A Measure of Mutual Complete Dependence

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften der
Technischen Universität Dortmund

Der Fakultät Statistik der
Technischen Universität Dortmund
vorgelegt von

Pavel A. Stoimenov

aus Sofia

Tag der mündlichen Prüfung: 5. März 2008

Erstgutachter: Professor Dr. Walter Krämer
Zweitgutachter: Professor Dr. Joachim Kunert

Acknowledgements

This dissertation was written during my time at the Institute of Business and Social Statistics at the Department of Statistics of the TU Dortmund.

First, I would like to thank the Ruhr Graduate School in Economics (RGS Econ) and, in particular, the Alfried Krupp von Bohlen and Halbach Foundation for granting me a three year scholarship, which enabled the research presented here.

There are two people I am particularly grateful to. The first one is my supervisor Professor Dr. Walter Krämer, who drew my attention to the interesting field of copulas and who has consistently and wholeheartedly aided and supported me in the years since. The second one is Professor Dr. Karl Friedrich Siburg, who has not only been my university mathematics professor and an invaluable help, but who also inspired me to pursue a dissertation in mathematical statistics.

I would also like to thank my colleagues at the Institute of Business and Social Statistics, many of whom contributed to this thesis in one way or another. In particular, I have to mention Matthias Arnold, Tobias Pröger and Gabi Totova.

Many thanks also to Professor Dr. Joachim Kunert for taking the effort to referee this thesis.

Contents

List of Figures

Chapter 1

Introduction and motivation

The concept of stochastic dependence plays a major role in probability theory and statistics. Its importance stems from the fact that without an appropriate modelling of the dependence in a multivariate setting, no meaningful statistical model can be developed. In particular, one often encounters the problem of measuring, e.g., by a scalar value, the strength of dependence between random variables.

In the following consider two random variables defined on a common probability space. The nature of their dependence can take a variety of forms, of which, undoubtedly, the most prominent one is actually a "lack of dependence", i.e., stochastic independence. Naturally, the question of the "strongest" dependence arises. Since independence is equivalent to complete unpredictability of either random variable from the other, the strongest dependence should correspond to complete predictability, i.e., almost sure bijective functional dependence. This form of extreme dependence, introduced by Lancaster (1982), is known as "mutual complete dependence". Since stochastic independence and mutual complete dependence are, obviously, exactly opposite in character, it follows that a natural requirement on a bivariate measure of dependence is to measure the degree of mutual complete dependence, with extreme values of 0 if and only if the two variables are independent and 1 if and only if they are mutually completely dependent.

However, despite abundant work on measures of dependence, especially for the bivariate case, several fundamental issues are still unsettled. In particular, a measure of dependence which captures adequately the extremes of stochastic dependence, in the above mentioned sense, does not exist. In fact, even the question of the "strongest" possible dependence is highly controversial in the literature. This is due to the counterintuitive fact that the joint distribution function of any two continuous random variables can be approximated uniformly by the joint distribution function of a pair of mutually completely dependent random variables with the same marginal distributions (Kimeldorf and Sampson, 1978; Mikusiński, Sherwood, and Taylor, 1992; Vitale, 1990). In particular, this means that, regardless of the type of the marginal distributions, one can find a sequence of pairs of mutually completely dependent random variables which converges in distribution to a pair of independent random variables. Thus, in terms of convergence in law, it would be impossible, experimentally, to distinguish between these two dependence concepts, although intuitively they are most opposite in character. This paradox led several authors to the conclusion that "mutual complete dependence is too broad a concept to be an antithesis of independence" (Kimeldorf and Sampson, 1978). The "defect" of mutual complete dependence motivated Kimeldorf and Sampson (1978) to consider a new concept of extreme dependence, called "monotone dependence", which also found considerable attention in the construction of measures of dependence.

In this dissertation, we argue that the inconsistency between mutual complete dependence and convergence in distribution neither weakens the concept of mutual complete dependence as the opposite of independence, nor does it imply that a measure of dependence should be restricted to monotone dependence. It rather suggests that convergence in law is an inappropriate concept for the construction of measures of dependence.

The main contribution of the dissertation consists in a new method to detect and measure mutual complete dependence of arbitrary form. The approach is based on copulas. As the next chapter provides an introduction to the theory of copulas, we only mention here that, by virtue of a fundamental result known as Sklar's theorem, the joint distribution function of any two random variables, defined on a common probability space, can be decomposed into the marginal distribution functions and a copula. If the marginal distribution functions are continuous, the copula is unique. In this case, it follows that the dependence between the random variables is fully captured by their copula. For example, they are independent if and only if their connecting copula is the so called product copula. Thus, a possible approach to measuring their stochastic dependence consists in measuring the distance between their copula and the product copula.

This method for constructing a measure of dependence is not new. We argue, however, that it yields, in general, a measure of independence only. While independence in the variables can be detected using any distance function, the type of dependence detected depends heavily on the type of the distance function employed. It follows that the choice of the distance function cannot be arbitrary, but is predetermined by the desired properties of the resulting measure of dependence.

We propose to measure the distance between two copulas by a (modified) Sobolev norm, introducing first a scalar product on the set of all two-dimensional copulas. This norm exploits the differentiability properties of copulas and turns out extremely advantageous since the degree of mutual complete dependence between two random variables with continuous distribution functions can be determined by analytical and algebraic properties of their copula. Furthermore, with respect to the Sobolev norm, a sequence of copulas corresponding to mutual complete dependence can only converge to a copula which itself links mutually completely dependent random variables. Thus, mutual complete dependence cannot approximate any other kind of stochastic dependence. This resolves the counterintuitive phenomenon described above and warrants the role of mutual complete dependence as the opposite of independence.

Using this Sobolev norm we define the first bivariate measure of mutual complete dependence for two random variables with continuous distribution functions, which is given by the (normalized) Sobolev distance between their unique copula and the product copula, corresponding to stochastic independence. We show that this measure has several appealing properties, e.g., it takes on its extreme values precisely at independence and mutual complete dependence. Furthermore, since the measure is based on copulas, it is nonparametric and remains invariant under strictly monotone transformations of the random variables.

The dissertation consists of four chapters. After this motivation, Chapter 2 provides an introduction to the theory of (bivariate) copulas, and thus establishes the main tool for modelling dependence in all other chapters to follow. The outline of that chapter is as follows. The first section introduces some notation and focuses on the notion of a "two-increasing function", which can be viewed as the two-dimensional analog of an increasing function of one variable. These preliminary concepts and results are used in the second section, where we define copulas and study their properties from the perspective of calculus. The third section introduces a product operation on the set of all copulas. The algebraic properties of copulas are less intuitive and not necessarily present in the standard literature. As will become clear in Chapter 4, however, they play a crucial role in establishing the main results of this dissertation. Altogether the first three sections in Chapter 2 introduce copulas and their properties from a purely mathematical perspective. The importance of copulas to statistics becomes clear in Section 2.4, where Sklar's theorem shows that a joint distribution function with continuous margins can be decomposed into the margins and a unique copula. Since distribution functions and random variables are interrelated, it follows that a unique copula can be associated to any pair of random variables with continuous distribution functions. A special focus should be laid upon the final Section 2.5, which clarifies the interpretation of copulas as dependence functions and, thus, justifies the use of copulas as a tool for modelling stochastic dependence.

Chapter 3 gives an overview of different dependence concepts and measures of dependence for the bivariate case. The exposition is based essentially on the theory of copulas and, in particular, on Section 2.5. There is no attempt to be exhaustive in mentioning all dependence concepts that have ever been proposed in the literature. The first section in that chapter deals with extreme stochastic dependence. On the one hand, it introduces the concept of mutual complete dependence; on the other hand, it elucidates the above mentioned counterintuitive phenomenon, which states that in the sense of convergence in law mutual complete dependence is indistinguishable from any other kind of dependence relation. In Section 3.2 we present and comment on the different axiomatic approaches for defining a bivariate measure of dependence which has been proposed in the literature. The last section in this chapter is concerned with different construction methods for a bivariate measure of dependence. Special attention is given to the existing measures defined in terms of copulas.

Chapter 4 contains the main results of the dissertation. It begins with a short summary of the main ideas and open issues presented in the preceding chapter. We argue in favour of the concept of mutual complete dependence as the opposite of stochastic independence and, thus, motivate the need for a new measure of bivariate dependence, namely one measuring the strength of mutual complete dependence. Section 4.2 establishes the necessary mathematical framework by introducing the Sobolev scalar product for copulas and its corresponding norm and metric. We show that the scalar product allows a representation via the product operation on the set of all copulas introduced in Section 2.3. In Section 4.3 we turn to the statistical interpretation of the Sobolev norm for copulas, which allows to detect mutual complete dependence of arbitrary form. This leads naturally, in Section 4.4, to a new nonparametric measure of dependence for two continuous random variables. The final Section 4.5 concludes the dissertation with some examples and comparisons.

Chapter 2

Copulas

2.1 Preliminaries

The focus of this section is the notion of a "two-increasing function", which can be viewed as the two-dimensional analog of an increasing function of one variable. We will see that, with some additional assumptions, this class of functions possesses several appealing properties. These preliminary concepts and results, which can be found in, e.g., Cherubini, Luciano, and Vecchiato (2004), Nelsen (2006) and Schweizer and Sklar (2005), will prove essential in the next section, where we define copulas and study their properties from the perspective of calculus.

First we need some notation. Let $\mathbb{R}$ denote the real line, $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ denote the extended real line, and $\overline{\mathbb{R}}^2 := \overline{\mathbb{R}} \times \overline{\mathbb{R}}$ denote the extended real plane. A (halfopen) rectangle $\mathcal{R} \subseteq \overline{\mathbb{R}}^2$ is the Cartesian product of two (halfopen) intervals: $\mathcal{R} := (x_1, x_2] \times (y_1, y_2]$. The vertices of $\mathcal{R}$ are the points $(x_1, y_1), (x_1, y_2), (x_2, y_1)$, and (x_2, y_2).

Definition 2.1. *Let $S_1, S_2 \subset \overline{\mathbb{R}}$ be nonempty, and let H be a function such that $H : S_1 \times S_2 \to \mathbb{R}$. Let $\mathcal{R} := (x_1, x_2] \times (y_1, y_2]$ be a rectangle all of whose vertices lie in $S_1 \times S_2$. Then the H-volume of $\mathcal{R}$ is given by*

$$V_H(\mathcal{R}) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \qquad (2.1)$$

Definition 2.2. *Let* $S_1, S_2 \subset \overline{\mathbb{R}}$ *be nonempty, and let* H *be a function such that* $H : S_1 \times S_2 \to \mathbb{R}$. H *is two-increasing if* $V_H(\mathcal{R}) \geq 0$ *for all rectangles* $\mathcal{R}$ *whose vertices lie in* $S_1 \times S_2$.

Note that the statement "H is two-increasing" neither implies nor is implied by the statement "H is increasing[1] in each argument." For example, let $H : [0,1]^2 \to \mathbb{R}$ with $(x, y) \mapsto max(x, y)$. Then H is increasing in each argument, however, $V_H([0,1]^2) = -1$, so that H is not two-increasing. In the other direction, let $H : [0,1]^2 \to \mathbb{R}$ with $(x, y) \mapsto (2x-1)(2y-1)$. Then H is two-increasing, but it is decreasing in x for $y \in (0, 1/2)$.

The following lemmas will be very useful in the next section, which establishes the continuity of copulas.

Lemma 2.3. *Let* $S_1, S_2 \subset \overline{\mathbb{R}}$, *and let* $H : S_1 \times S_2 \to \mathbb{R}$ *be a two-increasing function. Let* $x_1, x_2 \in S_1$ *with* $x_1 \leq x_2$ *and let* $y_1, y_2 \in S_2$ *with* $y_1 \leq y_2$. *Then the functions*

$$\begin{aligned} &t \mapsto H(t, y_2) - H(t, y_1) \\ &t \mapsto H(x_2, t) - H(x_1, t) \end{aligned}$$

are increasing on S_1 *and* S_2, *respectively.*

Proof. Since H is two-increasing, it follows from Definition 2.2 that

$$\begin{aligned} &H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \geq 0 \\ \Leftrightarrow\ &H(x_2, y_2) - H(x_2, y_1) \geq H(x_1, y_2) - H(x_1, y_1) \\ \Leftrightarrow\ &H(x_2, y_2) - H(x_1, y_2) \geq H(x_2, y_1) - H(x_1, y_1)\,. \end{aligned}$$

This proves the second statement. The first one follows analogously. □

[1] We call a function $f : M \subset \overline{\mathbb{R}} \to \mathbb{R}$ increasing [decreasing] if, for all $x, y \in M$, $x < y$ implies $f(x) \leq f(y)$ [$f(x) \geq f(y)$]. If strict inequalities hold, f is called strictly increasing [strictly decreasing]. The terms "nondecreasing" and "nonincreasing" are not used.

As an immediate application of this lemma, we can show that with an additional assumption a two-increasing function is increasing in each argument.

Definition 2.4. *Let* $S_1, S_2 \subset \overline{\mathbb{R}}$ *such that* $min\, S_1 =: a_1$ *and* $min\, S_2 =: a_2$ *exist, and let* $H : S_1 \times S_2 \to \mathbb{R}$. H *is grounded if*

$$H(x, a_2) = 0 = H(a_1, y) \quad \textit{for all} \quad (x, y) \in S_1 \times S_2 \,. \tag{2.2}$$

Lemma 2.5. *Let* $S_1, S_2 \subset \overline{\mathbb{R}}$*, and let* $H : S_1 \times S_2 \to \mathbb{R}$ *be a grounded, two-increasing function. Then* H *is increasing in each argument.*

Proof. Let $min\, S_1 =: a_1$ and $min\, S_2 =: a_2$, and set $x_1 = a_1$, $y_1 = a_2$ in Lemma 2.3. □

Definition 2.6. *Let* $S_1, S_2 \subset \overline{\mathbb{R}}$*, such that* $max\, S_1 =: b_1$ *and* $max\, S_2 =: b_2$ *exist, and let* $H : S_1 \times S_2 \to \mathbb{R}$*. Then* H *has margins, and the margins are the functions* F *and* G *given by:*

$$\begin{aligned} F : S_1 \to \mathbb{R} \quad &\textit{with} \quad x \mapsto H(x, b_2) \quad \textit{for all} \quad x \in S_1 \\ G : S_2 \to \mathbb{R} \quad &\textit{with} \quad y \mapsto H(b_1, y) \quad \textit{for all} \quad y \in S_2 \,. \end{aligned}$$

We close this section with an important lemma concerning grounded, two-increasing functions with margins.

Lemma 2.7. *Let* $S_1, S_2 \subset \overline{\mathbb{R}}$*, and let* $H : S_1 \times S_2 \to \mathbb{R}$ *be a grounded, two-increasing function with margins. Then*

$$|H(x_2, y_2) - H(x_1, y_1)| \le |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|$$

for all $(x_1, y_1), (x_2, y_2) \in S_1 \times S_2$.

Proof. From the triangle inequality, we have

$$|H(x_2, y_2) - H(x_1, y_1)| \le |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)| \,.$$

Now assume that $x_1 \leq x_2$. Lemma 2.3, Lemma 2.5 and Definition 2.6 imply $0 \leq H(x_2, y_2) - H(x_1, y_2) \leq F(x_2) - F(x_1)$. An analogous inequality holds when $x_2 \leq x_1$. Hence it follows that for any $x_1, x_2 \in S_1$, we have $|H(x_2, y_2) - H(x_1, y_2)| \leq |F(x_2) - F(x_1)|$. Similarly, for any $y_1, y_2 \in S_2$, $|H(x_1, y_2) - H(x_1, y_1)| \leq |G(y_2) - G(y_1)|$, which completes the proof. □

2.2 Definition and analytic properties

Equipped with the concepts and results from the preceding section, we are now in a position to define copulas. The approach presented here is essentially the same as in Nelsen (2006) and Schweizer and Sklar (2005).

Let I denote the closed unit interval $[0, 1]$; analogously I^2 denotes the closed unit square $[0, 1] \times [0, 1]$.

Definition 2.8. *A two-dimensional copula (or briefly, a copula) is a function $C : I^2 \to I$ satisfying the conditions:*

(i) C is grounded.

(ii) C has margins given by $C(u, 1) = u$ and $C(1, v) = v$ for all $u, v \in I$.

(iii) C is two-increasing.

The following definition is equivalent. It emphasizes, however, the three main properties characterizing a copula.

Definition 2.9. *A copula is a function $C : I^2 \to I$ satisfying the conditions:*

(i) $C(u, 0) = C(0, v) = 0$ for all $u, v \in I$.

(ii) $C(u, 1) = u$ and $C(1, v) = v$ for all $u, v \in I$.

