Chapter 1

Preliminaries

In this chapter, we recall several definitions and properties that will be used in the sequel. We begin with some notations on sets and functions (section 1.1) and, in particular, we present the construction of the pseudo–inverse of a monotone function. Section 1.2 is devoted to the presentation of the main concepts and results about the majorization ordering. Binary operations and, in particular, *triangular norms* are presented in sections 1.3 and 1.4.

After recalling some facts about distribution functions (section 1.5), we present the concept of *copula* and its applications to dependence concepts (sections 1.6–1.9). Two generalizations of the copula function are presented in the sections 1.10 and 1.11.

1.1 Sets and functions

We denote by $\mathbb{R}$ the ordinary set of real numbers $(-\infty, +\infty)$ and by $\overline{\mathbb{R}}$ its extension $[-\infty, +\infty]$. For every positive integer $n \geq 2$, $\mathbb{R}^n$ and $\overline{\mathbb{R}}^n$ denote, respectively, the cartesian product of n copies of $\mathbb{R}$ and $\overline{\mathbb{R}}$. We use vector notations for the points in $\overline{\mathbb{R}}^n$, e.g. $\mathbf{x} = (x_1, \dots, x_n)$, and we write $\mathbf{x} \leq \mathbf{y}$ when $x_i \leq y_i$ for all $i \in \{1, 2, \dots, n\}$.

A *n–box* B is a subset of $\overline{\mathbb{R}}^n$ given by the cartesian product of n closed intervals, $B = [x_1, y_1] \times \cdots \times [x_n, y_n]$, and we write it also in the form $[\mathbf{x}, \mathbf{y}]$, where we suppose $x_i < y_i$ for at least an index $i \in \{1, 2, \dots, n\}$. In particular, $[0, 1]^n$ indicates the cartesian product of n copies of the unit interval, i.e. the unit n–cube. The *vertices* of the n–box $B = [\mathbf{x}, \mathbf{y}]$ are the points $\mathbf{c} = (c_1, \dots, c_n) \in B$ such that $c_i \in \{x_i, y_i\}$ for all $i \in \{1, 2, \dots, n\}$. In every vertex $\mathbf{c}$, we can define the following function

$$sgn(\mathbf{c}) := \begin{cases} 1, & \text{if } card\{i \in \{1, 2, \dots, n\} \mid c_i = x_i\} \text{ is an even number;} \\ -1, & \text{if } card\{i \in \{1, 2, \dots, n\} \mid c_i = x_i\} \text{ is an odd number.} \end{cases}$$

An *n–place real function* H is a function whose domain, $DomH$, is a subset of $\overline{\mathbb{R}}^n$ and whose range, $RanH$, is a subset of $\overline{\mathbb{R}}$. As a convention, a 1–place real function is called simply *real function*. The *partial derivative* of H with respect to the i-th variable x_i is denoted by $\partial_{x_i} H$ or $\partial_i H$. If S is a subset of $\overline{\mathbb{R}}^n$, 1_S denote the *indicator function* of S defined by

$$1_S(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in S; \\ 0, & \text{if } \mathbf{x} \notin S. \end{cases}$$

A statement about the points of a set $S \subseteq \mathbb{R}^n$ is said to hold *almost everywhere* (briefly, a.e.) if the set of points of S where the statement fails to hold has Lebesgue measure zero.

Given a real function f and an accumulation point x_0 of $Dom f$, we denote the *left–hand limit* of f at x_0 (if it exists) by $f(x_0^-)$, and the *right–hand limit* of f at x_0 (if it exists) by $f(x_0^+)$. Analogously, $f'(x_0^-)$ and $f'(x_0^+)$ denote, resp., the *left–hand derivative* and *right–hand derivative* of f at x_0. Moreover, if $S \subseteq \mathbb{R}$, we will denote by id_S the *identity function* of S, i.e. $\mathrm{id}_S(x) := x$ for every $x \in S$.

A real function f is *increasing* (resp., *strictly increasing*) if, for every $x < y$, $f(x) \leq f(y)$ (resp., $f(x) < f(y)$). Similarly, f is *decreasing* (resp., *strictly decreasing*) if, for every $x < y$, $f(x) \geq f(y)$ (resp., $f(x) > f(y)$). A function f is (*strictly*) *monotone* if f is either (strictly) increasing or (strictly) decreasing.

Let $f : I \to \mathbb{R}$ be a real function whose domain I is an interval of $\mathbb{R}$. The function f is said to be *convex* on I if, for every $x, y \in I$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y).$$

The function f is called *Jensen–convex* on I (or *mid–convex*) if, for every $x, y \in I$,

$$f\left(\frac{x + y}{2}\right) \leq \frac{f(x) + f(y)}{2}.$$

A function f is said to be (*Jensen–*)*concave* on I if the function $-f$ is (Jensen–)convex.

Proposition 1.1.1 ([69])**.** *Let f be a continuous real function defined on an interval I of $\mathbb{R}$. Then f is convex if, and only if, f is Jensen–convex.*

In the same manner, we could define the convexity for an n–place real function whose domain is a convex subset of $\mathbb{R}^n$.

Some notations from lattice theory will be also necessary (see [25]). Let $(X, \leq)$ be a partially ordered set, $X \neq \emptyset$. For all $x, y \in X$, let $U(x, y) := \{z \in X : x \leq z, y \leq z\}$. If $U(x, y)$ has a unique smallest element $\widetilde{z}$ such that $\widetilde{z} \leq z$ for all $z \in U(x, y)$, then $\widetilde{z}$ is called the *supremum* of x and y, denoted by $x \vee y$ or $\sup\{x, y\}$. Similarly, if there is a unique greatest element z' smaller than x and y, then this is called the *infimum*, denoted by $x \wedge y$ or $\inf\{x, y\}$. If, for all $x, y \in X$, $x \wedge y$ and $x \vee y$ exist in X, then $(X, \leq)$ is called *lattice*. Moreover, for every $S \subseteq X$, we denote by $\bigvee S$ the supremum of the elements of S and by $\bigwedge S$ the infimum of the elements of S. If, for every $S \subseteq X$, $\bigvee S$ and $\bigwedge S$ exist in X, then $(X, \leq)$ is called *complete lattice*.

1.1.1 The pseudo–inverse of a real function

Definition 1.1.1. Let $[a, b]$ and $[c, d]$ be intervals of $\overline{\mathbb{R}}$ and let $f : [a, b] \to [c, d]$ be a monotone function. The *pseudo–inverse* of f is the function $f^{[-1]} : [c, d] \to [a, b]$ defined by

$$f^{[-1]}(y) = \begin{cases} \sup\{x \in [a, b] \mid f(x) < y\}, & \text{if } f(a) < f(b); \\ \sup\{x \in [a, b] \mid f(x) > y\}, & \text{if } f(a) > f(b); \\ a, & \text{if } f(a) = f(b). \end{cases}$$

Notice that, if f is a bijection, then the pseudo–inverse coincides with the inverse.

The graph of the pseudo–inverse of a non–constant monotone function f can be easily constructed by the following procedure:

(i) draw the graph of f and complete it, if it is necessary, by adding vertical line segments connecting the points $(x_0, f(x_0^-))$ and $(x_0, f(x_0^+))$ at each discontinuity point x_0 of f;

(ii) reflect the graph so obtained with respect to the graph of $\mathrm{id}_{\mathbb{R}}$, namely with respect to the bisector of the first quadrant;

(iii) remove all but the smallest point from any vertical line contained in the reflected graph.

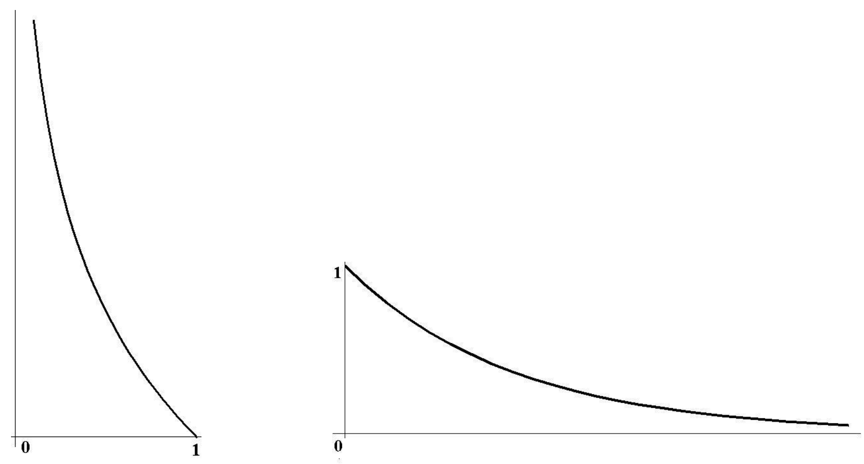

Figure 1.1: A function f and its inverse f^{-1}

Now, we consider a pseudo–inverse construction in two special cases.

Example 1.1.1. Let us consider a function $\varphi : [0,1] \to [0,+\infty]$ that is continuous and strictly decreasing with $\varphi(1) = 0$. The pseudo–inverse of φ is given by

$$\varphi^{[-1]}(t) := \begin{cases} \varphi^{-1}(t), & \text{if } t \in [0, \varphi(0)]\,; \\ 0, & \text{if } t \in [\varphi(0), +\infty]\,. \end{cases}$$

Note that $\varphi^{[-1]}$ is continuous and decreasing on $[0,+\infty]$ and strictly decreasing on $[0,\varphi(0)]$. Furthermore, for all $t \in [0,1]$,

$$\varphi^{[-1]}\left(\varphi(t)\right) = t \tag{1.1}$$

and, for all $t \in [0,+\infty]$,

$$\varphi\left(\varphi^{[-1]}(t)\right) = \min\{t, \varphi(0)\}. \tag{1.2}$$

Example 1.1.2. Given a function $h : [0,1] \to [0,1]$ that is continuous and strictly increasing with $h(1) = 1$, its pseudo–inverse $h^{[-1]} : [0,1] \to [0,1]$ is defined by

$$h^{[-1]}(t) := \begin{cases} h^{-1}(t), & \text{if } t \in [h(0), 1]; \\ 0, & \text{if } t \in [0, h(0)]. \end{cases}$$

Notice that $h^{[-1]}$ is continuous and increasing on $[0,1]$ and strictly increasing on $[h(0),1]$ and, for all $t \in [0,1]$

$$h^{[-1]}\left(h(t)\right) = t \qquad \text{and} \qquad h\left(h^{[-1]}(t)\right) = \max\{t, h(0)\}.$$

1.2 Majorization ordering

In this section we recall the concepts of majorization ordering on $\mathbb{R}^n$ and Schur–convexity, which can be found in the book by A.W. Marshall and I. Olkin (see [103]).

