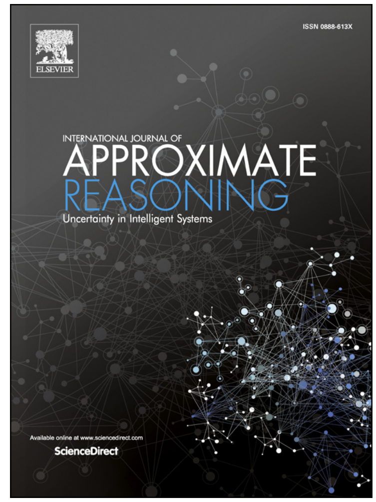
Journal Pre-proof



Stochastic dominance and statistical preference for random variables coupled by arbitrary copulas

Inés Couso and Luciano Sánchez

PII: S0888-613X(23)00244-X

DOI: https://doi.org/10.1016/j.ijar.2023.109113

Reference: IJA 109113

To appear in: *International Journal of Approximate Reasoning*

Received date: 19 July 2023

Revised date: 28 December 2023

Accepted date: 30 December 2023

Stochastic dominance and statistical preference for random variables coupled by arbitrary copulas

Inés Couso[a], Luciano Sánchez[b]

[a]Dep. of Statistics and Operational Research, University of Oviedo, Spain
[b]Dep. of Computer Science and Artificial Intelligence, University of Oviedo, Spain

Abstract

Recently, results have been published showing that first order stochastic dominance implies statistical preference and diff-stochastic dominance, when the copula relating the compared variables is either Archimedean, the product copula, or one of the Fréchet-Hoeffding bounds.

In the present paper, we rely on known results on multivariate stochastic orders to extend these results and simplify the proofs. The results are expanded in two directions: First, we show that it suffices for the copula to be symmetric. Second, we reveal that first stochastic dominance entails a wider range of stochastic preferences beyond statistical preference and diff-stochastic dominance.

We further analyze whether first stochastic dominance implies statistical preference for the case of asymmetric copulas. We observe that, when at least one of the marginal cumulative distribution functions has no discontinuity jumps, the family of asymmetric copulas for which the implication holds is at least as large as the one for which it does not.

Keywords: Decision under uncertainty, (Multivariate) stochastic ordering, First stochastic dominance, Statistical preference, Copula

1. Introduction

In the context of decision making under uncertainty, different binary relations between random variables have been introduced under terms such as "order", "preference" or "stochastic dominance". Some of these relationships can be expressed in terms of the marginal distributions of the random variables involved, so that they are binary relations between one-dimensional probability distributions, while others involve the joint distribution. Examples of the first case are "n-order stochastic dominance", for any $n \geq 1$, among others. Examples of the second are Condorcet preference [1] (also referred to as "statistical preference" [2] or "sign-preference" [3]) and "stochastic precedence" [4, 5, 6, 7].

Studies have examined the comparative strength between first order stochastic dominance and statistical preference, and the results suggest that the former is not necessarily stronger than the latter, as evidenced by De Santis et al. ([7]) and Montes and Montes ([8]) counterexamples.

However, under certain conditions concerning the copula that connects both marginal distributions, it has been established that first-order stochastic dominance is stronger than statistical preference. Thus De Schuymer et al. [9] showed that this implication holds when the two compared variables are independent. Later De Meyer et al. [10] showed that it also holds when they are coupled by the comotonicity copula and the marginals are Gaussian.

More recently, Montes and Montes [8] proved that it also holds when the copula connecting the two variables is either the Fréchet-Hoeffding upper bound or an Archimedean copula, and some additional conditions concerning the marginal distributions are fulfilled. In a later paper Montes et al. [11] introduced a new notion of preference called "diff-stochastic dominance", which is stronger that statistical preference. They showed that, for the same copulas, as well as for the Fréchet-Hoeffding lower bound, first-order stochastic dominance is also stronger than "diff-stochastic dominance". Very recently, Belzunce and Martínez-Riquelme [12], extended one of the results in [11] to a type of copulas that contains, as a particular case, the Archimedean copulas.

In this paper, we demonstrate that first stochastic dominance proves to be stronger than diff-stochastic dominance for any symmetric copula, with no additional requirements regarding the marginal distributions. Moreover, we generalize these results to encompass a broader family of preference relations, which includes diff-stochastic dominance (and consequently statistical preference) as specific instances. This broader family of preference relations was already considered by Shanthikumar and Yao [13] and bears similarities to SSD model [14, 15] and has been recently used in the context of certain machine learning tasks (intelligent condition monitoring of engines) [16]. Based on existing results on multivariate stochastic orders, the proof is significantly shorter and simpler than those offered in [11], [12] and [8].

We also investigate the conditions that non-symmetric copulas must meet for stochastic first dominance to imply statistical preference. Our study unveils that when we consider pairs of marginals, with at least one of them having a continuous cumulative distribution function (CDF), the set of asymmetric copulas that demonstrate this implication is at least as extensive as the set that does not.