(iii) $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for all rectangles $(u_1, u_2] \times (v_1, v_2] \subset I^2$.

Let $\mathfrak{C}$ denote the set of all (two-dimensional) copulas. The conditions in Definition 2.8 or, alternatively, Definition 2.9 together with the preliminary results presented in Section 2.1 imply directly the following properties of copulas.

Theorem 2.10. *The set $\mathfrak{C}$ is closed under convex combinations, i.e., for all $A, B \in \mathfrak{C}$ and for all $a, b \in I$ with $a + b = 1$, $aA + bB \in \mathfrak{C}$.*

The next theorem, which establishes the continuity of copulas, is an immediate consequence of Lemma 2.7.

Theorem 2.11. *Copulas are Lipschitz (and hence uniformly) continuous and satisfy*

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

for all $u_1, u_2, v_1, v_2 \in I$.

Definition 2.12. *Let C be a copula and let a be any number in I. The horizontal section of C at a is the function from I to I given by $t \mapsto C(t, a)$; the vertical section of C at a is the function from I to I given by $t \mapsto C(a, t)$; and the diagonal section of C is the function from I to I given by $t \mapsto C(t, t)$*

Since copulas are grounded, two increasing functions, it follows immediately from Lemma 2.5 that they are increasing in each argument. More formally, we have the following result, which is readily verified using Theorem 2.11.

Theorem 2.13. *The horizontal, vertical, and diagonal sections of a copula C are all increasing and Lipschitz continuous on I.*

Corollary 2.14. *The horizontal and vertical sections of a copula C are all absolutely continuous on I.*

Proof. Set $\epsilon > 0$ arbitrary. Choose $\delta = \epsilon$ and let $\{(x_i, y_i) : i = 1, \ldots, n\}$ be a finite collection of non overlapping bounded open intervals for which $\sum_{i=1}^{n}(y_i - x_i) < \delta$. Then $\sum_{i=1}^{n}(y_i - x_i) < \epsilon$, which immediately implies that $\sum_{i=1}^{n} |f(y_i) - f(x_i)| < \epsilon$ since the horizontal and vertical sections of a copula are Lipschitz continuous with a Lipschitz constant equal to 1. □

The next theorem concerns the differentiability properties of copulas. While these are well known, they are often not exploited. The main results of this dissertation, however, exploit the differentiability properties. For instance, they will prove essential in Chapter 4, where we introduce the Sobolev scalar product for copulas and its corresponding norm and distance.

Let $\partial_i C$, for $i = 1, 2$, denote the partial derivative of a copula C with respect to the i-th variable.

Theorem 2.15. *Let C be a copula. For any $v \in I$, $\partial_1 C(u, v)$ exists for almost all u, and for such u and v,*

$$0 \leq \partial_1 C(u, v) \leq 1 \, .$$

Similarly, for any $u \in I$, $\partial_2 C(u, v)$ exists for almost all v, and for such u and v,

$$0 \leq \partial_2 C(u, v) \leq 1 \, .$$

Furthermore, the functions $v \mapsto \partial_1 C(u, v)$ and $u \mapsto \partial_2 C(u, v)$ are defined and increasing almost everywhere on I.

Proof. The existence of the partial derivatives $\partial_1 C(u, v)$ and $\partial_2 C(u, v)$ follows immediately from Theorem 2.13 because monotone functions are differentiable almost everywhere. The two inequalities follow from Theorem 2.11 by setting $v_1 = v_2$ and $u_1 = u_2$, respectively. If $v_1 \leq v_2$, then, by Lemma 2.3 the function $u \mapsto C(u, v_2) - C(u, v_1)$ is increasing. Hence, $\partial_1(C(u, v_2) - C(u, v_1))$ is defined and nonnegative almost everywhere on I, from which it follows that $v \mapsto \partial_1 C(u, v)$ is defined and increasing almost everywhere on I. A similar result holds for $u \mapsto \partial_2 C(u, v)$. □

The next result shows that any copula C can be recovered from either of its first partial derivatives by integration. It follows immediately from the fact that, by Corollary 2.14, the horizontal and vertical sections of a copula are absolutely continuous.

Theorem 2.16. *Let C be a copula. For all $u, v \in I$, we have*

$$C(u,v) = \int_0^u \partial_1 C(t,v)\, dt = \int_0^v \partial_2 C(u,s)\, ds$$

Theorem 2.17. *In $\mathfrak{C}$, pointwise and L^∞-convergence are equivalent.*

Proof. The set $\mathfrak{C}$ is a compact and convex subset of the space of all continuous real valued functions defined on the unit square I^2 under the topology of uniform convergence. It follows that, in $\mathfrak{C}$, pointwise convergence implies uniform convergence. This yields the desired result. □

Theorem 2.18. *In $\mathfrak{C}$, L^∞-convergence implies L^p-convergence for each $p \in [1, \infty)$.*

Proof. Let $(C_n)_{n\in\mathbb{N}}$ be a sequence in $\mathfrak{C}$ with $\lim_{n\to\infty} \|C_n - C\|_{L^\infty} = 0$ for some $C \in \mathfrak{C}$. Now observe that for any L^∞-function f on I^2 with $\|f\|_{L^\infty} \leq 1$ we have

$$\|f\|_{L^p}^p = \int_{I^2} |f|^p \, d\lambda \leq \int_{I^2} \|f\|_{L^\infty}^p \, d\lambda \leq \int_{I^2} \|f\|_{L^\infty} \, d\lambda = \|f\|_{L^\infty}$$

where λ denotes the two-dimensional Lebesgue measure. It follows that, for all $n \in \mathbb{N}$, $\|C_n - C\|_{L^p}^p \leq \|C_n - C\|_{L^\infty}$, from which we immediately conclude that $\lim_{n\to\infty} \|C_n - C\|_{L^p} = 0$. □

Three copulas arise repeatedly:

$$C^-(u,v) = \max(u+v-1, 0), \tag{2.3}$$

$$C^+(u,v) = \min(u,v), \tag{2.4}$$

$$P(u,v) = uv. \tag{2.5}$$

A common property of these three copulas is that they all have important stochastic interpretation, as will become clear in Section 2.5. For now, we only mention that P is called the product copula and C^+ and C^- are called the Fréchet-Hoeffding upper and lower bound, respectively, as indicated by the following result.

Theorem 2.19. *For any copula C and any $(u, v) \in I^2$*

$$C^-(u, v) \leq C(u, v) \leq C^+(u, v).$$

Proof. For any $(u, v) \in I^2$, $C(u, v) \leq C(u, 1) = u$ and $C(u, v) \leq C(1, v) = v$ which yields $C(u, v) \leq \min(u, v)$. The first inequality follows from condition (iii) in Definition 2.9, taking the rectangle $(u, 1] \times (v, 1] \subset I^2$. □

Theorem 2.19 suggests a pointwise partial ordering on the set of all copulas $\mathfrak{C}$.

Definition 2.20. *For two copulas A and B, we say that A is smaller than B (or B is larger than A), and write $A \prec B$ (or $B \succ A$) if $A(u, v) \leq B(u, v)$ for all $(u, v) \in I^2$.*

It follows from Theorem 2.19 that the Fréchet-Hoeffding lower bound C^- is smaller than every copula, and the Fréchet-Hoeffding upper bound C^+ is larger than every copula. This pointwise partial ordering on $\mathfrak{C}$ is called the concordance ordering. Note that the ordering is not total because there are copulas which are not comparable.

2.3 Algebraic properties

In the previous section we introduced copulas and studied their properties from the perspective of calculus alone. The present section deals with the algebraic properties of the set of copulas $\mathfrak{C}$. These concepts, introduced by Darsow, Nguyen, and Olsen (1992), are less intuitive and not necessarily present in the standard literature. As will become clear in Chapter 4, however, they play a crucial role in establishing the main results of this dissertation. The presentation given here is essentially the same as the one in Darsow, Nguyen, and Olsen (1992), although some of the proofs are slightly modified and more detailed, while others are left out since they require additional concepts irrelevant for our purposes. We also refer to Darsow and Olsen (1995) and Olsen, Darsow, and Nguyen (1996).

Definition 2.21. *For any $A, B \in \mathfrak{C}$ and any $(u,v) \in I^2$, define*

$$(A * B)(u,v) = \int_0^1 \partial_2 A(u,t)\, \partial_1 B(t,v)\, dt.$$

First of all, the fact that the partial derivatives are bounded by Theorem 2.15 and integrable by Theorem 2.16 ensures that the integral in Definition 2.21 exists. The next theorem shows that $*$ defines a product operation on $\mathfrak{C}$.

Theorem 2.22. *Let A and B be in $\mathfrak{C}$. Then $A * B$ is in $\mathfrak{C}$.*

Proof. Properties (i) and (ii) in Definition 2.9 are easily verified. To show (iii) consider any rectangle $(u_1, u_2] \times (v_1, v_2] \subset I^2$.

$$\begin{aligned}
&A * B(u_1,v_1) + A * B(u_2,v_2) - A * B(u_1,v_2) - A * B(u_2,v_1) \\
&= \int_0^1 \partial_2 A(u_1,t)\, \partial_1 B(t,v_1) + \partial_2 A(u_2,t)\, \partial_1 B(t,v_2) \\
&\quad - \partial_2 A(u_1,t)\, \partial_1 B(t,v_2) - \partial_2 A(u_2,t)\, \partial_1 B(t,v_1)\, dt \\
&= \int_0^1 \partial_2 (A(u_2,t) - A(u_1,t))\, \partial_1 (B(t,v_2) - B(t,v_1))\, dt \\
&\geq 0
\end{aligned}$$

The last estimate follows from the fact that $u_1 \leq u_2$ and $v_1 \leq v_2$. Then, from Lemma 2.3 the functions $t \mapsto A(u_2,t) - A(u_1,t)$ and $t \mapsto B(t,v_2) - C(B,v_1)$ are increasing. Hence, $\partial_2(A(u_2,t) - A(u_1,t))$ and $\partial_1(B(t,v_2) - B(t,v_1))$ are defined and nonnegative almost everywhere on I. □

Let C be any copula. By direct calculation, the $*$-products of C with C^+ and P are as follows:

$$C^+ * C = C * C^+ = C, \tag{2.6}$$

$$P * C = C * P = P. \tag{2.7}$$

Thus, C^+ and P are unit and null elements with respect to the product operation.

It can be shown that the $*$-product is associative, i.e., for any copulas A, B, C, we have $(A*B)*C = A*(B*C)$. In algebraic terms, Theorem 2.22 together with the associativity property and (2.6) states that, under the $*$-product, the set $\mathfrak{C}$ becomes a monoid, i.e., a semi-group with a unit element. Although $\mathfrak{C}$ is not a group, some of its elements do possess inverses, which, however, are not necessarily commutative. It will turn out in Section 2.5 that the invertible elements of $\mathfrak{C}$ have important probabilistic interpretation and, thus, will play a key role in Chapter 4. Therefore, the rest of this section focuses on them, but first we need two definitions.

Definition 2.23. *For any copula C, the copula $C^\top$ defined by*

$$C^\top(u, v) = C(v, u),$$

for all $(u, v) \in I^2$, is called the transposed copula of C.

Definition 2.24. *Let C be a copula. C is called symmetric if $C = C^\top$.*

It is readily verified that for any $A, B \in \mathfrak{C}$ the following holds

$$(A * B)^\top = B^\top * A^\top. \tag{2.8}$$

Definition 2.25. *Let C be in $\mathfrak{C}$.*

(i) C is called left invertible if there is a copula A, called a left inverse, such that $A * C = C^+$.

(ii) C is called right invertible if there is a copula B, called a right inverse, such that $C * B = C^+$.

(iii) C is called invertible if it is both left and right invertible and, in this case, $A = B$ is called the inverse of C.

On analyzing Definition 2.25, several questions arise. First of all, part (iii) implicitly assumes that the left and right inverses of an invertible copula necessarily coincide. We will address this issue at the end of this section. Another point is the existence of invertible copulas. A simple calculation yields $C^+ * C^+ = C^+$, which shows that the Fréchet-Hoeffding upper bound C^+, defined in (2.4), is invertible. The next result yields a necessary and sufficient condition for invertibility.

Theorem 2.26. *Let C be in $\mathfrak{C}$.*

(i) C is left invertible if and only if for each $v \in I$, $\partial_1 C(u, v) \in \{0, 1\}$ for almost all $u \in I$.

(ii) C is right invertible if and only if for each $u \in I$, $\partial_2 C(u, v) \in \{0, 1\}$ for almost all $v \in I$.

(iii) C is invertible if and only if for each $v \in I$, $\partial_1 C(u, v) \in \{0, 1\}$ for almost all $u \in I$ and for each $u \in I$, $\partial_2 C(u, v) \in \{0, 1\}$ for almost all $v \in I$.

Proof. We prove (i) only, since (ii) follows from (i) by taking transposes, and (iii) follows per definition from (i) and (ii). Suppose that for each $v \in I$, $\partial_1 C \in \{0, 1\}$ for almost all $u \in I$. Since by, Theorem 2.15, for almost all u the function $v \mapsto \partial_1 C(u, v)$ is increasing, it follows that for $u \leq v$,

$\partial_1 C(u,u)\partial_1 C(u,v) = \partial_1 C(u,u)$ for almost all u. Hence

$$\begin{aligned} C^\top * C(u,v) &= \int_0^1 \partial_2 C^\top(u,t)\,\partial_1 C(t,v)\,dt \\ &= \int_0^1 \partial_1 C(t,u)\,\partial_1 C(t,v)\,dt \\ &= \int_0^1 \partial_1 C(t, min(u,v))\,dt \\ &= min(u,v) \\ &= C^+(u,v) \end{aligned}$$

Thus, C is left invertible, and $C^\top$ is a left inverse of C.

In the other direction, suppose $L * C = C^+$. Then for all v

$$\begin{aligned} v &= \int_0^1 \partial_2 L(v,t)\,\partial_1 C(t,v)\,dt \\ &\le \left(\int_0^1 \partial_2 L(v,t)^2\,dt\right)^{1/2} \left(\int_0^1 \partial_1 C(t,v)^2\,dt\right)^{1/2} \\ &\le v^{1/2} \left(\int_0^1 \partial_1 C(t,v)^2\,dt\right)^{1/2} \\ &\le v^{1/2} \left(\int_0^1 \partial_1 C(t,v)\,dt\right)^{1/2} \\ &= v^{1/2} v^{1/2} \\ &= v. \end{aligned}$$

This uses Schwartz's inequality and the fact that, by Theorem 2.15, the first partial derivatives of a copula lie between 0 and 1 almost certainly. It follows that equality must hold at each step in the foregoing chain, so that, from lines 3 and 4, for all $v > 0$,

$$\int_0^1 \left(\partial_1 C(t,v) - \partial_1 C(t,v)^2\right)\,dt = 0\,.$$

Since the integrand in this expression is almost certainly positive, it follows that for all $v > 0$, $\partial_1 C(u,v) \in \{0,1\}$ for almost all u as required. When $v = 0$, $\partial_1 C(u,v) = 0$ for all u, by the boundary condition satisfied by C. This completes the proof. □

Remark 2.27. The proof of Theorem 2.26 shows that the statement "for each $v \in I$, $\partial_1 C(u, v) \in \{0, 1\}$ for almost all $u \in I$" in (i) could be replaced by the weaker statement "$\partial_1 C(u, v) \in \{0, 1\}$ for almost all $(u, v) \in I^2$" without affecting the equivalence. Analogous simplifications hold for (ii) and (iii), and we will take advantage of these in Chapter 4

The proof of Theorem 2.26 shows that if a copula C is left invertible, right invertible, or invertible, the transposed copula $C^\top$ (see Definition 2.23) is a left inverse, a right inverse, or an inverse, respectively. More formally, we have the following result, whose proof we omit in order to simplify the exposition; instead, the reader is referred to Theorems 7.2 and 7.3 in Darsow, Nguyen, and Olsen (1992).

Theorem 2.28. *Left and right inverses in $\mathfrak{C}$ are unique and correspond to the transposed copula.*

Theorem 2.28 shows that if a copula has both left and right inverses, they necessarily coincide, which proves the statement implicitly assumed in Definition 2.25 (iii). Also note that the example $C^+ * C^+ = C^+$ given above to show the existence of invertible copulas by no means contradicts Theorem 2.28 since, by Definition 2.24, C^+ is symmetric, i.e., $C = C^\top$.

Remark 2.29. A copula invertible on one side need not be invertible on the other. Consider, for example, the following one-parameter family of copulas with $\theta \in (0, 1)$, and let

$$C(u, v) = \begin{cases} u & \text{if } u \le \theta v, \\ \theta v & \text{if } \theta v < u < 1 - (1 - \theta)v, \\ u + v - 1 & \text{if } 1 - (1 - \theta)v \le u. \end{cases}$$

It follows that, for any θ, $\partial_1 C \in \{0, 1\}$ almost everywhere in I^2 and therefore by Theorem 2.26 (i) and Remark 2.27 C is left invertible. However, for $\theta v < u < 1 - (1 - \theta)v$, we have $\partial_2 C(u, v) = \theta \notin \{0, 1\}$ and thus by Theorem 2.26 (ii) C is not right invertible.