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be two points in $\mathbb{R}^n$ and denote by

$$x_{[1]}, x_{[2]}, \ldots, x_{[n]} \qquad \text{and} \qquad y_{[1]}, y_{[2]}, \ldots, y_{[n]}$$

the components of $\mathbf{x}$ and $\mathbf{y}$ rearranged in decreasing order, and by

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)} \qquad \text{and} \qquad y_{(1)}, y_{(2)} \cdots, y_{(n)}$$

the components of $\mathbf{x}$ and $\mathbf{y}$ rearranged in increasing order.

Definition 1.2.1. The point $\mathbf{x}$ is *majorized* by $\mathbf{y}$ (and we write $\mathbf{x} \prec \mathbf{y}$) if

(i) $\displaystyle\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}$ for every $k \in \{1, 2, \ldots, n-1\}$;

(ii) $\displaystyle\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$.

Definition 1.2.2. The point $\mathbf{x}$ is said to be *weakly submajorized* by $\mathbf{y}$ (and we write $\mathbf{x} \prec_w \mathbf{y}$) if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \quad \text{for every } k \in \{1, 2, \ldots, n\}.$$

The point $\mathbf{x}$ is said to be *weakly supermajorized* by $\mathbf{y}$ (and we write $\mathbf{x} \prec^w \mathbf{y}$) if

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)} \quad \text{for every } k \in \{1, 2, \ldots, n\}.$$

In the case $n = 2$, the previous definitions take these forms.

$$(x_1, x_2) \prec (y_1, y_2) \iff \begin{cases} \max\{x_1, x_2\} \leq \max\{y_1, y_2\} \\ x_1 + x_2 = y_1 + y_2 \end{cases}$$

$$(x_1, x_2) \prec_w (y_1, y_2) \iff \begin{cases} \max\{x_1, x_2\} \leq \max\{y_1, y_2\} \\ x_1 + x_2 \leq y_1 + y_2 \end{cases}$$

$$(x_1, x_2) \prec^w (y_1, y_2) \iff \begin{cases} \min\{x_1, x_2\} \geq \min\{y_1, y_2\} \\ x_1 + x_2 \geq y_1 + y_2. \end{cases}$$

The following theorems characterize the majorization orderings ([68, 69, 103]).

Theorem 1.2.1 (Hardy, Littlewood and Pólya). *Given two points* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$*, the following statements are equivalent:*

(i) $\mathbf{x} \prec \mathbf{y}$*;*

(ii) *a doubly stochastic matrix* P *exists such that* $\mathbf{x} = P\mathbf{y}$*.*

Corollary 1.2.1. *Given two points* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^2$*, the following statements are equivalent:*

(i) $\mathbf{x} \prec \mathbf{y}$*;*

(ii) *there exists* $\alpha \in [0, 1]$ *such that*

$$x_1 = \alpha y_1 + (1 - \alpha) y_2 \quad \textit{and} \quad x_2 = (1 - \alpha) y_1 + \alpha y_2.$$

Theorem 1.2.2 (Hardy, Littlewood and Pólya). *Given two points* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$*, the following statements are equivalent:*

(a) $\mathbf{x} \prec \mathbf{y}$*;*

(b) *for every continuous convex function* $g : \mathbb{R} \to \mathbb{R}$

$$\sum_{i=1}^{n} g(x_i) \leq \sum_{i=1}^{n} g(y_i).$$

The following result, which extends Theorem 1.2.2 to the weak majorization ordering and which will be necessary in the sequel, can be found in [103] since it was published in a journal of difficult access ([155]).

Theorem 1.2.3 (**Tomić**). *Given two points* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$*, the following statements are equivalent:*

(a) $\mathbf{x} \prec_w \mathbf{y}$*;*

(b) *for every continuous, increasing and convex function* $g : \mathbb{R} \to \mathbb{R}$

$$\sum_{i=1}^{n} g(x_i) \leq \sum_{i=1}^{n} g(y_i).$$

Similarly, the following statements are equivalent

(a) $\mathbf{x} \prec^w \mathbf{y}$*;*

(b) *for every continuous, decreasing and convex function* $g : \mathbb{R} \to \mathbb{R}$

$$\sum_{i=1}^{n} g(x_i) \leq \sum_{i=1}^{n} g(y_i).$$

Definition 1.2.3. A function $\varphi : A \subset \mathbb{R}^n \to \mathbb{R}$ is said to be *Schur–convex* on A if it is increasing with respect to the majorization order $\prec$, namely if, for all $\mathbf{x}, \mathbf{y} \in A$, $\mathbf{x} \prec \mathbf{y}$ implies $\varphi(\mathbf{x}) \leq \varphi(\mathbf{y})$. If, in addition, $\varphi(\mathbf{x}) < \varphi(\mathbf{y})$ whenever $\mathbf{x} \prec \mathbf{y}$ but $\mathbf{x}$ is not a permutation of $\mathbf{y}$, then φ is said to be *strictly Schur–convex* on A.
Similarly, φ is said to be *Schur–concave* on A if, for all $\mathbf{x}, \mathbf{y} \in A$, $\mathbf{x} \prec \mathbf{y}$ implies $\varphi(\mathbf{x}) \geq \varphi(\mathbf{y})$. Moreover, φ is said to be *Schur–constant* if it is, at same time, Schur–convex and Schur–concave.

The next result characterizes continuously differentiable Schur–concave functions ([137, 126]).

Theorem 1.2.4 (**Schur, Ostrowski**). *Let* I *be an open interval in* $\mathbb{R}$ *and let* $\varphi : I^n \to \mathbb{R}$ *be a continuously differentiable function. Then* φ *is Schur–concave on* I^n *if, and only if,*

(i) φ *is symmetric, viz.* $\varphi(\mathbf{x}) = \varphi(\mathbf{x}\Pi)$ *for every permutation* Π*;*

(ii) *for all* $\mathbf{z} = (z_1, z_2, \ldots, z_n) \in I^n$ *and* $i \neq j$

$$(z_i - z_j)\,(\partial_i \varphi(\mathbf{z}) - \partial_j \varphi(\mathbf{z})) \leq 0.$$

1.3 Binary operations

Definition 1.3.1. A 2–place real function H is *binary operation* on a nonempty set $S \subseteq \mathbb{R}$ if $Dom\ H = S \times S$ and $Ran\ H \subseteq S$.

Let H be a binary operation on $[0, 1]$.

Definition 1.3.2. The *horizontal section of* H *at* $b \in [0, 1]$ is the function $h_b : [0, 1] \to [0, 1]$ defined by $h_b(x) := H(x, b)$; the *vertical section of* H *at* $a \in [0, 1]$ is the function $v_a : [0, 1] \to [0, 1]$ defined by $v_a(y) := H(a, y)$. The sections h_0, h_1, v_0 and v_1 are also called *margins* of H.

The *diagonal section of* H is the function $\delta_H : [0, 1] \to [0, 1]$ defined by $\delta_H(t) := H(t, t)$; the *opposite diagonal section of* H is the function $\delta_H^* : [0, 1] \to [0, 1]$ defined by $\delta_H^*(t) := H(t, 1 - t)$.

Definition 1.3.3. An element 0_H of $[0, 1]$ is said to be *annihilator* of H (or *zero, null element* of H) if $H(0_H, s) = 0_H = H(s, 0_H)$ for every s in $[0, 1]$.

An element 1_H of $[0, 1]$ is said to be *neutral element* of H if $H(1_H, s) = s = H(s, 1_H)$ for every s in $[0, 1]$.

Definition 1.3.4. An element a of $[0, 1]$ is said to be *idempotent* under H if $H(a, a) = \delta_H(a) = a$, namely if a is a fixed point for δ_H.

Definition 1.3.5. The *transpose* of H is the function H^T given by

$$H^T(x, y) = H(y, x) \qquad \text{for every } x, y \in [0, 1].$$

A binary operation H is said to be *commutative* (or *symmetric*) if

$$H(x, y) = H(y, x) \qquad \text{for every } x, y \in [0, 1], \tag{1.3}$$

viz. $H = H^T$.

Definition 1.3.6. A binary operation H is said to be *associative* if

$$H(H(x, y), z) = H(x, H(y, z)) \qquad \text{for every } x, y, z \in [0, 1]. \tag{1.4}$$

Definition 1.3.7. Let H be a binary operation on $[0, 1]$ and let x be an element of $[0, 1]$. The *H–powers* of x are the elements of $[0, 1]$ given recursively by

$$x_H^1 = x \qquad \text{and} \qquad x_H^{n+1} = H(x_H^n, x)$$

for all positive integers n.

1.4 Triangular norms

A *triangular norm* (briefly, *t*–norm) is a distinguished type of binary operation on the unit interval $[0, 1]$ that has been introduced (in a simplified form) by K. Menger

([106]) in order to extend the triangle inequality from the setting of metric spaces to probabilistic metric spaces. Since then, triangular norms were largely studied in this context and B. Schweizer and A. Sklar provided the axioms of t–norms as they are commonly used today (see the book [141] for an extended bibliography); but they also are widely used in statistics ([62, 65]) and in fuzzy logic, as a generalization of the classical logic connectives (see [160, 83]). For a complete discussion also on the recent developments of the theory of triangular norm, we refer to [139, 82, 3].

Definition 1.4.1. A binary operation T on $[0,1]$ is a *triangular norm* (briefly, *t–norm*) if it satisfies the following properties:

(T1) T is associative;

(T2) T is commutative;

(T3) T is increasing in each place;

(T4) T has neutral element 1.

The following functions are examples of t–norms:

$$M(x,y) := \min\{x,y\}; \qquad W(x,y) := \max\{x+y-1,0\};$$
$$\Pi(x,y) := xy; \qquad Z(x,y) = \begin{cases} 0, & \text{if } (x,y) \in [0,1[^2; \\ \min\{x,y\}, & \text{otherwise.} \end{cases}$$

They are called, resp., *minimum, Łukasiewicz, product and drastic t–norm* and are also denoted by T_M, T_L, T_P and T_D.

These four basic t–norms are remarkable for several reasons. For every t–norm T, we have

$$Z(x,y) \le T(x,y) \le M(x,y) \qquad \text{for all } (x,y) \in [0,1]^2.$$

The t–norms Π and W are prototypical examples of two important subclasses of t–norms called, respectively, *strict* and *nilpotent* t–norms ([83]). Moreover, M, Π and W play an important role in the theory of copulas, as we shall underline in the sequel.

An example of parametrized family of t–norm is the *Yager family* $\{T_\alpha\}_{\alpha \ge 0}$ (see [157]), given by

$$T_\alpha(x,y) = \begin{cases} Z(x,y), & \text{if } \alpha = 0; \\ M(x,y), & \text{if } \alpha = +\infty; \\ \max\{1 - [(1-x)^\alpha + (1-y)^\alpha]^{1/\alpha}\}, & \text{otherwise.} \end{cases}$$

Now, we present a simple way of constructing a new t–norm beginning from already known ones. This method goes back to some investigations by A.H. Clifford ([17]) on the theory of semigroups (see [141, 83] for more details).