The rest of the paper is divided into three sections: Section 2 shows some preliminary definitions and results on stochastic preferences and copulas. Section 3 studies the relationships between first-order stochastic dominance and other preference relationships based on the joint distribution for the case of symmetric copulas. Section 4 focuses on the study of the relationships between first-order stochastic dominance and statistical preference for the case of non-symmetric copulas.

2. Preliminaries

Throughout the paper, we will often refer to bivariate random vectors $(X, Y) : \Omega \to \mathbb{R}^2$ defined on a probability space $(\Omega, \mathcal{A}, P)$. We denote the respective (marginal) cumulative distribution functions (CDFs) of X and Y as F_X and F_Y, and their joint cumulative distribution function as $F_{(X,Y)}$. The induced joint probability measure will be denoted as $P_{(X,Y)}$.

2.1. Copulas

An $n-$dimensional copula is a multivariate cumulative distribution function over $[0, 1]^n$ with uniform marginals. Copulas are useful for capturing the dependence structure between random variables and constructing multivariate distributions from their marginal distributions ([17, 18, 19]). In this paper, we will restrict ourselves to the bivariate case.

Definition 1. *([20]) A two-dimensional copula is a mapping* $C : [0, 1]^2 \to [0, 1]$*, satisfying the following conditions*

(i) $C(u_1, v_1) + C(u_2, v_2) \geq C(u_1, v_2) + C(u_2, v_1)$ *for all* $(u_1, v_1), (u_2, v_2) \in [0, 1]^2$ *such that* $u_1 \leq u_2, v_1 \leq v_2$.

(ii) $C(u, 0) = C(0, u) = 0$ *and* $C(u, 1) = C(1, u) = u$ *for any* $u \in [0, 1]$.

(As a consequence, C *is increasing in both components.)*

The *transpose* C^T of a copula $C : [0,1]^2 \to [0,1]$ is the copula defined as $C^T(u,v) = C(v,u),\ \forall\,(u,v) \in [0,1]^2$. A copula $C : [0,1]^2 \to [0,1]$ is said to be *symmetric* if it coincides with its transpose. If a copula associated to the random vector (X,Y) is symmetric then the random vector (Y,X) is associated to the same copula.

Some special types of symmetric copulas are the product copula, the Fréchet-Hoeffding upper and lower bounds and the family of Archimedean copulas. The interested reader is referred to [21] for further details.

Any multivariate (in our case, bivariate) cumulative distribution function can be obtained by combining a copula with the corresponding marginal CDFs. Formally, we have:

Theorem 1. *([20], Sklar Theorem) Given a bivariate random vector (X,Y) there exists a copula C such that $F_{(X,Y)}(x,y) = C(F_X(x), F_Y(y)),\ \forall\,(x,y)$.*

2.2. Stochastic preferences and orders

Some well-known stochastic orders and preferences include:

- First stochastic dominance ([22]): we say that $X \leq_{FSD} Y$ if $P(X > t) \leq P(Y > t),\ \forall\, t \in \mathbb{R}$ or equivalently if $F_X \geq F_Y$.
- Statistical preference ([1, 2]). $X \leq_{sp} Y$ if $P(X > Y) \leq P(Y > X)$.
- Diff-stochastic dominance ([11]) or weak joint stochastic dominance ([12]). $X \leq_{\text{diff}} Y$ if $X - Y \leq_{FSD} Y - X$. This relation is clearly stronger than statistical preference.

We will refer to a family that contains the three above stochastic preferences as particular cases. It is based on the following family of mappings already considered by Shanthikumar and Yao ([13]):

$$\mathcal{G}^2_{st} = \{f : \mathbb{R}^2 \to \mathbb{R}\ :\ f(x_1, y_2) \leq f(x_2, y_1),\ \forall\, x_1 \leq x_2, y_1 \leq y_2\}.$$

From now on, we will use the term $\mathcal{G}$-expectation dominance, to refer to the stochastic preferences generated by subsets of that family, as shown below:

Definition 2. *Consider a subfamily $\mathcal{G}$ of $\mathcal{G}^2_{st}$. X is said to be $\mathcal{G}-$*expectation dominated *by Y if*

$$E[f(X,Y)] \leq E[f(Y,X)],\ \forall\, f \in \mathcal{G},$$

$$\textit{or equivalently if } E_{P_{(X,Y)}}[f] \leq E_{P_{(Y,X)}}[f],\ \forall f \in \mathcal{G},$$

for any f for which the corresponding expectations exist.

In the particular case where the subfamily is a singleton $\mathcal{G} = \{f\}$, we will simply refer to f-expectation dominance. This preference relation will be denoted $X \leq_{\mathcal{G}} Y$.

This formulation is reminiscent of the skew-symmetric bilinear (SSB) model of Kreweras ([14]) and Fishburn ([15]), who consider a preference relation between probability distributions, that can be reformulated as follows in terms of random variables:

$$X \leq_{\mathcal{G}}^{F} Y \text{ if } E_{P_X \otimes P_Y}[f] \leq E_{P_Y \otimes P_X}[f],\ \forall f \in \mathcal{G}^2_{st}, \tag{1}$$

where $P_X \otimes P_Y$ denotes the product probability[1].