2.4 Copulas and random variables

The preceding sections in this chapter introduced copulas and their properties from a purely mathematical perspective. The present section establishes the connection between copulas and random variables. The importance of copulas to the theory of statistics stems from a fundamental result known as Sklar's theorem. Sklar's theorem is the foundation of many, if not almost all, of the statistical applications of copulas. It reveals the role which copulas play in the relationship between multivariate distribution functions and their univariate margins.

Theorem 2.30 (Sklar's theorem)**.** *Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all $(x, y) \in \mathbb{R}^2$,*

$$H(x, y) = C(F(x), G(y)). \tag{2.9}$$

If F and G are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Range}\, F \times \mathrm{Range}\, G$.

Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (2.9) *is a joint distribution function with margins F and G.*

This theorem first appeared in Sklar (1959). For a proof we refer also to Nelsen (2006) and Schweizer and Sklar (1974).

Concisely put, a copula "couples" two univariate distribution functions to a joint distribution function. Since distribution functions and random variables are interrelated, it follows that a copula can be associated to any pair of random variables. In particular, if the marginal distribution functions of the random variables are continuous, then the corresponding copula is unique. Therefore, in terms of random variables, Sklar's theorem can be restated as follows.

Theorem 2.31. *Let X and Y be two random variables, on the same probability space, with univariate distribution functions F_X and F_Y, respectively, and joint distribution function $F_{X,Y}$. Then there exists a copula C such that for all $(x, y) \in \mathbb{R}^2$,*

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)). \tag{2.10}$$

If F_X and F_Y are continuous, then C is unique. Otherwise, C is uniquely determined on Range F_X × Range F_Y.

When F_X and F_Y are continuous, the unique copula C in Theorem 2.31 will be called the copula of X and Y and sometimes denoted by $C_{X,Y}$ to emphasize its identification with the random variables X and Y.

Sklar's theorem, however, is not the only way in which copulas relate to random variables. Observe that, with an appropriate extension of its domain to $\mathbb{R}^2$, every copula can be associated with a joint distribution function whose margins are uniform on I. More precisely, we can make the following definition.

Definition 2.32. *Let C be a copula. The function $H_C : \mathbb{R}^2 \to I$ defined as*

$$H_C(u, v) = \begin{cases} 0 & \text{if } u < 0 \text{ or } v < 0\,, \\ C(u, v) & \text{if } (u, v) \in I^2\,, \\ u & \text{if } v > 1 \text{ and } u \in I\,, \\ v & \text{if } u > 1 \text{ and } v \in I\,, \\ 1 & \text{if } u > 1 \text{ and } v > 1\,. \end{cases}$$

will be called the distribution function associated with C.

It is clear that a copula C defines the function H_C uniquely, and vice versa. It is also readily verified that H_C is a bivariate distribution function whose marginal distribution functions are uniform on I.

Observe that all distribution functions associated to copulas coincide on the set $\mathbb{R}^2 \setminus I^2$ and the probability of any Borel set which is contained in $\mathbb{R}^2 \setminus I^2$ is 0. Thus, without loss of information, every copula can be viewed as a restriction to I^2 of a joint distribution function whose margins are uniform on I. Therefore, in the sequel, we omit the word "restriction" and simply refer to copulas as joint distribution functions with uniform margins on I.

It follows that any copula C induces a probability measure P_C on the Borel subsets of I^2 via

$$P_C((u_1, u_2] \times (v_1, v_2]) = V_C((u_1, u_2] \times (v_1, v_2]),$$

where $V_C((u_1, u_2] \times (v_1, v_2])$ is the C-volume of the rectangle $(u_1, u_2] \times (v_1, v_2]$; see Definition 2.1. An extension of the P_C-measure to arbitrary Borel subsets of I^2 can be achieved by standard measure-theoretic techniques. Note that since copulas are continuous functions, the P_C-measure of an individual point in I^2 is 0.

Remark 2.33. P_C-measures are often called doubly stochastic measures, as for any Borel subset S of I,

$$P_C(S \times I) = P_C(I \times S) = \lambda(S),$$

where λ denotes the one-dimensional Lebesgue measure.

The next theorem shows that the copula of two random variables with continuous distribution functions is the joint distribution function of their probability integral transformations.

Theorem 2.34. *Let X and Y be two random variables on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, with continuous marginal distribution functions F_X and F_Y, respectively, joint distribution function $F_{X,Y}$, and copula C. Then C is the joint distribution function of the probability integral transformations $F_X(X)$ and $F_Y(Y)$.*

Proof. From Theorem 2.31 we have the identity:

$$\begin{aligned} C(F_X(x), F_Y(y)) &= F_{X,Y}(x, y) \\ &= \mathcal{P}[X \le x, Y \le y] \\ &= \mathcal{P}[F_X(X) \le F_X(x), F_Y(Y) \le F_Y(y)] \,. \end{aligned}$$

Since F_X and F_Y are continuous, we have $\mathit{Range}\, F_X = \mathit{Range}\, F_Y = I$, from which the desired result follows. □

As a joint distribution function, a copula admits a decomposition into an absolutely continuous component and a singular component; see Nelsen (2006). For any copula C, let

$$C(u, v) = A_C(u, v) + S_C(u, v),$$

where

$$A_C(u, v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} C(s, t)\, dt\, ds$$

and

$$S_C(u, v) = C(u, v) - A_C(u, v).$$

Note that, by virtue of Theorem 2.15, $\partial^2 C(s, t)/\partial s \partial t$ exists almost everywhere in I^2 and is non-negative. If $C = A_C$ on I^2, i.e., if considered as a joint distribution function, C has a joint density given by $\partial^2 C(s, t)/\partial s \partial t$, then C is absolutely continuous. If $C = S_C$ on I^2, i.e., if $\partial^2 C(s, t)/\partial s \partial t = 0$ almost everywhere on I^2, then C is singular. Otherwise, C has an absolutely continuous component A_C and a singular component S_C.

The support of a copula is defined as the complement of the union of all open Borel subsets of I^2 whose C-measure is zero. When the support of C is I^2, we say that C has "full support". In this case, C need not be absolutely continuous. However, when C is singular, its support has Lebesgue measure zero and conversely.

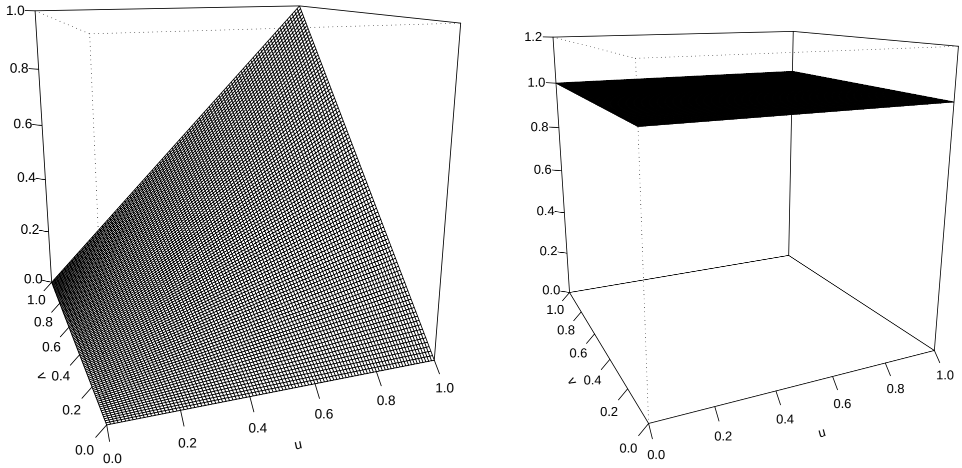

Figure 2.1: The graphs of the product copula P (left) and its density

The rest of the section provides some examples. For instance, the product copula $P(u, v) = uv$, whose graph and density are shown in Figure 2.1 is absolutely continuous, because for all $(u, v) \in I^2$,

$$A_P(u,v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} P(s,t)\, dt\, ds = \int_0^u \int_0^v 1\, dt\, ds = P(u,v)\,. \quad (2.11)$$

Another example of an absolutely continuous copula is the Gaussian copulas C_ρ^{Ga} with parameter $\rho \in [-1, 1]$. The Gaussian copula is defined as follows:

$$C_\rho^{\mathrm{Ga}}(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))\,, \quad (2.12)$$

where Φ_ρ is the joint distribution function of a bivariate standard normal vector, with linear correlation ρ, and Φ is the standard normal distribution function. Therefore (Cherubini, Luciano, and Vecchiato, 2004),

$$C_\rho^{\mathrm{Ga}}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(\frac{2\rho st - s^2 - t^2}{2(1-\rho^2)}\Big)\, ds\, dt \quad (2.13)$$

$$= \int_0^u \int_0^v \frac{1}{\sqrt{1-\rho^2}} \exp\Big(\frac{2\rho mn - m^2 - n^2}{2(1-\rho^2)} + \frac{m^2+n^2}{2}\Big)\, ds\, dt\,, \quad (2.14)$$

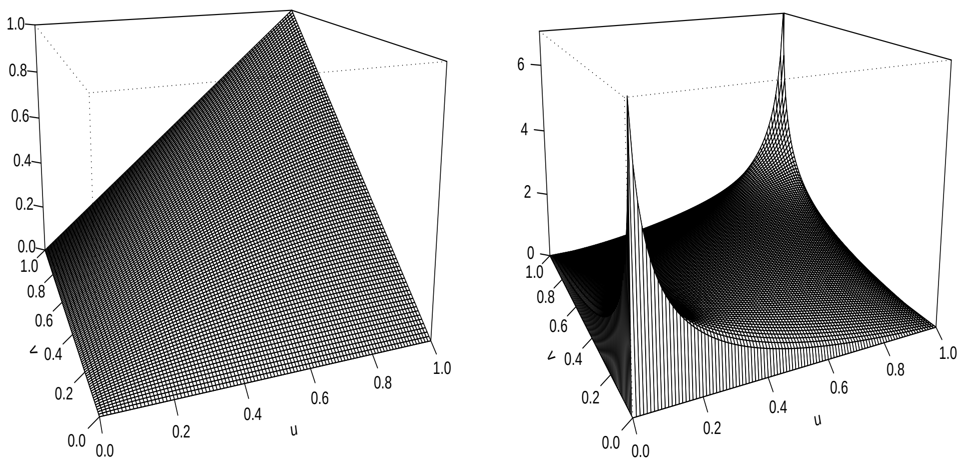

Figure 2.2: The graphs of the Gaussian copula C_ρ^{Ga} with $\rho = 0.5$ (left) and its density

where $m = \Phi^{-1}(s)$ and $n = \Phi^{-1}(t)$. The graph of the Gaussian copula and its density, which is given by the integrand in (2.14), are plotted in Figure 2.2. The next result relates the Gaussian copula to the normal distribution.

Theorem 2.35. *Let X and Y be two random variables defined on a common probability space, with normal marginal distributions and copula C. Then X and Y are jointly normal with correlation coefficient ρ if and only if $C = C_\rho^{Ga}$.*

Proof. If X and Y both have standard normal margins Φ, the result is an immediate consequence of (2.10) and (2.12). For arbitrary normal margins, it follows from (2.10) by (2.13) applying the transformation formula for the Lebesgue integral. □

The support of the Fréchet-Hoeffding upper bound $C^+(u, v) = \min(u, v)$ is the main diagonal of I^2, i.e., the set $\{(u, u) \mid u \in I\}$, so that C^+ is singular; see Figure 2.3. Also note that $\partial^2 C(s,t)/\partial s \partial t = 0$ everywhere in I^2 except on the main diagonal. Similarly the support of the Fréchet-Hoeffding lower

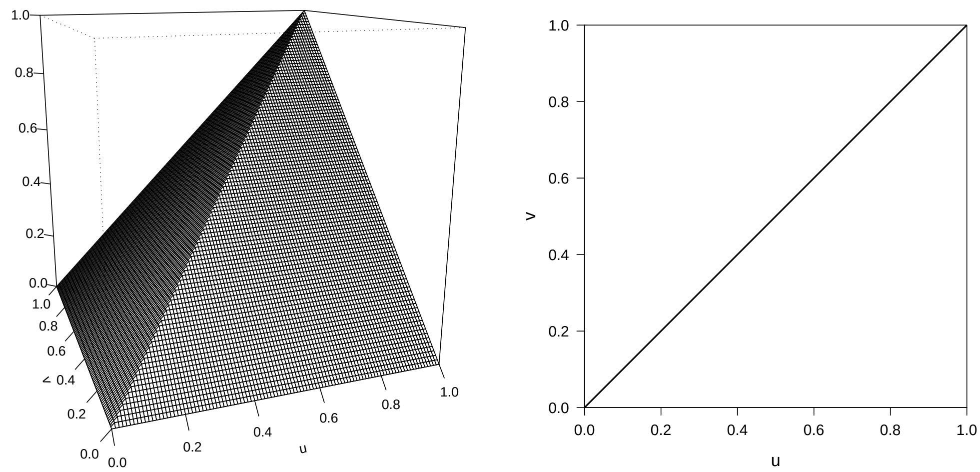

Figure 2.3: The graphs of the Fréchet-Hoeffding upper bound C^+ (left) and its support

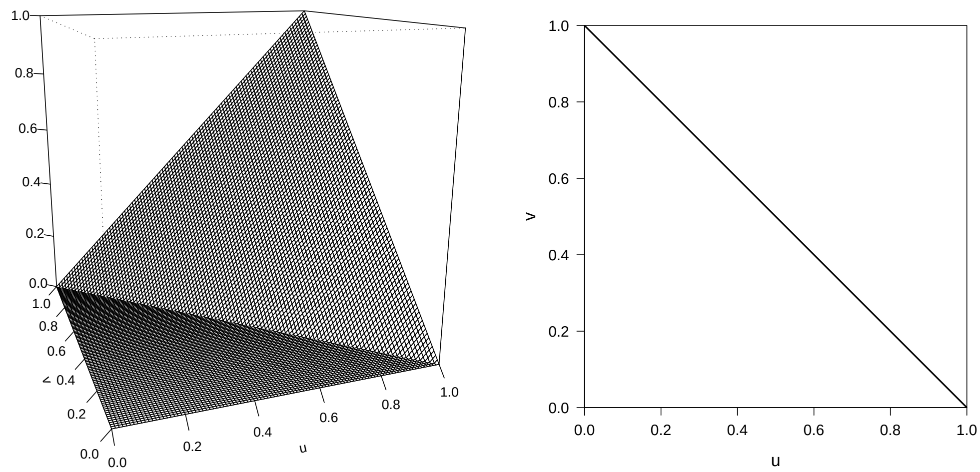

Figure 2.4: The graphs of the Fréchet-Hoeffding lower bound C^- (left) and its support

bound $C^-(u, v) = \max(u + v - 1, 0)$ is the secondary diagonal of I^2, i.e., the set $\{(u, 1 - u) \,|\, u \in I\}$, so C^- is singular as well; see Figure 2.4.

Other examples of singular copulas which are of special interest in this dissertation are the class of copulas called "shuffles of Min" introduced

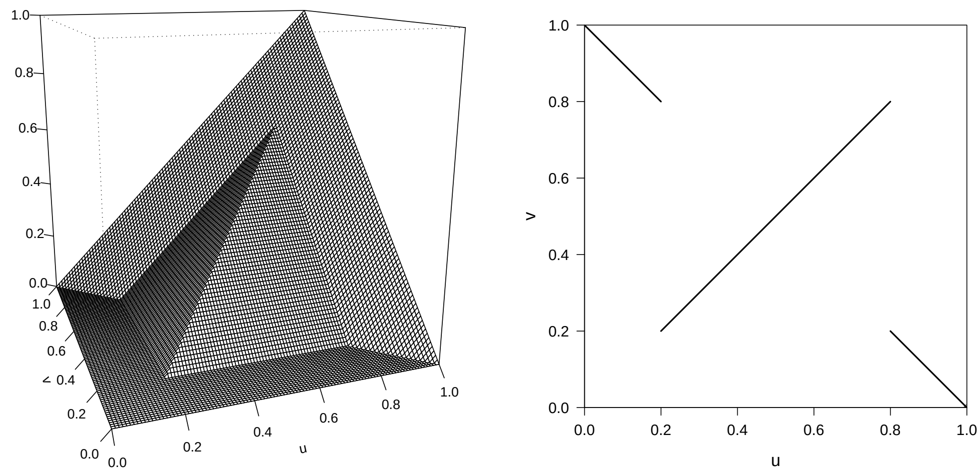

Figure 2.5: The graphs of a shuffle of Min given by $C(u,v) = C^+(u,v) - 0.2$ if $(u,v) \in [0.2, 0.8]^2$ and $C^-(u,v)$ otherwise (left) and its support

by Mikusiński, Sherwood, and Taylor (1992).