Let $\{T_i : i \in \mathfrak{I}\}$ be a (possibly countable) collection of binary operations on $[0,1]$ that are increasing and bounded from above by M, namely $T_i(x,y) \le M(x,y)$ for every $i \in \mathfrak{I}$ and all $(x,y) \in [0,1]^2$. Let $\{J_i := [a_i, b_i]\}_{i\in\mathfrak{I}}$ be a family of closed, non overlapping (except at the end points), non degenerate subintervals of $[0,1]$. Then the function T, given by

$$T(x,y) := \begin{cases} a_i + (b_i - a_i)\; T_i\left(\dfrac{x-a_i}{b_i-a_i}, \dfrac{y-a_i}{b_i-a_i}\right), & \text{if } (x,y) \in [a_i,b_i]^2; \\ \min\{x,y\}, & \text{otherwise;} \end{cases}$$

is a binary operation on $[0,1]$, called the *ordinal sum* of the *summands* $\langle a_i, b_i, T_i\rangle$, $i \in \mathfrak{I}$, and we shall write $T = (\langle a_i, b_i, T_i\rangle)_{i\in\mathfrak{I}}$.

Theorem 1.4.1 (Theorem 5.3.8, [141])**.** *An ordinal sum of t–norms is a t–norm.*

Clearly, every t–norm T can be viewed as a trivial ordinal sum with only one summand $\langle 0,1,T\rangle$ only, viz. $T = (\langle 0,1,T\rangle)$. Moreover, the t–norm M can be viewed as an empty ordinal sum of t–norms, when the index set $\mathfrak{I}$ is the empty set. Notice that, for an ordinal sum of the above type, the points a_i and b_i ($i \in \mathfrak{I}$) are the idempotent elements of T.

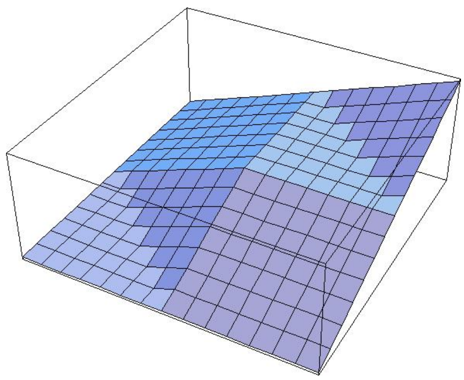

Figure 1.2: The ordinal sum $T = (\langle 0, 1/2, W\rangle, \langle 1/2, 1, W\rangle)$

Using ordinal sums, parametric families of t–norms can be easily constructed.

Example 1.4.1 (**Mayor–Torrens family**)**.** Given $\alpha \in [0,1]$, consider the following family

$$T_\alpha(x,y) := \begin{cases} \max\{0, x+y-\alpha\}, & \text{if } (x,y) \in [0,\alpha]^2; \\ \min\{x,y\}, & \text{otherwise.} \end{cases} \tag{1.5}$$

This family is known as the *Mayor–Torrens* family of t–norms and every T_α is an ordinal sum, $T = (\langle 0, \alpha, W\rangle)$.

An important property that a t–norm can have is the Archimedean one.

Definition 1.4.2. A t–norm T is called *Archimedean* if, for each $(x, y) \in]0, 1[^2$ there is an $n \in \mathbb{N}$ such that $x_T^n < y$.

For continuous Archimedean t–norms, we have the following representation (see [97, 3]).

Theorem 1.4.2. *For a binary operation $T : [0, 1]^2 \to [0, 1]$ the following statements are equivalent:*

(i) *T is a continuous Archimedean t–norm;*

(ii) *there exists a mapping $\varphi : [0, 1] \to [0, +\infty]$ continuous and strictly decreasing with $\varphi(1) = 0$ such that, for every $(x, y) \in [0, 1]$,*

$$T(x, y) = \varphi^{[-1]} (\varphi(x) + \varphi(y)) . \tag{1.6}$$

The function φ is said to be *additive generator* of T. A continuous and Archimedean t–norm T is said to be *strict* if it has an additive generator φ such that $\varphi(0) = +\infty$.

Theorem 1.4.3 (Representation of continuous t–norms). *Let T be a binary operation on $[0, 1]$ such that:*

(i) *T has annihilator element 0;*

(ii) *$T(1, 1) = 1$;*

(iii) *T is associative;*

(iv) *T is jointly continuous.*

Then T admits one of the following representations:

(a) *$T = M$;*

(b) *$T(x, y) = \varphi^{[-1]} (\varphi(x) + \varphi(y))$, where $\varphi : [0, 1] \to [0, +\infty]$ is a continuous and strictly decreasing function with $\varphi(1) = 0$;*

(c) *T is an ordinal sum of t–norms T_i, each of them representable in the form* (b).

1.5 Distribution Functions

Let n be a natural number, $n \in \mathbb{N}$.

Definition 1.5.1. Let H be an n–place real function and let $B = [\mathbf{x}, \mathbf{y}]$ be an n–box whose vertices belong to $DomH$. The *H–volume* of B is given by

$$V_H(B) = \sum sgn(\mathbf{c}) H(\mathbf{c}),$$

where the sum is taken over all the vertices $\mathbf{c}$ of B.

Definition 1.5.2. Let $S_1, \dots, S_n$ be nonempty subsets of $\overline{\mathbb{R}}$ and let H be an n–place real function such that $DomH = S_1 \times \cdots \times S_n$. The function H is said to be *n–increasing* if $V_H(B) \geq 0$ for every n–box B whose vertices lie in $DomH$.

In particular:

▷ a 1–increasing function is an increasing function in the classical sense;

▷ a 2–increasing function H satisfies the following condition

$$H(x_1, y_1) + H(x_2, y_2) \geq H(x_1, y_2) + H(x_2, y_1), \tag{1.7}$$

for every $x_1 \leq x_2$ and $y_1 \leq y_2$.

Definition 1.5.3. A function $H : \overline{\mathbb{R}}^n \to [0,1]$ is an *n–dimensional distribution function* (briefly n–d.f.) if

(i) H is n–increasing;

(ii) H is left–continuous in each place;

(iii) $H(+\infty, \dots, +\infty) = 1$;

(iv) $H(\mathbf{x}) = 0$, whenever $x \in \overline{\mathbb{R}}^n$ and $\min\{x_1, x_2, \dots, x_n\} = -\infty$.

The class of all n–dimensional d.f.'s will be denoted by Δ^n.

Specifically:

▷ $F : \overline{\mathbb{R}} \to [0,1]$ is a (unidimensional) d.f. it it is increasing and left–continuous with $F(-\infty) = 0$ and $F(+\infty) = 1$;

▷ $H : \overline{\mathbb{R}}^2 \to [0,1]$ is a bivariate d.f. if it is 2–increasing and left–continuous in each place, with $H(+\infty, +\infty) = 1$ and $H(x, -\infty) = 0 = H(-\infty, y)$ for all $x, y \in \overline{\mathbb{R}}$.

Definition 1.5.4. Let $n \in \mathbb{N}$, $n \geq 2$. Let $\{i_1, i_2, \dots, i_k\}$ be a nonempty set of k indices in $\{1, 2, \dots, n\}$ $(1 \leq k < n)$ and let H be an n–distribution function. The *k–margins* of H $(1 \leq k < n)$ are the $\binom{n}{k}$ functions $H_{i_1,\dots,i_k} : \overline{\mathbb{R}}^k \to [0,1]$ defined, for every $\mathbf{y} \in \overline{\mathbb{R}}^k$ by

$$H_{i_1,\dots,i_k}(\mathbf{y}) = H(\mathbf{x}),$$

where $\mathbf{x}$ is a point in $\overline{\mathbb{R}}^n$ defined by

$$x_j = \begin{cases} y_j, & \text{if } j \in \{i_1, \dots, i_k\}; \\ +\infty, & \text{if } j \notin \{i_1, \dots, i_k\}. \end{cases}$$

Proposition 1.5.1. *Given an n–dimensional d.f. H, every k–margin of H $(1 \leq k < n)$ is a k–dimensional distribution function.*

In particular, we shall generally denote the 1–margins of an n–d.f. H by $F_1, \dots, F_n$ instead of $H_1, \dots, H_n$ and we shall refer to them briefly as *margins* or *marginal d.f.'s*.

Remark 1.5.1. Given a probability space $(\Omega, \mathcal{F}, P)$ and a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$, namely $\mathbf{X} : \Omega \to \overline{\mathbb{R}}^n$ is a measurable mapping with respect to the σ–algebra $\mathcal{F}$ and the Borel σ–algebra over $\overline{\mathbb{R}}^n$, the function

$$H(\mathbf{x}) := P\left(\bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) < x_i\}\right) \tag{1.8}$$

is a n–d.f.. Conversely, in view of the classical *Kolmogorov's compatibility Theorem* (see [94]), given an n–dimensional d.f. H it is possible to construct a probability space $(\Omega, \mathcal{F}, P)$ and a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$, such that equation (1.8) holds for every $\mathbf{x}$ in $\mathbb{R}^n$.

Remark 1.5.2. In many applications on reliability theory, the random variables of interest represent lifetimes of individuals or objects and then it is very important to study the *survival d.f.* instead of the d.f.. For a r.v. X, its survival d.f. is defined by $\overline{F}(t) := P(X \geq t) = 1 - F_X(t)$. In general, the *joint survival d.f.* of the vector $(X_1, X_2, \dots, X_n)$ is defined by

$$\overline{H}(x_1, x_2, \dots, x_n) := P(X_1 \geq x_1, X_2 \geq x_2, \dots, X_n \geq x_n).$$

For a random pair (X, Y) with joint d.f. H and margins F_1 and F_2, the survival d.f. is given by

$$\overline{H}(x, y) = 1 - F_1(x) - F_2(y) + H(1 - x, 1 - y).$$

Finally, we recall the concept of Fréchet class, introduced in [55].

Definition 1.5.5. The *Fréchet class* determined by the univariate d.f.'s $F_1, F_2, \dots, F_n$ is the set $\Gamma[F_1, F_2, \dots, F_n]$ of all n–d.f.'s whose margins are $F_1, F_2, \dots, F_n$.

Notice that, for every choice of a set of n univariate d.f.'s, the corrisponding Fréchet class is not empty, because it contains the independence d.f. given by the product of the margins.

1.6 Copulas

In this section, we introduce the concept of copula. For simplicity's sake, first, we limit ourselves to consider two–dimensional copulas; the multivariate case ($n \geq 3$) will be, instead, considered briefly in section 1.9. For a deeper discussion of this topic, we refer to the book by R.B. Nelsen ([114]) and to chap. 6 of the book by B. Schweizer and A. Sklar ([141]) (see also the recent papers [128, 116]).

