1**, pages 99-116.

[12] Chakraborti, S. and Van der Laan, P. (1996). Precedence tests and confidence bounds for complete data: An overview and some results. *The Statistician* **45**, pages 351-369.

[13] Chakraborti, S. and Van der Laan, P. (2001). Precedence tests and Lehmann alternatives. *Statistical Papers* **42**, pages 301-312.

[14] Chang-Jo F.C. (1986). Confidence bands for quantile function and percentile residual lifetime under random censorship. *Ph.D. Thesis*, Carleton University.

[15] Cohen, A. and Sackrowitz, H.B. (1998). Directional tests for one-sided alternatives in multivariate models. *The Annals of Statistics* **26**, pages 2321-2338.

[16] Csörgo, M. and Horváth, L. (1988). A note on strong approximations of multivariate empirical processes. *Stochastic Processes and their Applications* **27**, pages 101-109.

[17] Efron, B. (1979). The Rietz Lecture. Bootstrap methods: another look at the jacknife. *The Annals of Statistics* **7**, pages 1-26.

[18] Eilbott, J. and Nadler, J. (1965). On precedence life testing. *Technometrics* **7**, pages 359-377.

[19] Einmahl, J.H.J. (1987). *Multivariate empirical processes.* First Edition, CWI Tract 32, The Netherlands.

[20] Emerson, J.D. and Moses, L.E. (1985). A note on the Wilcoxon-Mann-Whitney test for $2 \times k$ ordered tables. *Biometrics* **41**, pages 303-309.

[21] Epstein, B. (1954). Tables for the distribution of the number of exceedances. *Annals of Mathematical Statistics* **25**, pages 762-768.

[22] Fermanian, J.D., Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, pages 847-860.

[23] Fisher, N.I. (1983). Graphical methods in nonparametric statistics: a review and annotated bibliography. *International Statistical Review* **51**, pages 25-58.

[24] Gaenssler, P. and Stute, W. (1979). Empirical processes: A survey of results for independent and identically distributed random variables. *The Annals of Probability* **7**, pages 193-243.

[25] Gaenssler, P. and Stute, W. (1987). *Seminar on Empirical Processes.* DMV Sem **9**. Basel:Birkäuser.

[26] Gastwirth, J.L. (1968). The first-median test: a two-sided version of the control median test. *Journal of the American Statistical Association* **63**, pages 692-706.

[27] Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the Americal Statistical Association* **88**, pages 1034-1043.

[28] Ivanoff, B.G. and Merzbach, E. (2000). Set-indexed Markov processes. *Canadian Mathematical Society Conference Proceedings* **20**, pages 217-232.

[29] Ivanoff, B.G. and Merzbach, E. (2000). *Set-Indexed Martingales.* First Edition, Chapman & Hall/CRC, U.S.A.

[30] Lin, C.H. and Sukhatme, S. (1992). On the choice of precedence tests. *Commun. Statist.-Theory Meth.* **21**, pages 2949-2968.

[31] Liu, J. (1992). Precedence probabilities and their applications. *Commun. Statist.-Theory Meth.* **21**, pages 1667-1682.

[32] Liu, R.Y., Parelius, J.M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics* **27**, pages 783-858.

[33] Massey, F.J. (1951). Anote on a two-sample test. *Annals of Mathematical Statistics* **22**, pages 304-306.

[34] Mood, A.M. (1954). On the asymptotic relative efficency of certain non-parametric two-sample tests. *Annals of Mathematical Statistics* **25**, pages 514-522.

[35] Nelsen, R.B. (1999). *An Introduction to Copulas.* First Edition, Springer-Verlag, U.S.A.

[36] Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A. and Úbeda-Flores, M. (2001). Distribution functions of copulas: a class of bivariate probability integral transforms. *Statistics and Probability Letters.* **54**, pages 277-282.

[37] Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A. and Úbeda-Flores, M. (2003). Kendall distribution functions. *Statistics and Probability Letters.* **65**, pages 263-268.

[38] Nelson, L.S. (1963). Tables for a precedence life test. *Technometrics* **5**, pages 491-499.

[39] Nelson, L.S. (1993). Technical aids. *Journal of Quality Technology* **25**, pages 140-142.

[40] Rosenbaum, S. (1954). Tables for a non-parametric test of location. *Annals of Mathematical Statistics* **25** pages 146-150.

[41] Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications.* First Edition, Academic Press, Inc., U.S.A.

[42] Shorack, R.A. (1967). On the power of precedence life tests. *Technometrics* **9**, pages 154-158.

[43] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with applications to Statistics.* First Edition, John Wiley & Sons, U.S.A.

[44] Slud, E.V. (1992). Best precedence tests for censored data. *Journal of Statistical Planning and Inference* **31**, pages 283-293.

[45] Tukey, J.W. (1959). A quick, compact, two-sample test to Duckworth's specifications. *Technometrics* **1**, pages 31-48.

[46] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* First Edition, Springer-Verlag, U.S.A.

[47] van Lieshoot, M.N.M. (2000). *Markov Point Processes and Their applications.* First Edition, Imperial College Press, England.

[48] Westenberg, J. (1948). Significance test for median and interquartile range in samples from continuous populations of any form. *Proc. Konin. Neder. Akad. Wetens.* **51**, pages 252-261.

[49] Young, D.H. (1973). Distributions of some censored rank statistics under Lehmann alternatives for the two-sample case. *Biometrika* **60**, pages 543-549.