Definition 2.36. *A copula is called a shuffle of Min if its support is obtained by*

(i) placing the support for $C^+(u,v) = min(u,v)(= Min)$ on I^2,

(ii) cutting I^2 vertically into a finite number of strips,

(iii) shuffling the strips with perhaps some of them flipped around their vertical axes of symmetry, and then

(iv) reassembling them to form the square again.

Observe that by definition any shuffle of Min is a singular copula. The graph of a shuffle of Min given by

$$C(u,v) = \begin{cases} C^+(u,v) - 0.2 & \text{if } (u,v) \in [0.2, 0.8]^2 \\ C^-(u,v) & \text{if } (u,v) \in I^2 \backslash [0.2, 0.8]^2 \end{cases}$$

and its corresponding support are shown in Figure 2.5.

In the next section, we will see that the preceding examples of copulas, namely P, C^+, C^- and the class called shuffles of Min have very important probabilistic interpretation.

2.5 Copulas as dependence functions

Let X and Y be two random variables on a common probability space, with marginal distribution functions F_X and F_Y, respectively, and joint distribution function $F_{X,Y}$. Then, $F_{X,Y}$ contains the whole information about the distribution of the random vector (X, Y). Intuitively, this information consists of two parts – knowledge of the marginal distributions and knowledge of the dependence structure. By Sklar's theorem 2.31, if F_X and F_Y are continuous, there exists a unique copula $C_{X,Y}$ such that

$$F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y)). \tag{2.15}$$

Thus, in this case, the joint distribution function can be decomposed into the marginal distribution functions and the (unique) copula. It follows that the dependence between X and Y is fully captured by their copula. For this reason, copulas can be interpreted as dependence functions.

For example, the next theorem shows that the product copula $P(u, v) = uv$ corresponds to stochastic independence.

Theorem 2.37. *Let X and Y be two random variables on a common probability space with continuous marginal distribution functions and (unique) copula C. Then $C = P$ if and only if X and Y are independent.*

Proof. This is an immediate consequence of Theorem 2.31 and the observation that X and Y are independent if and only if $F_{X,Y}(x, y) = F_X(x)F_Y(y)$, i.e., the joint distribution function $F_{X,Y}$ is equal to the product of the marginal distribution functions F_X and F_Y. □

Theorem 2.37 motivates the question of the probabilistic interpretations of other copulas, for example, the Fréchet-Hoeffding bounds, C^+ and C^-, or the shuffles of Min, which we encountered in Section 2.4.

Theorem 2.38. *Let X and Y be two random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions and copula C. Then the following assertions hold:*

(i) $C = C^+$ if and only if $\mathcal{P}[Y = f(X)] = 1$, where f is a strictly increasing Borel-measurable function.

(ii) $C = C^-$ if and only if $\mathcal{P}[Y = f(X)] = 1$, where f is a strictly decreasing Borel-measurable function.

Theorem 2.39. *Let X and Y be two random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions and copula C. Then C is a shuffle of Min if and only if $\mathcal{P}[Y = f(X)] = 1$, where f is an invertible Borel-measurable function with a finite number of discontinuities.*

Theorem 2.38 is due to Hoeffding (1994) and Fréchet (1951). For a proof we also refer to Mikusiński, Sherwood, and Taylor (1991/92) and Nelsen (2006). The result in Theorem 2.39 is proved in Mikusiński, Sherwood, and Taylor (1992). We omit both proofs since the verifications are long and rather technical.

Remark 2.40. Mikusiński, Sherwood, and Taylor (1992) refer to the Borel-measurable function f in Theorem 2.38 as strongly piecewise monotone. This is clear because the finitely many discontinuity points of f form a partition of its domain and since f is invertible, it follows that on each of the partition intervals it has to be either strictly increasing or strictly decreasing.

The preceding examples justify the interpretation of copulas as dependence functions. Representation (2.15) also makes clear that the study of stochastic dependence via copulas has the advantage of being independent of the particular type of marginal distributions. Thus, copulas provide a nonparametric tool for assessing and measuring dependence, allowing a direct comparison of two random vectors with arbitrary marginal distributions.

Furthermore, much of the usefulness of copulas in nonparametric statistics derives from the fact that for strictly monotone transformations of the random variables, copulas are either invariant or change in a predictable way. The next theorem shows that it is precisely the copula of a bivariate random vector which captures those properties of the joint distribution function which are invariant under strictly increasing transformations of the univariate margins – so-called scale invariance (Schweizer and Wolff, 1981).

Theorem 2.41. *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions and (unique) copula $C_{X,Y}$, and let f and g be real-valued Borel-measurable functions. Then, for all $u, v \in I$, the following statements are true:*

(i) If f and g are strictly increasing, then

$$C_{f(X),g(Y)}(u, v) = C_{X,Y}(u, v).$$

(ii) If f is strictly increasing and g is strictly decreasing, then

$$C_{f(X),g(Y)}(u, v) = u - C_{X,Y}(u, 1 - v).$$

(iii) If f is strictly decreasing and g is strictly increasing, then

$$C_{f(X),g(Y)}(u, v) = v - C_{X,Y}(1 - u, v).$$

(iv) If f and g are strictly decreasing, then

$$C_{f(X),g(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v).$$

Proof. Let $F_X, F_Y, F_{f(X)}$ and $F_{g(Y)}$ denote the distribution functions of $X, Y, f(X)$ and $g(Y)$, respectively.

If f and g are strictly increasing,

$$F_{f(X)}(x) = P[f(X) \leq x] = \mathcal{P}[X \leq f^{-1}(x)] = F_X(f^{-1}(x))$$

and likewise $F_{g(Y)}(y) = F_Y(g^{-1}(y))$. Thus, for any $x, y \in \mathbb{R}$,

$$\begin{aligned} C_{f(X),g(Y)}(F_{f(X)}(x), F_{g(Y)}(y)) &= \mathcal{P}[f(X) \leq x, g(Y) \leq y] \\ &= \mathcal{P}[X \leq f^{-1}(x), Y \leq g^{-1}(y)] \\ &= C_{X,Y}(F_X(f^{-1}(x)), F_Y(g^{-1}(y))) \\ &= C_{X,Y}(F_{f(X)}(x), F_{g(Y)}(y)). \end{aligned}$$

Because X and Y are continuous, $\mathit{Range}\, F_{f(X)} = \mathit{Range}\, F_{g(Y)} = I$, whence it follows that $C_{f(X),g(Y)}(u, v) = C_{X,Y}(u, v)$ on I^2. This proves (i). (ii)-(iv) can be proved in a similar way. □

Chapter 3

Dependence properties and measures

3.1 Extreme stochastic dependence

Dependence relations between random variables is one of the most widely studied subjects in probability and statistics. The nature of the dependence can take a variety of forms of which, undoubtedly, the most prominent one is actually stochastic independence. Naturally, the question of the "strongest" type of dependence arises. Concentrating on the bivariate case, the following definition by Lancaster (1963) provides an answer.

Definition 3.1. *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Y is called completely dependent on X if there exists a Borel-measurable function f such that*

$$\mathcal{P}[Y = f(X)] = 1. \tag{3.1}$$

X and Y are called mutually completely dependent if Y is completely dependent on X, and X is completely dependent on Y.

Remark 3.2. In other words, X and Y are mutually completely dependent if and only if there is a Borel-measurable bijection f satisfying (3.1).

Obviously, stochastic independence and mutual complete dependence are most opposite in character. The former case entails complete unpredictability of either random variable from the other, whereas the latter corresponds to complete predictability.

It is also clear that if a sequence $(X_n, Y_n)_{n\in\mathbb{N}}$ of pairs of independent random variables converges in distribution to a pair (X, Y) of random variables, then X and Y must be independent. However, if a sequence $(U_n, V_n)_{n\in\mathbb{N}}$ of pairs of mutually completely dependent random variables converges in distribution to a pair (U, V) of random variables, then U and V need not be mutually completely dependent. In fact, it is even possible to construct a sequence of pairs of mutually completely dependent random variables, all having uniform marginal distributions on $I = [0, 1]$, which converges in law to a pair of independent random variables, each having a uniform distribution on I; see Kimeldorf and Sampson (1978). This can be achieved in the following way.

Partition the unit square into n^2 congruent squares and denote by (i, j) the square whose upper right corner is the point with coordinates $x = i/n$, $y = j/n$. Similarly, partition each of these n^2 squares into n^2 subsquares and let (i, j, p, q) denote subsquare (p, q) of square (i, j). Now let the bivariate random variable (U_n, V_n) distribute probability mass n^{-2} uniformly on either one of the diagonals of each of the n^2 subsquares of the form (i, j, j, i) for $1 \le i \le n$, $1 \le j \le n$ (Kimeldorf and Sampson, 1978). Figure 3.1 illustrates the cases $n = 1, \dots, 6$.

In particular, Kimeldorf and Sampson (1978) prove the following result.

Theorem 3.3. *Each of the random variables U_n, V_n has a uniform distribution on I. For each n the random variables U_n and V_n are mutually completely dependent. The sequence $(U_n, V_n)_{n\in\mathbb{N}}$ converges in law to a pair (U, V) of independent random variables with uniform distributions on I.*

Proof. For each n, it is clear that U_n and V_n are mutually completely dependent since the support of their joint distribution function is, by construction, the graph of a bijection. Also, since U_n and V_n each assign probability mass n^{-1} uniformly to each interval $((i-1)/n, i/n)$, it is clear that U_n and V_n have uniform distributions on I. Finally, since (U_n, V_n) assigns total probability mass n^{-2} to each of the n^2 large squares,

$$\lim_{n\to\infty} P(U_n \leq u, V_n \leq v) = uv$$

for each point $(u, v) \in I^2$. □

In the following we will see that the phenomenon described by Kimeldorf and Sampson (1978) can be expressed easily in terms of copulas. Recall from Section 2.4 that copulas are bivariate distribution functions with uniform marginal distributions on I. In fact, it is easy to show that, for all $n \in \mathbb{N}$, the support of the distribution of the random vector (U_n, V_n), which, for $n = 1, \ldots, 6$, is shown in Figure 3.1, can be obtained by the procedure described in Definition 2.36. It follows that, for all n, the joint distribution of (U_n, V_n) is a shuffle of Min. In view of Theorem 2.37, Kimeldorf and Sampson (1978) prove essentially that the product copula, $P(u, v) = uv$, can be approximated pointwise and therefore, by Corollary 2.17, uniformly by certain shuffles of Min.

The next theorem shows that any copula, $C \in \mathfrak{C}$, can be approximated uniformly, arbitrarily closely by shuffles of Min.

Theorem 3.4. *Shuffles of Min are dense in* $\mathfrak{C}$ *endowed with the* L^∞*-norm.*

For a proof see Theorem 3.1 in Mikusiński, Sherwood, and Taylor (1992).

Corollary 3.5. *Shuffles of Min are dense in* $\mathfrak{C}$ *endowed with any* L^p*-norm,* $p \in [1, \infty]$.

Proof. The result follows immediately from Corollary 2.18. □

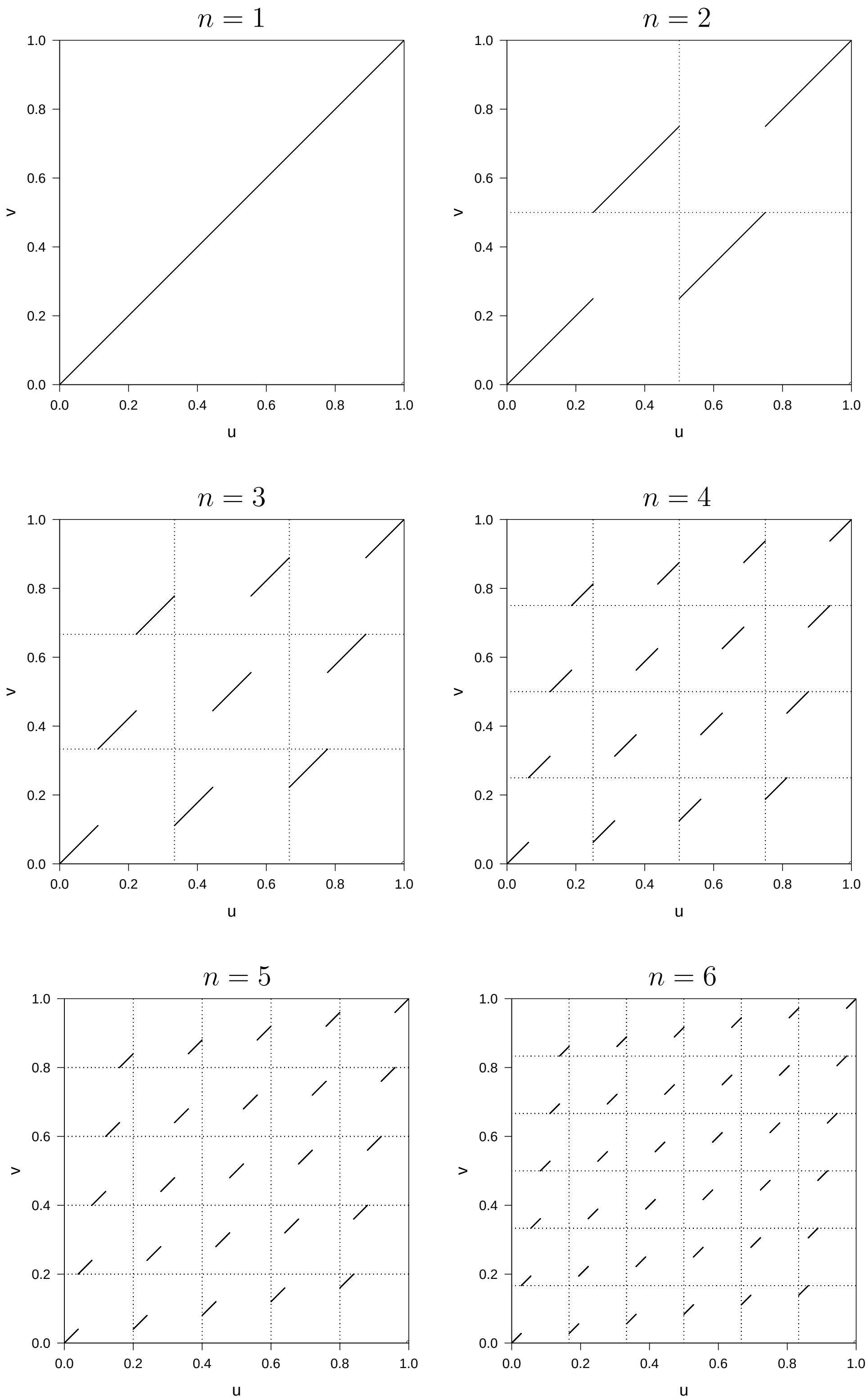

Figure 3.1: Support of the distribution of (U_n, V_n) for $n = 1, \ldots, 6$.

A similar result like the one in Theorem 3.4 was also proved by Vitale (1990) and Vitale (1991) in the context of stochastic processes.

Recall from Theorem 2.39 that the copula of two random variables X and Y with continuous marginal distribution functions is a shuffle of Min if and only if $\mathcal{P}[Y = f(X)] = 1$, where f is an invertible Borel-measurable function with a finite number of discontinuities. In view of Definition 3.1 this results implies immediately the following theorem.

Theorem 3.6. *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions and (unique) copula C and let C be a shuffle of Min. Then X and Y are mutually completely dependent, i.e., $\mathcal{P}[Y = f(X)] = 1$ for some Borel-measurable bijection f.*

Remark 3.7. As noted in Nelsen (2006), however, the converse implication in Theorem 3.6 is not true – there are mutually completely dependent random variables with more complex copulas. In Section 4.3 we will introduce a necessary and sufficient condition for a copula to link mutually completely dependent random variables.

Theorem 3.6 together with Theorem 3.4 yields the following result, which is due to Mikusiński, Sherwood, and Taylor (1992).