Definition 1.6.1. A function $C\colon [0, 1]^2 \to [0, 1]$ is a (bivariate) *copula* if it satisfies:

(C1) the *boundary conditions*,

$$\forall x \in [0,1] \quad C(x,0) = C(0,x) = 0 \quad \text{and} \quad C(x,1) = C(1,x) = x;$$

(C2) the 2–*increasing property*, i.e. for all x, x', y, y' in $[0,1]$, with $x \leq x'$ and $y \leq y'$,

$$V_C\left([x,x'] \times [y,y']\right) := C(x',y') - C(x,y') - C(x',y) + C(x,y) \geq 0.$$

In particular, every copula is *increasing in each place*, viz.

$$C(x,y) \leq C(x',y) \qquad \text{and} \qquad C(x,y) \leq C(x,y') \quad \text{for } x \leq x',\, y \leq y', \tag{1.9}$$

and satisfies the 1–*Lipschitz condition*, i.e. for all $x, x', y, y' \in [0,1]$

$$|C(x,y) - C(x',y')| \leq |x - x'| + |y - y'|. \tag{1.10}$$

Moreover, if $C\colon [0,1]^2 \to [0,1]$ is twice continuously differentiable, condition (C2) is equivalent to

$$\frac{\partial^2 C(x,y)}{\partial_x \partial_y} \geq 0 \qquad \text{for all } (x,y) \in [0,1]^2. \tag{1.11}$$

In order to prove that a function $F : [0,1]^2 \to [0,1]$ satisfies the so–called *rectangular inequality* (C2), the following technical Proposition will be useful. But, first, we denote by Δ_+ and Δ_- the subsets of the unit square given by:

$$\Delta_+ := \{(x,y) \in [0,1]^2 : x \geq y\}, \qquad \Delta_- := \{(x,y) \in [0,1]^2 : x \leq y\}, \tag{1.12}$$

and we prove

Lemma 1.6.1. *For every $F : [0,1]^2 \to [0,1]$, the F–volume $V_F(R)$ of any rectangle $R \subseteq [0,1]^2$ can be expressed as the sum $\sum_i V_F(R_i)$ of at most three terms, where the rectangles R_i may have a side in common and belong to one of the following types:*

(a) $R_i \subseteq \Delta_+$;

(b) $R_i \subseteq \Delta_-$;

(c) $R_i = [s,t] \times [s,t]$.

Proof. Let a rectangle $R \subseteq [0,1]^2$ be given; if it belongs to one of the three types (a), (b) or (c) there is nothing to prove. Then, consider the other possible cases: R may have one, two or three vertices in Δ_-.

If $R = [x_1, x_2] \times [y_1, y_2]$ has one vertex in Δ_+ and three vertices in Δ_-, then, since $y_2 > x_2 > y_1 > x_1$, we can write

$$R = \left([x_1, y_1] \times [y_1, y_2]\right) \cup \left([y_1, x_2] \times [y_1, x_2]\right) \cup \left([y_1, x_2] \times [x_2, y_2]\right);$$

of these rectangles, the first and the third one are of type (b), while the second one is of type (c). Now

$$\begin{aligned} V_F([x_1, y_1] \times [y_1, y_2]) &= F(y_1, y_2) - F(y_1, y_1) - F(x_1, y_2) + F(x_1, y_1), \\ V_F([y_1, x_2] \times [y_1, x_2]) &= F(x_2, x_2) - F(x_2, y_1) - F(y_1, x_2) + F(y_1, y_1), \\ V_F([y_1, x_2] \times [x_2, y_2]) &= F(x_2, y_2) - F(x_2, x_2) - F(y_1, y_2) + F(y_1, x_2). \end{aligned}$$

Therefore, summing these equalities we have

$$\begin{aligned} &V_F([x_1, y_1] \times [y_1, y_2]) + V_F([y_1, x_2] \times [y_1, x_2]) + V_F([y_1, x_2] \times [x_2, y_2]) \\ &\qquad = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) = V_F([x_1, x_2] \times [y_1, y_2]), \end{aligned}$$

which proves the assertion in this case. The other cases can be proved in a similar manner. □

Proposition 1.6.1. *A binary operation* $F : [0,1]^2 \to [0,1]$ *is* 2*–increasing if, and only if, the three following conditions hold:*

(a) $V_F(R) \geq 0$ *for every rectangle* $R \subseteq \Delta_+$*;*

(b) $V_F(R) \geq 0$ *for every rectangle* $R \subseteq \Delta_-$*;*

(c) $V_F(R) \geq 0$ *for every rectangle* $R = [s,t] \times [s,t] \subseteq [0,1]^2$*.*

Proof. If F is 2–increasing, (a), (b) and (c) follow easily. Conversely, let R be a rectangle of $[0,1]^2$. Then, because of the previous Lemma, R can be decomposed into the union of at most three sub–rectangles R_i of type (a), (b) and (c); and for each of them $V_F(R_i) \geq 0$ holds. Therefore $V_F(R) = \sum V_F(R_i) \geq 0$. □

For every $(x,y) \in [0,1]^2$ and for every copula C

$$W(x,y) \leq C(x,y) \leq M(x,y); \tag{1.13}$$

this inequality is known as the *Fréchet–Hoeffding bounds inequality* ([109]), and W and M are copulas, called also *Fréchet–Hoeffding bounds*, in honour of the pioneering works of Hoeffding ([71]) and Fréchet ([55]). Hence the graph of a copula is a surface within the unit cube $[0,1]^3$ that lies between the graphs of the copulas W and M.

A third important copula is the *product copula* Π.

Notice that a copula is the restriction to $[0,1]^2$ of the bivariate d.f. H_C, given by

$$H_C(x,y) := \begin{cases} 0, & \text{if } \min\{x,y\} < 0; \\ C(x,y), & \text{if } (x,y) \in [0,1]^2; \\ x, & \text{if } x \in [0,1] \text{ and } y > 1; \\ y, & \text{if } x > 1 \text{ and } y \in [0,1]; \\ 1, & \text{if } x > 1 \text{ and } y > 1; \end{cases}$$

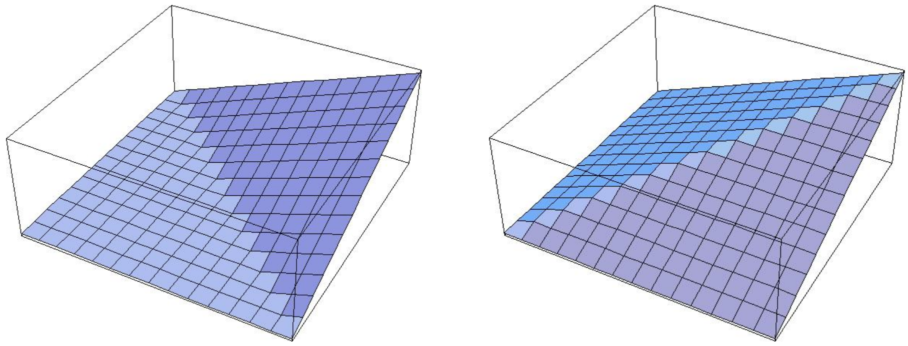

Figure 1.3: The copulas W and M

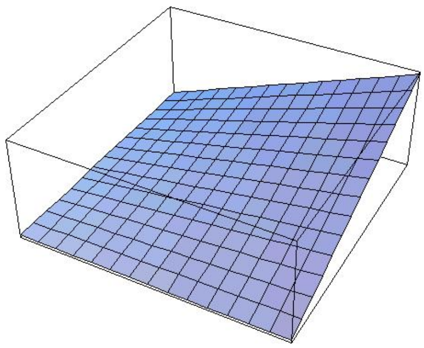

Figure 1.4: The copula Π

whose margins are uniformly distributed on $[0, 1]$.

Every copula C induces a probability measure P_C on $[0, 1]^2$ given, for every rectangle R, by $P_C(R) := V_C(R)$. In particular, such a probability measure P_C is *doubly stochastic*, namely $P(J \times [0, 1]) = P([0, 1] \times J) = \lambda(J)$, where J is a Borel set in $[0, 1]$ and λ is the the Lebesgue measure on $[0, 1]$. The *support of a copula* C is the complement of the union of all open subsets of $[0, 1]^2$ with P_C measure equal to zero. If a Borel set $R \subseteq [0, 1]^2$ has P_C–measure equal to $m \in \,]0, 1]$, we said that the *probability mass* of C on R is m (or C *spreads* a mass m on R). For every copula C, we have the decomposition

$$C(x, y) = C_A(x, y) + C_S(x, y),$$

where

$$C_A(x, y) := \int_0^x \int_0^y \frac{\partial^2}{\partial s \partial t} C(s, t)\, ds dt, \quad C_S(x, y) = C(x, y) - C_A(x, y).$$

The function C_A is the *absolutely continuous component* of C and C_S is the *singular component* of C. If $C = C_A$, then it is called *absolutely continuous* (e.g. Π) and the mixed second derivative of C, $\partial_{12}C$ is called *density* of C. If $C = C_S$, then it is called *singular* (e.g. M and W). If one of the first derivatives of C has a jump discontinuity, then C has a singular component (see [74, page 15]).

When a copula C is singular, then its support has Lebesgue measure zero, and conversely. For example, the support of M is the main diagonal of $[0,1]^2$, $\{(x,y) \in [0,1]^2 \mid x = y\}$, namely M is singular. Also W is singular and its support is the opposite diagonal of $[0,1]^2$, $\{(x,y) \in [0,1]^2 \mid x + y = 1\}$.

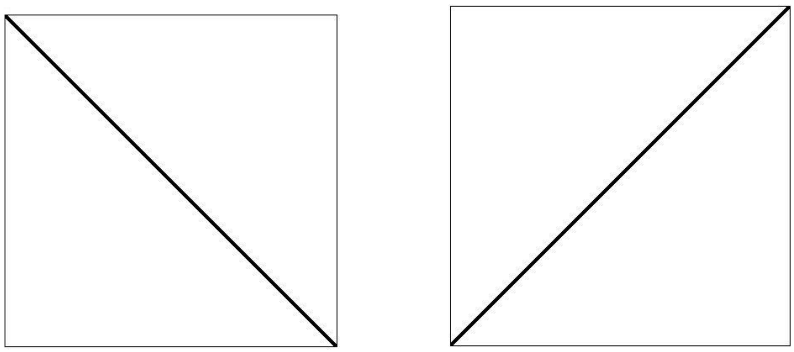

Figure 1.5: Supports of the copulas W and M

We shall denote by $\mathcal{C}$ (or $\mathcal{C}_2$) the class of all the (bivariate) copulas. The set $\mathcal{C}$ is convex and compact under the topology induced by the norm $\| \ \|_\infty$, given for every A in $\mathcal{C}$ by

$$\|A\|_\infty := \max\left\{|A(x,y)| : (x,y) \in [0,1]^2\right\}.$$

Moreover, pointwise convergence in $\mathcal{C}$ is equivalent to uniform convergence, in the sense that, if a sequence $\{C_n : n \in \mathbb{N}\}$ of copulas converges pointwise to a copula C, then it converges also uniformly.

Notice that, since the set $\mathcal{C}$ of copulas is a convex and compact subset of the class of real–valued continuous functions defined on $[0,1]^2$, equipped with the $\| \ \|_\infty$ norm, from the classical Krein–Milman's Theorem (see, e.g., [32]) it follows that $\mathcal{C}$ is the convex hull of its extremal points.