Since the set of joint probability distributions over $\mathbb{R}^2$ strictly contains the set of product distributions, Equation 1 determines a (strict) subset of the family of preferences characterized by Definition 2. Fishburn ([15]) characterized the set of rules of reasoning (also called "coherence rules" by some authors) that must be satisfied when ordering alternatives so that the ranking can be expressed in terms of Equation 1. These reasoning rules constitute a subset of the list of conditions proposed by Savage in his subjective expected utility (SEU) model ([23]). One of the conditions that is eliminated from this SEU model is the transitivity property, already questioned earlier by Kreweras ([24]), and later discussed by Fishburn ([15, 25]). Fishburn formulation also bears some relation to Loomes-Sugden's "regret theory" ([26]).

Example 1. *We easily observe that the three examples of stochastic preferences and orders mentioned at the beginning of Subsection 2.2 are part of the family of $\mathcal{G}$-expectation dominance relations. Indeed:*

- *First stochastic dominance coincides with $\mathcal{G}_{FSD}$-expectation dominance for $\mathcal{G}_{FSD} = \{f_c : c \in \mathbb{R}\}$ with $f_c(x,y) = 1_{\{(x,y)\in\mathbb{R}^2\,:\,x>c\}}$, for all $c \in \mathbb{R}$, (where 1_A denotes the indicator function of $A \subseteq \mathbb{R}^2$).*

[1] I.e. the one whose marginals are the probability distributions respectively induced by X and Y, and its copula is the product copula.

- *Statistical preference coincides with f_{sp}-expectation dominance for $f_{sp}(x,y) = 1_{\{(x,y)\in\mathbb{R}^2\,:\,x>y\}}$.*

- *Diff-stochastic dominance coincides with $\mathcal{G}_{diff}$-expectation dominance for $\mathcal{G}_{diff} = \{1_{\{(x,y)\in\mathbb{R}^2\,:\,x-y>c\}}\ :\ c\in\mathbb{R}\}$.*

Remark 1. *Some additional stochastic orders and preferences from recent literature not named in Example 1 such as second-order dominance [27, 28] or Savage dominance [23] are special cases of $\mathcal{G}$-expectation dominance. Further well known preference relations also fit the formulation:*

$$Y \text{ is preferred to } X \text{ if } E[f(X,Y)] \leq E[f(Y,X)],\ \forall f \in \mathcal{F},$$

but for some family of functions $\mathcal{F}$ not necessarily included in $\mathcal{G}^2_{st}$. This is the case, for instance, of "less dangerous than" order ([29]), "tail dispersive ordering" ([30]) or "bidirectional ordering" ([31]). These two examples do not fit the idea of "taking larger values" but with other ideas such as taking values farther away from some point (like the value 0) or simply taking more dispersed values. All these examples could be alternatively expressed in terms of a non-necessarily increasing function $h:\mathbb{R}\to\mathbb{R}$ as "$h(Y)$ $\mathcal{F}'$-expectation dominates $h(X)$", for some $\mathcal{F}'\subseteq\mathcal{G}^2_{st}$.

3. FSD and $\mathcal{G}^2_{st}$-expectation preference for symmetric copulas

In the introduction, we referred to the existence of results in the literature that formally connect the concepts of first stochastic dominance, statistical preference, and diff-stochastic dominance. As a compilation of the most recent studies on this matter, the following two results are cited:

Theorem 2. *([8, Theorem 45]) Let X and Y be two random variables. $X \leq_{FSD} Y$ implies $X \leq_{sp} Y$ under any of the following conditions:*

1. *X and Y are independent.*
2. *X and Y are absolutely continuous and comonotone.*
3. *X and Y are absolutely continuous and countermonotone.*
4. *X and Y are discrete with finite supports and comonotone.*
5. *X and Y are discrete with finite supports and countermonotone.*
6. *X and Y are absolutely continuous random variables coupled by a strict Archimedean copula with a twice differentiable generator.*

7. *X and Y are absolutely continuous random variables coupled by a nilpotent Archimedean copula with a twice differentiable generator.*

Theorem 3. *([11, Theorem 15]) $X \leq_{FSD} Y$ implies $X \leq_{\text{diff}} Y$ whenever X and Y satisfy any of the following properties:*

1. *X and Y are independent.*
2. *X and Y are comonotone or countermonotone.*
3. *X and Y are coupled by an Archimedean copula with a twice differentiable generator.*

Now, we will show that existing results in the literature on multivariate stochastic orders allow us to extend all these findings, and to considerably shorten the proofs. On one hand, we will relax the conditions required to the copula, as we only need it to be symmetric. Furthermore we will replace diff-stochastic dominance by $\mathcal{G}^2_{st}$-expectation dominance.

Concretely, the following equivalence holds when the marginals are connected by a symmetric copula:

$$X \leq_{FSD} Y \text{ if and only if } X \leq_{\mathcal{G}^2_{st}} Y.$$

Let the reader notice that the necessity is immediate, since $\mathcal{G}_{\text{FSD}}$ is included in $\mathcal{G}^2_{st}$, as showed in Example 1. Theorem 4 deals with the proof of the sufficiency. We will first remind the definition of usual stochastic order between random vectors, which relies on the use of "upper sets".