Theorem 3.8. *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions F and G, respectively, and joint distribution function H. Then it is possible to find a sequence $(X_n, Y_n)_{n\in\mathbb{N}}$ of pairs of mutually completely dependent random variables with marginal distribution functions $F_n = F$ and $G_n = G$, for all n, and joint distribution functions H_n such that*

$$\lim_{n\to\infty} \|H_n - H\|_{L^\infty} = 0.$$

Proof. Let C be the connecting copula of X and Y. Then by Theorem 3.4 for any $\epsilon > 0$ there exists a shuffle of Min, denoted by $C^{'}$, which uniformly approximates C, i.e.,

$$\|C(u,v) - C^{'}(u,v)\|_{L^\infty} < \epsilon. \tag{3.2}$$

By Theorem 2.31 (Sklar's theorem) the function $H^{'}$ defined on $\mathbb{R}^2$ via

$$H^{'}(x,y) = C^{'}(F(x), G(y)) \tag{3.3}$$

is the joint distribution function of two random variables $X^{'}$ and $Y^{'}$, with marginal distribution functions $F^{'} = F$ and $G^{'} = G$, respectively. Moreover, by Theorem 3.6, $X^{'}$ and $Y^{'}$ are mutually completely dependent. It follows from (3.2) and (3.3) that

$$\sup_{x,y\in\mathbb{R}} |H(x,y) - H^{'}(x,y)| = \sup_{x,y\in\mathbb{R}} |C(F(x),G(y)) - C^{'}(F(x),G(y))| < \epsilon\,.$$

This completes the proof. □

Theorem 3.8 expresses the astonishing fact that the joint distribution function of any two continuous random variables can be approximated uniformly by the joint distribution function of a pair of mutually completely dependent random variables with the same marginal distributions. In particular, this means that, regardless of the type of the marginal distributions, one can pass continuously from mutual complete dependence to stochastic independence in the sense of weak convergence of distribution functions. Thus, in terms of convergence in law, it would be impossible, experimentally, to distinguish between these two dependence concepts, although intuitively they are most opposite in character.

This disturbing and counterintuitive phenomenon led several authors to the conclusion that "mutual complete dependence is too broad a concept to be an antithesis of independence" (Kimeldorf and Sampson, 1978). The 'defect' of mutual complete dependence motivated Kimeldorf and Sampson

(1978) to consider a new concept of extreme dependence, called monotone dependence.

Definition 3.9. *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions. Then Y is called monotone dependent on X if there exists a strictly monotone Borel-measurable function g such that*

$$\mathcal{P}[Y = g(X)] = 1. \tag{3.4}$$

It is obvious that Y is monotone dependent on X if and only if X is monotone dependent on Y. Therefore, Kimeldorf and Sampson (1978) make the following definition.

Definition 3.10. *Two random variables X and Y on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions are monotone dependent if there exists a strictly monotone Borel-measurable function g for which*

$$\mathcal{P}[Y = g(X)] = 1. \tag{3.5}$$

If g is strictly increasing, X and Y are said to be increasing monotone dependent. If g is strictly decreasing, X and Y are said to be decreasing monotone dependent.

Remark 3.11. In view of Definitions 3.1 and 3.10, monotone dependence implies mutual complete dependence, but the converse implication is not true since the Borel-measurable bijection f in (3.1) need not be continuous and therefore monotone. For instance, if the copula of two random variables X and Y with continuous distribution functions is a shuffle of Min, X and Y are mutually completely dependent by Theorem 3.6, but it is clear that they are not monotone dependent; see, for example, Figure 2.5.

Definition 3.10 and Theorem 2.38 immediately yield a necessary and sufficient condition that two continuous random variables be monotone dependent. The result emphasizes again the role of copulas as dependence functions and provides a statistical interpretation of the Fréchet-Hoeffding bounds, $C^+(u,v) = min(u,v)$ and $C^-(u,v) = max(u+v-1,0)$.

Theorem 3.12. *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with continuous marginal distribution functions and (unique) copula C. Then the following assertions hold:*

(i) X and Y are increasing monotone dependent if and only if $C = C^+$.

(ii) X and Y are decreasing monotone dependent if and only if $C = C^-$.

Kimeldorf and Sampson (1978) argue that monotone dependence could be interpreted as the opposite of stochastic independence because, in contrast to mutual complete dependence, the property of monotone dependence is preserved under weak convergence. In particular they prove the following result.

Theorem 3.13. *If $(X_n, Y_n)_{n\in\mathbb{N}}$ is a sequence of pairs of monotone dependent continuous random variables which converges in law to a pair (X,Y) of continuous random variables, then X and Y are monotone dependent.*

Proof. Denote by H_n and H the respective joint distribution functions of (X_n, Y_n) and (X,Y), and denote by F_n, G_n, F, and G the marginal distribution functions of X_n, Y_n, X and Y, respectively. Since $(X_n, Y_n)_{n\in\mathbb{N}}$ converges in law to (X,Y), it follows that, for all $x, y \in \mathbb{R}$, $\lim_{n\to\infty} H_n(x,y) = H(x,y)$, $\lim_{n\to\infty} F_n(x) = F(x)$, and $\lim_{n\to\infty} G_n(y) = G(y)$. Furthermore, there exists a subsequence (X_{n_k}, Y_{n_k}) such that either X_{n_k} and Y_{n_k} are increasing monotone dependent for all k or decreasing monotone dependent for all k. It follows in the former case by Theorem 3.12 and Sklar's theorem that $H_{n_k}(x,y) = min(F_{n_k}(x), G_{n_k}(y))$, which converges to $H(x,y) = min(F(x), G(y))$ where we have used the continuity of the min-function.

Therefore, X and Y are increasing monotone dependent. A similar argument holds if X_{n_k} and Y_{n_k} are decreasing monotone dependent for all k. $\square$

Summarizing the main results of this section it turns out that, in the literature, the question of the "strongest" possible type of dependence is highly controversial. On the one hand, it is intuitively clear that if two random variables are mutually completely dependent, then there exists complete determination of either random variable from the other, due to the almost sure bijective functional relation between them. Consequently, mutual complete dependence can be considered most opposite to stochastic independence. On the other hand, the phenomenon discovered by Kimeldorf and Sampson (1978) and studied among others by Mikusiński, Sherwood, and Taylor (1992) and Vitale (1990) shows that, in the sense of weak convergence of distribution functions, mutual complete dependence is indistinguishable from any other kind of dependence relation, in particular from independence. In the words of Vitale (1990), "this obviously weakens complete dependence as a foil for independence". The search for a solution to this counterintuitive phenomenon led to the concept of monotone dependence, which, as we will see in the next section, also found considerable attention in the construction of measures of dependence.

We argue that the inconsistency between mutual complete dependence and weak convergence by no means weakens the concept of mutual complete dependence as the opposite of independence. In Chapter 4 we introduce a new method which allows to detect mutual complete dependence of arbitrary form. This leads naturally to a measure of mutual complete dependence, but before that the next section provides an overview of the existing scalar measures of dependence.

3.2 Definitions and properties of measures of dependence

In almost every field of application of statistics one often encounters the problem of characterizing by a scalar value the strength of dependence between two random variables X and Y defined on a common probability space. Here, as is often done in the literature, such a scalar value will be referred to as a measure of dependence for X and Y. Obviously, measures of dependence and dependence properties are interrelated and, consequently, it is not surprising that the problem of determining the extremes of stochastic dependence, discussed in the preceding section, is also reflected in the question of how to measure stochastic dependence. In fact, many different measures of dependence have been proposed in the literature, although it is not even clear what a measure of dependence is.

One possible approach to define a measure of dependence is to state a set of desirable properties which the measure should satisfy. For the first time, this was done by Rényi (1959), who proposed a set of seven axioms for an appropriate measure of dependence.

Definition 3.14 (Rényi (1959)). *Let X and Y be random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$. A scalar quantity $\delta(X,Y)$ is a measure of dependence for X and Y if it satisfies the following conditions:*

(i) $\delta(X,Y)$ *is defined for any X and Y, neither of them being constant with probability 1.*

(ii) $\delta(X,Y) = \delta(Y,X)$.

(iii) $0 \leq \delta(X,Y) \leq 1$.

(iv) $\delta(X,Y) = 0$ *if and only if X and Y are independent.*

(v) $\delta(X, Y) = 1$ *if either* Y *is completely dependent on* X *or* X *is completely dependent on* Y*, i.e.,* $\mathcal{P}[Y = f(X)] = 1$ *or* $\mathcal{P}[X = g(Y)] = 1$ *where* f *and* g *are Borel-measurable functions.*

(vi) If $\alpha : \mathbb{R} \to \mathbb{R}$ *and* $\beta : \mathbb{R} \to \mathbb{R}$ *are Borel-measurable bijections, then* $\delta(\alpha(X), \beta(Y)) = \delta(X, Y)$.

(vii) If X *and* Y *are jointly normal with correlation coefficient* ρ*, then* $\delta(X, Y) = |\rho|$.

In the following, we present some comments on the above conditions.

- Condition (i) is to avoid trivialities (Lancaster, 1982).
- Condition (ii) requires that the measure δ is symmetric in X and Y. Lancaster (1982) points out that although independence is a symmetrical property, complete dependence (see Definition 3.1) is not. For example, X could be completely dependent on Y, but, certainly, Y need not be completely dependent on X. In our view, that does not pose a problem since it is impossible to construct a single quantity which measures both the overall dependence and the dependence of either random variable on the other, as alluded to by Lancaster (1982). Therefore, (ii) seems rather intuitive.
- With respect to (iii), Rényi (1959) argued that it is natural to choose the range $[0, 1]$. However, it is clear that if (iv) and (v) are suitably modified, any interval $[a, b] \subset [0, \infty]$ would serve equally well. Therefore, we point out that the true importance of this condition lies in the implication that one is interested in the "strength" rather than the "sign" of dependence. This suggests that the most widely known indices such as Pearson's linear correlation, Spearman's rank correlation, or Kendall's rank correlation, whose range is the interval $[-1, 1]$,

do not qualify as measures of dependence, unless their absolute value is taken.

In fact, Spearman's and Kendall's rank correlations measure the degree of concordance between X and Y, a concept introduced by Scarsini (1984) and discussed, e.g., by Joe (1997) and Nelsen (1991, 2002, 2006). Therefore, they are called measures of concordance. We refer to Scarsini (1984) for an axiomatic definition of a measure of concordance. Informally, two random variables are concordant if "large" values of one tend to be associated with "large" values of the other and "small" values of one with "small" values of the other.

- Condition (iv) states that a measure of dependence should reach its lower bound 0 if and only if the random variables are independent. This condition has two important implications. First, the absolute values of all three correlation coefficients mentioned above cannot serve as measures of dependence since it is well-known that if X and Y are independent, their correlation is 0, but that the converse, in general, is not true. Second, postulate (iv) hints at the definition of a measure of dependence in terms of a measure of distance between the joint distribution and the distribution representing independence of the random variables.

- While there is hardly any disagreement in the literature on desiderata (i)-(iv), (v) is probably the most controversial one. It states that a measure of dependence should reach its highest value 1 if X is completely dependent on Y or vice versa. This postulate has been considered as too strong (Schweizer and Wolff, 1981). Even Renyi himself said that it seems at first sight natural to require that $\delta(X, Y) = 1$ if and only if X is completely dependent on Y or vice versa, but that this condition was rather restrictive, "and it is better to leave it out".

We argue that, even with the converse implication, (v) is not strong enough and even counterintuitive. It hints at complete dependence between X and Y as the strongest type of dependence, although Definition 3.1 makes clear that mutual complete dependence is a much stronger requirement. Therefore, in our view, (v) should read, $\delta(X,Y) = 1$ if and only if X and Y are mutually completely dependent.

- Condition (vi) requires that δ remain invariant under measure-preserving transformations of either or both variables, so δ is a function of Ω.

- (vii) states that if δ is to have general validity, it should coincide in absolute value with Pearson's linear correlation coefficient ρ, when X and Y are jointly normal. This postulate is, obviously, motivated by the fact that the strength of the dependence in a bivariate normal distribution is completely captured by $|\rho|$. However, it is clear, that any strictly increasing function of $|\rho|$ would serve equally well. Therefore, in our view, (vii) is unnecessarily restrictive since it suffices that δ is a strictly increasing function of $|\rho|$.

The axiomatic framework introduced by Rényi (1959) has enjoyed the utmost attention of researchers, e.g., Schweizer and Wolff (1981), Lancaster (1982), Joe (1989), Mikusiński, Sherwood, and Taylor (1992), Zografos (2000), Micheas and Zografos (2006). Several authors have criticized his postulates and tried to extend and enrich them, e.g., Schweizer and Wolff (1981) and Lancaster (1982). One of the major criticisms is that these axioms are too strong. In fact, Rényi himself showed that among various well-known measures of dependence, the only one which satisfies all of his axioms is the maximal correlation coefficient introduced by Gebelein (1941). The maximal correlation is defined by

$$\tilde{\rho}(X,Y) = \sup_{g,h} \rho(g(X), h(Y)), \tag{3.6}$$

where the supremum is taken over all Borel-measurable functions g, h such that $\operatorname{Var} g(X), \operatorname{Var} h(Y) \in (0, \infty)$, and ρ denotes Pearson's correlation coefficient.

However, as pointed out by Hall (1970), the maximal correlation has a number of major drawbacks. For example, it is not effectively computable unless certain regularity conditions are assumed; otherwise nothing is known about the evaluation of $\tilde{\rho}$, and it need not even be attained, i.e., it may exceed the correlation between every pair of random variables $g(X)$ and $h(Y)$. Another drawback of the maximal correlation is that it too easily equals unity, as it suffices that some function $g(X)$ equals some function $h(Y)$ with probability 1.

As already mentioned, we argue that it is a natural desideratum for a measure of dependence between X and Y to measure the strength of mutual complete dependence, with extreme values of 0 if and only if X and Y are independent, and 1 if and only if X and Y are mutually completely dependent. Therefore, in our view, the main drawback of the maximal correlation, $\tilde{\rho}$, is that it does not measure the strength of mutual complete dependence. $\tilde{\rho}$ equals 1 too easily since two random variables with maximum correlation 1 need not be mutually completely dependent.

However, it must be said that for the majority of researchers this does not pose a problem. On the contrary, recall from Theorem 3.8 the astonishing fact that the joint distribution function of any two continuous random variables can be approximated uniformly by the joint distribution function of a pair of mutually completely dependent random variables with the same marginal distributions. This counterintuitive phenomenon led to the concept of monotone dependence (see Definition 3.10), which was suggested as the opposite of stochastic independence, since, by Theorem 3.13, it is preserved under weak convergence; see Kimeldorf and Sampson (1978); Schweizer and Wolff (1981); Vitale (1990, 1991); Mikusiński, Sherwood, and Taylor (1992).

As a natural consequence of these considerations, it was argued that a measure of dependence should measure the strength of monotone dependence. For the first time, this was done by Kimeldorf and Sampson (1978), who proposed the monotone correlation, given by

$$\rho^*(X,Y) = \sup_{g,h} \rho(g(X), h(Y)), \tag{3.7}$$

where the supremum is taken only over strictly monotone Borel-measurable functions g, h such that $\operatorname{Var} g(X), \operatorname{Var} h(Y) \in (0, \infty)$. As noted in Kimeldorf and Sampson (1978), however, two random variables which are monotone dependent have monotone correlation 1, but the converse implication fails.

As a further step in this direction, Schweizer and Wolff (1981) modified Rényi's postulates and proposed a new set of reasonable properties for a measure of dependence. Their definition, which was restricted to continuous random variables, explicitly makes clear that a measure of dependence should measure the strength of monotone dependence.

Definition 3.15 (Schweizer and Wolff (1981))**.** *Let X and Y be two continuous random variables on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$. A scalar quantity $\delta(X,Y)$ is a measure of dependence for X and Y if it satisfies the following conditions:*

(i) $\delta(X,Y)$ *is defined for any X and Y.*

(ii) $\delta(X,Y) = \delta(Y,X)$.

(iii) $0 \leq \delta(X,Y) \leq 1$.

(iv) $\delta(X,Y) = 0$ *if and only if X and Y are independent.*

(v) $\delta(X,Y) = 1$ *if and only if X and Y are monotone dependent, i.e., $\mathcal{P}[X = f(Y)] = 1$ where f is a strictly monotone Borel-measurable function.*

(vi) *If f and g are strictly monotone Borel-measurable functions, then* $\delta(f(X), g(Y)) = \delta(X, Y)$.

(vii) *If X and Y are jointly normal with correlation coefficient ρ, then* $\delta(X, Y)$ *is a strictly increasing function ϕ of $|\rho|$.*

(viii) *If $(X_n, Y_n)_{n\in\mathbb{N}}$ is a sequence of pairs of continuous random variables and if $(X_n, Y_n)_{n\in\mathbb{N}}$ converges in distribution to the pair (X, Y), then* $\lim_{n\to\infty} \delta(X_n, Y_n) = \delta(X, Y)$.

The differences between Definition 3.15 and Rényi's original postulates given in Definition 3.14 can be summarized as follows.

- Conditions (i)-(iv) are identical, except that Rényi's axioms were not restricted to continuously distributed random variables.