Given two copulas C and D, D is said to be *more concordant* (or *more PQD*) than C ($C \leq D$, for short) if $C(x,y) \leq D(x,y)$ for every x, y in $[0,1]$ (see [74]). The concordance order is only a partial ordering; however, some parametric families of copulas are totally ordered. In particular, we say that a family $\{C_\theta : \theta \in I \subseteq R\}$ is *positively ordered* (resp., *negatively*) if $C_\alpha \leq C_\beta$ whenever $\alpha \leq \beta$ (resp., $\alpha \geq \beta$).

1.6.1 Copulas and random variables

Sklar's Theorem (see [149, 150, 151]) is surely the most important result in the theory of copulas and it is the foundation of many of the applications of copulas to statistics. From that, it is clear in which sense we say that "*a copula is a function which joins or couples a bivariate distribution function to its one–dimensional margins*".

Theorem 1.6.1 (**Sklar, 1959**). *If X and Y are random variables with unidimensional d.f.'s F and G, respectively, and joint d.f. H, then there exists a copula C (uniquely determined on $\operatorname{Ran} F \times \operatorname{Ran} G$, and hence unique when X and Y are continuous) such that*

$$\forall (x, y) \in \overline{\mathbb{R}}^2 \qquad H(x, y) = C(F(x), G(y)). \tag{1.14}$$

Conversely, given a copula C and two univariate d.f.'s F and G, the function H defined in (1.14) *is a bivariate d.f. with margins F and G.*

Given a joint d.f. H with continuous margins F and G, it is easy to construct the corresponding copula is given by:

$$C(x, y) = H(F^{[-1]}(x), G^{[-1]}(y)),$$

where $F^{[-1]}(t) = \sup\{x : F(x) \leq t\}$ is the pseudo–inverse of F (and similarly for $G^{[-1]}$). Conversely, given a copula C and two univariate d.f.'s F and G, the equality (1.14) allows us to construct a bivariate d.f. H.

Note as well that, if X and Y are continuous r.v.'s with d.f.'s F and G, C is the joint d.f. of the r.v.'s $U = F(X)$ and $V = G(Y)$.

The following result gives an interesting probabilistic interpretation of the three basic copulas M, Π and W.

Theorem 1.6.2. *For continuous r.v.'s X and Y with copula C the following statements hold:*

- ▹ *X and Y are independent if, and only if, $C = \Pi$;*
- ▹ *Y is almost surely an increasing function of X if, and only if, $C = M$;*
- ▹ *Y is almost surely a decreasing function of X if, and only if, $C = W$.*

In general, Sklar's Theorem allows us to study the dependence properties of a random vector by examination of the copula alone, *if the r.v.'s are continuous*. This last assumption is essential because, for discontinuous r.v.'s, the copula is not unique and many problems arise, as discussed, e.g., in [100, 146, 124].

Example 1.6.1. Let X and Y be r.v.'s with d.f.'s $F_X = 1_{]a,+\infty]}$ and $F_Y = 1_{]b,+\infty]}$, with $a < b$. Then the joint d.f. of X and Y is

$$H(x,y) = \begin{cases} 1, & \text{if } (x,y) \geq (a,b); \\ 0, & \text{otherwise.} \end{cases}$$

Notice that, in view of Sklar's Theorem, there exists a (not uniquely determined) copula C such that (1.14) holds. In this case, C has to satisfy only the assumptions

$$C(1,1) = 1, \qquad C(0,1) = C(1,0) = C(0,0) = 0.$$

Therefore, every copula can be associated with the random pair (X,Y).

In the sequel, when we speak about "the copula of the random pair (X,Y)", we assume that X and Y are continuous and, therefore, the copula is unique and it will also be denoted by C_{XY}.

Remark 1.6.1. The first–order derivatives of a copula have a nice interpretation. If C is the copula of the random pair (U,V) of two r.v.'s uniformly distributed on $[0,1]$, then

$$\frac{\partial C(u,v)}{\partial u} = P(V \leq v \mid U = u) \quad \text{and} \quad \frac{\partial C(u,v)}{\partial v} = P(U \leq u \mid V = v).$$

Now, we express the copula of a random vector obtained from another one by strictly monotone transformations.

Theorem 1.6.3. *Let X and Y be continuous r.v.'s with copula C. Let α and β be two functions strictly monotone on $RanX$ and $RanY$, respectively.*

(i) *If α and β are both strictly increasing, then*

$$C_{\alpha(X)\beta(Y)} = C_{XY}.$$

(ii) *If α is strictly increasing and β is strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(x,y) = x - C_{XY}(x, 1-y).$$

(iii) *If α is strictly decreasing and β is strictly increasing, then*

$$C_{\alpha(X)\beta(Y)}(x,y) = y - C_{XY}(1-x, y).$$

(iii) *If α and β are both strictly decreasing, then*

$$C_{\alpha(X)\beta(Y)}(x,y) = x + y - 1 + C_{XY}(1-x, 1-y).$$

From the above result we have that, given a copula C, the following function are copulas (see [84]):

$$C_{0,1}(x,y) := x - C(x,1-y), \tag{1.15}$$

$$C_{1,0}(x,y) := y - C(1-x,y), \tag{1.16}$$

$$C_{1,1}(x,y) := x + y - 1 + C(1-x,1-y). \tag{1.17}$$

In particular, $C_{1,1}$ is called *survival copula* and it is denoted more frequently by $\hat{C}$. It has a large use in reliability theory, where Sklar's Theorem can be reformulated under the following form:

Theorem 1.6.4. *Let X and Y be two continuous r.v.'s with copula C. Let $\overline{H}$ be the joint survival d.f. of (X,Y) and let $\overline{F}$ and $\overline{G}$ be the univariate survival d.f.'s. Then*

$$\overline{H}(x,y) = \hat{C}\left(\overline{F}(x), \overline{G}(y)\right),$$

where $\hat{C}$ is the survival copula of C.

Remark 1.6.2. Notice that the survival copula $\hat{C}$ is not the joint survival d.f. $\overline{C}$ of two r.v.'s unifomly distributed on $[0,1]$ whose joint d.f. is the copula C. In such a case, in fact, we have $\overline{C}(x,y) := 1 - x - y + C(x,y)$.

The symmetry properties of a random pair can also be expressed in terms of the associated copula (see [114, 84] for more details).

Definition 1.6.2. Two r.v.'s X and Y are *exchangeable* if, and only if, (X,Y) and (Y,X) are identically distributed.

Proposition 1.6.2. *Let X and Y be continuous r.v.'s with margins d.f.'s F and G, respectively, and copula C. Then X and Y are exchangeable if, and only if, $F = G$ and C is symmetric.*

Definition 1.6.3. Let X and Y be r.v.'s and let (a,b) be a point in $\mathbb{R}^2$.

- ▷ (X,Y) is *radially symmetric* about (a,b) if the joint d.f. of $(X-a)$ and $(Y-b)$ is the same as the joint d.f. of $(a-X)$ and $(b-Y)$.
- ▷ (X,Y) is *jointly symmetric* about (a,b) if the following four pairs of r.v.'s have a common joint d.f.: $(X-a, Y-b)$, $(X-a, b-Y)$, $(a-X, Y-b)$ and $(a-X, b-Y)$.

Note that the joint symmetry implies the radial symmetry.

Proposition 1.6.3. *Let X and Y be continuous r.v.'s with marginal d.f.'s F and G, respectively, and copula C. Given a point $(a,b) \in \mathbb{R}^2$, assume that $(X-a)$ has the same d.f. as $(a-X)$, and $(Y-b)$ has the same d.f. as $(b-Y)$. Then:*

- ▷ *(X,Y) is radially symmetric about (a,b) if, and only if, $C = \widehat{C}$;*
- ▷ *(X,Y) is jointly symmetric about (a,b) if, and only if, $C = C_{0,1}$ and $C = C_{1,0}$ (and then also $C = \widehat{C}$).*

1.6.2 Families of copulas

For many years, statisticians have been fascinated by the following problem: given two univariate d.f.'s F and G, find a bivariate d.f. H having F and G as its margins, and having useful properties such as a simple analytic expression, a simple stochastic representation, some desirable dependence properties, and a suitable number of parameters. Many methods and procedures for constructing such joint distributions have been introduced and studied in the literature (see, for example [75, 73]). As noted in subsection 1.6.1, thanks to Sklar's Theorem, we can decompose this problem into two easier steps: the construction of a copula and the construction of two univariate margins.

Having several families of bivariate distributions at disposal is of great importance in statistical applications. In fact, for many years, multivariate models have been often constructed either under the assumption of the independence of their components or by assuming the components are connected by a multivariate normal distribution (see, e.g., [58]). Copulas, instead, allow to study models with a more flexible and wide range of dependence.

In [74, 77], some criteria are given in order to ensure that a family of copulas is a "good" family, in the sense that it can be useful in certain statistical applications. Here we list some desirable properties for a parametric class of copulas C_α, where α belongs to an interval of the real line:

- ▷ *interpretability*, which means having a probabilistic interpretation;
- ▷ *flexible and wide range of dependence*, which implies that the copula Π and at least one of the Fréchet–Hoeffding bounds W and M belong to the class;
- ▷ *closed form*, in the sense that every copula of the class is absolutely continuous or has a simple representation;
- ▷ *ordering*, with respect, for example, to concordance.

Now, we present some families of copulas (see [114] for more details).

Example 1.6.2 (Fréchet family). For all $x, y \in [0,1]$ and $\alpha, \beta \in [0,1]$ such that $\alpha + \beta \leq 1$, the family

$$C_{\alpha,\beta}(x,y) = \alpha\, M(x,y) + (1-\alpha-\beta)\,\Pi(x,y) + \beta\, W(x,y)$$

is a family of copulas, known as the *Fréchet family.* A slight modification of this family is the so–called *linear Spearman copula* (see [72] and [74, family B11]), given, for every $\alpha \in [-1,1]$, by

$$C_\alpha(x,y) = (1-|\alpha|)\cdot \Pi(x,y) + |\alpha| \cdot C_{sgn(\alpha)}(x,y),$$

where $C_{sgn(\alpha)} = M$, if $\alpha \geq 0$, and $C_{sgn(\alpha)} = W$, otherwise.

Example 1.6.3 (**FGM family**)**.** For all $x, y \in [0,1]$ and $\alpha \in [-1,1]$

$$C_\alpha(x,y) = xy + \alpha xy\,(1-x)\,(1-y)$$

is a family of copulas, known as the *Farlie-Gumbel-Morgenstern family* (often abbreviated FGM) and contains as its members copulas with sections that are quadratic in both x and y.