Definition 3. *([32]) A set $U \subseteq \mathbb{R}^d$ is called an* upper set *if, for any pair of vectors $(u_1, \ldots, u_d) \in U$ and $(v_1, \ldots, v_d) \in \mathbb{R}^d$ satisfying the inequalities $v_i \geq u_i,\ \forall\, i = 1, \ldots, d$, $(v_1, \ldots, v_d) \in U$.*

Definition 4. *([32]) Consider two d-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$. $\mathbf{X}$ is said to be smaller than $\mathbf{Y}$ in the usual stochastic (multivariate) order if $P(\mathbf{X} \in U) \leq P(\mathbf{Y} \in U)$, for every measurable upper set $U \subseteq \mathbb{R}^d$ It will be denoted by $\mathbf{X} \leq_{st} \mathbf{Y}$.*

Theorem 4. *If $X \leq_{FSD} Y$, and they are connected by at least one symmetric copula, then $X \leq_{\mathcal{G}^2_{st}} Y$.*

Proof: *First, we recall that, for every copula C associated with (X, Y), the copula $C'(u, v) = u - C(u, 1 - v)$ is associated with $(X, -Y)$ (see [33,* *Theorem 2.4.4]). Moreover, since C is symmetric, it is also a copula for (Y, X). Thus, C' is a copula for $(Y, -X)$ too.*

Now, if $X \leq_{FSD} Y$, then $-Y \leq_{FSD} -X$ (see [32, Theorem 1.A.3]). Thus, it follows from [34, Theorem 4.1], that $(X, -Y) \leq_{st} (Y, -X)$ since $(X, -Y)$ and $(Y, -X)$ have at least one copula in common. But $(X, -Y) \leq_{st} (Y, -X)$ is equivalent to $X \leq_{\mathcal{G}^2_{st}} Y$ in view of [13, Theorem 4.8], which is the desired assertion.

Since the set of functions $\mathcal{G}^2_{st}$ contains the set $\mathcal{G}_{\text{diff}}$, which in turn contains the singleton $\{1_{\{(x,y)\in\mathbb{R}^2 \,:\, x>y\}}\}$, we can state the following:

Corollary 1. *If $X \leq_{FSD} Y$, and they are connected by at least one symmetric copula, then $X \leq_{\text{diff}} Y$ and $X \leq_{SD} Y$.*

In other words, Corollary 1 (and thus also Theorem 4) generalize the entirety of the results stated in [11] and [8], which have been encompassed in the above Theorems 2 and 3.

On the other hand, Belzunce and Martínez-Riquelme [12] have very recently extended the third part of Theorem 3 ([11, Theorem 15]) to a more general case of copulas, as follows:

Theorem 5. *([12, Theorem 2.9]) Let (X, Y) be a bivariate random vector with absolutely continuous distribution and copula C. If $X \leq_{FSD} Y$ and*

$$\frac{\partial}{\partial p} C(u, v_1) \leq \frac{\partial}{\partial q} C(v_2, u), \ \forall\, u, v_1, v_2 \in (0, 1) \text{ such that } v_1 \leq v_2, \tag{2}$$

then $X \leq_{\text{diff}} Y$.

Remark 2. *Given a copula C, consider the family of mappings $f_{C;\,(v_1,v_2)} : \mathbb{R} \to [-1, 1]$ defined as follows:*

$$f_{C;\,(v_1,v_2)}(u) = C(u, v_1) - C(v_2, u), \ \forall\, u \in (0, 1), \ \forall\, 0 \leq v_1 \leq v_2 \leq 1. \tag{3}$$

Using this notation, Equation 2 can be formally rewritten as follows:

$$f'_{C;\,(v_1,v_2)}(u) \leq 0, \ \forall\, u \in (0, 1), \forall\, 0 < v_1 \leq v_2 < 1. \tag{4}$$

We show below that the requirement of Equation 2 (i.e. the requirement of Equation 4) implies the symmetry of the copula:

Proposition 1. *Consider a copula* $C : [0,1]^2 \to [0,1]$ *and the family of mappings defined from it in Equation 3. If they satisfy Equation 4, then* C *is symmetric.*

Proof: *Taking into account the properties of copulas, we can easily check that* $f_{C;(v_1,v_1)}(0) = f_{C;(v_1,v_1)}(1) = 0$ *for every* $v_1 \in [0,1]$, $f_{C;(0,0)}(u) = f_{C;(1,1)}(u) = 0, \forall\, u \in [0,1]$, *and* $f_{C;(v_1,v_1)}$ *is continuous in* $u = 0$. *Furthermore, if Equation 4 holds,* $f_{C;(v_1,v_1)}$ *is monotone decreasing (non increasing) in* $(0,1)$ *for every* $v_1 \in (0,1)$. *Thus,* $f_{C;(v_1,v_1)}(u) = C(u,v_1) - C(v_1,u) = 0$ *for every* $u \in [0,1]$ *and every* $v_1 \in [0,1]$, *which means that the copula is symmetric.* ■

Thus, according to Proposition 1, the copulas considered by Belzunce and Martínez-Riquelme in [12, Theorem 2.9] are particular instances of symmetric copulas, and therefore Corollary 1 also generalizes their result.