- Condition (v) states that a measure of dependence should reach its highest value 1 if and only if X and Y are monotone dependent. It differs from the original condition, which was $\delta(X, Y) = 1$ if X completely dependent on Y or vice versa. In fact, as pointed out by Mikusiński, Sherwood, and Taylor (1992), in the presence of conditions (i) and (iv) and in light of Theorem 3.8 one must choose between Rényi's axiom (v) and condition (viii) in Definition 3.15; they cannot both be true. To see this, let X and Y be independent. We may by Theorem 3.8 construct a sequence $(X_n, Y_n)_{n\in\mathbb{N}}$ which converges in law to the pair (X, Y), such that, for all n, X_n and Y_n are mutually completely dependent, i.e., $\mathcal{P}[X_n = f(Y_n)] = 1$ for a Borel-measurable bijection f. If the measure satisfies both condition (iv) and Rényi's axiom (v), then $\delta(X, Y) = 0$ while, for each n, $\delta(X_n, Y_n) = 1$. This contradicts (viii).

- Condition (vi) is much weaker than Rényi's original postulate. It requires the invariance of $\delta(X, Y)$ to strictly monotone transformations of X and Y only and not to arbitrary ones, as required by Rényi. In addition, Mikusiński, Sherwood, and Taylor (1992) show that if a measure of dependence satisfies (i), (iv) and (v), then it cannot remain invariant under arbitrary measure-preserving transformations and satisfy condition (viii) simultaneously.

- (vii) is weaker than the original condition, but without loss of information. Therefore, as mentioned previously, Rényi's axiom is unnecessarily restrictive and thus (vii) is an improvement.

- Axiom (viii) is a new requirement, introduced probably to avoid inconsistencies with respect to convergence in distribution. However, while the property of monotone dependence is preserved under convergence in law, the property of mutual complete dependence is not; see Theorem 3.13 and Theorem 3.8. Thus, (viii) seems a natural requirement for a measure of monotone dependence, but not for a measure of mutual complete dependence.

An axiomatic approach for defining a measure of dependence has also been suggested by Lancaster (1982). We do not list all of his axioms as they are very similar to the ones suggested by Rényi (1959) with one important exception. Namely, a measure of dependence for two random variables X and Y should take on its highest value 1 if and only if X and Y are mutually completely dependent.

In conclusion, several axiomatic definitions of a measure of dependence for two random variables X and Y have been proposed in the literature, the most prominent of which are the original one proposed by Rényi (1959) and its modifications by Schweizer and Wolff (1981) and Lancaster (1982). In all three definitions most of the conditions for a measure $\delta(X, Y)$ to be useful are identical or very similar. In particular, they all require that a measure of dependence be symmetric and take on a value belonging to the interval $[0, 1]$. Thus, all these measures are nonnegative and do not reveal whether the dependence is "positive" or "negative". In other words they aim at measuring the strength of functional relationship between two random variables rather than their concordance. The latter property can be assessed by a measure of concordance like Spearman's or Kendall's rank correlation. Another property common for all definitions is that a measure of dependence reach its lower bound 0 if and only if the random variables are independent. However, with respect to the behavior of δ at the upper bound 1, the three definitions differ substantially. Rényi's axiom states that $\delta(X, Y) = 1$ if X is completely dependent on Y or vice versa, Schweizer and Wolff's condition requires that $\delta(X, Y) = 1$ if and only if X and Y are monotone dependent, and Lancaster postulates that $\delta(X, Y) = 1$ if and only if X and Y are mutually completely dependent.

As discussed in Section 3.1, we argue that stochastic independence and mutual complete dependence are exactly opposite in character. Therefore, we claim that Lancaster's axiom is a natural and, in fact, an indispensable requirement for a measure of dependence to be able to capture adequately the extremes of stochastic dependence.

3.3 Methods for constructing measures of dependence

In the previous section we saw that a measure of dependence for two random variables X and Y can be defined via a set of desirable properties which it should satisfy. Although different sets of requirements have been proposed in the literature a common axiom of all of them is that the measure equals 0 if and only if X and Y are stochastically independent. This requirement hints at the construction of a measure of dependence in terms of a metrical distance or, in a broad sense, dissimilarity between the joint distribution and the distribution representing independence. Indeed, there is an extensive literature on measures of dependence based on this idea. We refer to Rényi (1959), Joe (1989), Hoeffding (1994), Zografos (1998), Micheas and Zografos (2006) and the references therein for an exhaustive list.

If X and Y have continuous marginal distributions functions F_X and F_Y, respectively, and joint distribution function $F_{X,Y}$, a similar approach to construct a measure of dependence for X and Y is to measure the distance between the joint distribution of their probability integral transformations $F_X(X), F_Y(Y)$ and the respective distribution representing independence. By Theorem 2.34, the joint distribution function of $F_X(X)$ and $F_Y(Y)$ is given by the unique copula of X and Y denoted by $C_{X,Y}$, where the uniqueness of $C_{X,Y}$ follows from Sklar's theorem 2.31. It follows that $F_X(X)$ and $F_Y(Y)$ are independent if and only if $C_{X,Y} = P$, where $P(u, v) = uv$ is the product copula. Thus, this modified method for the construction of a measure of dependence, introduced by Schweizer and Wolff (1981) (see also Schweizer (1991)), is tantamount to measuring the distance between $C_{X,Y}$ and the product copula P.

In fact, the usefulness of this approach for constructing measures of dependence derives primarily from the interpretation of copulas as dependence functions, discussed in detail in Section 2.5. The main idea behind this interpretation stems from Sklar's theorem 2.31, which allows a decomposition of the joint distribution function into the marginal distribution functions and a copula. When the marginal distribution functions are continuous, the copula is unique and it follows that the dependence between X and Y is fully captured by their copula. In particular, Theorem 2.37 shows that $C_{X,Y} = P$ is also equivalent to the stochastic independence of X and Y. It turns out that for continuous random variables, the modified method for the construction of measures of dependence is comparable with the direct measuring of the distance between $F_{X,Y}$ and $F_X \cdot F_Y$.

Moreover, measures of dependence defined in terms of the distance between copulas have several advantages. For instance, they are nonparametric since the particular type of the marginal distributions is irrelevant for determining the dependence structure. This enables a direct comparison between the dependence of two random vectors with arbitrary marginal distributions. In addition, such measures benefit from the fact that under strictly monotone transformations of the random variables, copulas are either invariant or change in a predictable way (see Theorem 2.41).

Exploiting these facts, Schweizer and Wolff (1981) argued that *any* suitably normalized distance between $C_{X,Y}$ and P, in particular, any L^p-distance, should yield a symmetric nonparametric measure of dependence for X and Y. Specifically, they studied the L^1, L^2 and L^∞ distances and denoted the resulting measures by $\sigma(X,Y)$, $\gamma(X,Y)$, and $\kappa(X,Y)$, respectively. These are given by

$$\sigma(X,Y) = 12\|C_{X,Y} - P\|_{L^1}, \tag{3.8}$$

$$\gamma(X,Y) = \sqrt{90}\|C_{X,Y} - P\|_{L^2}, \tag{3.9}$$

$$\kappa(X,Y) = 4\|C_{X,Y} - P\|_{L^\infty}. \tag{3.10}$$

In particular, Schweizer and Wolff (1981) prove the following result.

Theorem 3.16. *Let X and Y be random variables with continuous marginal distribution functions and copula $C_{X,Y}$. Then the quantity $\sigma(X,Y)$ given by* (3.8) *satisfies all conditions in Definition 3.15, with the function ϕ in (vii) given by*

$$\phi(|\rho|) = \frac{6}{\pi}\arcsin(|\rho|/2)\,. \tag{3.11}$$

Proof. It is clear from (3.8) that $\sigma(X,Y)$ is well-defined.

Next, it follows from Theorem 2.34 that $C_{X,Y}(u,v) = C_{Y,X}(v,u)$ which yields (ii).

(iii) follows from Theorem 2.19 and the fact that, for any copula C,

$$\|C - P\|_{L^1} \le 1/12\,. \tag{3.12}$$

(iv) follows from Theorem 2.37 and the fact that the set $\mathfrak{C}$ of copulas, which are continuous functions, endowed with any L^p-distance, $p \in [1, \infty]$, is a metric space.

(v) follows from Theorem 2.38 and Theorem 3.12 and the fact that equality holds in (3.12) if and only if $C_{X,Y} = C^+$ or $C_{X,Y} = C^-$, which is a consequence of Theorem 2.19.

As regards (vi) we distinguish four cases. (a) If f and g are strictly increasing, then by Theorem 2.41 (i) we have $C_{f(X),g(Y)}(u, v) = C_{X,Y}(u, v)$ for all $(u, v) \in I^2$, and thus $\sigma(f(X), g(Y)) = \sigma(X, Y)$. (b) If f is strictly increasing and g is strictly decreasing, then, by Theorem 2.41 (ii), we have $C_{f(X),g(Y)}(u, v) = u - C_{X,Y}(u, 1 - v)$ for all $(u, v) \in I^2$. It follows that $(C_{f(X),g(Y)} - P)(u, v) = (P - C_{X,Y})(u, 1 - v)$ which, by the transformation formula for the Lebesgue measure, implies $\sigma(f(X), g(Y)) = \sigma(X, Y)$. (c) If f is strictly decreasing and g is strictly increasing, the result follows from (b) and (ii). (d) If f and g are strictly decreasing, then, by Theorem 2.41 (iv), we have $C_{f(X),g(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v)$ for all $(u, v) \in I^2$. It follows that $(C_{f(X),g(Y)} - P)(u, v) = (C_{X,Y} - P)(1 - u, 1 - v)$ which, again by the transformation formula for the Lebesgue measure, implies $\sigma(f(X), g(Y)) = \sigma(X, Y)$.

Turning to (vii), 3.11 can be established by exploiting the relationship between σ and Spearman's rank correlation ρ_S, which can be expressed in terms of copulas as (Schweizer and Wolff, 1981)

$$\rho_S(X, Y) = 12 \int_{I^2} (C_{X,Y} - P)\, d\lambda\,. \tag{3.13}$$

As a first step in this direction, note that when X and Y are jointly normal with correlation coefficient ρ, then by Theorem 2.35 $C_{X,Y} = C^{Ga}_{\rho}$. As a consequence of the fact that it is parametrized by ρ, which respects the partial order on the set of copulas $\mathfrak{C}$ (see Definition 2.20), C^{Ga}_{ρ} is positively ordered with respect to ρ (Cherubini, Luciano, and Vecchiato, 2004):

$$C^{Ga}_{\rho=-1} \prec C^{Ga}_{\rho<0} \prec C^{Ga}_{\rho=0} \prec C^{Ga}_{\rho>0} \prec C^{Ga}_{\rho=1}\,. \tag{3.14}$$

This implies immediately that when X and Y are jointly normal, then $C_{X,Y} \succ P$, $C_{X,Y} = P$ or $C_{X,Y} \prec P$ according as $\rho > 0$, $\rho = 0$ or $\rho < 0$. Thus in this case, $\sigma(X,Y) = \rho_S(X,Y)$. But it is well-known (McNeil, Frey, and Embrechts, 2005, Theorem 5.31) that for the bivariate normal distribution

$$\rho_S = \frac{6}{\pi} \arcsin(|\rho|/2),$$

whence (vii) follows.

Lastly, let $(X_n, Y_n)_{n\in\mathbb{N}}$ be a sequence of pairs of continuous random variables with joint distribution functions H_n and copulas C_n. If $(H_n)_{n\in\mathbb{N}}$ converges pointwise to the joint distribution function $H_{X,Y}$ of the pair (X,Y), then $(C_n)_{n\in\mathbb{N}}$ converges pointwise to $C_{X,Y}$ and, by Theorem 2.17 uniformly. This establishes (viii) and completes the proof. □

As noted in Schweizer and Wolff (1981), using completely analogous arguments, it can be shown that the quantity γ defined in (3.9), as well as any other normalized L^p-distance, $p \in [1,\infty)$, satisfies all conditions in Definition 3.15. However, the explicit form of ϕ in (vii) remains to be determined.

When one considers the normalized L^∞-distance κ given by (3.10), the situation changes slightly. The above arguments show that κ does not satisfy condition (v). If X and Y are monotone dependent, i.e., by Theorem 3.12 $C_{X,Y} = C^+$ or $C_{X,Y} = C^-$, then $\kappa(X,Y) = 1$. However, the converse is false since there exist other copulas C, distinct from C^+ and C^-, for which $\|C - P\|_{L^\infty} = 1/4$.

Chapter 4

A measure of mutual complete dependence

4.1 Motivation

By Definition 3.1, two random variables X and Y, defined on a common probability space, are mutually completely dependent if there exists an invertible functional relation between them with probability 1. As argued in Section 3.1, obviously, stochastic independence and mutual complete dependence are exactly opposite in character. In the first case, neither variable provides any information about the other, whereas in the second case there is complete predictability. Therefore, we claim that a measure of dependence for X and Y should measure the degree of mutual complete dependence with extreme values satisfying the following: (i) the measure equals 0 if and only if X and Y are independent, and (ii) it equals 1 if and only if X and Y are mutually completely dependent.

On analyzing the implications of these requirements, several things become apparent. Condition (i) is common for all three axiomatic definitions of a measure of dependence, which were introduced and discussed in detail in Section 3.2. Furthermore, as argued in Section 3.3, (i) hints at the construction of a measure of dependence in terms of a metrical distance or, in a broad sense, dissimilarity between the joint distribution and the distribution representing independence. If X and Y have continuous marginal distribution functions, an alternative method for constructing measures of dependence is to measure the distance between their (unique) copula $C_{X,Y}$ and the product copula P, which corresponds to stochastic independence. This modified approach, introduced by Schweizer and Wolff (1981), has two important advantages. First, since the resulting measure of dependence is a function of the copula only, it is nonparametric as it does not depend on the particular type of the marginal distributions. Second, under strictly increasing transformations of the random variables, the copula, and therefore the dependence measure, are either invariant or change in predictable ways. Exploiting these facts, Schweizer and Wolff (1981) argued that *any* suitably normalized distance between $C_{X,Y}$ and P, in particular, any L^p-distance, $p \in [1, \infty]$, should yield a symmetric nonparametric measure of dependence.

It should be noted, however, that both construction methods described above yield, in general, a measure of independence only, since the measure always satisfies (i), but not necessarily (ii). In other words, while any distance guarantees that, at its lower bound, such a measure can capture independence in the variables, the type of the "highest" dependence, detected at the upper bound, depends crucially on the type of the distance function employed. Therefore, the choice of the distance function cannot be arbitrary, but is predetermined by the desired properties of the resulting measure of dependence.

Since mutual complete dependence is the opposite of stochastic independence, a measure of dependence should take this into account by satisfying (ii). This condition, which is also required by Lancaster (1982), is much stronger than both Renyi's original postulate formulated in Definition 3.14 (v) and its modification by Schweizer and Wolff in Definition 3.15 (v). Actually, no measure exists satisfying (ii), probably due to the disturbing fact that mutual complete dependence seems incompatible with the concept of convergence in distribution; see Section 3.1 and, in particular, Theorem 3.8. Instead Kimeldorf and Sampson (1978) suggested monotone dependence as the opposite of stochastic independence because it is preserved under convergence in law.

In fact, the quantities $\sigma(X,Y)$ and $\gamma(X,Y)$ given by (3.8) and (3.9), respectively, as well as any other normalized L^p-distance, $p \in [1,\infty)$, between $C_{X,Y}$ and the independence copula P are examples of measures of monotone dependence. By Theorem 3.16 they satisfy all conditions of Definition 3.15 and, thus, attain their maximum of 1 if and only if X and Y are monotone dependent. However, if X and Y are mutually completely dependent, the measures can attain any value in $(0,1]$. This follows from the fact that, by Corollary 3.5, the set of copulas linking mutually completely dependent random variables is dense in the set of all copulas $\mathfrak{C}$ with respect to any L^p-distance. Thus, none of the L^p-distances is capable of detecting mutual complete dependence, which substantiates our claim that the choice of the metrical distance function used in the construction of a measure of dependence is crucial for its resulting properties.

We argue that the inconsistency between mutual complete dependence and convergence in distribution neither weakens the concept of mutual complete dependence as the opposite of independence, nor does it imply that a measure of dependence should be restricted to monotone dependence. It rather suggests that convergence in law, or, alternatively, L^p-convergence of the corresponding copulas, is an inappropriate concept for the construction of measures of dependence.

Instead of the L^p-norm, we propose to measure the distance between two copulas by a modified Sobolev norm $\| \ \|$ given by

$$\|C\| = \Big(\int_{I^2} |\nabla C|^2 \, d\lambda \Big)^{1/2}. \tag{4.1}$$

This norm derives from a scalar product which, among other things, allows a straightforward representation via the $*$-product for copulas, introduced in Definition 2.21. Furthermore, this Sobolev norm turns out extremely advantageous since the degree of dependence between two continuous random variables X and Y, and, in particular, mutual complete dependence, can be determined by analytical properties of their copula. It follows that, in contrast to the L^p-distance, with respect to the Sobolev norm, mutual complete dependence cannot approximate any other kind of stochastic dependence.

Using this Sobolev norm we define a new nonparametric measure of dependence for two continuous random variables X and Y with copula C by

$$\omega(X,Y) = \left(3\|C\|^2 - 2\right)^{1/2} = \sqrt{3}\|C - P\| , \tag{4.2}$$

which represents the normalized Sobolev distance between C and the independence copula P. We show that $\omega(X,Y)$ has several appealing properties, e.g., its extremes are precisely at independence and mutual complete dependence.

4.2 The Sobolev scalar product for copulas

Denote by $\cdot$ the Euclidean scalar product, and by $|\ |$ the Euclidean norm on $\mathbb{R}^2$. As in the previous sections, λ denotes the two-dimensional Lebesgue measure, I^2 the closed unit square, and $\mathfrak{C}$ the set of copulas. Let $\operatorname{span}(\mathfrak{C})$ be the vector space generated by the linear span of $\mathfrak{C}$, i.e., the set of all finite real linear combinations of copulas.

Lemma 4.1 (Darsow and Olsen (1995))**.** *Any element $S \in span(\mathfrak{C})$ can be written in the form*

$$S = aA - bB$$

where a and b are nonnegative real numbers, and A and B are copulas.

Proof. Observe that if

$$S = \sum_{k=1}^{n} c_k C_k$$

where, for all k, $C_k \in \mathfrak{C}$ and $c_k \in \mathbb{R}$, then we can write

$$S = a \sum_{c_k \geq 0} \frac{c_k}{a} C_k - b \sum_{c_k \leq 0} \frac{-c_k}{b} C_k = aA - bB$$

where the last equality defines A and B and

$$a = \sum_{c_k \geq 0} c_k \quad \text{and} \quad b = -\sum_{c_k \leq 0} c_k \, .$$

(In both of the foregoing equations, the sum is taken to be zero if the set summed is empty.) Observe that A and B are copulas, since, by Theorem 2.10, $\mathfrak{C}$ is closed under convex combinations. □

Remark 4.2. It follows immediately from Lemma 4.1 and Theorem 2.15 that if $S \in \operatorname{span}(\mathfrak{C})$, then for any $v \in I$, $\partial_1 S(u, v)$ exists for almost all u, and for any $u \in I$, $\partial_2 S(u, v)$ exists for almost all v. For such u and v, where S is partially differentiable, we have

$$\nabla S = a \nabla A - b \nabla B, \tag{4.3}$$

where a, b, A and B are as in Lemma 4.1 and ∇A denotes the gradient of A.

For $S, T \in \operatorname{span}(\mathfrak{C})$, set

$$\langle S, T \rangle = \int_{I^2} \nabla S \cdot \nabla T \, d\lambda, \tag{4.4}$$

$$\|S\| = \Big(\int_{I^2} |\nabla S|^2 \, d\lambda \Big)^{1/2}, \tag{4.5}$$

$$d(S, T) = \Big(\int_{I^2} |\nabla S - \nabla T|^2 \, d\lambda \Big)^{1/2}. \tag{4.6}$$

Theorem 4.3. $\langle \ , \ \rangle$, $\| \ \|$ *and* d *define a scalar product, a norm and a metric on span*($\mathfrak{C}$)*, respectively.*

Proof. We need only prove the first statement. Clearly, by definition we have $\langle S, T \rangle = \langle T, S \rangle$.

To show that $\langle\ ,\ \rangle$ is bilinear, let $R, S, T \in \operatorname{span}(\mathfrak{C})$ and $r, s \in \mathbb{R}$. Then

$$\begin{aligned}\langle rR + sS, T\rangle &= \int_{I^2} \nabla(rR + sS) \cdot \nabla T \, d\lambda \\ &= \int_{I^2} (r\nabla R + s\nabla S) \cdot \nabla T \, d\lambda \\ &= \int_{I^2} r(\nabla R \cdot \nabla T) + s(\nabla S \cdot \nabla T) \, d\lambda \\ &= r\int_{I^2} \nabla R \cdot \nabla T + s\int_{I^2} \nabla S \cdot \nabla T \, d\lambda \\ &= r\langle R, T\rangle + s\langle S, T\rangle \,.\end{aligned}$$

Furthermore, $\langle S, S\rangle = \int_{I^2} |\nabla S|^2 \, d\lambda \geq 0$.

Lastly, if $S(u, v) = 0$ for all $(u, v) \in I^2$, then $\langle S, S\rangle = 0$. For the converse implication, observe that $\langle S, S\rangle = 0$ is equivalent to $\int_{I^2} |\nabla S|^2 \, d\lambda = 0$ which implies that for each $v \in I$, $\partial_1 S(u, v) = 0$ for almost all u. This in turn implies, by Theorem 2.16, that for each $v \in I$, $S(u, v)$ is constant for almost all u, but since S is continuous (S is a finite linear combination of copulas) this means for all u. Finally, since any element $S \in \operatorname{span}(\mathfrak{C})$ satisfies $S(0, 0) = 0$, the continuity of S implies that $S(u, v) = 0$ for all $(u, v) \in I^2$. This shows that $\langle\ ,\ \rangle$ is nondegenerate and completes the proof.

□

Thus, with a slight abuse of notation (because $\mathfrak{C}$ is not a vector space itself), we can make the following definition.

Definition 4.4. *The restrictions of $\langle\ ,\ \rangle$, $\|\ \|$ and d to $\mathfrak{C}$ are called the Sobolev scalar product, the Sobolev norm and the Sobolev distance function on $\mathfrak{C}$, respectively.*

Remark 4.5. The designation "Sobolev" derives from the fact that the set of copulas $\mathfrak{C} \subset W^{1,p}(I^2, \mathbb{R})$ for every $p \in [1, \infty]$ where $W^{1,p}(I^2, \mathbb{R})$ is the standard Sobolev space. This fact has also been noticed in Darsow and Olsen (1995). However, it has not been exploited in this context that

$W^{1,2}(I^2, \mathbb{R})$ is a Hilbert space with respect to the usual $W^{1,2}$-scalar product

$$\langle f, g\rangle_{W^{1,2}} = \int_{I^2} fg \, d\lambda + \int_{I^2} \nabla f \cdot \nabla g \, d\lambda$$

so that $\mathfrak{C}$ comes equipped with a scalar product structure. Thus, our definitions of the Sobolev scalar product, the Sobolev norm and the Sobolev distance in (4.4), (4.5) and (4.6), respectively, can be viewed as modifications deriving from the $W^{1,2}$-scalar product.

Remark 4.6. Darsow and Olsen (1995) show that $(\mathfrak{C}, d)$ is a complete metric space, and that the $*$-product is jointly continuous with respect to d.

We have seen that the Sobolev scalar product for copulas appears very naturally from an analytical point of view. However, it also allows a representation via the $*$-product given in Definition 2.21.

Theorem 4.7. *For all $A, B \in \mathfrak{C}$ we have the identity*

$$\begin{aligned}\langle A, B\rangle &= \int_0^1 (A^\top * B + A * B^\top)(t,t)\, dt \\ &= \int_0^1 (A^\top * B + B * A^\top)(t,t)\, dt.\end{aligned}$$

Proof. It follows from Definition 2.23 that

$$\begin{aligned}\partial_1 A^\top(u,v) &= \partial_2 A(v,u) \\ \partial_2 A^\top(u,v) &= \partial_1 A(v,u)\end{aligned} \tag{4.7}$$

Using Definition 2.23 and (4.7) we can write

$$\begin{aligned}\int_0^1\int_0^1 \partial_1 A(u,v)\,\partial_1 B(u,v)\, du\, dv &= \int_0^1 \Big(\int_0^1 \partial_2 A^\top(v,u)\,\partial_1 B(u,v)\, du\Big) dv \\ &= \int_0^1 (A^\top * B)(v,v)\, dv \\ \int_0^1\int_0^1 \partial_2 A(u,v)\,\partial_2 B(u,v)\, du\, dv &= \int_0^1 \Big(\int_0^1 \partial_2 A(u,v)\,\partial_1 B^\top(v,u)\, dv\Big) du \\ &= \int_0^1 (A * B^\top)(u,u)\, du.\end{aligned}$$

Adding up both terms we obtain the first identity.

The second equation in Theorem 4.7 is equivalent to

$$\int_0^1 (A * B^\top)(t,t)\,dt = \int_0^1 (B * A^\top)(t,t)\,dt$$

which follows from $(A * B^\top)(t,t) = (A * B^\top)^\top(t,t) = (B * A^\top)(t,t)$, where we have used (2.8). □

The representation in Theorem 4.7 becomes particularly simple for symmetric copulas.

Corollary 4.8. *If $A, B \in \mathfrak{C}$ are symmetric, then*

$$\langle A, B\rangle = 2\int_0^1 (A * B)(t,t)\,dt.$$

Theorem 4.7 yields upper and lower bounds for the scalar product of two copulas. More precisely, we have the following result:

Theorem 4.9. *Let $A, B \in \mathfrak{C}$. Then*

$$\frac{1}{2} \le \langle A, B\rangle \le 1,$$

where both bounds are sharp.

Proof. Theorem 4.7, in connection with Theorem 2.19, implies that

$$2\int_0^1 C^-(t,t)\,dt \le \langle A, B\rangle \le 2\int_0^1 C^+(t,t)\,dt\,.$$

Simple calculations yield $\int_0^1 C^-(t,t)\,dt = 1/4$ and $\int_0^1 C^+(t,t)\,dt = 1/2$.

Finally, one easily computes that

$$\begin{aligned} \langle C^-, C^-\rangle &= \langle C^+, C^+\rangle = 1 \\ \langle C^-, C^+\rangle &= \frac{1}{2}\,. \end{aligned} \tag{4.8}$$

This shows that the bounds in the statement are sharp, and the proof is complete. □

Remark 4.10. The diameter of $(\mathfrak{C}, d)$ is 1. To prove this, consider the identity

$$d(A,B)^2 = \|A - B\|^2 = \|A\|^2 + \|B\|^2 - 2\,\langle A, B\rangle. \tag{4.9}$$

It follows from Theorem 4.9 and (4.8) that $d(A,B) \leq 1 = d(C^-, C^+)$, which proves our claim.

Theorem 4.11. *For all $C \in \mathfrak{C}$, the following hold:*

$$\langle C, P\rangle = 2/3, \tag{4.10}$$

$$\|C - P\|^2 = \|C\|^2 - \frac{2}{3}. \tag{4.11}$$

Proof. For (4.10), we remark that $P = P^\top$, so Theorem 4.7 and (2.7) imply

$$\begin{aligned}\langle P, C\rangle &= \int_0^1 (P * C + C * P)(t,t)\,dt \\ &= 2\int_0^1 P(t,t)\,dt \\ &= \frac{2}{3}.\end{aligned}$$

To show (4.11), observe that by (4.9)

$$\begin{aligned}\|C - P\|^2 &= \|C\|^2 + \|P\|^2 - 2\langle C, P\rangle \\ &= \|C\|^2 - \frac{2}{3}\end{aligned}$$

where we have used (4.10). □

4.3 Statistical interpretation of the Sobolev norm for copulas

We now turn to the probabilistic interpretation of the Sobolev norm for copulas.

Lemma 4.12. *Let X and Y be random variables on the same probability space with continuous distribution functions and (unique) copula C. The following statements are equivalent:*

(i) Y is completely dependent on X.

(ii) C is left invertible.

(iii) $\partial_1 C \in \{0,1\}$ almost everywhere.

Consequently, the following are also equivalent:

(i) X and Y are mutually completely dependent.

(ii) C is invertible.

(iii) $\partial_1 C, \partial_2 C \in \{0,1\}$ almost everywhere.

Proof. Darsow, Nguyen, and Olsen (1992) prove in Theorem 11.1 that Y is completely dependent on X if and only if C is left invertible. Moreover, Theorem 2.26 states that C has a left inverse if and only if for each $v \in I$, $\partial_1 C(u,v) \in \{0,1\}$ for almost all $u \in I$. Actually, as noted in Remark 2.27 the latter is tantamount to assuming that $\partial_1 C(u,v) \in \{0,1\}$ almost everywhere. This proves the first part.

Analogous statements hold for right invertible copulas, from which the second part of the lemma follows. □

The next theorem describes one of the main results of this dissertation.

Theorem 4.13. *For any copula C, the Sobolev norm satisfies*

$$\frac{2}{3} \leq \|C\|^2 \leq 1.$$

Moreover, if X and Y are random variables on the same probability space with continuous marginal distribution functions and (unique) copula C, the following statements hold:

(i) $\|C\|^2 = 2/3$ if and only if X and Y are independent.

(ii) $\|C\|^2 \in [5/6, 1]$ if Y is completely dependent on X (or vice versa).

(iii) $\|C\|^2 = 1$ *if and only if* X *and* Y *are mutually completely dependent.*

In terms of the algebraic properties of C*, these statements read as follows:*

(i) $\|C\|^2 = 2/3$ *if and only if* $C = P$.

(ii) $\|C\|^2 \in [5/6, 1]$ *if* C *is left (or right) invertible.*

(iii) $\|C\|^2 = 1$ *if and only if* C *is invertible.*

Proof. The foremost statement follows from (4.11) and Theorem 4.9.

The assertion that $\|C\|^2 = 2/3$ if and only if $C = P$ is an immediate consequence of (4.11). Furthermore, by Theorem 2.37, $C = P$ is equivalent to the independence of X and Y.

As for statement (ii), it follows from the Definition of the Sobolev norm in (4.5) that

$$\|C\|^2 = \int_0^1 \int_0^1 (\partial_1 C(u,v))^2 \, du \, dv + \int_0^1 \int_0^1 (\partial_2 C(u,v))^2 \, du \, dv. \qquad (4.12)$$

If Y is completely dependent on X we know from Lemma 4.12 that $\partial_1 C \in \{0, 1\}$ almost everywhere in I^2, which implies that $(\partial_1 C)^2 = \partial_1 C$ almost everywhere in I^2, so the first summand in (4.12) is equal to

$$\int_0^1 \int_0^1 \partial_1 C(u,v) \, du \, dv = \int_0^1 v \, dv = \frac{1}{2}.$$

To estimate the second term in (4.12), consider the inequality

$$\begin{aligned} 0 &\leq \int_0^1 \int_0^1 (\partial_2 C(u,v) - u)^2 \, du \, dv \\ &= \int_0^1 \int_0^1 (\partial_2 C(u,v))^2 \, du \, dv - 2 \int_0^1 u \int_0^1 \partial_2 C(u,v) \, dv \, du + \int_0^1 \int_0^1 u^2 \, du \, dv \\ &= \int_0^1 \int_0^1 (\partial_2 C(u,v))^2 \, du \, dv - \frac{1}{3}. \end{aligned}$$

Hence, the second term in (4.12) is at least $1/3$, which proves $\|C\|^2 \geq 5/6$. Equality holds if and only if $\partial_2 C(u,v) = u$ almost everywhere in I^2, which, by Theorem 2.16, is equivalent to $C = P$. But this contradicts the assumption that Y is completely dependent on X since, by Lemma 4.12, the latter is equivalent to C being left invertible. It follows that $\|C\|^2 > 5/6$. Analogous arguments hold for right invertible copulas. This, together with Lemma 4.12, proves the second pair of statements.

Finally, in view of Theorem 2.15, we have $(\partial_i C)^2 \leq \partial_i C$, for $i = 1, 2$, with equality if and only if $\partial_i C \in \{0, 1\}$. Consequently, (4.12) implies that

$$\|C\|^2 \leq \int_0^1 \int_0^1 \partial_1 C(u, v)\, du\, dv + \int_0^1 \int_0^1 \partial_2 C(u, v)\, du\, dv = \frac{1}{2} + \frac{1}{2} = 1$$

with equality if and only if $\partial_1 C, \partial_2 C \in \{0, 1\}$ almost everywhere in I^2. By Lemma 4.12, the latter is equivalent to X and Y being mutually completely dependent, which is also equivalent to C being invertible. This proves the third pair of assertions and completes the proof. □

Corollary 4.14. *Let X and Y be continuous random variables on the same probability space with copula C. The following are equivalent:*

(i) *X and Y are mutually completely dependent.*

(ii) $\|C\| = 1$.

(iii) $\partial_1 C, \partial_2 C \in \{0, 1\}$ *almost everywhere in* I^2.

(iv) *C is invertible, i.e.,* $C * C^\top = C^\top * C = C^+$.

(v) $\int_0^1 (C * C^\top + C^\top * C)(t, t)\, dt = 1$.