Example 1.6.4 (**Cuadras–Augé family**)**.** For every $\alpha \in [0,1]$, the following function

$$C_\alpha(x,y) := \begin{cases} xy^{1-\alpha}, & \text{if } x \le y; \\ x^{1-\alpha}y, & \text{if } x \ge y; \end{cases}$$

is a copula, belonging to the family introduced by Cuadras and Augé ([18]). Notice that C_α is the weighted geometric mean of M and Π; in particular, $C_0 = \Pi$ and $C_1 = M$.

Example 1.6.5 (**Marshall–Olkin family**)**.** For every α and β in $[0,1]$, the following function

$$C_{\alpha,\beta}(x,y) := \begin{cases} x^{1-\alpha}y, & \text{if } x^\alpha \ge y^\beta; \\ xy^{1-\beta}, & \text{if } x^\alpha \le y^\beta; \end{cases}$$

is a copula, belonging to the family introduced by Marshall and Olkin ([101, 102]), which contains the family given in Example 1.6.4 for $\alpha = \beta$.

Example 1.6.6 (**BEV Copula**)**.** Let $A : [0,1] \to [1/2,1]$ be a convex function such that $\max\{t, 1-t\} \le A(t) \le 1$ for every $t \in [0,1]$. The following function

$$C_A(x,y) := \exp\left[(\ln x + \ln y) A\left(\frac{\ln x}{\ln x + \ln y}\right)\right]$$

is a copula, known as *bivariate extreme value copula* (briefly, BEV) (see [74, chap. 6]). This copula satisfies the equality $C^n(x,y) = C(x^n, y^n)$ for every $n \in \mathbb{N}$. The name of this class arises from the theory of extreme statistics. In fact, let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from bivariate distribution H, define $M_n := \max\{X_1, \dots, X_n\}$ and $N_n := \max\{Y_1, \dots, Y_n\}$ and suppose that there exist constants a_{1n}, a_{2n}, b_{1n} and b_{2n}, with $a_{1n} > 0$ and $a_{2n} > 0$, for which the pair

$$\left(\frac{M_n - b_{1n}}{a_{1n}}, \frac{N_n - b_{2n}}{a_{2n}}\right)$$

has a non-degenerate joint limiting distribution H^*. Then the copula associated with H^* is a BEV copula (see [59, 129]).

Example 1.6.7 (**Normal copula**)**.** Let $N_\rho(x,y)$ denote the standard bivariate normal joint d.f. with correlation coefficient ρ. Then the corresponding copula is

$$C_\rho(x,y) = N_\rho(\Phi^{-1}(x), \Phi^{-1}(y)),$$

where Φ denotes the standard normal d.f.. Because Φ^{-1} does not have a closed form, there is no closed form for C_ρ, which can be only evaluated approximately.

Example 1.6.8 (Shuffle of Min). The copulas known as *shuffles of* M were introduced in [110] and do not have a simple explicit expression. However, the procedure to obtain their mass distribution can be easily described:

1. spread uniformly the mass on the main diagonal of $[0,1]^2$,
2. cut $[0,1]^2$ vertically into a finite number of strips,
3. shuffling the strips with perhaps some of them flipped around their vertical axes of symmetry,
4. reassembling them to form the square again.

The resulting mass distribution corresponds to a copula called *shuffle of* M. Formally, a shuffle of M is determined by a partition $\{J_i\}_{i=1,2,\dots,n}$, a permutation of $(1,2,\dots,n)$ and an orientation n–ple $(i_1, i_2, \dots, i_n)$ such that $i_k = -1$ or 1 according to whether or not the strip $J_i \times [0,1]$ is flipped.

For instance, the shuffle given by $\{[0,1/2],[1/2,1]\}$, permutation $(2,1)$ and orientation $(-1,-1)$ is W. Moreover, the shuffle of M with partition $\{[0,a],[a,1-a],[1-a,1]\}$, $(a \in [0,1/2])$, permutation $(3,2,1)$ and orientation $(-1,+1,-1)$ is the copula $C_\alpha(x,y) = \max\{W(x,y), M(x,y) - \alpha\}$.

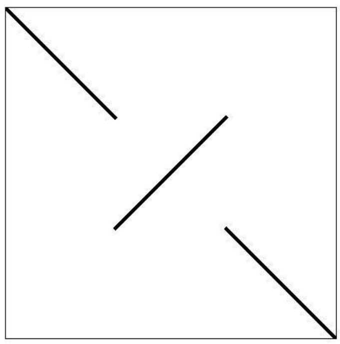

Figure 1.6: Support of the copula C_α for $\alpha = 1/3$

A way of constructing new copulas is given by the *ordinal sum* construction, a method already presented in section 1.4, and reproduced here.

Theorem 1.6.5. *Let $C = (\langle a_i, b_i, C_i \rangle)_{i\in\mathcal{I}}$ be an ordinal sum such that C_i is a copula for every $i \in \mathcal{I}$. Then C is a copula.*

1.6.3 Diagonal sections of copulas

Given a copula C, it is easily proved that its diagonal δ satisfies the following properties:

(D1) $\delta(1) = 1$;

(D2) $\delta(t) \le t$ for all $t \in [0,1]$;

(D3) δ is increasing;

(D4) $|\delta(t) - \delta(s)| \le 2|t - s|$ for all $t, s \in [0,1]$.

The set of functions $\delta : [0,1] \to [0,1]$ satisfying (D1)–(D3) will be denoted by $\mathcal{D}$, instead $\mathcal{D}_2$ will denote the subset of $\mathcal{D}$ of the functions satisfying also (D4).

For each function $\delta \in \mathcal{D}_2$, there is always a copula whose diagonal section coincides with δ. Consider, for example, the *diagonal copula*

$$K_\delta(x,y) := \min\left\{x, y, \frac{\delta(x)+\delta(y)}{2}\right\}, \tag{1.18}$$

introduced in [117, 56]. Another example is given by the *Bertino copula* ([9, 57])

$$B_\delta(x,y) := \min\{x,y\} - \min\{t - \delta(t) : t \in [x \wedge y, x \vee y]\}. \tag{1.19}$$

In particular, a Bertino copula is called *simple* if it can be expressed in the form

$$B_\delta(x,y) := \min\{x,y\} - \min\{x - \delta(x), y - \delta(y)\}. \tag{1.20}$$

From a probabilistic point of view, investigations on diagonal sections of copulas are of interest because, if X and Y are random variables with the same distribution function F and copula C, then the distribution function of $\max\{X,Y\}$ is $\delta_C(F(t))$. Moreover, copulas with given diagonal section have important consequences in finding the bounds on arbitrary subsets of joint d.f.'s (see [121]). An absolutely continuous copula with given diagonal section is also given in the recent paper [52].

1.6.4 Archimedean copulas

From a general point of view, copulas are special type of binary operations on $[0,1]$, and many important copulas are also t–norms. In particular, the class of Archimedean copulas (i.e. associative copulas with the Archimedean property as defined in section 1.4), is a very useful subclass of copulas, both in the statistical context (see [62, 63, 113, 112]) and in applications, especially in finance, actuarial science ([58, 70]) and hydrology ([134]), due to their simple form and nice properties. Archimedean copulas are characterized here.

Theorem 1.6.6. *A function C is an Archimedean copula if, and only if, it admits the representation*

$$C(x,y) := \varphi^{[-1]}\left(\varphi(x)+\varphi(y)\right), \qquad \text{for all } x,y \in [0,1], \tag{1.21}$$

where $\varphi : [0,1] \to [0,+\infty]$ is continuous, strictly decreasing and convex with $\varphi(1) = 0$.

The function φ is said to be an *additive generator* of C and, therefore, C is also denoted as C_φ. Notice that, by setting $h(t) := \exp(-\varphi(t))$ for every $t \in [0,1]$, C_φ may be represented in the form

$$C_\varphi(x,y) = h^{[-1]}\left(h(x)\cdot h(y)\right) \qquad \text{for all } x,y \in [0,1]. \tag{1.22}$$

This function h is a *multiplicative generator* of C_φ and Theorem 1.6.6 may be rephrased in the following (multiplicative) form.

Theorem 1.6.7. *A function C is an Archimedean copula if, and only if, it admits the representation*

$$C(x,y) := h^{[-1]}\left(h(x)\cdot h(y)\right), \qquad \text{for all } x,y \in [0,1], \tag{1.23}$$

where $h : [0,1] \to [0,1]$ is continuous, strictly increasing and log–concave, viz. for every α, s and t in $[0,1]$, it satisfies the inequality

$$h^\alpha(s)\,h^{1-\alpha}(t) \le h\left(\alpha s + (1-\alpha)t\right).$$

Notice that, neither the additive nor the multiplicative generator of an Archimedean copula are unique. In fact, if φ is an additive generator of C, then every additive generator of C has the form $\varphi_1 := k\,\varphi$, for $k > 0$. Analogously, if h is a multiplicative generator of a copula D, then $h_1(t) := h(t^\alpha)$ $(\alpha > 0)$ is also a multiplicative generator for D. The next result yields a technique for finding generators of Archimedean copulas ([62]).

Theorem 1.6.8. *Let C be an Archimedean copula with generator φ. Then*

$$\varphi'(x)\cdot \partial_y C(x,y) = \varphi'(y)\cdot \partial_x C(x,y) \qquad \text{a.e. on } [0,1]^2.$$

In Table 1.1 we list some known families of Archimedean copulas and their additive generators.

In the spirit of the representation of continuous t–norms (see 1.4.3), Archimedean copulas allow us to give a full characterization of associative copulas.

Theorem 1.6.9 (Representation of associative copulas). *Let C be an associative copula with diagonal section δ_C. Then:*

- ▷ *$C = M$ if, and only if, $\delta_C = id_{[0,1]}$;*

Family	Copula $C_\theta(x,y)$	$\theta \in$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta x}-1)(e^{-\theta y}-1)}{e^{-\theta}-1}\right)$	$[-\infty,+\infty]$
Clayton	$\max\left\{(x^{-\theta}+y^{-\theta}-1)^{-1/\theta},0\right\}$	$[-1,+\infty]$
Gumbel–Hougaard	$\exp\left(-\left((-\ln x)^{\theta}+(-\ln y)^{\theta}\right)^{1/\theta}\right)$	$[1,+\infty]$
Ali–Mikhail–Haq	$\frac{xy}{1-\theta(1-x)(1-y)}$	$[-1,1]$

Table 1.1: Families of Archimedean copulas

- ▷ *C is Archimedean if, and only if, $\delta_C(t) < t$ on $]0,1[$;*
- ▷ *C is an ordinal sum of Archimedean copulas if, and only if, $\delta_C(t) < t$ for some, but not all, t in $]0,1[$.*

In [14], the following generalization of an Archimedean copula is studied.

Example 1.6.9. [**Archimax copula**] Let $A : [0,1] \to [1/2,1]$ be a convex function such that $\max\{t, 1-t\} \le A(t) \le 1$ for every $t \in [0,1]$. Let φ be an additive generator of an Archimedean copula. The following function

$$C_{\varphi,A}(x,y) := \varphi^{[-1]}\left[(\varphi(x)+\varphi(y))A\left(\frac{\varphi(x)}{\varphi(x)+\varphi(y)}\right)\right]$$

is a copula, known as *Archimax*. The family of Archimax copulas includes both Archimedean copulas and BEV copulas. The functions A and φ, which uniquely determine $C_{\varphi,A}$, are called, respectively, *dependence function* and *Archimedean generator*.

1.7 Dependence Properties

Here we recall some dependence properties between random variables that will be expressed in terms of copulas. For more details on this topic, see [114, chap. 5] and [74].

Definition 1.7.1. Let X and Y be random variables.

- ▷ X and Y are *positively quadrant dependent* (briefly, *PQD*) if, for every (x,y) in $\mathbb{R}^2$, $P(X \le x, Y \le y) \ge P(X \le x)P(Y \le y)$.
- ▷ X and Y are *negatively quadrant dependent* (briefly, *NQD*) if, for every (x,y) in $\mathbb{R}^2$, $P(X \le x, Y \le y) \le P(X \le x)P(Y \le y)$.

Proposition 1.7.1. *Let X and Y be continuous r.v.'s with copula C. X and Y are PQD (resp. NQD) if, and only if, $C \ge \Pi$ (resp. $C \le \Pi$).*

Definition 1.7.2. Let X and Y be random variables.

- ▷ Y is *left tail decreasing* in X (briefly, $LTD(Y|X)$) if, and only if, the mapping $x \mapsto P(Y \le y \mid X \le x)$ is a decreasing function for all y.
- ▷ X is *left tail decreasing* in Y (briefly, $LTD(X|Y)$) if, and only if, the mapping $y \mapsto P(X \le x \mid Y \le y)$ is a decreasing function for all x.
- ▷ Y is *right tail increasing* in X (briefly, $RTI(Y|X)$) if, and only if, the mapping $x \mapsto P(Y > y \mid X > x)$ is an increasing function for all y.
- ▷ X is *right tail increasing* in Y (briefly, $RTI(X|Y)$) if, and only if, the mapping $y \mapsto P(X > x \mid Y > y)$ is an increasing function for all x.

Proposition 1.7.2. *Let X and Y be continuous r.v.'s with copula C.*

- ▷ *$LTD(Y|X)$ if, and only if, for every $y \in [0,1]$,*

$$x \mapsto C(x,y)/x \qquad \text{is decreasing.}$$

- ▷ *$LTD(X|Y)$ if, and only if, for every $x \in [0,1]$,*

$$y \mapsto C(x,y)/y \qquad \text{is decreasing.}$$

- ▷ *$RTI(Y|X)$ if, and only if, for every $y \in [0,1]$,*

$$x \mapsto [y - C(x,y)]/(1-x) \qquad \text{is decreasing.}$$

- ▷ *$RTI(Y|X)$ if, and only if, for every $x \in [0,1]$,*

$$y \mapsto [x - C(x,y)]/(1-y) \qquad \text{is decreasing.}$$

Definition 1.7.3. Let X and Y be random variables.

- ▷ Y is *stochastically increasing* in X (briefly, $SI(Y|X)$) if, and only if, the mapping $x \mapsto P(Y > y \mid X = x)$ is an increasing function for all y.
- ▷ X is *stochastically increasing* in Y (briefly, $SI(X|Y)$) if, and only if, the mapping $y \mapsto P(X > x \mid Y = y)$ is an increasing function for all x.
- ▷ Y is *stochastically decreasing* in X (briefly, $SD(Y|X)$) if, and only if, the mapping $x \mapsto P(Y > y \mid X = x)$ is a decreasing function for all y.
- ▷ X is *stochastically decreasing* in Y (briefly, $SD(X|Y)$) if, and only if, the mapping $y \mapsto P(X > x \mid Y = y)$ is a decreasing function for all x.

Proposition 1.7.3. *Let X and Y be continuous r.v.'s with copula C.*

- ▷ $SI(Y|X)$ *if, and only if,* $x \mapsto C(x, y)$ *is concave for every* $y \in [0, 1]$.
- ▷ $SI(X|Y)$ *if, and only if,* $y \mapsto C(x, y)$ *is concave for every* $x \in [0, 1]$.
- ▷ $SD(Y|X)$ *if, and only if,* $x \mapsto C(x, y)$ *is convex for every* $y \in [0, 1]$.
- ▷ $SD(X|Y)$ *if, and only if,* $y \mapsto C(x, y)$ *is convex for every* $x \in [0, 1]$.

Definition 1.7.4. Let X and Y be random variables

- ▷ X and Y are *left corner set decreasing* (briefly , $LCSD(X, Y)$) if, and only if, $P(X \leq x, Y \leq y \mid X \leq x', Y \leq y')$ is decreasing in x' and in y' for all x and y.
- ▷ X and Y are *left corner set increasing* (briefly , $LCSI(X, Y)$) if, and only if, $P(X \leq x, Y \leq y \mid X \leq x', Y \leq y')$ is increasing in x' and in y' for all x and y.
- ▷ X and Y are *right corner set increasing* (briefly , $RCSI(X, Y)$) if, and only if, $P(X > x, Y > y \mid X > x', Y > y')$ is increasing in x' and in y' for all x and y.
- ▷ X and Y are *right corner set decreasing* (briefly , $RCSD(X, Y)$) if, and only if, $P(X > x, Y > y \mid X > x', Y > y')$ is increasing in x' and in y' for all x and y.

Proposition 1.7.4. *Let* X *and* Y *be r.v.'s uniformly distributed on* $[0, 1]$ *with associated copula* C.

- ▷ $LCSD(X, Y)$ *if, and only if,*

$$C(x, y)C(x', y') \geq C(x, y')C(x', y)$$

  for every x, x', y, y' *in* $[0, 1]$, $x \leq x'$, $y \leq y'$.

- ▷ $RCSI(X, Y)$) *if, and only if,*

$$\widehat{C}(x, y)\widehat{C}(x', y') \geq \widehat{C}(x, y')\widehat{C}(x', y)$$

  for every x, x', y, y' *in* $[0, 1]$, $x \leq x'$, $y \leq y'$.

The scheme of implications among the various dependence concepts is presented in Table 1.2.

For the study of dependence between extreme values, the concept of *tail dependence* is useful and can be also expressed in terms of copula (see [74, 113]).

Definition 1.7.5. Let X and Y be continuous r.v.'s with d.f.'s, resp., F and G. If the following limits exist in $[0, 1]$, then the *upper tail dependence parameter* λ_U of (X, Y) is defined by

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

$$\begin{array}{ccccc}
\mathrm{SI}(Y|X) & \Longrightarrow & \mathrm{RTI}(Y|X) & \Longleftarrow & \mathrm{RCSI}(X,Y) \\
\Downarrow & & \Downarrow & & \Downarrow \\
\mathrm{LTD}(Y|X) & \Longrightarrow & \mathrm{PQD}(X,Y) & \Longleftarrow & \mathrm{RTI}(X|Y) \\
\Uparrow & & \Uparrow & & \Uparrow \\
\mathrm{LCSD}(X,Y) & \Longrightarrow & \mathrm{LTD}(X|Y) & \Longleftarrow & \mathrm{SI}(X|Y)
\end{array}$$

Table 1.2: Implications among dependence concepts

and the *lower tail dependence parameter* λ_L of (X,Y) is defined by

$$\lambda_L = \lim_{t\to 0^+} P\left(Y \le G^{[-1]}(t) \mid X \le F^{[-1]}(t)\right).$$

In particular, if $\lambda_U = 0$ (resp. $\lambda_L = 0$), then X and Y are said to be *asymptotically independent in the upper tail* (resp. *in the lower tail*).

Proposition 1.7.5. *Let X and Y be continuous r.v.'s with copula C. If the following limits exist and take values in $]0,1]$, then*

$$\lambda_L = \lim_{u\to 0^+} \frac{C(u,u)}{u} \quad \textit{and} \quad \lambda_U = \lim_{u\to 1^-} \frac{1-2u+C(u,u)}{1-u}.$$

Moreover, if δ_C is the diagonal section of C, we have:

$$\lambda_L = \delta_C'(0^+) \quad \textit{and} \quad \lambda_U = 2 - \delta_C'(1^-).$$

1.8 Measures of Association

There are a variety of ways to measure the association (or dependence) between random variables and, as noted by Hoeffding, many such descriptions are "scale invariant" ([71]), that is they remain unchanged under stricly increasing transformations of r.v.'s. But, in the words of B. Schweizer and E.F. Wolff, "*it is precisely the copula which captures those properties of the joint distribution function which are invariant under almost surely strictly increasing transformations*" ([143]). Thus, Sklar's Theorem and Theorem 1.6.3(i) suggest that copulas are a powerful tool to measure dependence.

In this section, we give a representation of some known measures of association in terms of copula; for more details, see [114, chapter 5] and [143, 74, 50].

Theorem 1.8.1. *Let X and Y be continuous r.v.'s whose copula is C. Then the population version for Kendall's tau for X and Y is given by*

$$\tau_{X,Y} := 4\int\int_{[0,1]^2} C(u,v)\,dC(u,v) - 1 = 1 - 4\int\int_{[0,1]^2} \partial_1 C(u,v)\cdot\partial_2 C(u,v)\,du dv.$$

Theorem 1.8.2. *Let X and Y be continuous r.v.'s whose copula is C. Then the population version of Spearman's rho for X and Y is given by*

$$\rho_{X,Y} := 12 \int\int_{[0,1]^2} C(u,v)\, dudv - 3.$$

Theorem 1.8.3. *Let X and Y be continuous r.v.'s whose copula is C. Then the population version of Gini's measure of association for X and Y is given by*

$$\gamma_{X,Y} := 4 \left[\int_0^1 C(u, 1-u)\, du - \int_0^1 (u - C(u,u))\, du \right].$$

Theorem 1.8.4. *Let X and Y be continuous r.v.'s whose copula is C. Then the medial correlation coefficient of X and Y (called also Blomqvist coefficient) is given by*

$$\beta_{X,Y} := 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1.$$

Theorem 1.8.5. *Let X and Y be continuous r.v.'s whose copula is C. Then the Spearman's footrule coefficient of X and Y is given by*

$$\varphi_{X,Y} := 6 \int_0^1 C(u,u)\, du - 2.$$

On the definition of such measures for non–continuous random variables, we refer to the paper [124].

1.9 Multivariate Copulas

In this section, we consider copulas in the n–dimensional case ($n \geq 3$).

Definition 1.9.1. A function $C\colon [0,1]^n \to [0,1]$ is an n–copula if, and only if, it satisfies the following conditions:

(C1') $C(\mathbf{x}) = 0$ if at least one coordinate of $\mathbf{x}$ is 0, and $C(\mathbf{x}) = x_i$ if all coordinates of $\mathbf{x}$ are 1 except at most the i–th one;

(C2') C is n–increasing.