Remark 3. *We must emphasize that, if the copula is not symmetric, the implication mentioned in Theorem 4 and, in particular, the two implications of Corollary 1 are not fulfilled in general. The interested reader may note that, indeed, the examples of Montes and Montes [8] and De Santis et al. ([7]) in which pairs of variables are shown such that* $X \leq_{FSD} Y$ *and yet* $X \not\leq_{sp} Y$ *correspond to cases where* X *and* Y *are not connected by a symmetric copula.*

4. FSD and statistical preference for asymmetric copulas

Once we have seen the relationship between first-stochastic order dominance and statistical preference for the case in which the copula connecting X and Y is symmetric, it is worth asking whether any formal relation can be found in other cases. The work of De Santis et al. [7] makes some progress in that direction.

In the following, given two CDFs $F, G : \mathbb{R} \to [0,1]$ and a copula $C : [0,1]^2 \to [0,1]$, the notation $(X,Y) \equiv_d (C; F, G)$ will be used to indicate that the joint CDF associated with the random bivariate vector (X,Y) is:

$$F_{(X,Y)}(x,y) = C(F(x), G(y)),\ \forall\,(x,y) \in \mathbb{R}^2.$$

In other words, this notation means that the marginal CDFs of (X,Y) are respectively F and G and they are connected by copula C.

Given an arbitrary copula $C : [0,1]^2 \to [0,1]$ and a pair of CDFs F and G, Santis et al. [7] introduce the notation $\eta(C; F, G)$ to refer to the following quantity

$$\eta(C; F, G) = P(X \leq Y),$$

for any $(X, Y) \equiv_d (C; F, G)$, and they prove the following equalities:

$$\eta(C; F, F) = \eta(C; G, G),\ \forall\, F, G \in \mathcal{H}, \forall\, C \in \mathcal{C},$$

where $\mathcal{H}$ denotes the family of CDFs that are continuous and strictly increasing in the subset of the domain where they take values in $(0,1)$. Thus, they can univocally define the quantity $\eta(C) = \eta(C; F, F)$, where F is an arbitrary CDF in $\mathcal{H}$. In particular, $\eta(C)$ can be expressed as $P(U_1 \leq U_2)$, for any pair of variables U_1 and U_2, both with uniform marginal distribution on $[0,1]$ and connected by the copula C. In other words, $\eta(C) = P(U_1 \leq U_2)$ for any bivariate random vector (U_1, U_2) whose joint cumulative distribution function coincides with C on $[0,1]^2$. The authors prove that the following equality holds for any copula C:

$$\eta(C) = \inf\{P(X \leq Y)\ :\ X \leq_{FSD} Y,\ X \text{ and } Y \text{ are coupled by } C\}.$$

The authors conclude that, if $(X, Y) \equiv_d (C; F, G)$ with $F \geq G$ and $\eta(C) \geq \gamma$, then $P(X \leq Y) \geq \gamma$. In particular, the following implication can be derived from their results:

$$\text{If } X \leq_{FSD} Y \text{ and } \eta(C) \geq 0.5, \text{ then } P(X \leq Y) \geq 0.5.$$

However, as they remark, we cannot infer that Y is statistical preferred to X, as the condition $P(X \leq Y) \geq 0.5$ does not imply in general the inequality $P(X < Y) \geq P(X > Y)$.

Notwithstanding, and with these precedents in mind, we will investigate the formal relationship between stochastic dominance and statistical preference for non necessarily symmetric copulas. In particular, we will focus on the copulas that satisfy the following property:

Definition 5. *A copula C is said to be FSD-sp consistent if*

$$X \leq_{FSD} Y \Rightarrow X \leq_{sp} Y,\ \forall\, (X, Y) \in \mathcal{X}_C,$$

where $\mathcal{X}_C$ denotes the following family of bivariate random vectors:

$$\mathcal{X}_C = \{(X,Y) \ : \ F_X \text{ or } F_Y \text{ is continuous on } \mathbb{R} \text{ and } F_{(X,Y)}(\cdot,\cdot) = C(F_X(\cdot), F_Y(\cdot))\}. \quad (5)$$

In our analysis, in addition to $\eta(C)$, we need to consider the quantities $\nu(C)$ and $I(C)$ defined as follows:

$$\nu(C) := P(U_1 < U_2), \quad I(C) = P(U_1 = U_2),$$

where (U_1, U_2) represents any bivariate vector whose CDF coincides with C on $[0,1]^2$. Furthermore, we will denote by $\mathcal{C}_D$ the set of those copulas C satisfying the condition $I(C) = 0$.

Now we will prove a result connecting the above values for any copula and its transpose. This result will serve as a support in a later discussion.

Lemma 6. *Consider a copula $C : [0,1]^2 \to [0,1]$ Then:*

(a) $\nu(C) + I(C) = \eta(C)$

(b) $\mu(C) = \eta(C)$ *if and only if* $I(C) = 0$.

(c) $I(C) = I(C^T)$.