Proof. This follows immediately from Lemma 4.12, Theorem 4.13 and Theorem 4.7. □

Theorem 4.13, together with the identity

$$d(C, P)^2 = \|C - P\|^2 = \|C\|^2 - 2/3$$

expresses the astonishing fact that the Sobolev norm itself measures stochastic dependence, with extremes exactly at independence and mutual complete dependence. In addition, the Sobolev norm is able to detect that two random variables are not completely dependent.

Lemma 4.15. *If $(C_k)_{k\in\mathbb{N}}$ is a sequence of left invertible copulas in $\mathfrak{C}$ with*

$$\lim_{k\to\infty} \|C_k - C\| = 0$$

for some $C \in span(\mathfrak{C})$, then C is in $\mathfrak{C}$ and is left invertible. Analogous statements hold for right invertible and invertible copulas.

Proof. By Remark 4.6, the Sobolev limit of a sequence of copulas is a copula and the $*$-product on $\mathfrak{C}$ is (jointly) continuous with respect to d. Moreover, $\lim_{k\to\infty} \|C_k - C\| = 0$ implies $\lim_{k\to\infty} \|C_k^\top - C^\top\| = 0$. Thus, if each C_k is left invertible, then

$$C^\top * C = \lim_{k\to\infty} C_k^\top * \lim_{k\to\infty} C_k = \lim_{k\to\infty} (C_k^\top * C_k) = C^+,$$

which proves that C is left invertible. The case where each C_k is right invertible is shown analogously. Thus, it follows that the Sobolev limit of a sequence of invertible copulas is again an invertible copula. □

As an immediate consequence of Lemma 4.15 we have the following result.

Theorem 4.16. *Let $(X_n, Y_n)_{n\in\mathbb{N}}$ and (X, Y) be, respectively, a sequence of pairs and a pair of continuous random variables on a common probability space with copulas $(C_n)_{n\in\mathbb{N}}$ and C. Then the following assertions hold:*

(i) If, for all n, Y_n is completely dependent on X_n and

$$\lim_{n\to\infty} \|C_n - C\| = 0,$$

then Y is completely dependent on X.

(ii) If, for all n, X_n and Y_n are mutually completely dependent and

$$\lim_{n\to\infty} \|C_n - C\| = 0,$$

then X and Y are mutually completely dependent.

Theorem 4.16 emphasizes the advantage of the Sobolev distance over any L^p-distance with $p \in [1,\infty]$, as mentioned in the Introduction. While, with respect to any L^p-distance, any copula, in particular, the independence copula P, can be approximated by copulas of mutually completely dependent random variables, the Sobolev convergence preserves the property of mutual complete dependence. Hence, with respect to the Sobolev distance, mutual complete dependence cannot approximate any other kind of stochastic dependence. In fact, the Sobolev convergence preserves even the property of complete dependence.

In summary, measuring the distance between copulas with the Sobolev norm resolves the disturbing phenomenon observed in Kimeldorf and Sampson (1978) and Mikusiński, Sherwood, and Taylor (1992).

4.4 The measure ω and its properties

The remarkable statistical properties of the Sobolev norm lead immediately to the following definition:

Definition 4.17. *Given two continuous random variables X, Y with copula C, we define*

$$\omega(X,Y) = \big(3\|C\|^2 - 2\big)^{1/2}.$$

In view of Theorem 4.11, the quantity $\omega(X,Y)$ represents a normalized Sobolev distance of C from the independence copula P:

$$\omega(X,Y) = \sqrt{3}\,\|C - P\| = \frac{\|C-P\|}{\|\widehat{C} - P\|}, \tag{4.13}$$

where $\widehat{C}$ is any copula of mutually completely dependent random variables. The normalization guarantees that $\omega(X,Y) \in [0,1]$. Definition 4.17, however, makes clear that the Sobolev norm of C itself serves as a measure of dependence.

For symmetric C we may use Corollary 4.8 to write

$$\omega(X,Y) = \Big(6 \int_0^1 (C * C)(t,t)\,dt - 2\Big)^{1/2}. \tag{4.14}$$

Theorem 4.18. *Let X and Y be continuous random variables on the same probability space with copula C. The quantity $\omega(X,Y)$ has the following properties:*

(i) $\omega(X,Y)$ *is defined for any* X *and* Y.

(ii) $\omega(X,Y) = \omega(Y,X)$.

(iii) $0 \le \omega(X,Y) \le 1$.

(iv) $\omega(X,Y) = 0$ *if and only if* X *and* Y *are independent.*

(v) $\omega(X,Y) = 1$ *if and only if* X *and* Y *are mutually completely dependent.*

(vi) $\omega(X,Y) \in [\sqrt{1/2}, 1]$ *if* Y *is completely dependent on* X *(or vice versa).*

(vii) If f *and* g *are strictly monotone functions on Range*(X) *and Range*(Y)*, respectively, then* $\omega(f(X), g(Y)) = \omega(X,Y)$.

(viii) If $(X_n, Y_n)_{n\in\mathbb{N}}$ *is a sequence of pairs of continuous random variables with copulas* C_n*, and if* $\lim_{n\to\infty} \|C_n - C\| = 0$*, then* $\lim_{n\to\infty} \omega(X_n, Y_n) = \omega(X,Y)$.

Proof. (i) is obvious since, by Sklar's theorem 2.31, the copula C exists and is unique.

(ii) is clear by Definition 4.17 since the copula of Y and X is given by the transposed copula $C^\top$ (see Definition 2.23) and thus by (4.7) and (4.12) we have $\|C\| = \|C^\top\|$.

(iii) follows from Definition 4.17 and Theorem 4.11 (iii).

(iv) follows from Definition 4.17 and Theorem 4.13 (i).

(v) follows from Definition 4.17 and Theorem 4.13 (iii).

(vi) follows from Definition 4.17 and Theorem 4.13 (ii).

With respect to (vii) we distinguish four different cases. For the sake of clarity, let $C_{X,Y}$ denote the copula of X and Y. If both f and g are increasing it follows from Theorem 2.41 (i) that $C_{f(X),g(Y)} = C_{X,Y}$ which implies

$$\omega(f(X), g(Y)) = \sqrt{3}\,\|C_{f(X),g(Y)} - P\| = \sqrt{3}\,\|C_{X,Y} - P\| = \omega(X, Y).$$

If f is increasing and g is decreasing then, by Theorem 2.41 (ii),

$$C_{f(X),g(Y)}(u, v) = u - C_{X,Y}(u, 1 - v).$$

Therefore,

$$(C_{f(X),g(Y)} - P)(u, v) = (P - C_{X,Y})(u, 1 - v)$$

which, by the transformation formula for the Lebesgue measure, again implies $\omega(f(X), g(Y)) = \omega(X, Y)$. If f is decreasing and g is increasing, the result follows from interchanging f and g in the previous case. The case when f and g are both decreasing can be shown similarly.

Finally, (viii) follows immediately from Definition 4.17. □

Remark 4.19. If X and Y are jointly normal with correlation coefficient ρ, then $\omega(X, Y)$ is a strictly increasing function of $|\rho|$ whose graph is shown in Figure 4.1.

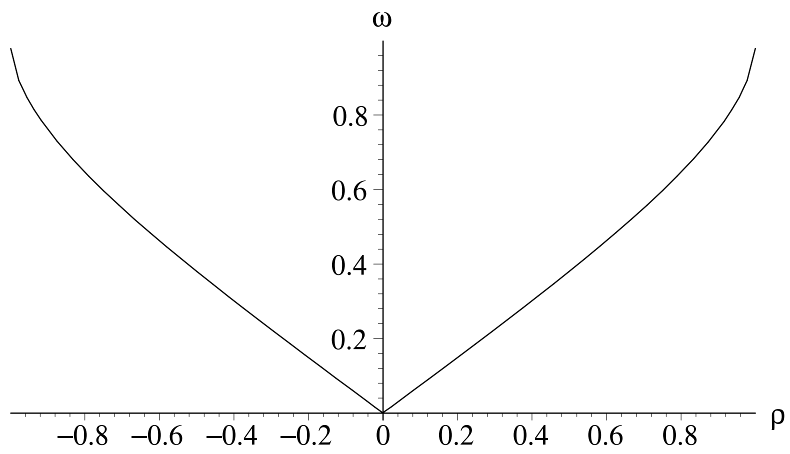

Figure 4.1: $\omega(X, Y)$ as a function of ρ for jointly normal X, Y

4.5 Some examples

We conclude the dissertation with some examples clarifying the relationship between the measure of dependence $\omega(X, Y)$ and the quantity $\sigma(X, Y)$, as defined in (3.8).

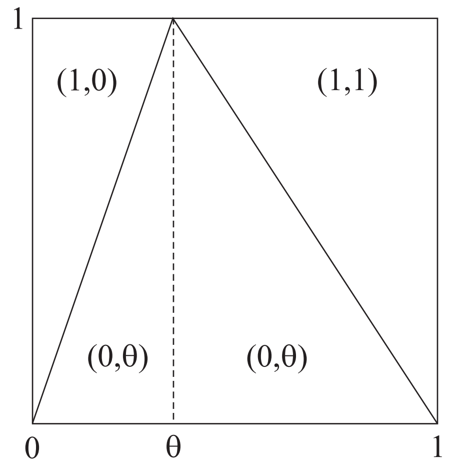

Figure 4.2: The gradient ∇C of the copula C in Example 4.20

Example 4.20. Let $\theta \in [0,1]$, and consider the singular copula C whose support consists of two line segments in I^2, one joining $(0,0)$ and $(\theta,1)$, and the other joining $(\theta,1)$ and $(1,0)$ (Nelsen, 2006, Example 3.3). It follows that

$$C(u,v) = \begin{cases} u & \text{if } u \leq \theta v \\ \theta v & \text{if } \theta v < u < 1-(1-\theta)v \\ u+v-1 & \text{if } 1-(1-\theta)v \leq u. \end{cases}$$

Clearly, Y is completely dependent on X, but not vice versa. Since probability mass θ and $1-\theta$ is uniformly distributed on the first and second line segments, respectively, it is heuristically clear that the value $\theta = 1/2$ describes the least dependent situation, whereas the limiting cases $\theta = 0$ and $\theta = 1$, when $C = C^-$ and $C = C^+$, respectively, correspond to mutual complete dependence.

This is perfectly reflected in the behavior of $\omega(X,Y)$. Indeed, a straightforward calculation (compare Fig. 4.2) shows that

$$\|C\|^2 = \frac{1}{2}\Big(\theta - \frac{1}{2}\Big)^2 + \frac{7}{8} \in \Big[\frac{7}{8}, 1\Big]$$

with the lowest and highest values attained precisely for $\theta = 1/2$ and $\theta \in \{0,1\}$, respectively. Consequently, $\omega(X,Y)$ takes on its smallest value $\sqrt{10}/4 \approx .79$ for $\theta = 1/2$.

The quantity $\sigma(X,Y)$ shows the same qualitative behavior, however, its minimal value is .5.

Example 4.21. Let $\theta \in [0,1]$, and consider the singular copula C whose support consists of the two segments $\{(u, 1-u) \mid u \in [0,\theta] \cup [1-\theta, 1]\}$ and the segment $\{(u,u) \mid u \in [\theta, 1-\theta]\}$ (Nelsen, 2006, Exercise 3.15). It follows that

$$C(u,v) = \begin{cases} C^+(u,v) - \theta & \text{if } (u,v) \in [\theta, 1-\theta]^2 \\ C^-(u,v) & \text{otherwise}. \end{cases}$$

Now X and Y are mutually completely dependent, so $\omega(X,Y) = 1$, regardless of the value of θ.

In contrast, $\sigma(X, Y)$ varies between 1 (for $\theta \in \{0, 1\}$) and values around .46 (for $\theta \approx .12$), indicating a definite degree of independence when, actually, there is none. Note that the copula from Example 4.20 with $\theta = 1/2$ yields almost the same value for σ.

Bibliography

CHERUBINI, U., E. LUCIANO, AND W. VECCHIATO (2004): *Copula Methods in Finance*. John Wiley & Sons Ltd., Chichester.

DARSOW, W., B. NGUYEN, AND E. OLSEN (1992): "Copulas and Markov processes," *Illinois J. Math.*, 36(4), 600–642.

DARSOW, W., AND E. OLSEN (1995): "Norms for copulas," *Int. J. Math. and Math. Sci.*, 18(3), 417–436.

FRÉCHET, M. (1951): "Sur les tableaux de corrélation dont les marges sont données," *Ann. Univ. Lyon. Sect. A. (3)*, 14, 53–77.

GEBELEIN, H. (1941): "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung," *Z. Angew. Math. Mech.*, 21, 364–379.

HALL, W. (1970): "On characterizing dependence in joint distributions," in *Essays in probability and statistics*, ed. by R. Bose, I. Chakravarti, P. Mahalanobis, C. Rao, and K. Smith, pp. 339–376. The University of North Carolina Press, Chapel Hill, N.C.

HOEFFDING, W. (1994): *The collected works of Wassily Hoeffding*. Springer-Verlag, New York, Edited and with a preface by N. I. Fisher and P. K. Sen.

JOE, H. (1989): "Relative entropy measures of multivariate dependence," *J. Amer. Stat. Assoc.*, 84(405), 157–164.

——— (1997): *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, London.

Kimeldorf, G., and A. Sampson (1978): "Monotone dependence," *Ann. Stat.*, 6(4), 895–903.

Lancaster, H. (1963): "Correlation and complete dependence of random variables," *Ann. Math. Stat.*, 34, 1315–1321.

Lancaster, H. O. (1982): "Measures and indices of dependence," in *Encyclopedia of Statistical Sciences*, ed. by M. Kotz, and N. L. Johnson, vol. 2, pp. 334–339. Wiley, New York.

McNeil, A., R. Frey, and P. Embrechts (2005): *Quantitative risk management*. Princeton University Press, Princeton, NJ, Concepts, techniques and tools.

Micheas, A., and K. Zografos (2006): "Measuring stochastic dependence using ϕ-divergence," *J. Multivar. Analysis*, 97(3), 765–784.

Mikusiński, P., H. Sherwood, and M. Taylor (1992): "Shuffles of min," *Stochastica*, 13(1), 61–74.

Mikusiński, P., H. Sherwood, and M. D. Taylor (1991/92): "The Fréchet bounds revisited," *Real Anal. Exchange*, 17(2), 759–764.

Nelsen, R. (1991): "Copulas and association," in *Advances in probability distributions with given marginals*, ed. by G. Dall'Aglio, S. Kotz, and G. Salinetti, pp. 51–74. Kluwer Academic Publishers, Dordrecht.

——— (2002): "Concordance and copulas: a survey," in *Distributions with given marginals and statistical modelling*, ed. by C. Cuadras, J. Fortiana, and J. Rodriguez-Lallena, pp. 169–177. Kluwer Academic Publishers, Dordrecht.

——— (2006): *An Introduction to Copulas.* Springer, New York, 2nd edn.

Olsen, E., W. Darsow, and B. Nguyen (1996): "Copulas and Markov operators," in *Distributions with fixed marginals and related topics*, ed. by L. Rüschendorf, B. Schweizer, and M. Taylor, pp. 244–259. Institute of Mathematical Statistics, Hayward, CA.

Rényi, A. (1959): "On measures of dependence," *Acta Math. Acad. Sci. Hungar.*, 10, 441–451.

Scarsini, M. (1984): "On measures of concordance," *Stochastica*, 8(3), 201–218.

Schweizer, B. (1991): "Thirty years of copulas," in *Advances in probability distributions with given marginals*, ed. by G. Dall'Aglio, S. Kotz, and G. Salinetti, pp. 13–50. Kluwer Academic Publishers, Dordrecht.

Schweizer, B., and A. Sklar (1974): "Operations on distribution functions not derivable from operations on random variables," *Studia. Math.*, 52, 43–52.

——— (2005): *Probabilistic metric spaces.* Dover Publications, New York.

Schweizer, B., and E. Wolff (1981): "On nonparametric measures of dependence for random variables," *Ann. Stat.*, 9(4), 879–885.

Sklar, M. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Vitale, R. (1990): "On stochastic dependence and a class of degenerate distributions," in *Topics in statistical dependence*, ed. by H. Block, A. Sampson, and T. Savits, pp. 459–469. Institute of Mathematical Statistics, Hayward, CA.

——— (1991): "Approximation by mutually completely dependent processes.," *J. Approx. Theory*, 66(2), 225–228.

Zografos, K. (1998): "On a measure of dependence based on Fisher's information matrix," *Comm. Stat. Theory and Methods*, 27(7), 1715–1728.

——— (2000): "Measures of multivariate dependence based on a distance between Fisher information matrices," *J. Stat. Planning Inference*, 89(1–2), 91–107.