As a consequence, every copula is increasing in each place and satisfies the 1–Lipschitz condition, viz.

$$|C(x_1, x_2, \ldots, x_n) - C(x_1', x_2', \ldots, x_n')| \leq \sum_{i=1}^{n} |x_i - x_i'|$$

for all $(x_1, x_2, \ldots, x_n)$ and $(x_1', x_2', \ldots, x_n')$ in $[0,1]^n$.

For every n–copula C, we have

$$W_n(\mathbf{x}) \leq C(\mathbf{x}) \leq M_n(\mathbf{x}) \quad \text{for all } \mathbf{x} \in [0,1]^n,$$

where

$$W_n(\mathbf{x}) := \max\left\{\sum_{i=1}^{n} x_i - n + 1, 0\right\}, \qquad M_n(\mathbf{x}) := \min\{x_1, x_2, \dots, n_n\}.$$

These bounds are the best–possible. Notice that, for $n \geq 3$, W_n is not a copula. Another important n–copula is the product

$$\Pi_n(\mathbf{x}) := \Pi_{i=1}^{n} x_i.$$

The set of all n copulas will be denoted by $\mathcal{C}_n$.

For sake of completeness, we give the analogous of Sklar's Theorem.

Theorem 1.9.1. *Let $X_1, X_2, \dots, X_n$ be r.v.'s with joint d.f. H and marginal d.f.'s $F_1, F_2, \dots, F_n$. Then there exists an n–copula C such that, for all $x \in \overline{\mathbb{R}}^n$*

$$H(\mathbf{x}) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)). \tag{1.24}$$

If $F_1, F_2, \dots, F_n$ are continuous, then C is unique; otherwise C is uniquely determined on $Ran\, F_1 \times Ran\, F_2 \times \cdots \times Ran\, F_n$.

Conversely, if C is an n–copula and $F_1, F_2, \dots, F_n$ are univariate d.f.'s, then the function H given by (1.24) is an n–d.f. with margins $F_1, F_2, \dots, F_n$.

In the case $n \geq 3$, Theorems 1.6.2 and 1.6.3 can be partially reformulated in this way:

Theorem 1.9.2. *Let $X_1, X_2, \dots, X_n$ be continuous r.v.'s with copula C.*

- ▷ *$X_1, X_2, \dots, X_n$ are independent if, and only if, $C = \Pi_n$.*
- ▷ *each of the r.v.'s $X_1, X_2, \dots, X_n$ is almost surely a strictly increasing function of any of the others if, and only if, $C = M_n$.*
- ▷ *If $\alpha_1, \alpha_2, \dots, \alpha_n$ are strictly increasing mapping, respectively, on $RanX_1$, $RanX_2$, $\dots$, $RanX_n$, then $C_{\alpha_1(X_1)\dots\alpha_n(X_n)} = C_{X_1 \dots X_n}$.*

The following result gives an important class of multivariate copulas, called *multivariate Archimedean copulas* for their analogy with the bivariate case (see [114, 112]).

Theorem 1.9.3. *Let $\varphi : [0, 1] \to [0, +\infty]$ be continuous and strictly decreasing function with $\varphi(0) = +\infty$ and $\varphi(1) = 0$. Let C be the function defined by*

$$C_\varphi(\mathbf{x}) := \varphi^{-1}\left(\varphi(x_1) + \varphi(x_2) + \cdots + \varphi(x_n)\right).$$

If, for all $t \in \,]0, +\infty[$ and $k \in \mathbb{N} \cup \{0\}$

$$(-1)^k \frac{d^k}{dt^k}(\varphi^{-1}(t)) \geq 0,$$

then C_φ is an n–copula, called Archimedean copula.

1.10 Quasi-copulas

Quasi–copulas were introduced by Alsina, Nelsen and Schweizer ([4]) in order to characterize operations on distribution functions that can, or cannot, be derived from operations on random variables (see [122] and [116]). The concept of quasi–copula, which will be defined shortly, is closely connected to that of copula.

Definition 1.10.1. An *n–track* is any subset B of $[0,1]^n$ that can be written in the form

$$B = \{(F_1(t), F_2(t), \dots, F_n(t)) : t \in [0,1]\},$$

where $F_1, F_2, \dots, F_n$ are some continuous and increasing functions such that $F_i(0) = 0$ and $F_i(1) = 1$ for $i = 1, 2, \dots, n$.

Definition 1.10.2. An *n–quasi–copula* is a function $Q : [0,1]^n \to [0,1]$ such that for any n–track B there exists an n–copula C_B that coincides with Q on B, namely, for all $\mathbf{x} \in B$, $Q(\mathbf{x}) = C_B(\mathbf{x})$.

Such a definition of quasi–copula is, however, of little practical use because it is hard to tell whether a function $Q : [0,1]^n \to [0,1]$ is, or is not, a quasi–copula according to it. In view of this purpose, quasi–copulas were characterized in a different way: see [64] for the bivariate case and [21] for the multivariate case.

Theorem 1.10.1. *A function $Q : [0,1]^n \to [0,1]$ is an n–quasi–copula if, and only if, it satisfies the following conditions:*

(Q1) $Q(\mathbf{x}) = x_i$ *if all coordinates of* $\mathbf{x}$ *are* 1 *except at most the i–th one;*

(Q2) Q *is increasing in each variable;*

(Q3) Q *satisfies the* 1*–Lipschitz condition, viz.*

$$|Q(x_1, x_2, \dots, x_n) - Q(x'_1, x'_2, \dots, x'_n)| \le \sum_{i=1}^{n} |x_i - x'_i|$$

for all $(x_1, x_2, \dots, x_n)$ *and* $(x'_1, x'_2, \dots, x'_n)$ *in* $[0,1]^n$.

The set of all n–quasi–copulas will be denoted by $\mathcal{Q}_n$. Since an n–copula is obviously also an n–quasi–copula, the set $\mathcal{C}_n$ of all n–copulas is (strictly) included in $\mathcal{Q}_n$. If Q belongs to $\mathcal{Q}_n \setminus \mathcal{C}_n$, then we say that it is a *proper* n–quasi–copula.

For every n–quasi–copula Q, we have

$$W_n(\mathbf{x}) \le C(\mathbf{x}) \le M_n(\mathbf{x}) \quad \text{for all } \mathbf{x} \in [0,1]^n,$$

and W_n is a quasi–copula.

The concept of quasi–copulas has important applications on finding of best–possible bounds on arbitrary sets of d.f.'s (see [121, 131]). In particular, if we restrict to the bivariate case, we have:

Theorem 1.10.2 ([123]). *A function* $Q : [0,1]^2 \to [0,1]$ *is a quasi–copula if, and only if, there exists a nonempty set* $\mathcal{B}$ *of copulas such that, for every* $(x, y) \in [0,1]^2$, $Q(x, y) := \sup\{C(x, y) : C \in \mathcal{B}\}$.

1.11 Aggregation operators

The aggregation of several input values into a single output is an indispensable tool not only in mathematics, but also in any other disciplines where the fusion of different pieces of information is of vital interest (see [12]). In a very intuitive sense, an *aggregation operator* associates a single value to a list of values, where a value is simply an element of a given class (e.g., numbers, functions, sets, etc.). Therefore, from a mathematical point of view, an aggregation operator is simply a function that, *a priori*, has a varying number of variables. Here, following [10], we restrict ourselves to aggregations of a finite number of input values that belong to the unit interval $[0, 1]$ into an output value belonging to the same interval and we consider aggregation operators according to the following

Definition 1.11.1. Let $n \in \mathbb{N}$, $n \geq 2$. An *n–ary aggregation operator* (briefly, *n–agop*) is a function $A : [0,1]^n \to [0,1]$ satisfying

(A1) $A(0, 0, \ldots, 0) = 0$ and $A(1, 1, \ldots, 1) = 1$;

(A2) A is increasing in each variable.

We note that the above conditions seem quite natural with respect to the intuitive idea of aggregation: (A1) states that if we have only minimal (respectively, maximal) possible inputs, then we should obtain the minimal (respectively, maximal) possible output; (A2) ensures that the aggregation preserves the cartesian ordering on the inputs. The assumptions that inputs and outputs belong to $[0, 1]$ is not restrictive: in fact, if they belong to some interval $[a, b] \subset \overline{\mathbb{R}}$, it is always possible to re–scale them on $[0, 1]$.

Definition 1.11.2. A (*global*) *aggregation operator* is a family $\mathbf{A} = \{A_{(n)}\}_{n\in\mathbb{N}}$ of n–agops, with the convention that $\mathrm{id}_{[0,1]}$ is the only 1–agop.

Such a definition of global aggregation operator is very useful because, in general, the number of input values to be aggregated is not known. Notice that, given a global aggregation operator A, $A_{(n)}$ and $A_{(m)}$ need not be related for $n \neq m$.

Remark 1.11.1. In 2005, during the Summer School on Aggregation Operators, E.P. Klement suggested to use the term “aggregation function” instead of “aggregation operator”, when we aggregate real numbers and not complex quantities. We agree with this point of view, but it is not adopted here for the sake of uniformity with the literature of this field.

As it is easily seen, copulas and quasi–copulas are special types of n–agops. In particular, they are in the class of 1–stable n–agops, as stated in the following

Definition 1.11.3. Let $n \in \mathbb{N}$, $n \geq 2$ and $p \in [1, +\infty]$. An n–agop A is *p–stable* if, for all $\mathbf{x}$ and $\mathbf{y}$ in $[0,1]^n$

$$|A(\mathbf{x}) - A(\mathbf{y})| \leq \|x - y\|_p, \tag{1.25}$$

where $\|\cdot\|_p$ is the standard L^p norm on $\mathbb{R}^n$.

The class of p–stable aggregation operators was introduced in [11] for controlling output errors in aggregation processes. In particular, a 1–stable 2–agop A, also called 1–*Lipschitz* 2–*agop* ([90]), satisfies

$$|A(x,y) - A(x',y')| \leq |x - x'| + |y - y'|, \quad \text{for every } x, x'y, y' \in [0,1];$$

and a ∞–stable 2–agop A, also called *kernel* 2–*agop* ([93]), satisfies

$$|A(x,y) - A(x',y')| \leq \max\{|x - x'|, |y - y'|\}, \quad \text{for every } x, x'y, y' \in [0,1].$$

In the sequel, if no confusion arises, we use the term agop to denote simply a binary aggregation operators.

For every agop $A : [0,1]^2 \to [0,1]$, we have

$$A_S(x,y) \leq A(x,y) \leq A_G(x,y) \qquad \text{for every } (x,y) \in [0,1]^2,$$

where

$$A_S(x,y) = \begin{cases} 1, & \text{if } (x,y) = (1,1); \\ 0, & \text{otherwise}; \end{cases} \qquad A_G(x,y) = \begin{cases} 0, & \text{if } (x,y) = (0,0); \\ 1, & \text{otherwise}; \end{cases}$$

are called, respectively, *the smallest* and *the greatest* agop .

Given an A, the *dual* of A is defined, for every point (x,y) in $[0,1]^2$, by $A^d(x,y) := 1 - A(1-x, 1-y)$.