(d) $\nu(C) + \nu(C^T) = 1 - I(C)$

(e) $\nu(C^T) = 1 - \nu(C)$ *if and only if* $C \in \mathcal{C}_D$.

Proof: *Consider a bivariate random vector (U_1, U_2) whose CDF is C. The proof of the above results is immediate if we just take into account that C coincides the joint CDF associated with (U_1, U_2) and C^T coincides the joint CDF associated with the (U_2, U_1) on $[0,1]^2$.* ■

Lemma 7. *[35, Prop 1, Section 7] If a cumulative distribution function $F : \mathbb{R} \to [0,1]$ is continuous on $\mathbb{R}$ then the generalized inverse defined on $[0,1]$ as $F^{-1}(p) = \inf\{x \ : \ F(x) \geq p\}$ is strictly increasing.*

Lemma 8. *Consider an arbitrary copula C and two CDFs $F, G : \mathbb{R} \to [0,1]$, and suppose that at least one of them is a continuous function on $\mathbb{R}$. Assume* *that $F \geq G$ (in other words, F is stochastically dominated by G). Consider a bivariate vector $(X, Y) \equiv_d (C; F, G)$ and suppose that $X \leq_{FSD} Y$. Then:*

$$P(X < Y) \geq \nu(C) \text{ and } P(X > Y) \leq \nu(C^T).$$

Proof: *We will assume that F is a continuous function (an analogous proof will serve for the case where G is continuous). Consider a bivariate random vector (U_1, U_2) with joint CDF equal to C on $[0,1]^2$, and define the random variables $X^* = F^{-1} \circ U_1$, $(X')^* = F^{-1} \circ U_2$ and $Y^* = G^{-1} \circ U_2$. We see that $(X^*, (X')^*) \equiv_d (C; F, F)$, $(X^*, Y^*) \equiv_d (C; F, G)$ and $(X')^* \leq Y^*$ (a.s.) Thus we observe that:*

$$P(X < Y) = P(X^* < Y^*) \geq P(X^* < (X')^*) \text{ and}$$

$$P(X > Y) = P(X^* > Y^*) \leq P(X^* > (X')^*).$$

Furthermore, as F is continuous, its generalized inverse F^{-1} is strictly increasing (see Lemma 7). Therefore, as $(X^*, (X')^*) = (F^{-1} \circ U_1, F^{-1} \circ U_2)$, $P(X^* < (X')^*) = P(U_1 < U_2)$ and $P(X^* > (X')^*) = P(U_1 > U_2)$. If we join all the above inequalities we arrive at the thesis of the lemma:

$$P(X < Y) \geq P(U_1 < U_2) = \nu(C) \text{ and } P(X > Y) \leq P(U_1 > U_2) = \nu(C^T).$$

If G is continuous, we can consider $(Y')^* = G^{-1} \circ U_1$ and proceed with an analogous proof. ■

Concerning the relationship with the results presented by Santis et al. ([7]), in which the authors characterize "γ-stochastic precedence" ($P(Y \geq X) \geq \gamma$), we can highlight the following points. Lower bounding $P(Y \geq X)$ does not imply lower bounding $P(Y > X)$ and consequently, we cannot replicate the argument employed in the proof of [7, Theorem 6], where they consider the closure of the set $A = \{(x, y) : x \leq y\}$, in order to lower bound its probability. A different line of reasoning is therefore necessary in our proof for Lemma 8, where we find a lower bound for the probability of the (open) set $A' = \{(x, y) : x < y\}$. In this context, we need to require the continuity of at least one marginal CDF to ensure the strict monotonicity of the corresponding generalized-inverse. Furthermore, as we do not impose the condition $P(X = Y) = 0$, Lemma 8 does not follow from the results established in [7]. Lemma 8 is instrumental for Corollary 2 and Theorem 9.

Corollary 2. *Consider an arbitrary copula C. Then:*

(a) C is FSD-sp consistent if and only if $\nu(C) \geq \nu(C^T)$.

(b) C is FSD-sp consistent if and only if $\nu(C) \geq \frac{1-I(C)}{2}$.

(c) C is FSD-sp consistent if and only if $\nu(C^T) \leq \frac{1-I(C)}{2}$.

(d) If C is symmetric then $\nu(C) = \frac{1-I(C)}{2} = \nu(C^T)$.

From Corollary 2 and Lemma 6, we can derive the following result:

Theorem 9. *Consider an arbitrary copula C. If C is not FSD-sp consistent, then C^T is.*

In other words, if we restrict ourselves to the set $\mathcal{X}_C$ of bivariate random vectors (X, Y) considered in Equation 5, one of the following situations may occur:

- $C = C^T$ (i.e., C is symmetric), and then $\nu(C) = \nu(C^T)$, so that C is FSD-sp consistent (something we already knew by virtue of Corollary 1 proved in the previous section).

- $C \neq C^T$. In that case, either $\nu(C) \geq \nu(C^T)$, and then the copula is FSD-sp consistent, or, $\nu(C) < \nu(C^T)$ and then the transpose copula is FSD-sp consistent.

In short, the family of consistent asymmetric copulas is at least as large as the family of non-consistent ones. Moreover, it can be verified that the family of nonconsistent copulas is nonempty. In fact, Santis et al. [7] recall an indexed family of copulas such that $\eta(C_\theta) = \theta,\ \forall, \theta \in (0, 1]$. It is the family of copulas, $\mathcal{N} = \{C_\theta\ :\ \theta \in [0, 1]\}$ defined by Nelsen [21] and Siburg & P. Stoimenov [36] as follows:

$$C_\theta(u, v) = \min\{u, v, (u - \theta)^+ + (v - (1 - \theta))^+\}, \forall\, (u, v) \in [0, 1]^2,$$

where $x^+ = \max\{x, 0\}$, $\forall\, x \in \mathbb{R}$.

According to [21, 36], C_θ is singular, with probability mass θ uniformly distributed on the line segment from point $(0, 1-\theta)$ to $(\theta, 1)$, and mass $1-\theta$

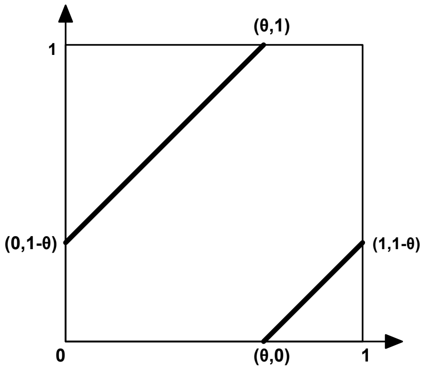


Figure 1: The support of C_θ.

uniformly distributed on the line segment joining $(\theta, 0)$ to $(1, 1-\theta)$. See Figure 1 for further clarification.

According to the location of these two segments with respect to the sets $A = \{(u,v) \ : \ u \leq v\}$ and $A^T = \{(u,v) \ : \ u \geq v\}$, we see that, if (U_1^θ, U_2^θ) has CDF C_θ, then

$$\eta(C_\theta) = P(U_1^\theta \leq U_2^\theta) = \theta$$

and

$$\eta(C_\theta^T) = P(U_2^\theta \leq U_1^\theta) = 1 - \theta,$$

$\forall\, \theta \in (0,1]$. (The interested reader can consult [7, Proposition 7] for more explicit calculations). Moreover $C_{1-\theta} = C_\theta^T, \ \forall\, \theta \in [0,1]$. For every $\theta \notin \{0, 0.5, 1\}$ the copula C_θ is asymmetric and included in $\mathcal{C}_D$. This means that not only $\eta(C_\theta) = \theta$, but also $\nu(C_\theta) = \theta, \ \forall\, \theta \in (0,1)$.

If we consider the family $\mathcal{N}^* = \mathcal{N} \setminus \{C_0, C_{0.5}, C_1\}$, and according to Corollary 2, there exists a one-to-one correspondence between the subfamilies of FSD-sp consistent and nonconsistent copulas, associating each copula C_θ for $\theta \in (0.5, 1)$ to its transpose $C_\theta^T = C_{1-\theta}$.

For the (asymmetric) copulas included in $\mathcal{N}^{**} = \{C_\theta \ : \ \theta \in (0, 0.5)\}$, first stochastic dominance does not imply statistical preference, nor diff-stochastic dominance, nor any other more restrictive $\mathcal{G}$-expectation preference. In particular, if we choose a pair of random variables X and Y, both uniform in the same interval, and with a copula included in $\mathcal{N}^{**}$, we will have that Y stochastically dominates X and yet is not statistically preferred to it.

5. Conclusion

The entirety of the extensive results contained in [11] and [8], in addition to Theorem 2.9 from [12], have been generalized in a single result (Theorem 4), whose proof is very short. From this, we deduce that none of the specific examples of copulas considered in those prior works (such as comonotone, antitone, Archimedean, etc.) exhibits any distinctive characteristics in the relationship between first stochastic dominance and diff-stochastic dominance, beyond their inherent symmetry.

Additionally, in Section 4, we demonstrate that when we restrict our analysis to random vectors with at least one continuous marginal CDF, first stochastic dominance implies statistical preference for at least as many asymmetric copulas as those for which it does not. It should be noted that we do not require any of the marginals to have a density function. We just require that at least one of them has no singletons with strictly positive probability.

Furthermore, we have identified an infinite and indexed set of asymmetric copulas for which first stochastic dominance does not entail statistical preference. Consequently, it does not either imply diff-stochastic dominance or any other stronger form of $\mathcal{G}$-expectation dominance.

Acknowledgements

This work is partially supported by Projects SV-PA-21-AYUD/2021/50994 (Principality of Asturias) and PID2020-112726RB-I00 (Spanish Government).

References

[1] N. d. C. Marquis de Condorcet, Essai sur l'application de l'analyse à la probabilité des decisions rendues à la pluralité des voix, l'Imprimerie Royale, 1785.

[2] B. De Schuymer, H. De Meyer, B. De Baets, S. Jenei, On the cycle-transitivity of the dice model, Theory and Decision 54 (2003) 261–285.

[3] I. Couso, S. Moral, L. Sánchez, The behavioral meaning of the median, Information Sciences 294 (2015) 127–138.

[4] M. A. Arcones, P. H. Kvam, F. J. Samaniego, Nonparametric estimation of a distribution subject to a stochastic precedence constraint, Journal of the American Statistical Association 97 (457) (2002) 170–182.

[5] P. J. Boland, H. Singh, B. Cukic, The stochastic precedence ordering with applications in sampling and testing, Journal of Applied Probability 41 (1) (2004) 73–82.

[6] J. Navarro, R. Rubio, Comparisons of coherent systems using stochastic precedence, Test 19 (3) (2010) 469–486.

[7] E. De Santis, F. Fantozzi, F. Spizzichino, Relations between stochastic orderings and generalized stochastic precedence, Probability in the Engineering and Informational Sciences 29 (3) (2015) 329–343.

[8] I. Montes, S. Montes, Stochastic dominance and statistical preference for random variables coupled by an Archimedean copula or by the Fréchet–Hoeffding upper bound, Journal of Multivariate Analysis 143 (2016) 275–298.

[9] B. De Schuymer, H. De Meyer, B. De Baets, Cycle-transitive comparison of independent random variables, Journal of Multivariate Analysis 96 (2005) 352–373.

[10] H. De Meyer, B. De Schuymer, B. De Baets, On the transitivity of the comonotonic and countermonotonic comparison of random variables, Journal of Multivariate Analysis 98 (2007) 177–193.

[11] I. Montes, J. J. Salamanca, S. Montes, A modified version of stochastic dominance involving dependence, Statistics & Probability Letters 165 (2020) 108848.

[12] F. Belzunce, C. Martínez-Riquelme, A new stochastic dominance criterion for dependent random variables with applications, Insurance: Mathematics and Economics 108 (2023) 165–176.

[13] J. G. Shanthikumar, D. D. Yao, Bivariate characterization of some stochastic order relations, Advances in Applied Probability 23 (3) (1991) 642–659.

[14] G. Kreweras, Sur une possibilité de rationaliser les intransitivités, Editions du CNRS, Colloques Internationaux du CNRS, "La Decision". Paris (1961) 27–32.

[15] P. C. Fishburn, Nontransitive measurable utility, Journal of Mathematical Psychology 26 (1) (1982) 31–67.

[16] L. Sánchez, N. Costa, I. Couso, Simplified models of remaining useful life based on stochastic orderings, Reliability Engineering & System Safety (2023) 109321.

[17] F. Durante, C. Sempi, Principles of Copula Theory, CRC press, 2015.

[18] F. Durante, A. Gianfreda, F. Ravazzolo, L. Rossini, A multivariate dependence analysis for electricity prices, demand and renewable energy sources, Information Sciences 590 (2022) 74–89.

[19] G. Puccetti, Measuring linear correlation between random vectors, Information Sciences 607 (2022) 1328–1347.

[20] M. Sklar, Fonctions de repartition an dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

[21] R. B. Nelsen, Extremes of nonexchangeability, Statistical Papers 48 (2007) 329–336.

[22] J. Hadar, W. R. Russell, Rules for ordering uncertain prospects, The American Economic Review 59 (1) (1969) 25–34.

[23] L. J. Savage, The Foundations of Statistics, Courier Corporation, 1972.

[24] G. Kreweras, Aggregation of preference orderings, in: Mathematics and Social Sciences I: Proceedings of the seminars of Menthon-Saint-Bernard, France (1–27 July 1960) and of Gösing, Austria (3–27 July 1962), 1965, pp. 73–79.

[25] P. C. Fishburn, Nontransitive preferences in decision theory, Journal of Risk and Uncertainty 4 (1991) 113–134.

[26] G. Loomes, R. Sugden, Regret theory: An alternative theory of rational choice under uncertainty, The Economic Journal 92 (368) (1982) 805–824.

[27] S. Ross, Stochastic processes, Wiley, New York, 1983.

[28] D. Stoyan, D. J. Daley, Comparison Methods for Queues and Other Stochastic Models., Wiley, New York, 1983.

[29] S. Karlin, A. Novikov, Generalized convex inequalities, Pacific J Math 13 (1963) 1251–1279.

[30] M. A. Sordo, M. C. de Souza, A. Suárez-Llorens, A new variability order based on tail- heaviness, Statistics: A Journal of Theoretical and Applied Statistics (2014) DOI: 10.1080/02331888.2014.946930.

[31] M. López-Díaz, A stochastic ordering for random variables with applications, Australian & New Zealand Journal of Statistics 52 (1) (2010) 1–16.

[32] M. Shaked, J. G. Shanthikumar, Stochastic orders, Springer, 2007.

[33] R. B. Nelsen, An introduction to copulas, Springer, 2006.

[34] A. Müller, M. Scarsini, Stochastic comparison of random vectors with a common copula, Mathematics of Operations Research 26 (4) (2001) 723–74.

[35] P. Embrechts, M. Hofert, A note on generalized inverses, Mathematical Methods of Operations Research 77 (2013) 423–432.

[36] K. Siburg, P. Stoimenov, Symmetry of functions and exchangeability of random variables, Statistical Papers 52 (2011) 1–15.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: