A Primer on Copulas for Count Data

Journal Article

Author(s):
Genest, Christian; Nešlehová, Johanna

A PRIMER ON COPULAS FOR COUNT DATA

BY

CHRISTIAN GENEST AND JOHANNA NEŠLEHOVÁ

ABSTRACT

The authors review various facts about copulas linking discrete distributions. They show how the possibility of ties that results from atoms in the probability distribution invalidates various familiar relations that lie at the root of copula theory in the continuous case. They highlight some of the dangers and limitations of an undiscriminating transposition of modeling and inference practices from the continuous setting into the discrete one.

KEYWORDS

Bernoulli distribution; Copulas; Count data; Discrete distribution; Geometric distribution; Kendall's tau; Poisson distribution; Spearman's ρ; Subcopula.

1. INTRODUCTION

Count data abound in actuarial science, and the issue of how to model the joint behavior of random variables X and Y taking values on the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of integers is a common problem in analyzing the impact of dependence on loss distributions, among others.

Given the recent surge of interest for copula modeling of multivariate continuous data and its successful application in many contexts, one may be tempted to adopt this approach in the discrete case also. Following Sklar (1959), this would entail writing the joint cumulative distribution function H of the pair (X, Y) as

$$H(x, y) = C(F(x), G(y)), \quad x, y \in \mathbb{R} \tag{1}$$

where F and G are the marginal distributions of X and Y, respectively. A parametric class (C_θ) of copulas could then be postulated for C, and inference techniques developed in the context of copula models with continuous margins might then be used, either to develop a descriptive model for H or to make predictions.

Assuming for example that X and Y are Poisson random variables with rate μ and v respectively, one could suppose that the copula C in (1) belongs to Clayton's family of copulas with dependence parameter $\theta > 0$, viz.

$$C_{\theta}(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad u, v \in (0, 1). \tag{2}$$

Thus if the Poisson distribution function with intensity λ is denoted

$$F_{\lambda}(k) = e^{-\lambda}\left(\frac{\lambda^0}{0!} + \cdots + \frac{\lambda^k}{k!}\right), \quad k \in \mathbb{N}$$

the model for (X, Y) would be of the form

$$\mathrm{P}(X \le i,\ Y \le j) = \{F_{\mu}^{-\theta}(i) + F_{v}^{-\theta}(j) - 1\}^{-1/\theta}, \quad i, j \in \mathbb{N}.$$

Based on a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (X, Y), one could then estimate μ, v and θ, say by the method of maximum likelihood. Alternatively, μ and v could be estimated individually, and to guard against the possibility that the margins are not Poisson, a nonparametric rank-based approach to the estimation of θ could be adopted. Following Oakes (1982), for example, one might estimate θ by $\hat{\theta}_n = 2\tau_n/(1 - \tau_n)$, where τ_n represents the sample value of Kendall's τ. Confidence intervals for $\hat{\theta}_n$ could then be computed, leading to predictions and margins of error for the probability of events of interest.

Is this alright? Yes and no. The purpose of this paper is to review in the simplest terms possible the issues that lead to this mitigated answer. As explained in Section 2, the primary source of difficulty arises from the lack of uniqueness of Sklar's representation (1) in the discrete case. A problem of statistical unidentifiability ensues whose extent is delineated in Section 3 using results from Carley (2002).

Despite the unidentifiability issue, copula models for discrete distributions are valid constructions. They are helpful, e.g., in the context of simulation and robustness studies. However, their use is subject to caution, because many of their convenient properties do not carry over from the continuous to the discrete case. Some crucial differences are highlighted in Sections 4 and 5. Their consequences on the inference program outlined above are then illustrated in Section 6. A summary of the main findings is given in the Conclusion.

Most of the facts reported herein are known, but scattered throughout the literature; some are new. It is hoped that by bringing them together in one source, those who wish to model count data through copulas can rapidly identify the advantages and limitations of this approach. While the following discussion is limited to the bivariate case, most conclusions easily extend to a general multivariate setting.

2. Lack of uniqueness of the copula and consequences thereof

Throughout this paper, (X, Y) denotes a pair of random variables with joint distribution function H having margins F and G, i.e., for all $x, y \in \mathbb{R}$,

$$H(x, y) = P(X \le x, Y \le y), \quad F(x) = P(X \le x), \quad G(y) = P(Y \le y).$$

The inverses of F and G are defined in the usual way by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\}, \quad G^{-1}(v) = \inf\{y \in \mathbb{R} : G(y) \ge v\},$$

so that they are left-continuous for all $u, v \in (0, 1]$. As they may not be right-continuous, however, it is useful to introduce

$$F^{-1}(u_-) = \lim_{s \downarrow u} F^{-1}(s), \quad G^{-1}(v_-) = \lim_{t \downarrow v} G^{-1}(t).$$

Similarly, the left limits of discontinuous distribution functions F and G are denoted

$$F(x_-) = \lim_{s \uparrow x} F(s), \quad G(y_-) = \lim_{t \uparrow y} G(t).$$

The following definition introduces additional notation which is helpful in clarifying the distinctions between the cases where F and G are either continuous or discrete.

Definition 1. *Let (X, Y) be a pair of random variables such that $P(X \le x, Y \le y) = H(x, y)$, $P(X \le x) = F(x)$ and $P(Y \le y) = G(y)$ for all $x, y \in \mathbb{R}$.*

(A) *$\mathcal{A}$ is the set of "sub-copulas" associated with H, i.e., the collection of functions $A : [0, 1]^2 \to [0, 1]$ such that for all $x, y \in \mathbb{R}$,*
$$H(x, y) = A(F(x), G(y)).$$

(B) *$B : [0, 1]^2 \to [0, 1]$ is the function defined for all $u, v \in [0, 1]$ by*
$$B(u, v) = H(F^{-1}(u), G^{-1}(v)).$$

(C) *$C : [0, 1]^2 \to [0, 1]$ is the function defined for all $u, v \in [0, 1]$ by*
$$C(u, v) = H(F^{-1}(u_-), G^{-1}(v_-)).$$

(D) *D is the distribution function of the random pair $(F(X), G(Y))$.*

(E) *E is the distribution function of the random pair $(F(X_-), G(Y_-))$.*

When both F and G are continuous, the various objects introduced in Definition 1 are known to coincide by Sklar's Theorem, i.e.,

$$\mathcal{A} = \{B\} = \{C\} = \{D\} = \{E\}, \tag{3}$$

and $\mathcal{A}$ consists of the *unique* copula associated with H. See Chapter 1 of Nelsen (1999) for a proof.

This fact, which has important consequences, is the corner stone of inference in copula modeling as outlined, e.g., by Genest and Favre (2007) or Genest et al. (2007):

(a) Given H, C is a uniquely defined (functional) parameter that can be estimated consistently on the basis of a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from H.
(b) If F and G are known, a random sample $(U_1, V_1), \ldots, (U_n, V_n)$ obtains for C upon setting $U_i = F(X_i)$ and $V_i = G(Y_i)$ for each $i \in \{1, \ldots, n\}$.
(c) If F and G are unknown, inference about C can still be performed using pairs of pseudo observations $\hat{U}_i = \hat{F}(X_i)$, $\hat{V}_i = \hat{G}(Y_i)$ derived from estimates $\hat{F}$ and $\hat{G}$ of F and G (the empirical distribution functions are often used to this end).
(d) Given a parametric family (C_θ) for C, one can also estimate the dependence parameter θ or perform goodness-of-fit testing using the pairs $(\hat{U}_1, \hat{V}_1), \ldots, (\hat{U}_n, \hat{V}_n)$.

The source of difficulties in the discrete case is that when F and G have jumps, their inverses have plateaus. When this happens, Sklar's Theorem still guarantees that there exists a copula representation for H in $\mathcal{A}$, but the latter is no longer unique. Thus, an identifiability issue arises and the above program for inference breaks down, because one can no longer rely on the string of identities (3) on which it is based. In fact, the functions defined through (B)–(E) then represent different objects; neither of them is a copula, and some but not all of them belong to $\mathcal{A}$.

The following example illustrates these various findings in an extreme case. The extent of the identifiability issue is qualified in Section 3.

Example 1. *Let X and Y be Bernoulli random variables with* $P(X = 0) = p$ *and* $P(Y = 0) = q$. *Let also* $P(X = 0, Y = 0) = r \le \min(p, q)$.

In that case, $\mathcal{A}$ consists of all functions $A : [0,1]^2 \to [0,1]$ such that

$$
\begin{array}{lll}
A(0,0) = 0, & A(0,q) = 0, & A(0,1) = 0, \\
A(p,0) = 0, & A(p,q) = r, & A(p,1) = p, \\
A(1,0) = 0, & A(1,q) = q, & A(1,1) = 1.
\end{array}
$$

It is also straightforward to see that

$$
B(u,v) = \begin{cases}
0 & \text{if } u = 0 \text{ or } v = 0, \\
r & \text{if } (u,v) \in (0,p] \times (0,q], \\
q & \text{if } (u,v) \in (p,1] \times (0,q], \\
p & \text{if } (u,v) \in (0,p] \times (q,1], \\
1 & \text{if } (u,v) \in (p,1] \times (q,1],
\end{cases}
$$

while

$$
C(u,v) = \begin{cases}
r & if \ (u,v) \in [0,p) \times [0,q), \\
q & if \ (u,v) \in [p,1] \times [0,q), \\
p & if \ (u,v) \in [0,p) \times [q,1], \\
1 & if \ (u,v) \in [p,1] \times [q,1].
\end{cases}
$$

The difference between them may seem small, but it cannot be neglected. For, C is a distribution function while B isn't. In addition, $B \in \mathcal{A}$ while $C \notin \mathcal{A}$. Furthermore, it can be checked that $C = E$ is the distribution function of $(F(X_-), G(Y_-))$, which is different from that of $(F(X), G(Y))$. The latter, which belongs to $\mathcal{A}$, is

$$
D(u,v) = \begin{cases}
0 & if \ (u,v) \in [0,p) \times [0,1], \\
0 & if \ (u,v) \in [0,1] \times [0,q), \\
r & if \ (u,v) \in [p,1) \times [q,1), \\
p & if \ (u,v) \in [0,p) \times \{1\}, \\
q & if \ (u,v) \in \{1\} \times [q,1), \\
1 & if \ (u,v) = (1,1).
\end{cases}
$$

The observations made in Example 1 hold in general, as stated below.

Proposition 1. Let $\mathcal{A}$, B, C, D and E be as introduced in Definition 1. If the joint distribution of the pair (X,Y) is discrete, then

(i) $B \in \mathcal{A}$ is not a distribution function.
(ii) $D \in \mathcal{A}$ is a distribution function but not a copula.
(iii) $C = E \notin \mathcal{A}$ is a distribution function but not a copula.

Proof. The arguments in support of the above assertions are as follows:

(i) B cannot be a distribution function because it is not right-continuous.
(ii) That $B \in \mathcal{A}$ is an immediate consequence of Sklar's Theorem.
(iii) D and E are distribution functions by definition, but they cannot be copulas because F and G have jumps, so that $F(X)$, $F(X_-)$, $G(Y)$ and $G(Y_-)$ are not uniformly distributed on the interval $(0,1)$.
(iv) That $C = E$ follows from the fact that in view of the definition of F^{-1}, the equivalence $u \geq F(x_-) \Leftrightarrow F^{-1}(u_-) \geq x$ holds for any $u \in (0,1)$ and $x \in \mathbb{R}$. Of course, a similar statement holds for G.
(v)   Finally, to see that $D \in \mathcal{A}$, one must show that
$$
P\{F(X) \leq F(x), G(Y) \leq G(y)\} = H(x,y).
$$

This stems from the fact that for any $x \in \mathbb{R}$, one has $\mathrm{P}\{F(X) \le F(x)\} = \mathrm{P}(X \le x)$, even though the equivalence $u \le x \Leftrightarrow F(u) \le F(x)$ is not valid. Again, a similar statement is valid for G. □

3. HOW IMPORTANT IS THE UNIDENTIFIABILITY ISSUE?

Given a bivariate distribution function H with discrete margins, let $C_H \subset \mathcal{A}$ be the set of copulas for which identity (1) holds. Should this class be reasonably small, the copula models that are compatible with H would mimic fairly closely the dependence characteristics of the latter. The purpose of this section is to investigate to what extent such is the case.

To assess the importance of the unidentifiability issue, one must somehow measure the size of C_H. This can be done in two steps:

(1) Identify the smallest and largest elements in the class C_H.
(2) Measure their associated degree of dependence, e.g., via Kendall's τ or Spearman's ρ, respectively defined for every copula C by

$$\tau(C) = -1 + 4 \int_0^1 \int_0^1 C(u,v) \, dC(u,v), \qquad (4)$$

$$\rho(C) = -3 + 12 \int_0^1 \int_0^1 C(u,v) \, dv du. \qquad (5)$$

Sections 3.1 and 3.2 address these problems in turn. As will be seen, the class C_H is generally quite large. The unidentifiability of the copula is thus a real issue, whose importance is shown in Section 3.3 to be related to the flatness of the margins.

3.1. Carley's bounds on C_H

When the variables X and Y take only finitely many values, the best pointwise bounds for elements of the class C_H have been identified by Carley (2002). Her construction is presented here in the slightly more general case of a distribution H that specifies the copula at all grid points $(F(i), G(j))$ with $i, j \in \mathbb{N}$.

Assume without loss of generality that X and Y take integer values and let

$$h_{ij} = \mathrm{P}(X = i, Y = j), \quad H(i,j) = \mathrm{P}(X \le i, Y \le j),$$

so that h_{ij} is the mass assigned by C to the rectangle $(F(i-1), F(i)] \times (G(j-1), G(j)]$ with $F(-1) = G(-1) = 0$ by convention. Let also

$$h_{i+} = \sum_{j=0}^{\infty} h_{ij} = \mathrm{P}(X = i), \quad h_{+j} = \sum_{i=0}^{\infty} h_{ij} = \mathrm{P}(Y = j)$$

stand for the marginal probability functions of X and Y, respectively.

Proposition 2. *Let $\mathcal{C}_H$ be the set of copulas which fulfill equation* (1) *for a fixed discrete distribution H on $\mathbb{N}^2$. The upper Carley bound for the class $\mathcal{C}_H$ is given by*

$$C_H^+(u,v) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \max\{0, \min(u - \alpha_{ij}, v - \beta_{ij}, h_{ij})\}$$

and the lower Carley bound is equal to

$$C_H^-(u,v) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \max\{0, -h_{ij} + \min(u - \gamma_{ij}, h_{ij}) + \min(v - \delta_{ij}, h_{ij})\},$$

where

$$\alpha_{ij} = \sum_{k=0}^{i-1} h_{k+} + \sum_{\ell=0}^{j-1} h_{i\ell}, \quad \beta_{ij} = \sum_{\ell=0}^{j-1} h_{+\ell} + \sum_{k=0}^{i-1} h_{kj},$$

$$\gamma_{ij} = \sum_{k=0}^{i-1} h_{k+} + \sum_{\ell=j+1}^{\infty} h_{i\ell}, \quad \delta_{ij} = \sum_{\ell=0}^{j-1} h_{+\ell} + \sum_{k=i+1}^{\infty} h_{kj},$$

and by convention, an empty sum is equal to zero.

This result is essentially the same as that given by Carley (2002), and hence the proof is omitted. The reader will derive far more benefit from the following illustration, which reveals the nature of Carley's construction.

Example 2. *Consider a bivariate distribution on $\{0,1,2,3\} \times \{0,1,2\}$ whose frequencies $f_{ij} = 24 h_{ij}$ are given in Table 1. It can be seen from Fig. 1 that the corresponding Carley bounds C_H^- and C_H^+ are "shuffles of min", in the sense of Mikusiński et al. (1992). In the lower bound (on the left) all the segments have slope -1, while in the upper bound (on the right) their slope is $+1$. On each segment, the probability mass is uniformly distributed and equal to the corresponding value of h_{ij}.*

TABLE 1

CONTINGENCY TABLE DEFINING THE GENERIC DISTRIBUTION FUNCTION H FOR EXAMPLE 2.

	$X = 0$	$X = 1$	$X = 2$	$X = 3$	Total
$Y = 2$	1	2	3	0	6
$Y = 1$	1	3	6	2	12
$Y = 0$	1	1	3	1	6
	3	6	12	3	24

The location of segment (i,j) is best explained by reference to Fig. 2. Note that the shadowed regions in that figure contain no mass. For added clarity, the locations of the x- and y-divisions are given in Tables 2 and 3 for C_H^- and C_H^+, respectively.

TABLE 2

COORDINATES OF THE SEGMENTS OF C_H^- FOR THE DISTRIBUTION H DEFINED IN TABLE 1.

$24\,\gamma_{ij}$	$i=0$	$i=1$	$i=2$	$i=3$	$24\,\delta_{ij}$	$i=0$	$i=1$	$i=2$	$i=3$
$j=2$	0	3	9	21	$j=2$	23	21	18	18
$j=1$	1	5	12	21	$j=1$	17	14	8	6
$j=0$	2	8	18	23	$j=0$	5	4	1	0

TABLE 3

COORDINATES OF THE SEGMENTS OF C_H^+ FOR THE DISTRIBUTION H DEFINED IN TABLE 1.

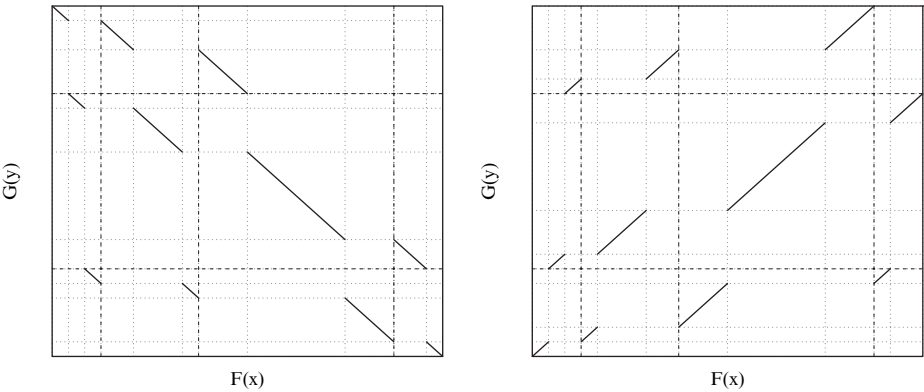| $24\,\alpha_{ij}$	$i=0$	$i=1$	$i=2$	$i=3$	$24\,\beta_{ij}$	$i=0$	$i=1$	$i=2$	$i=3$
$j=2$	2	7	18	24	$j=2$	18	19	21	24
$j=1$	1	4	12	22	$j=1$	6	7	10	16
$j=0$	0	3	9	21	$j=0$	0	1	2	5

FIGURE 1: Carley's bounds C_H^- and C_H^+ corresponding to the distribution function H defined in Table 1; left panel: lower bound; right panel: upper bound.

Comparing the two panels of Fig. 2, one can detect a form of symmetry between the bounds. What is being perceived is in fact a reorientation and repositioning of each segment within the limits of its confining box. Just how the repositioning is achieved is the object of the following proposition, also excerpted from Carley (2002). Although this fact is only valid for the distributions of count data with finite support, it is stated here formally, as it has useful implications.
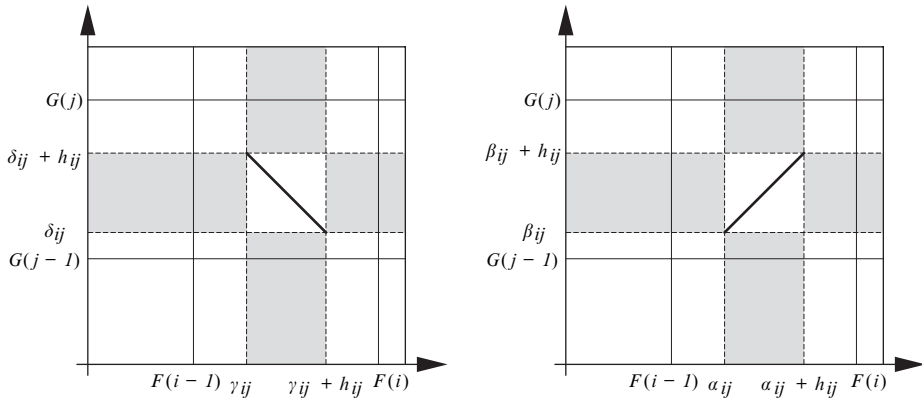
FIGURE 2: Intersections of the supports of C_H^- (left panel) and C_H^+ (right panel) with the rectangle $(F(i-1), F(i)] \times (G(j-1), G(j)]$.

Proposition 3. *Let H be a distribution function on $\mathbb{N}^2$ whose support is comprised in $\{0, \ldots, m\} \times \mathbb{N}$. If H^* is another distribution on $\mathbb{N}^2$ with probabilities $h_{ij}^* = h_{(m-i)j}$ for all $i \in \{0, \ldots, m\}$ and $j \in \mathbb{N}$, then for all $u, v \in (0,1)$,*

$$C_{H^*}^-(u,v) = v - C_H^+(1-u,v) \ \text{ and } \ C_{H^*}^+(u,v) = v - C_H^-(1-u,v).$$

In the above, H^* is the distribution of the pair $(m - X, Y)$. An extension of this result to the case where the support of H is unbounded may be found in Appendix A.

3.2. Extreme values for τ and ρ in $\mathcal{C}_H$

The Carley bounds make it possible to measure the breadth of degrees of dependence spanned by the set $\mathcal{C}_H$ of copulas that are compatible with a given pair (X, Y) of counting random variables with joint distribution function H. If κ stands for any measure of concordance in the sense of Scarsini (1984), a standard result of Tchen (1980) implies that

$$\kappa(C_H^-) \leq \kappa(C) \leq \kappa(C_H^+).$$

These bounds are best possible, given that both C_H^- and C_H^+ belong to $\mathcal{C}_H$. Furthermore, all points in $[\kappa(C_H^-), \kappa(C_H^+)]$ are possible values for $\kappa(C)$, because for every $\theta \in [0,1]$

$$C_\theta = \theta C_H^- + (1 - \theta) \, C_H^+ \in \mathcal{C}_H$$

and $\kappa(C_\theta)$ is a continuous function of θ.

For Kendall's τ and Spearman's ρ, the computation of the bounds is facilitated by the following propositions, whose proofs are relegated to Appendices B and C.

Proposition 4. *Let H be a distribution function on $\mathbb{N}^2$ with probabilities h_{ij} for all $i, j \in \mathbb{N}$. Then*

$$\tau(C_H^-) = -1 + 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{i-1} \sum_{\ell=0}^{j-1} h_{ij}\, h_{k\ell}$$

and

$$\tau(C_H^+) = 1 - 4 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{i-1} \sum_{\ell=j+1}^{\infty} h_{ij}\, h_{k\ell}.$$

Proposition 5. *Let H be a distribution function on $\mathbb{N}^2$ with probabilities h_{ij} for all $i, j \in \mathbb{N}$. Then*

$$\rho(C_H^-) = -1 - 6 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij} \{\delta_{ij} + h_{ij} - (1 - \gamma_{ij})\} \{2(1 - \gamma_{ij}) - h_{ij}\}$$

and

$$\rho(C_H^+) = 1 + 6 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_{ij} (\beta_{ij} - \alpha_{ij}) (2\alpha_{ij} + h_{ij}),$$

where α_{ij}, β_{ij}, γ_{ij} and δ_{ij} are as defined in Proposition 2.

The following examples show that the Carley bounds on Kendall's τ and Spearman's ρ can sometimes be fairly wide.

Example 3. *Let X and Y be Bernoulli random variables with $P(X = 0) = p$ and $P(Y = 0) = q$. Let also $r = P(X = 0, Y = 0) \in [\max(0, p + q - 1), \min(p, q)]$. Then C is a possible copula model for the relation between X and Y if and only if $C(p, q) = r$. Simple substitution into the bounds from Proposition 4 implies that*

$$4rs - 1 \leq \tau(C) \leq 1 - 4(p - r)(q - r),$$

where $s = 1 - p - q + r = P(X = 1, Y = 1)$. Similarly, Proposition 5 yields

$$6rs(r + s) - 1 \leq \rho(C) \leq 1 - 6(p - r)(q - r)(1 - r - s).$$

The bounds on τ are displayed in Fig. 3 along with their difference, in the special case $p = q$. One can see from the graphs (this is easily checked algebraically) that

(i) *the lower bound equals 0 precisely when $p = q = r = 1/2$, i.e., when $X = Y$ almost surely;*
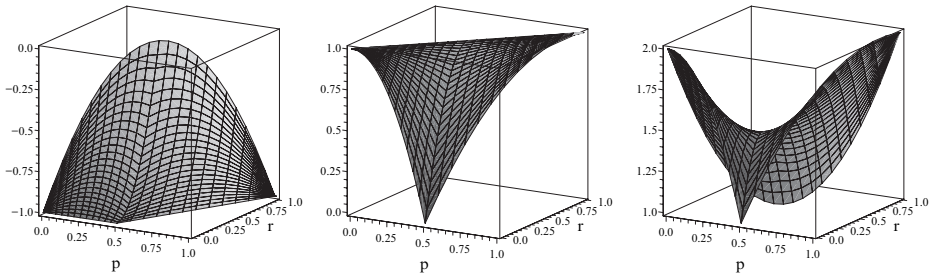
FIGURE 3: Plot of the lower bound $\tau(C_H^-)$ (left panel) and upper bound $\tau(C_H^+)$ (middle panel) on τ as a function of $p = q$ and r for the bivariate Bernoulli model of Example 3. The difference $\tau(C_H^+) - \tau(C_H^-)$ is represented in the right panel.

(ii) *the upper bound vanishes precisely when $p = q = 1/2$ and $r = 0$, i.e., when $Y = 1 - X$ almost surely;*

(iii) *the difference between the two bounds is never smaller than 1.*

The graphs for the bounds on ρ (not displayed here) are similar in shape, but the range of values is larger. Specifically, the lower bound on ρ varies in $[-1, 0.5]$, the upper bound lies in $[-0.5, 1]$, and hence the difference spans the interval $[0.5, 2]$.

In the light of Example 3, one can see that the unidentifiability of the copula can sometimes be a severe problem. The following example explores the extent of this phenomenon and makes additional observations in another context.

Example 4. *Suppose X and Y are independent geometric random variables on $\mathbb{N}$ with success probabilities $1 - p$ and $1 - q$, respectively. If C is a possible copula for the pair (X, Y), then*

$$\left| \tau(C) \right| \le 1 - 4 \frac{p}{1 + p} \frac{q}{1 + q}.$$

This follows from simple algebraic manipulations, upon setting $h_{ij} = (1 - p)p^i (1 - q)q^j$ in formulas from Proposition 4. It also follows from Proposition 5 that

$$\rho(C) \le 1 - 6pq \frac{p^2 q + pq^2 + p^2 + q^2 + p + q}{(1 + p)(1 + q)\left(p^2 + p + 1\right)\left(q^2 + q + 1\right)}$$

and

$$\rho(C) \ge -1 + 6pq \frac{p^2 q + pq^2 + p^2 + q^2 + pq + 1}{(1 + p)(1 + q)\left(p^2 + p + 1\right)\left(q^2 + q + 1\right)}.$$

The differences $T_H = \tau(C_H^+) - \tau(C_H^-)$, $R_H = \rho(C_H^+) - \rho(C_H^-)$ and $T_H - R_H$ are shown in Fig. 4 as a function of p and q. Among the features of the graphs, observe that
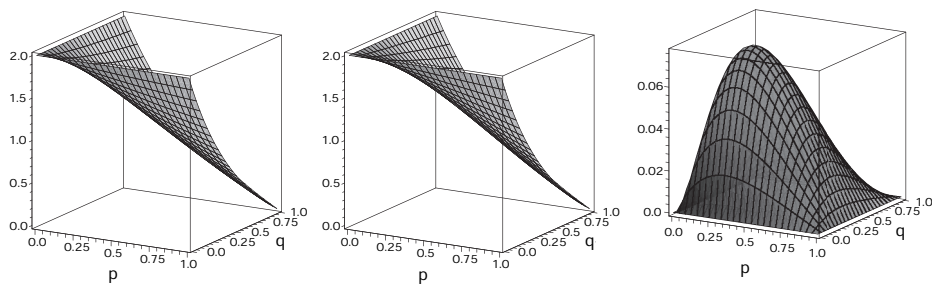
FIGURE 4: Plot of the differences $T_H = \tau(C_H^+) - \tau(C_H^-)$ (left panel), $R_H = \rho(C_H^+) - \rho(C_H^-)$ (middle panel) and $T_H - R_H$ (right panel) as a function of p and q for the bivariate geometric model of Example 4.

(i) both T_H and R_H are decreasing in p and q; they reach 0 in the limiting case where $p = q = 1$, i.e., when the probabilities of success vanish;

(ii) the fact that $T_H \geq R_H$ everywhere shows that the bounds for ρ are wider than those for τ for all parameter values. In fact, it can be verified (not shown here) that $\rho(C_H^+) \geq \tau(C_H^+)$ everywhere, but that $\rho(C_H^-) - \tau(C_H^-)$ is not always of the same sign.

3.3. Unidentifiability as a function of the flatness of the margins

Another way of quantifying the size of the class C_H is given below as a function of the "degree of flatness" of random variables X and Y jointly distributed as H.

Proposition 6. Let (X, Y) be a pair of integer-valued random variables whose joint distribution function H has margins F and G. For all $i, j \in \mathbb{N}$, let $R_{ij} = (F(i-1), F(i)] \times (G(j-1), G(j)]$. Then for arbitrary $C, D \in C_H$ and $u, v \in (0,1)$,

$$\left| C(u,v) - D(u,v) \right| \leq 2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left\{ P(X=i) + P(Y=j) \right\} \mathbf{1}_{R_{ij}}(u,v)$$

$$\leq 2 \left\{ \max_{i \in \mathbb{N}} P(X=i) + \max_{j \in \mathbb{N}} P(Y=j) \right\}.$$

Proof. For arbitrary $(u,v) \in R_{ij}$, one has

$$\left| C(u,v) - D(u,v) \right| \leq \left| C(u,v) - C(F(i), G(j)) \right| + \left| D(F(i), G(j)) - D(u,v) \right|.$$

Indeed, C and D coincide on $\mathrm{Ran}(F) \times \mathrm{Ran}(G)$, where $\mathrm{Ran}(F)$ and $\mathrm{Ran}(G)$ denote the ranges of F and G, respectively. Consequently, $C(F(i), G(j)) = D(F(i), G(j))$

for all $i, j \in \mathbb{N}$. Now any copula is Lipschitz, as shown in Chapter 2 of Nelsen (1999), among others. Therefore, it follows from the above inequality that

$$|C(u,v) - D(u,v)| \leq 2\,|F(i) - u| + 2\,|G(j) - v| \leq 2\,\{P(X = i) + P(Y = j)\}$$

and also that

$$|C(u,v) - D(u,v)| \leq 2\left\{\max_{i \in \mathbb{N}} P(X = i) + \max_{j \in \mathbb{N}} P(Y = j)\right\}.$$

Thus the proof is complete. $\square$

It is clear from this proposition that $\mathcal{C}_H$ shrinks to a singleton when H is continuous. At the same time, the result suggests that the unidentifiability issue could be more damaging when either X or Y takes on a small number of values. To confirm this conjecture, however, a lower bound on the supremum norm of the difference between the Carley bounds would be needed. This issue is left unresolved.

4. WHAT IMPACT DOES THIS HAVE ON COPULA MODELS FOR DISCRETE DATA?

A copula model for a pair (X, Y) is a class of bivariate distributions of the form (1), where C, F and G are assumed to belong to parametric classes, viz., $C \in (C_\theta)$, $F \in (F_\mu)$, $G \in (G_\nu)$. That many copulas are compatible with the joint distribution H when the latter is discrete does not make such models invalid *prima facie*.

For one thing, many common copula models are easy to simulate and thus provide useful tools for scenario analysis or robustness studies, even when F and G are discrete. See, e.g., Choulakian and de Tibeiro (2000), Genest et al. (2003) and Pfeifer and Nešlehová (2004), as well as Chapter 8 of McNeil et al. (2005).

Several examples of copula models for multivariate discrete data can also be found in the literature; see, e.g., Meester and MacKay (1994), Trégouët et al. (2004) and Cameron et al. (2004) for illustrations in biostatistics, genetics and econometrics, respectively.

In actuarial science and finance, the survey papers by Frees and Valdez (1998) and Embrechts et al. (2002) heralded the current wave of interest for modeling dependence through copulas. Although this approach does not yet appear to have been applied to discrete data in peer-reviewed journals, it is clearly imminent. A careful consideration of the limitations inherent to this methodology thus seems urgent.

4.1. The copula alone does not characterize the dependence between X and Y

When random variables X and Y are continuous, their unique underlying copula C is often depicted as characterizing the dependence between them. For example, it is easy to see that X and Y are independent if and only if $C = \Pi$, viz.

$$\Pi(u,v) = uv, \quad u,v \in (0,1).$$

This fact has been used, e.g., by Genest and Rémillard (2004) to derive powerful nonparametric tests of independence based on the empirical copula process.

Similarly, a familiar fact of copula theory is that continuous variables X and Y are in perfect monotone functional dependence if and only if their copula C is one of the Fréchet-Hoeffding bounds. In other words,

$$G(Y) = 1 - F(X) \text{ a.s.} \Leftrightarrow C = W,$$
$$G(Y) = F(X) \text{ a.s.} \Leftrightarrow C = M,$$

where

$$W(u,v) = \max(0, u+v-1) \quad \text{and} \quad M(u,v) = \min(u,v), \quad u,v \in (0,1).$$

In contrast, suppose that the joint distribution function of discrete random variables X and Y is modeled via relation (1) for some copula C. While it is still true that $C = \Pi$ implies independence between X and Y, the reverse implication is no longer valid, as the following example shows.

Example 5. *Let X and Y be independent Bernoulli random variables with* $P(X=0) = p$ *and* $P(Y=0) = q$. *Then C is a possible copula model for the pair* (X,Y) *if and only if* $C(p,q) = P(X=0, Y=0) = pq$. *Suppose in particular that* $p = q = 1/2$ *and consider the copula* $C = (W+M)/2$. *Then* $C(1/2, 1/2) = 1/4 = pq$ *and hence this copula induces independence between X and Y, yet* $C \neq \Pi$.

Similarly, it is possible to construct examples of discrete random variables X and Y that are in perfect monotone functional dependence, even though the copula used to model their joint distribution is neither W nor M.

More surprisingly, perhaps, although W and M continue to induce the best pointwise bounds on H, viz.

$$H^-(x,y) = W(F(x), G(y)), \quad H^+(x,y) = M(F(x), G(y)), \quad x,y \in \mathbb{R} \qquad (6)$$

the pairs (X^-, Y^-) and (X^+, Y^+) distributed respectively as H^- and H^+ are not necessarily in perfect monotone functional dependence! An example to this effect is given below.

Example 6. *For the bivariate discrete distribution defined in Table 1, the joint probability functions of the pairs (X^-, Y^-) and (X^+, Y^+) are given in Tables 4 and 5, respectively. These contingency tables, which were computed using formulas provided by Nelsen (1987), are reproduced in Appendix D.*

TABLE 4

CONTINGENCY TABLE DESCRIBING THE LOWER FRÉCHET-HOEFFDING BOUND
FOR THE BIVARIATE DISCRETE DISTRIBUTION DEFINED IN TABLE 1.

	$X = 0$	$X = 1$	$X = 2$	$X = 3$	Total
$Y = 2$	3	3	0	0	6
$Y = 1$	0	3	9	0	12
$Y = 0$	0	0	3	3	6
	3	6	12	3	24

TABLE 5

CONTINGENCY TABLE DESCRIBING THE UPPER FRÉCHET-HOEFFDING BOUND
FOR THE BIVARIATE DISCRETE DISTRIBUTION DEFINED IN TABLE 1.

	$X = 0$	$X = 1$	$X = 2$	$X = 3$	Total
$Y = 2$	0	0	3	3	6
$Y = 1$	0	3	9	0	12
$Y = 0$	3	3	0	0	6
	3	6	12	3	24

While it is plain from Tables 4 and 5 that neither of the pairs (X^-, Y^-) and (X^+, Y^+) has functionally dependent components, it should be stressed that it remains possible to generate *jointly* the two components of either (X^-, Y^-) or (X^+, Y^+) with a single uniform random variable. This result, which is stated formally below, is proved, e.g., by Embrechts et al. (2002). Accordingly, the pairs (X^-, Y^-) and (X^+, Y^+) may still be construed as counter- or co-monotonic, respectively, in the sense given to those terms by Yaari (1987).

Proposition 7. *Let (X, Y) be a random pair with distribution function H and margins F and G. Then $W \in \mathcal{C}_H$ if and only if there exists a random variable U uniformly distributed on $(0, 1)$ such that $X = X^- = F^{-1}(U)$ and $Y = Y^- = G^{-1}(1 - U)$ almost surely. Similarly, $M \in \mathcal{C}_H$ if and only if $X = X^+ = F^{-1}(U)$ and $Y = Y^+ = G^{-1}(U)$ almost surely.*

4.2. Concordance measures are margin-dependent

When dealing with continuous random variables, measures of concordance such as Kendall's τ or Spearman's ρ are convenient. Because they provide margin-free measures of the level of dependence in the bivariate distribution, they can be used to construct reliable estimation procedures when the underlying copula is assumed to belong to a specific parametric family.

When the random variables are discrete, however, all but one measure of concordance in the sense of Scarsini (1984) share with Pearson's correlation a dependence on the marginal distributions. The following result, due to Marshall (1996), shows that the only exception is not a particularly useful one.

Proposition 8. *Let $\mathcal{H}$ be the class of bivariate distribution functions whose support is contained in $\mathbb{N}^2$. Assume that κ is a dependence measure for which the implication $C \in \mathcal{C}_H \Rightarrow \kappa(H) = \kappa(C)$ holds for all $H \in \mathcal{H}$. Then κ is a scalar.*

The following example illustrates this dependence concretely in the case of Kendall's τ and Spearman's ρ.

Example 7. *Let X and Y be Bernoulli random variables with $\mathrm{P}(X=0)=p$ and $\mathrm{P}(Y=0)=q$. Let also $r = \mathrm{P}(X=0, Y=0) \in [\max(0, p+q-1), \min(p,q)]$. Then*

$$\tau(X,Y) = \rho(X,Y) = r - pq.$$

4.3. The probabilistic and analytical definitions of τ and ρ do not coincide

In Example 7 as in other cases where variables X and Y are discrete, $\tau(X,Y)$ and $\rho(X,Y)$ refer to the probabilistic definitions of Kendall's τ and Spearman's ρ, i.e.,

$$\tau(X,Y) = \mathrm{P}\{(X_1-X_2)(Y_1-Y_2) > 0\} - \mathrm{P}\{(X_1-X_2)(Y_1-Y_2) < 0\}$$

and

$$\rho(X,Y) = 3[\mathrm{P}\{(X_1-X_3)(Y_1-Y_3) > 0\} - \mathrm{P}\{(X_1-X_3)(Y_1-Y_3) < 0\}],$$

where (X_1, Y_1) and (X_2, Y_2) are two independent copies of (X,Y) while X_3 and Y_3 are independent and distributed as X and Y, respectively.

When X and Y are continuous random variables with unique underlying copula C, the above definitions are known to be equivalent to formulas (4) and (5); see, e.g., Chapter 5 of Nelsen (1999). Expressed symbolically, this means that

$$\tau(X,Y) = \tau(C) \text{ and } \rho(X,Y) = \rho(C).$$

When X and Y are discrete, however, this is no longer true. For, each possible choice of $C \in \mathcal{C}_H$ leads to different values for $\tau(C)$ and $\rho(C)$. In fact,

$$\{\tau(C) : C \in \mathcal{C}_H\} = [\tau(C_H^-), \tau(C_H^+)], \quad \{\rho(C) : C \in \mathcal{C}_H\} = [\rho(C_H^-), \rho(C_H^+)].$$

Nevertheless, $\tau(X,Y) = \tau(H)$ and $\rho(X,Y) = \rho(H)$ remain uniquely defined as functions of H. The following result, which makes this dependence explicit, is stated in total or in parts by Conti (1993), Kowalczyk and Niewiadomska-Bugaj (2001), Mesfioui and Tajar (2005) and Nešlehová (2007).

Proposition 9. *Let (X,Y) be a random pair with joint distribution H. Then*

$$\tau(X,Y) = -1 + \mathrm{E}\{H(X,Y) + H(X_-,Y) + H(X,Y_-) + H(X_-,Y_-)\}$$

and

$$\rho(X,Y) = -3 + 3\mathrm{E}\{H(X^*,Y^*) + H(X_-^*,Y^*) + H(X^*,Y_\rightarrow^*) + H(X_\rightarrow^*,Y_\rightarrow^*)\},$$

where X^* and Y^* are independent and distributed as X and Y, respectively. When H is continuous, the formulas immediately reduce to

$$\tau(X,Y) = -1 + 4\mathrm{E}\{H(X,Y)\}, \quad \rho(X,Y) = -3 + 12\mathrm{E}\{H(X^*,Y^*)\},$$

which further reduce to (4) *and* (5)*, upon making the margins uniform.*

Interestingly, there exists a copula $C_H^{\circledast} \in \mathcal{C}_H$ such that

$$\tau(X,Y) = \tau(C_H^{\circledast}) \quad \text{and} \quad \rho(X,Y) = \rho(C_H^{\circledast}). \tag{7}$$

As shown, e.g., in Schweizer and Sklar (1974) or Whitt (1976), this copula can be constructed in a unique way for any pair (X,Y) with margins F and G. It is obtained by linear interpolation of the distribution D of the pair $(F(X), G(Y))$ in such a way that $C_H^{\circledast}(u,v) = D(u,v)$ for all pairs $(u,v) \in \mathrm{Ran}(F) \times \mathrm{Ran}(G)$.

An algebraic description of $C_H^{\circledast}$ is given in Lemma 2.3.5 of Nelsen (1999). The following alternative definition stems from the work of Denuit and Lambert (2005) and Mesfioui and Tajar (2005).

Definition 2. *The bilinear extension $C_H^{\circledast} \in \mathcal{C}_H$ of a bivariate distribution function H on $\mathbb{N}^2$ is the unique copula associated with the pair $(X^{\circledast}, Y^{\circledast})$ defined by $X^{\circledast} = X + U$, $Y^{\circledast} = Y + V$ where U, V are independent uniform random variables on $(0,1)$.*

For various proofs of identity (7), see Denuit and Lambert (2005), Mesfioui and Tajar (2005) or Nešlehová (2007).

4.4. Perfect monotone dependence does not imply $|\tau| = |\rho| = 1$

It was already mentioned in Section 4.1 that when X and Y are discrete random variables, the Fréchet-Hoeffding bounds do not necessarily induce perfect monotone functional dependence. However, one might wonder whether the following equivalences remain valid in that case:

$$G(Y) = 1 - F(X) \iff \tau(X,Y) = \rho(X,Y) = -1 \text{ a.s.},$$
$$G(Y) = F(X) \iff \tau(X,Y) = \rho(X,Y) = 1 \text{ a.s.}$$

While these equivalences hold true when X and Y are continuous random variables, the following example shows that they break down in the discrete case. As a result, not only do $\tau(X,Y)$ and $\rho(X,Y)$ depend on the marginal distributions when the latter are discrete, but their range is also a function thereof.

Example 8. *Let X and Y be Bernoulli random variables such that* $\mathrm{P}(X=0) = \mathrm{P}(Y=0) = \mathrm{P}(X=0, Y=0) = p \in (0,1)$. *Then $Y = X$ almost surely and* $\tau(X,Y) = \rho(X,Y) = p(1-p) < 1$. *Similarly, if* $\mathrm{P}(X=0) = p = 1 - q = \mathrm{P}(Y=1)$ *and* $\mathrm{P}(X=0, Y=0) = r = 0$, *then $Y = 1 - X$ almost surely and* $\tau(X,Y) = \rho(X,Y) = -p(1-p) > -1$.

4.5. Rescaled versions of τ and ρ may be more relevant

The dependence of Kendall's τ and Spearman's ρ on the margins may be regarded as a consequence of two facts. On one hand, if (X_1, Y_1) and (X_2, Y_2) are two independent observations from H, there is a non-zero probability that a tie occurs, i.e., $X_1 = X_2$ or $Y_1 = Y_2$ or both. On the other hand, the probabilistic definitions of $\tau(H)$ and $\rho(H)$ do not account for ties.

As a further manifestation of the presence of ties, the range of τ and ρ is variable in the discrete case. This makes their interpretation much less straightforward than when the variables X and Y are continuous. This problem was recognized very early on; see, e.g., Hoeffding (1940) or Kendall (1945). The following rescaled versions of Kendall's τ and Spearman's ρ were introduced in an attempt to correct it.

Definition 3. *Let (X,Y) be a pair of discrete random variables with margins $F(x) = \mathrm{P}(X \leq x)$ and $G(y) = \mathrm{P}(Y \leq y)$ for all $x, y \in \mathbb{R}$. Kendall's "tau-b" is then given by*

$$\tau_b(X,Y) = \frac{\tau(X,Y)}{\sqrt{\mathrm{P}(X_1 \neq X_2)\,\mathrm{P}(Y_1 \neq Y_2)}},$$

where X_1 and X_2 are two independent copies of X, while Y_1 and Y_2 are two independent copies of Y. Furthermore, Spearman's "grade correlation coefficient" may be defined as

$$\rho_s(X,Y) = \frac{\rho(X,Y)}{12\sqrt{\mathrm{var}(\bar{X})\,\mathrm{var}(\bar{Y})}},$$

where $\bar{X} = \{F(X) + F(X_-)\}/2$ and $\bar{Y} = \{G(Y) + G(Y_-)\}/2$.

Note that when X and Y are continuous, $\mathrm{P}(X_1 \neq X_2) = \mathrm{P}(Y_1 \neq Y_2) = 1$ and hence $\tau_b(X,Y) = \tau(X,Y)$. Furthermore, the unique copula C associated with H is then the joint distribution of the pair $(\bar{X}, \bar{Y})$, so that $\rho_s(X,Y) = \rho(X,Y)$.

Remark 1. *Nešlehová (2007) shows that one can also write $\rho_s(X,Y) = \mathrm{corr}(\bar{X},\bar{Y})$.*

The normalizations used in Definition 3 are obviously not the only possible ones; see, e.g., Denuit and Lambert (2005) for an alternative. As argued by Nešlehová (2007), however, an argument in favor of the above is that they coincide with the empirical definitions of τ_b and ρ_s when the sample is treated as a finite population.

4.6. Perfect monotone dependence does not always imply $|\tau_b| = |\rho_s| = 1$ either

While rescaling Kendall's τ and Spearman's ρ broadens their range of values by correcting for ties, it may still happen that $|\tau_b(X,Y)| < 1$ and $|\rho_s(X,Y)| < 1$ when discrete variables X and Y are in perfect monotone dependence. The following example illustrates this fact.

Example 9. *Let X and Y be uniform distributions on the sets $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$, respectively. The upper and lower bounds on $\tau(X,Y)$ are displayed in Fig. 5*
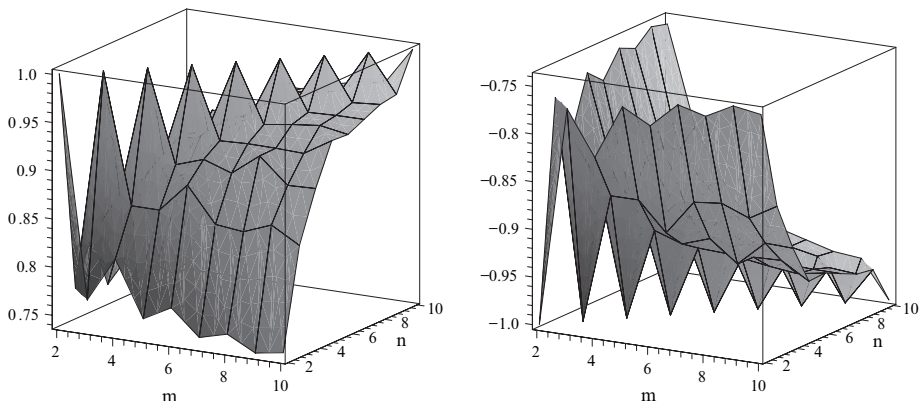

FIGURE 5: Value of τ_b as a function of m and n for co- (left panel) and counter-monotonic (right panel) uniform random variables $X \in \{1, \ldots, m\}$ and $Y \in \{1, \ldots, n\}$.

for all values of $m, n \in \{2, \ldots, 10\}$. In the graphs, the values at non-integer values of m and n were obtained by linear interpolation; they do not correspond to anything real but help to interpret the overall pattern.

Note that in the special case $m = n$, the bounds on τ_b are exactly ± 1. This is in accordance with a result of Mesfioui and Tajar (2005), who show that in this case, $\tau(X, Y) \in [-1 + 1/n, 1 - 1/n]$. The upper and lower bounds are reached when $Y = X$ and $Y = n + 1 - X$, respectively. This clearly implies that $|\tau_b(X, Y)| = 1$ when X and Y are in perfect monotone dependence, because $\mathrm{P}(X_1 = X_2) = \mathrm{P}(Y_1 = Y_2) = 1/n$ and hence the denominator of τ_b equals $(n-1)/n$ in that case. Similar results hold for ρ_s, as can readily be checked.

For additional illustrations of this phenomenon, refer to Denuit and Lambert (2005) and Nešlehová (2007), respectively. The final result of this section, excerpted from the latter paper, delineates the conditions under which $|\tau_b| = |\rho_s| = 1$.

Proposition 10. *Let (X, Y) be a random pair with distribution function H. Then $|\tau_b| = |\rho_s| = 1$ if and only if $X = \varphi(Y)$ almost surely, where φ is a strictly monotone mapping on $\mathrm{Ran}(Y)$.*

In a further attempt to correct for ties, Vandenhende and Lambert (2003) recently proposed several new indices of association whose functional form distinguishes between positive and negative degrees of dependence. For example, the index of concordance that they propose is based on a normalization of

$$\tau^*(X, Y) = \mathrm{E}\{H(X, Y)\} - \mathrm{E}\{F(X)\}\,\mathrm{E}\{G(Y)\}.$$

More specifically, they define

$$\tau_{VL} = \begin{cases} \dfrac{\tau^*(X, Y)}{\mathrm{E}\{F(X)\}\,\mathrm{E}\{G(Y)\} - \mathrm{E}\{H^-(X, Y)\}} & \text{if } \tau^*(X, Y) < 0, \\[3ex] \dfrac{\tau^*(X, Y)}{\mathrm{E}\{H^+(X, Y)\} - \mathrm{E}\{F(X)\}\,\mathrm{E}\{G(Y)\}} & \text{if } \tau^*(X, Y) > 0. \end{cases}$$

where $H^-(x, y) = W(F(x), G(y))$ and $H^+(x, y) = M(F(x), G(y))$, as in (6).

The downside is that $\tau^*(X, Y) \neq \tau(X, Y)$, i.e., τ_{VL} is no longer interpretable as a scaled difference between the probabilities of concordance and discordance.

An alternative measure that preserves this interpretation while achieving the bounds ± 1 in cases of co- and counter-monotonicity is offered below.

Definition 4. *Let X and Y be random variables with distributions F and G, respectively. Let H be the joint distribution of the pair (X, Y). Let also (X^-, Y^-) and*

(X^+, Y^+) *be random pairs with distributions* H^- *and* H^+, *respectively. A concordance measure is then given by*

$$\tau_{CJ} = \begin{cases} -\tau(X, Y)/\tau(X^-, Y^-) & \text{if } \tau(X, Y) < 0, \\ \tau(X, Y)/\tau(X^+, Y^+) & \text{if } \tau(X, Y) > 0. \end{cases}$$

It is immediate that $|\tau_{CJ}| = 1$ if X and Y are either co- or counter-monotonic. It also follows from equation (7) that τ_{CJ} can be alternatively expressed in terms of copulas as follows:

$$\tau_{CJ} = \begin{cases} -\tau\left(C_H^{\circledast}\right)/\tau\left(C_{H^-}^{\circledast}\right) & \text{if } \tau\left(C_H^{\circledast}\right) < 0, \\ \tau\left(C_H^{\circledast}\right)/\tau\left(C_{H^+}^{\circledast}\right) & \text{if } \tau\left(C_H^{\circledast}\right) > 0. \end{cases}$$

Needless to say, a similar strategy could be used to construct alternatives to Spearman's rho, etc. The properties of such measures may be the object of future work.

5. ARE COPULA MODELS FOR DISCRETE DATA INTERESTING AT ALL?

The accumulation of evidence in Section 4 may conduce the belief that models of the form (1) with $C \in (C_\theta)$ are totally inappropriate for describing the dependence between two discrete random variables X and Y. This view is misguided on many accounts, as highlighted below.

   In addition to the fact that construction (1) yields *bona fide* bivariate distributions with given margins F and G, it is seen in Section 5.1 that H inherits most of the common dependence properties of the copula $C \in (C_\theta)$ from which it is derived. As explained in Section 5.2, the dependence parameter θ continues to govern the association between the variables X and Y. However, Section 5.3 emphasizes that this parameter is not a natural object to interpret.

5.1. The dependence properties of C_θ are often inherited by H

Beyond their specific algebraic form, copula families are often chosen because of the nature of association that they induce between variables X and Y. In the continuous case, say, a class of copulas might be deemed appropriate because when $(U, V) = (F(X), G(Y))$ is distributed as $C \in (C_\theta)$,

$$4P(X \le m_X, Y \le m_Y) - 1 = 4P(U \le 1/2, V \le 1/2) - 1 \ge 0,$$

where m_X and m_Y are the medians of X and Y, respectively. This amounts to requiring that Blomqvist's median correlation coefficient between X and Y is positive whatever $C \in (C_\theta)$, a fact that might be denoted $\text{DEP}_0(X, Y)$.

Now suppose that X and Y are integer-valued variables and that their joint distribution H is given by a model of the form (1). One might then wonder whether

$$\mathrm{DEP}_0(U,V) \;\Rightarrow\; \mathrm{DEP}_0(X,Y),$$

where the right-hand side now means that

$$4\mathrm{P}(X \le m_X,\, Y \le m_Y) - 1 \ge 0,$$

with m_X and m_Y being the smallest integers such that $\mathrm{P}(X \le m_X) \ge 1/2$ and $\mathrm{P}(Y \le m_Y) \ge 1/2$.

The answer is trivially yes, because

$$\mathrm{P}(X \le m_X,\, Y \le m_Y) \ge C\left(\frac{1}{2}, \frac{1}{2}\right) \ge \frac{1}{4}.$$

The reverse implication is not valid, however, as the following simple example shows.

Example 10. *Let X and Y be Bernoulli random variables such that $\mathrm{P}(X=0) = \mathrm{P}(Y=0) = 0.7$ and $\mathrm{P}(X=0, Y=0) = 0.4$. Then $m_X = m_Y = 0$ and $\mathrm{P}(X=0, Y=0) > 1/4$, so that $\mathrm{DEP}_0(X,Y)$ holds true. However, $H(0,0) = W(0.7, 0.7)$ and hence H can be regarded as being generated via the copula model (1) with $C = W$, the Fréchet-Hoeffding lower bound. Unfortunately, $W(1/2, 1/2) = 0 < 1/4$, i.e., one does not have $\mathrm{DEP}_0(U,V)$ for a pair (U,V) distributed as W.*

In practice, of course, copula models satisfy much stronger concepts of positive dependence. Most common are the following notions, introduced by Lehmann (1966), as well as Esary and Proschan (1972).

Definition 5. *Let (X,Y) be a pair of random variables. Then*

(a) *X and Y are said to be in positive quadrant dependence, denoted $\mathrm{DEP}_1(X,Y)$, if and only if $\mathrm{P}(X \le x, Y \le y) \ge \mathrm{P}(X \le x)\,\mathrm{P}(Y \le y)$ for all $x,y \in \mathbb{R}$.*

(b) *Y is said to be left-tail decreasing in X, denoted $\mathrm{DEP}_2(X,Y)$, if and only if $\mathrm{P}(Y \le y \mid X \le x)$ is a non-increasing function of x for every $y \in \mathbb{R}$.*

(c) *Y is said to be right-tail increasing in X, denoted $\mathrm{DEP}_3(X,Y)$, if and only if $\mathrm{P}(Y \le y \mid X > x)$ is a non-decreasing function of x for every $y \in \mathbb{R}$.*

(d) *Y is said to be stochastically increasing in X, denoted $\mathrm{DEP}_4(X,Y)$, if and only if $\mathrm{P}(Y \le y \mid X = x)$ is a non-increasing function of x for every $y \in \mathbb{R}$.*

(e) *X and Y are said to be in positive likelihood ratio dependence, denoted $\mathrm{DEP}_5(X,Y)$, if and only if for all $x_1 < x_2$, $y_1 < y_2$,*

$$h(x_1, y_1)\, h(x_2, y_2) \ge h(x_1, y_2)\, h(x_2, y_1),$$

where $h(x,y)$ stands for the probability function or the density of the pair (X,Y) as the case may be. (In the continuous case, the concept only makes sense if the density function exists.)

The following proposition shows that as in the case of Blomqvist's median correlation dependence, each of the above concepts is preserved in a copula model construction. As in Example 10, however, none of the implications can be reversed, except in the special case of the bilinear extension copula.

Proposition 11. *Let X and Y be random variables with marginal distribution functions F and G, respectively. Let C be a copula such that* $P(X \leq x, Y \leq y) = C(F(x), G(y))$ *for all* $x, y \in \mathbb{R}$.

(a) *If* (U, V) *denotes a random pair with joint distribution C, then* $\mathrm{DEP}_i(U, V) \Rightarrow \mathrm{DEP}_i(X, Y)$ *for* $i = 1, \ldots, 5$.

(b) *Let* $C_H^{\maltese} \in \mathcal{C}_H$ *be the bilinear extension of H from Definition 2. If* $(X^{\maltese}, Y^{\maltese})$ *is as in Definition 2, then* $\mathrm{DEP}_i(X, Y) \Leftrightarrow \mathrm{DEP}_i(X^{\maltese}, Y^{\maltese})$ *for* $i = 1, \ldots, 5$.

Proof. For DEP_1, statement (a) is an observation of Marshall (1996) and statement (b) is a special case of Proposition 13 below, due to Nešlehová (2004). For $i = 2, 3$, the claims concerning DEP_i are proved by Mesfioui and Tajar (2005). As for the results pertaining to DEP_4 and DEP_5, they are shown in Appendix E. □

Another important concept used in practice is the notion of tail dependence, introduced by Joe (1993). The following result, whose proof is given in Appendix F, shows that tail dependence is also preserved in a copula model construction, at least under the restriction that the marginal distributions coincide.

Proposition 12. *Let X and Y be random variables with margins F and G, respectively. Assume that* $H(x, y) = P(X \leq x, Y \leq y) = C(F(x), G(y))$ *for all* $x, y \in \mathbb{R}$ *and some copula C. Let* (U, V) *be a random pair distributed as C. If* $F = G$, *then*

$$\lim_{u \uparrow 1} P(V > u \,|\, U > u) = \lim_{t \uparrow 1} P\{Y > G^{-1}(t) \,|\, X > F^{-1}(t)\}, \tag{8}$$

provided that the limit on the left exists. In that case, therefore, H has upper tail dependence as soon as C does.

Whether the result extends to the general case is not clear, because after discretization, the right-hand side of (8) amounts to the limit of $C(u_n, v_n)$ for a sequence $(u_n, v_n) \to (1, 1)$ as $n \to \infty$, but for which $u_n \neq v_n$.

5.2. θ can still be interpreted as a dependence parameter

In order for a model of the form (1) to be statistically meaningful, a fundamental requirement is that the members of the copula family (C_θ) be ordered by positive quadrant dependence, i.e.,

$$\theta < \theta' \Rightarrow C_\theta(u, v) \leq C_{\theta'}(u, v) \tag{9}$$

holds true for all $u, v \in (0,1)$. Tchen (1980) showed that this condition, denoted $C_\theta \prec_{\mathrm{PQD}} C_{\theta'}$, implies

$$\kappa(C_\theta) \le \kappa(C_{\theta'})$$

for any measure of concordance in the sense of Scarsini (1984). The latter is a minimal requirement in order for θ to be interpreted as a dependence parameter.

Fortunately, this property is inherited by the parametric family (H_θ) defined by

$$H_\theta(x,y) = C_\theta(F(x), G(y)), \ x, y \in \mathbb{R}.$$

In other words,

$$C_\theta \prec_{\mathrm{PQD}} C_{\theta'} \ \Rightarrow \ H_\theta \prec_{\mathrm{PQD}} H_{\theta'}.$$

This implication holds whether the margins F and G are continuous or not. The following result, whose proof is given in Section 5.3 of Nešlehová (2004), provides a partial converse.

Proposition 13. *Let H and L be two bivariate distribution functions with the same margins. Let $C_H^{\maltese}$ and $C_L^{\maltese}$ be their bilinear extensions, as per Definition 2. Then*

$$H \prec_{\mathrm{PQD}} L \ \Leftrightarrow \ C_H^{\maltese} \prec_{\mathrm{PQD}} C_L^{\maltese}.$$

It is unclear whether similar statements hold for stronger stochastic orderings, e.g., the LTD and RTI partial orders of Avérous and Dortet-Bernadet (2000), let alone the monotone regression dependence ordering of Yanagimoto and Okamoto (1969) and its extensions considered by Capéraà and Genest (1990) or Fang and Joe (1992).

5.3. $\tau(C_\theta) \ne \tau(H)$ and $\rho(C_\theta) \ne \rho(H)$

Suppose that the joint distribution H of a random pair (X, Y) is of the form (1), where C belongs to a parametric class of copulas (C_θ). If the latter satisfies condition (9), there is then a one-to-one correspondence between θ and $\tau(C_\theta)$, as well as between θ and $\rho(C_\theta)$.

When the margins F and G of X and Y are continuous, it is well known that

$$\tau(C_\theta) = \tau(H), \rho(C_\theta) = \rho(H).$$

This, in turn, provides a natural interpretation for θ, either in terms of Kendall's τ or Spearman's ρ. In the event that X and Y are discrete random variables, however, it was mentioned in Section 4 that

$$\tau(X,Y) \ne \tau(C_\theta) \ne \tau_b(X,Y) \ \text{and} \ \rho(X,Y) \ne \rho(C_\theta) \ne \rho_s(X,Y).$$

Furthermore, there is no simple way of expressing the relation between Kendall's τ or Spearman's ρ for C_θ and either one of the dependence measures based on (X, Y). Recall from Section 4.2 that the latter all depend on the marginal distributions.

This lack of direct interpretability of the parameter θ is not a serious concern when the copula model for the pair (X, Y) is used, say, to investigate the robustness of a procedure under various scenarios of dependence. As seen in the next section, however, it calls for additional care when time comes to estimate the parameter θ.

6. DOING INFERENCE ON THE BASIS OF A DISCRETE COPULA MODEL

Let us now return to the initial estimation problem in the context of a copula model for count data. To be specific, suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are $n \geq 2$ independent copies of a pair (X, Y) of integer-valued random variables with joint distribution H. Assuming that H is of the form (1) for some copula $C \in (C_\theta)$, what are the implications of the accumulated wisdom on an appropriate course of action for estimating the dependence parameter θ?

When the variables X and Y are continuous, the estimation of θ from the data does not pose any major challenge. If one is willing to assume parametric forms for the margins F and G, one can either maximize the full likelihood or proceed in two steps, i.e., estimate the margins first, and subsequently consider them as given in the maximization of the likelihood for θ. The relative merits of these two approaches are described and compared by Joe (1997, 2005).

An alternative approach advocated by Genest et al. (1995) and Shih and Louis (1995) also proceeds in two steps but replaces parametric estimates of F and G by their empirical distribution functions; the latter are often multiplied by $n/(n + 1)$ to avoid evaluating the density $c_\theta(u, v)$ at the edges of the unit square. This approach is generally preferred in practice because in addition to being reasonably efficient, the resulting estimator for θ is rank-based and hence robust to the misspecification of the marginal distributions, as recently illustrated by Kim et al. (2007).

In many applications, however, a simpler technique is preferred. It is based on an interpretation of $\theta \in \mathbb{R}$ as a function of some concordance measure. For instance, simple formulas are available for Kendall's tau in Archimedean and meta-elliptical models; for the basic properties of these two classes of distributions, see, e.g., Chapter 4 of Nelsen (1999) and Chapter 5 of McNeil et al. (2005), respectively.

Assuming that the mapping $\tau : \theta \mapsto \tau(C_\theta)$ is one-to-one, a method-of-moment estimator of θ is obtained by finding the unique value $\hat{\theta}_n$ of θ such that

$$\tau(C_\theta) = \tau_n(X, Y).$$

Here, τ_n denotes the standard sample value of Kendall's τ, defined by

$$\tau_n(X,Y) = \frac{N_c - N_d}{\binom{n}{2}} \tag{10}$$

as a function of the numbers of concordant and discordant pairs in the sample, viz.

$$N_c = \sum_{1 \le i < j \le n} \mathbf{1}\big\{(X_i - X_j)(Y_i - Y_j) > 0\big\},$$

$$N_d = \sum_{1 \le i < j \le n} \mathbf{1}\big\{(X_i - X_j)(Y_i - Y_j) < 0\big\}.$$

For instance if (C_θ) is the Clayton family defined by (2), then $\tau(C_\theta) = \theta/(\theta + 2)$ and

$$\hat{\theta}_n = 2\,\frac{\tau_n(X,Y)}{1 - \tau_n(X,Y)}.$$

In the continuous case, the estimator $\hat{\theta}_n$ of θ based on the inversion of Kendall's tau is known to be consistent and asymptotically normal. Section 6.1 shows that the situation is radically different in the discrete case. As emphasized in Section 6.2, however, this is not to say that θ cannot be estimated.

6.1. Estimates of θ based on Kendall's τ are biased

When data $(X_1, Y_1), \ldots, (X_n, Y_n)$ are assumed to arise from a copula model of the form (1) for some $C \in (C_\theta)$, each observation (X_i, Y_i) can be construed as arising from a latent pair (U_i, V_i), viz.

$$X_i = F^{-1}(U_i),\ Y_i = G^{-1}(V_i),$$

where $(U_1, V_1), \ldots, (U_n, V_n)$ is a random sample from copula C. When F and G are continuous, it was seen in Section 4 that $\tau(H) = \tau(C)$. Furthermore,

$$\tau_n(X,Y) = \tau_n(U,V) \tag{11}$$

because the quantile functions F^{-1} and G^{-1} are strictly monotone. For this reason, $\tau_n(X,Y)$ is an unbiased estimate of $\tau(C_\theta)$, even though the pairs (U_i, V_i) are *not* observable. This holds true whether F and G are known or not.

When F and G are discrete, however, it is not at all clear whether relation (11) still holds. For one thing, given that the quantile functions are not strictly monotone, there is now a non-zero probability that $(X_i - X_j)(Y_i - Y_j) = 0$ for some $i \ne j$. At a more fundamental level, however, the relationship between (X_i, Y_i) and (U_i, V_i) is no longer one-to-one, even when F and G are known. As a result, the discretization process is irreversible, whatever the sample size!

Notwithstanding this fact, an optimistic user might be tempted to estimate θ by solving the equation

$$\tau(C_\theta) = \tau_n(X, Y).$$

Or perhaps, in the light of the discussion in Section 4, this person might attempt to improve the estimate of $\tau(H)$ by correcting for ties, considering that

(i) unless the definitions of N_c and N_d are modified, ties observed in the sample do not contribute to the numerator of (10);

(ii) treating each tie as a half concordance and a half discordance does not affect the numerator of τ_n either;

(iii) a random allocation of ties to the N_c or N_d counts would be unacceptable, because two analysts would not always get the same solution.

Standard remedies consist of correcting the denominator of $\tau_n(X, Y)$, e.g., by letting

$$\tau_{a,n}(X, Y) = \frac{N_c - N_d}{N_c + N_d} \quad \text{or} \quad \tau_{b,n}(X, Y) = \frac{N_c - N_d}{\sqrt{N_x \, N_y}},$$

where $N_x = \sum_{i < j} \mathbf{1}(X_i \neq X_j)$ and $N_y = \sum_{i < j} \mathbf{1}(Y_i \neq Y_j)$. The first coefficient is often referred to as "gamma" in the categorical analysis literature; it was introduced by Goodman and Kruskal (1954). As for $\tau_{b,n}$, which is a consistent estimate of τ_b, it is often attributed to Kendall.

While $\tau_n(X, Y)$, $\tau_{a,n}(X, Y)$ and $\tau_{b,n}(X, Y)$ all coincide with the usual empirical version of Kendall's τ in the absence of ties, the following examples show that the mere fact that ties *could* occur with probability greater than zero causes these estimators to have different behavior, both in finite samples and asymptotically. In particular, all of them lead to biased estimates of θ when the marginal distributions are discrete.

Example 11. *Ten thousand random samples of size $n = 100$ were generated from a pair (X, Y) of geometric random variables whose joint distribution H is of the*

TABLE 6

AVERAGE AND STANDARD ERROR OF FOUR ESTIMATES OF θ IN EXAMPLE 11.

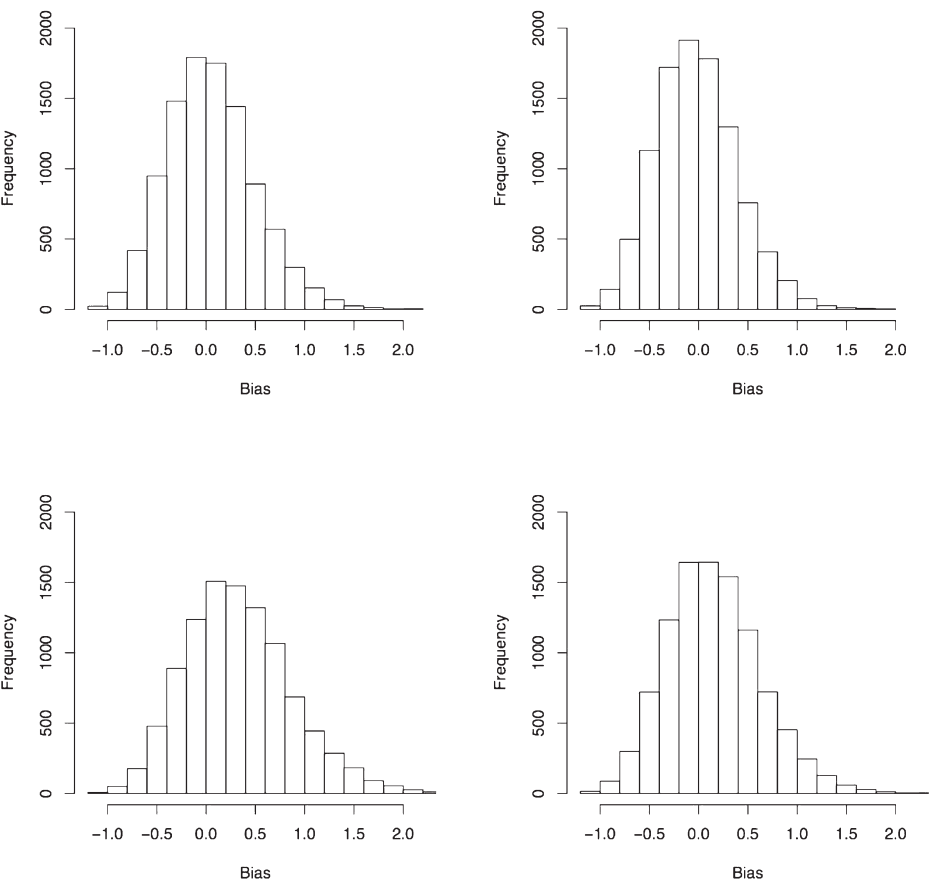|	Estimate of θ based on			
	$\tau_n(U, V)$	$\tau_n(X, Y)$	$\tau_{a,n}(X, Y)$	$\tau_{b,n}(X, Y)$
Average	2.054	1.979	2.331	2.153
Standard deviation	0.444	0.412	0.542	0.476

FIGURE 6: Histograms showing, from left to right and top to bottom, the bias of estimators of θ based on $\tau_n(U,V)$, $\tau_n(X,Y)$, $\tau_{a,n}(X,Y)$ and $\tau_{b,n}(X,Y)$ in 10,000 random samples of size $n = 100$ from the bivariate geometric distribution described in Example 11.

form (1). *It was assumed more specifically that* $P(X = 0) = 0.05$, $P(Y = 0) = 0.1$ *and that* $C \in (C_\theta)$, *the Clayton family of copulas defined by* (2). *In the simulation,* θ *was taken equal to* 2, *so that* $\tau(C_\theta) = 1/2$.

Displayed in panels (b)-(d) *of Fig. 6 are histograms of the observed bias of estimators of* θ *based on* $\tau_n(X,Y)$, $\tau_{a,n}(X,Y)$ *and* $\tau_{b,n}(X,Y)$. *For comparison purposes, panel* (a) *shows the bias of the estimator of* θ *derived from* $\tau_n(U,V)$, *the empirical value of Kendall's* τ *based on the continuous bivariate observations generated from copula C, prior to the transformation of their margins.*

In practice, of course, $\tau_n(U,V)$ *is not a possible estimator for* τ, *because the process by which the data were made discrete is irreversible. Complementary numerical information about the mean and standard deviation of each of the four estimators is provided in Table 6.*

In his/her anonymous comments about the first draft of this article, a referee suggested that one possible way of circumventing the discrete nature of the data would be to add independent random components to the coordinates of each pair (X_i, Y_i) by setting

$$\tilde{X}_i = X_i + U_i, \quad \tilde{Y}_i = Y_i + V_i,$$

for each $i \in \{1, \ldots, n\}$, where $U_1, \ldots, U_n$ and $V_1, \ldots, V_n$ are independent random samples from the uniform distribution on the interval $(0,1)$. An estimate of θ might then be given by

$$\theta_n = \tau^{-1}\{\tau_n(\tilde{X}, \tilde{Y})\}.$$

In recognition of the fact that the sample $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_n, \tilde{Y}_n)$ was obtained via randomization, the process might be replicated a large number of times, say N, and an average of the resulting values of τ_n might be used in the above inversion formula. As the following example shows, this randomization strategy does not yield any improvement over the previously considered moment estimators.

Example 12. *The Monte Carlo experiment of Example 11 was repeated in the case where (X, Y) is a pair of Poisson random variables with $E(X) = 1$ and $E(Y) = 2$. As can be seen from Fig. 7 and Table 7, the bias of the three estimators based on the observations of (X, Y) is larger here than in Example 11.*

In addition, Fig. 8 illustrates the behavior of the estimator of θ based on $\tau_n(\tilde{X}, \tilde{Y})$. The left panel displays its bias, when the randomization process is performed only once, i.e., $N = 1$. A comparison of this graph with the upper-right histogram in Fig. 7 reveals that the inversion of $\tau_n(X, Y)$ and $\tau_n(\tilde{X}, \tilde{Y})$ produces essentially the same bias. This observation is confirmed by Table 7.

Upon reflection, this is not surprising. For, when the process of generating a randomized sample is repeated indefinitely, the average of the resulting values of $\tau_n(\tilde{X}, \tilde{Y})$ becomes a consistent estimator of $\tau_n(X, Y)$. This fact is illustrated in the right panel of Fig. 8 for a single sample of size 100 from the bivariate Poisson model specified above. In that particular sample, $\tau_n(X, Y)$ was 0.3307, as compared to 0.3306 for the average value of $\tau_n(\tilde{X}, \tilde{Y})$.

TABLE 7

AVERAGE AND STANDARD ERROR OF FOUR ESTIMATES OF θ IN EXAMPLE 12

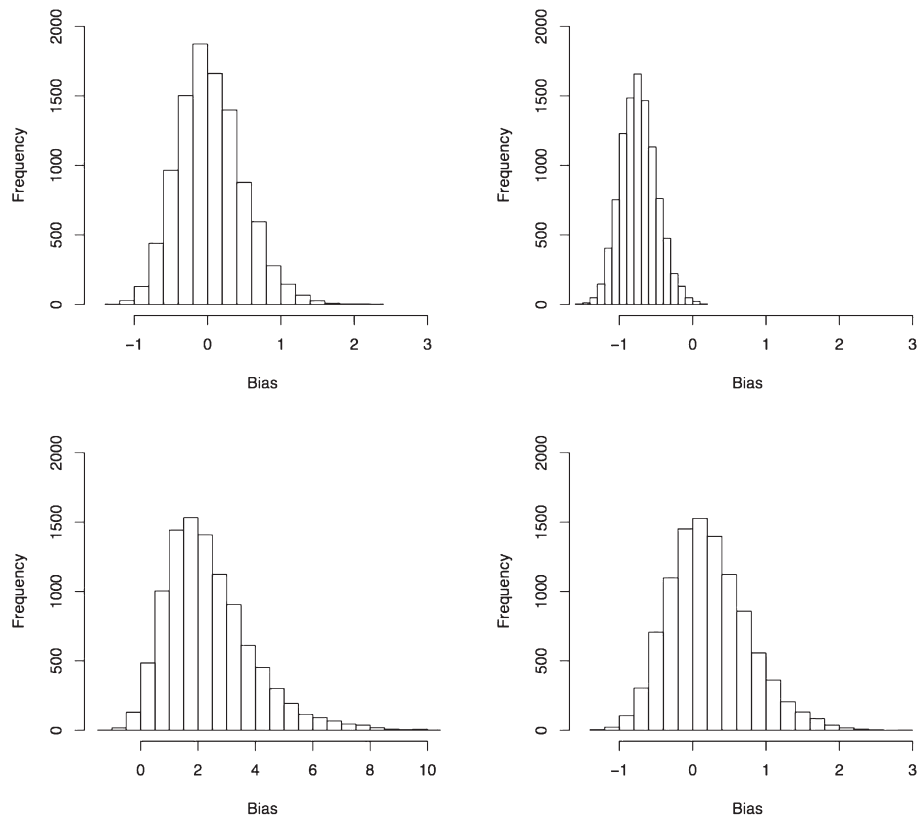|	Estimate of θ based on				
	$\tau_n(U, V)$	$\tau_n(X, Y)$	$\tau_{a,n}(X, Y)$	$\tau_{b,n}(X, Y)$	$\tau_n(\tilde{X}, \tilde{Y})$
Average	2.039	1.262	4.358	2.213	1.269
Standard deviation	0.446	0.243	1.495	0.537	0.285

FIGURE 7: Histograms showing, from left to right and top to bottom, the bias of estimators of θ based on $\tau_n(U,V)$, $\tau_n(X,Y)$, $\tau_{a,n}(X,Y)$ and $\tau_{b,n}(X,Y)$ in 10,000 random samples of size $n = 100$ from the bivariate Poisson distribution described in Example 12.
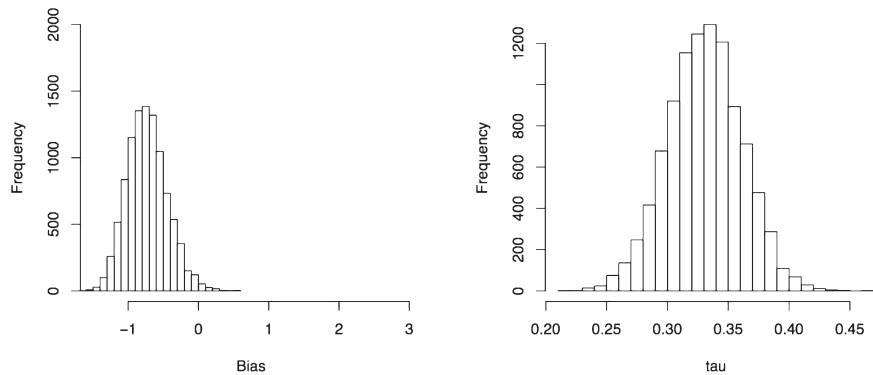

FIGURE 8: Histogram showing the bias of a randomized-based estimator of θ derived from $\tau_n(\tilde{X}, \tilde{Y})$ in 10,000 random samples of size $n = 100$ from the bivariate Poisson distribution described in Example 12.

Unless a large number of replicates is taken, however, it is also apparent from Table 7 that inversion of $\tau_n(\tilde{X}, \tilde{Y})$ is less efficient than inversion of $\tau_n(X, Y)$. Clearly, therefore, the randomization-based estimation procedure is not advisable.

As it turns out, this recommendation extends to other estimation procedures based on randomization. In the present simulation experiment, e.g., the estimator maximizing the pseudo-likelihood of the randomized sample $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_n, \tilde{Y}_n)$ exhibited a bias of 1.144 and a mean squared error of 1.339, as compared to $0.604 = (2 - 1.262)^2 + 0.243^2$ for the estimator based on $\tau_n(X, Y)$.

Remark 2. *The perceptive reader will of course have noticed that the formal reason for the poor performance of the randomization-based estimator of θ is that for n large, the pairs of normalized ranks associated with $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_n, \tilde{Y}_n)$ amount to a random sample from $C_H^{\circledast}$. As in general $C_H^{\circledast} \neq C_\theta$, rank-based estimation from such randomized samples is doomed to failure.*

6.2. Is θ estimable?

It should not be concluded from the illustrations in Section 6.1 that when the distribution of a pair (X, Y) of discrete random variables arises from a copula model of the form (1) with $C \in (C_\theta)$, the dependence parameter θ cannot be estimated. Rather, what these examples show is that rank-based methods in general (and inversion of Kendall's tau in particular) will not always lead to a consistent estimator.

In contrast, the following example shows that in the simplest possible case — that which involves two Bernoulli variables — standard maximum likelihood estimation works. However, the exact conditions under which θ is estimable in this context remain to be elucidated.

Example 13. *Let (X, Y) be a pair of Bernoulli random variables with $P(X = 0) = p$, $P(Y = 0) = q$ and $P(X = 0, Y = 0) = r$. For $i, j = 1, 2$, let n_{ij} represent the number of times that $X = i$ and $Y = j$ in a random sample of size n. The maximum likelihood estimates of the three parameters are then given respectively by*

$$\hat{p}_n = \frac{n_{00} + n_{01}}{n}, \quad \hat{q}_n = \frac{n_{00} + n_{10}}{n}, \quad \hat{r}_n = \frac{n_{00}}{n}.$$

When the joint distribution of (X, Y) is of the form (1) for some $C \in (C_\theta)$, one has $C_\theta(p, q) = r$, so that the maximum likelihood estimate $\hat{\theta}_n$ is the unique value of θ such that $C_{\hat{\theta}_n}(\hat{p}_n, \hat{q}_n) = \hat{r}_n$. Standard theory then implies that this estimation is consistent and asymptotically normal. In particular, θ is then estimable.

To illustrate this point concretely, suppose that

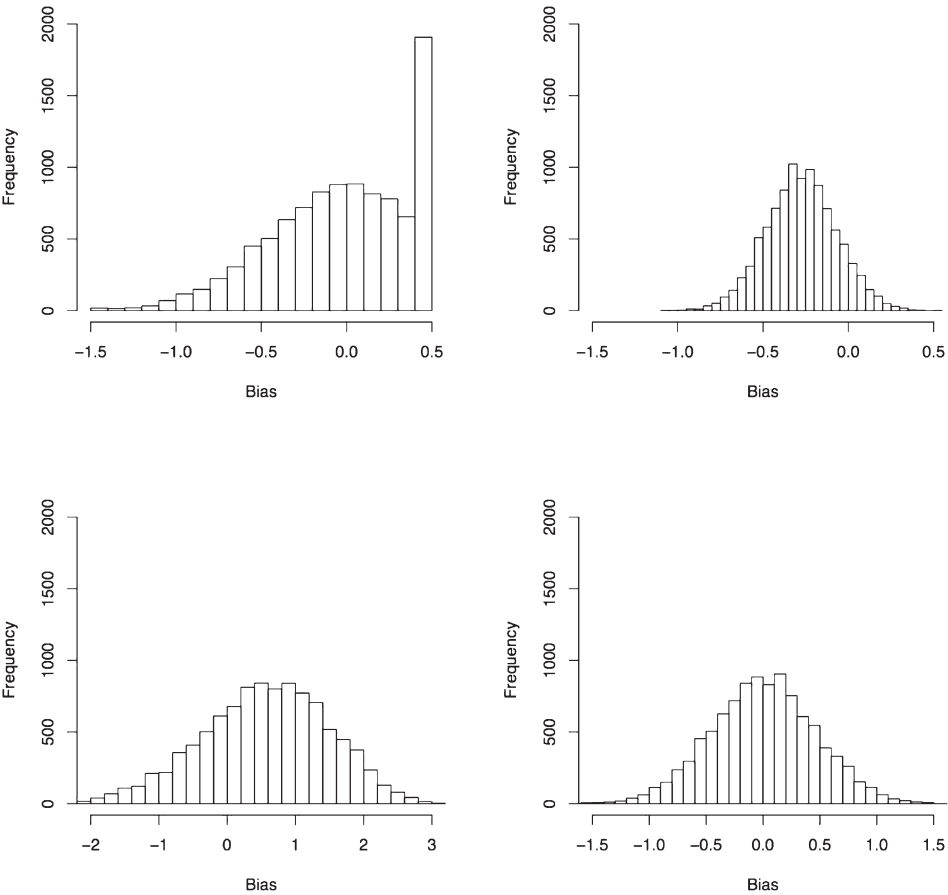$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v), \quad \theta \in [-1, 1]$$

FIGURE 9: Histograms showing, from left to right and top to bottom, the bias of estimators of θ based on maximization of the likelihood and inversion of $\tau_n(X,Y)$, $\tau_{a,n}(X,Y)$ and $\tau_{b,n}(X,Y)$ in 10,000 random samples of size $n = 100$ from the bivariate Bernoulli distribution described in Example 13.

the Farlie-Gumbel-Morgenstern family of copulas. Then

$$\hat{\theta}_n = \frac{\hat{r}_n - \hat{p}_n \hat{q}_n}{\hat{p}_n \hat{q}_n (1 - \hat{p}_n)(1 - \hat{q}_n)}.$$

Panel (a) of Fig. 9 shows the empirical distribution of the above estimate of θ, based on 10,000 random samples of size $n = 100$ from the pair (X,Y) with $P(X=0) = 0.3$, $P(Y=0) = 0.4$ and $P(X=0,Y=0) = r = 0.1452$; this corresponds to $C_\theta(p,q)$ with $\theta = 0.5$. For comparison purposes, panels (b)-(d) show the estimates of θ obtained by inversion of $\tau_n(X,Y)$, $\tau_{a,n}(X,Y)$ and $\tau_{b,n}(X,Y)$ using $\tau(C_\theta) = 2\theta/9$.

7. Conclusion

This paper has reviewed some of the pros and cons of modeling the joint distribution H of a random pair (X, Y) using copulas when the margins F and G are discrete. As explained in Section 2, the main source of difficulties is that while it is always possible to represent H in the form

$$H(x, y) = C(F(x), G(y)), \quad x, y \in \mathbb{R} \tag{12}$$

in terms of a copula C, the latter is not uniquely defined except on $\mathrm{Ran}(F) \times \mathrm{Ran}(G)$. Many copulas are compatible with this representation, and the distribution function of the transformed pair $(F(X), G(Y))$ is not one of them.

Here are, in a nutshell, some important consequences of this unidentifiability issue:

1. The set of copulas that are compatible with (12) can be quite large. This is illustrated in Section 3 in terms of the range of possible values for Kendall's tau and Spearman's rho. The unidentifiability issue is likely to be more severe when the variables X and Y put most of their mass on a few atoms.

2. The dependence between the variables X and Y is no longer a function of the copula alone and as seen in Section 4, the probabilistic and copula-based definitions of classical measures of concordance no longer coincide. Similar discrepancies arise for dependence concepts, as explained in Section 5.

3. Depending on the choice of margins, measures of concordance such as Kendall's tau and Spearman's rho may or may not span the entire interval $[-1, 1]$; in fact, it is even shown in Section 4 that monotone increasing functional dependence between X and Y does not necessarily imply $\tau = \rho = 1$.

4. To account for the discreteness of margins, several variants of Kendall's τ and Spearman's ρ have been proposed in the literature; a new proposal was made in Definition 4. None of them is margin-free, however, and some of them do not reach the bounds ± 1.

5. Finally, inference for the dependence parameter θ under a copula model $C \in (C_\theta)$ should not be founded on the inversion of Kendall's tau, Spearman's rho or their adjusted versions. As vividly illustrated in Section 6, this could induce serious bias in the estimation, not to speak of inconsistency.

This being said, copula modeling remains a valid option for constructing multivariate distributions with discrete margins. As emphasized in Section 5, many stochastic dependence properties of copulas are inherited by such models. Through the preservation of stochastic ordering relations such as $\prec_{\mathrm{PQD}}$, the interpretation of θ as a dependence parameter of a copula model $C \in (C_\theta)$ is also maintained, even when discrete margins are used.

In short, copula models provide a viable approach to the construction of multivariate distributions with given margins, even in the discrete case. They can be recommended without qualification for elaborating counterexamples

or generating alternatives in the context of simulations or robustness studies. When dealing with count data, however, modeling and interpreting dependence through copulas is subject to caution. Furthermore, inference (and particularly rank-based inference) for copula parameters from discrete data is fraught with difficulties.

Insofar as the dependence parameter θ is identifiable, however, its estimation remains possible via fully parametric maximum likelihood techniques, although exact conditions under which identifiability is guaranteed are yet to be delineated. Once obstacles posed by inference are resolved, and in view of their richness and flexibility, copula-based models are likely to become as attractive for discrete variables as they have grown to be for continuous data.

APPENDIX A: An extension of Proposition 3

This appendix gives a proof of the following extension of Proposition 3.

Proposition 14. *Let H be the distribution of a pair (X,Y) on $\mathbb{N}^2$. If H^* is the distribution of $(\alpha(X),Y)$ for some strictly decreasing function α on $\mathbb{N}$, then*

$$v - C_H^+(1-u,v) \leq C_{H^*}(u,v) \leq v - C_H^-(1-u,v)$$

for all $u,v \in (0,1)$ and all $C_{H^*} \in \mathcal{C}_{H^*}$. The bounds, however, are not those of Carley (2002), simply because α cannot possibly map $\mathbb{N}$ to $\mathbb{N}$.

By Sklar's theorem (Nelsen, 1999, Theorem 2.10.9), all elements of $\mathcal{C}_H$ coincide on the closure of $\mathrm{Ran}(F) \times \mathrm{Ran}(G)$. Let $S_H(u,v)$ be that common value for each $(u,v) \in \overline{\mathrm{Ran}(F)} \times \overline{\mathrm{Ran}(G)}$. For each such pair (u,v), let also $S_{H^*}(1-u,v) = v - S_H(u,v)$.

It is then a simple matter to show that a copula C^* is in $\mathcal{C}_{H^*}$ precisely when $C^*(u,v) = S_{H^*}(u,v)$ for all $(u,v) \in \overline{\mathrm{Ran}(F^*)} \times \overline{\mathrm{Ran}(G)}$, with F^* standing for the distribution of $\alpha(X)$. Thus $C^* \in \mathcal{C}_{H^*}$ if and only if there exists $C \in \mathcal{C}_H$ such that $C^*(u,v) = v - C(1-u,v)$ for all $(u,v) \in [0,1]^2$. This completes the proof. $\qquad\square$

APPENDIX B: Proof of Proposition 4

It is straightforward to check that C_H^- spreads mass uniformly along the line segments connecting the points $(\gamma_{ij}, \delta_{ij} + h_{ij})$ and $(\gamma_{ij} + h_{ij}, \delta_{ij})$. Therefore,

$$\int_0^1 \int_0^1 C_H^-(u,v)\, dC_H^-(u,v) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \int_{\gamma_{ij}}^{\gamma_{ij}+h_{ij}} C_H^-(u, \delta_{ij} + \gamma_{ij} + h_{ij} - u)\, du.$$

Furthermore,

$$C_H^-\left(u, \delta_{ij} + \gamma_{ij} + h_{ij} - u\right) = \sum_{k=0}^{i-1}\sum_{\ell=0}^{j-1} h_{k\ell}$$

for any $u \in [\gamma_{ij}, \gamma_{ij} + h_{ij}]$. Therefore,

$$\int_0^1\int_0^1 C_H^-(u, v)\, dC_H^-(u, v) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\int_{\gamma_{ij}}^{\gamma_{ij}+h_{ij}}\sum_{k=0}^{i-1}\sum_{\ell=0}^{j-1} h_{k\ell}\, du = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{i-1}\sum_{\ell=0}^{j-1} h_{ij}h_{k\ell},$$

which yields the desired result for $\tau(C_H^-)$.

Analogously, C_H^+ spreads its mass uniformly along the line segments connecting the points $(\alpha_{ij}, \beta_{ij})$ and $(\alpha_{ij} + h_{ij}, \beta_{ij} + h_{ij})$. By reference to Fig. 2, it is also easy to convince oneself that $C_H^+(u, u + \beta_{ij} - \alpha_{ij}) = u - A_{ij}$, where

$$A_{ij} = \alpha_{ij} - \sum_{k=0}^{i-1}\sum_{\ell=0}^{j-1} h_{k\ell} - \sum_{k=0}^{i-1} h_{kj} - \sum_{\ell=0}^{j-1} h_{i\ell}.$$

Further note that for fixed i, the non-decreasing sequence (α_{ij}) yields a partition of the interval $[F(i-1), F(i))$. Consequently,

$$\int_0^1\int_0^1 C_H^+(u, v)\, dC_H^+(u, v) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\int_{\alpha_{ij}}^{\alpha_{ij}+h_{ij}}(u - A_{ij})\, du = \frac{1}{2} - \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h_{ij}A_{ij}.$$

The announced formula for $\tau(C_H^+)$ then obtains after a mere substitution, because

$$A_{ij} = \sum_{k=0}^{i-1}\left(\sum_{\ell=0}^{\infty} h_{k\ell} - \sum_{\ell=0}^{j-1} h_{k\ell} - h_{kj}\right) = \sum_{k=0}^{i-1}\sum_{\ell=j+1}^{\infty} h_{ij}h_{k\ell}.$$

$\square$

APPENDIX C: Proof of Proposition 5

Using the identity

$$\int_0^1\int_0^1 (1-u)v\, dC(u, v) = \frac{1}{2} - \int_0^1\int_0^1 uv\, dC(u, v),$$

valid for any copula C, one can proceed as in the proof of Proposition 4 to see that

$$\int_0^1\int_0^1 uv\, dC_H^-(u, v) = \frac{1}{2} - \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\int_{\gamma_{ij}}^{\gamma_{ij}+h_{ij}}(1-u)(\delta_{ij} + \gamma_{ij} + h_{ij} - u)\, du$$

and

$$\int_0^1 \int_0^1 uv dC_H^+(u,v) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \int_{\alpha_{ij}}^{\alpha_{ij}+h_{ij}} u\left(u + \beta_{ij} - \alpha_{ij}\right) du.$$

Exploiting once again the fact that for fixed i, the nondecreasing sequence (α_{ij}) yields a partition of the interval $[F(i-1), F(i))$, one then gets

$$\int_0^1 \int_0^1 uv dC_H^+(u,v) = \frac{1}{3} + \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h_{ij}(\beta_{ij} - \alpha_{ij})(2\alpha_{ij} + h_{ij}).$$

This leads to the stated formula for $\rho(C_H^+)$.

Now for fixed i, the nondecreasing sequence (γ_{ij}) also yields a partition of the interval $[F(i-1), F(i))$. Consequently,

$$\int_0^1 \int_0^1 uv dC_H^-(u,v) = \frac{1}{2} - \frac{1}{3} - \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \int_{1-\gamma_{ij}-h_{ij}}^{1-\gamma_{ij}} v(\delta_{ij} + \gamma_{ij} + h_{ij} - 1)\, dv$$

$$= \frac{1}{6} - \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h_{ij}\left\{2(1-\gamma_{ij}) - h_{ij}\right\}\left\{\delta_{ij} - (1-\gamma_{ij} - h_{ij})\right\},$$

from which the formula for $\rho(C_H^-)$ ensues.   □

APPENDIX D: Nelsen's formulas for computing the Fréchet-Hoeffding bounds in the discrete case

Given the bivariate joint distribution function H of random variables X and Y taking values in $\mathbb{N}$, the following expressions are obtained by Nelsen (1987) for the computation of the probability functions h^- and h^+ corresponding to the Fréchet-Hoeffding bounds H^- and H^+. These are given as functions of the probability functions f and g of the margins F and G, respectively.

Let

$$D = \{(x,y) \in \mathbb{N}^2 : F(x) + G(y) - 1 > 0\},$$
$$\partial D = \{(x,y) \in D : (x-1,y), (x,y-1) \text{ or } (x-1,y-1) \notin D\}.$$

Then

$$h^-(x,y) = \begin{cases} f(x) & \text{if } (x,y) \in \partial D, (x,y-1) \notin D, (x-1,y) \in \partial D, \\ g(y) & \text{if } (x,y) \in \partial D, (x-1,y) \notin D, (x,y-1) \in \partial D, \\ F(x) + G(y) - 1 & \text{if } (x,y) \in \partial D, (x,y-1) \notin D, (x-1,y) \notin \partial D, \\ 1 - F(x-1) - G(y-1) & \text{if } (x,y) \in \partial D, (x,y-1) \in \partial D, (x-1,y) \in \partial D, \\ 0 & \text{otherwise}. \end{cases}$$

Furthermore, let

$$S = \{(x,y) \in \mathbb{N}^2 : F(x) \le G(y)\}, \quad T = \{(x,y) \in \mathbb{N}^2 : F(x) > G(y)\},$$
$$\partial S = \{(x,y) \in S : (x,y-1) \notin S\}, \quad \partial T = \{(x,y) \in T : (x-1,y) \notin T\}.$$

Then

$$h^+(x,y) = \begin{cases} f(x) & \text{if } (x,y) \in \partial S, (x-1,y-1) \in T \text{ or } y = 0, \\ F(x) - G(y-1) & \text{if } (x,y) \in \partial S, (x-1,y-1) \in S \text{ or } x = 0, y \neq 0, \\ g(y) & \text{if } (x,y) \in \partial T, (x-1,y-1) \in S \text{ or } x = 0, \\ G(y) - F(x-1) & \text{if } (x,y) \in \partial T, (x-1,y-1) \in T \text{ or } y = 0, x \neq 0, \\ 0 & \text{otherwise}. \end{cases}$$

APPENDIX E: Proof of Proposition 11

This appendix contains proofs of statements (a) and (b) of Proposition 11 for dependence concepts DEP_4 and DEP_5.

Proof for DEP_4. Recall from Chapter 5 of Nelsen (1999) that if V is stochastically increasing in U, then the mapping $u \mapsto C(u,v)$ is concave for arbitrary fixed $v \in (0,1)$. Accordingly if $u_1 < u_2 < u_3$, then

$$\frac{C(u_2,v) - C(u_1,v)}{u_2 - u_1} \ge \frac{C(u_3,v) - C(u_2,v)}{u_3 - u_2}.$$

To show statement (a) for $\text{DEP}_4(X,Y)$, pick any $y, x, x' \in \mathbb{N}$ such that $x < x'$. Then $F(x'-1) \ge F(x)$ and there is nothing to show in the case of equality. Otherwise, apply the above inequality with $u_1 = F(x-1)$, $u_2 = F(x)$ and $u_3 = F(x'-1)$. Then

$$\begin{aligned} P(Y \le y | X = x) &= \frac{H(x,y) - H(x-1,y)}{F(x) - F(x-1)} \\ &= \frac{C(F(x),G(y)) - C(F(x-1),G(y))}{F(x) - F(x-1)} \\ &\ge \frac{C(F(x'-1),G(y)) - C(F(x),G(y))}{F(x') - F(x)}. \end{aligned}$$

Next, set $u_1 = F(x)$, $u_2 = F(x'-1)$, $u_3 = F(x')$, and apply the same inequality on the right-hand most term of the above chain of inequalities. This yields

$$\frac{C(F(x'-1), G(y)) - C(F(x), G(y))}{F(x') - F(x)} \geq \frac{C(F(x'), G(y)) - C(F(x'-1), G(y))}{F(x') - F(x'-1)}$$

$$= \frac{H(x', y) - H(x'-1, y)}{F(x') - F(x'-1)}$$

$$= \mathrm{P}(Y \leq y \,|\, X = x').$$

In other words, $\mathrm{P}(Y \leq y \,|\, X = x)$ is a non-increasing function of x for all possible values of $y \in \mathbb{N}$, i.e., $\mathrm{DEP}_4(X, Y)$.

Turning to part (b), the above argument shows that if $(X^\circledast, Y^\circledast)$ is as in Definition 2, then $\mathrm{DEP}_4(X^\circledast, Y^\circledast) \Rightarrow \mathrm{DEP}_4(X, Y)$. To prove the reverse implication, one calls on the fact that for arbitrary $(u, v) \in (F(x-1), F(x)] \times (G(y-1), G(y)]$,

$$C_H^\circledast(u, v) = u_x v_y H(x-1, y-1) + (1 - u_x) v_y H(x, y-1)$$
$$+ u_x (1 - v_y) H(x-1, y) + (1 - u_x)(1 - v_y) H(x, y),$$

where

$$u_x = \frac{F(x) - u}{\mathrm{P}(X = x)} \quad \text{and} \quad v_y = \frac{G(y) - v}{\mathrm{P}(Y = y)}.$$

Upon differentiation with respect to u, one finds

$$\frac{\partial}{\partial u} C_H^\circledast(u, v) = v_y \mathrm{P}(Y \leq y-1 \,|\, X = x) + w_y \mathrm{P}(Y \leq y \,|\, X = x),$$

where v_y and $w_y = \{v - G(y-1)\} / \mathrm{P}(Y = y)$ are non-negative constants for fixed $y \in \mathbb{N}$. Furthermore, the conditional probabilities are non-increasing in x by assumption. Consequently, the mapping $u \mapsto C_H^\circledast(u, v)$ is concave for any given value of $v \in (0, 1)$. In other words, $\mathrm{DEP}_4(U, V)$ or, equivalently, $\mathrm{DEP}_4(X^\circledast, Y^\circledast)$. $\square$

Proof for DEP₅. Note that if $C \in \mathcal{C}_H$, then for all $i, j \in \mathbb{N}$,

$$h_{ij} = \mathrm{P}(X = i, Y = j)$$
$$= C(F(i), G(j)) + C(F(i-1), G(j-1))$$
$$- C(F(i), G(j-1)) - C(F(i-1), G(j)) = \mathrm{P}(U \in A_i, V \in B_j),$$

where $A_i = (F(i-1), F(i)]$, $B_j = (G(j-1), G(j)]$ and the pair (U, V) is distributed as C. Thus if $i < i'$ and $j < j'$, then $h_{ij} h_{i'j'} \geq h_{ij'} h_{i'j}$ holds if and only if

$$\mathrm{P}(U \in A_i, V \in B_j) \, \mathrm{P}(U \in A_{i'}, V \in B_{j'})$$
$$\geq \mathrm{P}(U \in A_i, V \in B_{j'}) \, \mathrm{P}(U \in A_{i'}, V \in B_j). \qquad (13)$$

Further observe that if $x \in A_i$ and $x' \in A_{i'}$, then $x < x'$, a fact that can be denoted $A_i < A_{i'}$. Likewise, $B_j < B_{j'}$.

Now as observed by Kimeldorf and Sampson (1987), if a pair (U, V) is in positive likelihood ratio dependence, inequality (13) is then automatically verified for any intervals $A_i < A_{i'}$ and $B_j < B_{j'}$. This proves statement (a).

To establish part (b), it is sufficient to show that $\text{DEP}_5(X, Y) \Rightarrow \text{DEP}_5(X^{\circledast}, Y^{\circledast})$. To do this, first note that $C_H^{\circledast}$ is absolutely continuous and that whenever $(u, v) \in A_i \times B_j$, its density is given by

$$c_H^{\circledast}(u, v) = \frac{h_{ij}}{\text{P}(X = i)\,\text{P}(Y = j)}.$$

Now take $u' > u$ and $v' > v$ and suppose without loss of generality that $u' \in A_{i'}$ and $v' \in B_{j'}$ for some $i' > i$ and $j' > j$. It then follows from $\text{DEP}_5(X, Y)$ that

$$c_H^{\circledast}(u, v)\, c_H^{\circledast}(u', v') \geq c_H^{\circledast}(u, v')\, c_H^{\circledast}(u', v). \qquad \square$$

APPENDIX F: Proof of Proposition 12

Under the assumptions of Proposition 12, C has upper tail dependence

$$\lambda = 2 - \lim_{t \uparrow 1} \frac{1 - C(t, t)}{1 - t} \leq 1.$$

Given that $F = G$, one can easily see that

$$\lim_{t \uparrow 1} \text{P}\left\{ Y > G^{-1}(t) \,\middle|\, X > F^{-1}(t) \right\} = 2 - \lim_{t \uparrow 1} \frac{1 - C\left(F \circ F^{-1}(t),\, F \circ F^{-1}(t) \right)}{1 - F \circ F^{-1}(t)}.$$

Now for $t \in (F(n-1), F(n)]$, $F^{-1}(t) = n$ and hence $F \circ F^{-1}(t) = F(n)$. Therefore, if $a_n = F(n)$ for each $n \in \mathbb{N}$,

$$2 - \lim_{t \uparrow 1} \frac{1 - C\left(F \circ F^{-1}(t),\, F \circ F^{-1}(t) \right)}{1 - F \circ F^{-1}(t)} = 2 - \lim_{n \uparrow \infty} \frac{1 - C(a_n, a_n)}{1 - a_n} = \lambda$$

because $a_n \to 1$ as $n \to \infty$. $\qquad \square$

REFERENCES

AVÉROUS, J. and DORTET-BERNADET, J.-L. (2000) LTD and RTI dependence orderings. *Canad. J. Statist.*, **28**, 151-157.

CAMERON, A.C., LI, T., TRIVEDI, P.K. and ZIMMER, D.M. (2004) Modeling the differences in counted outcomes using bivariate copula models: With application to mismeasured counts. *Econom. J.*, **7**, 566-584.

CAPÉRAÀ, P. and GENEST, C. (1990) Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles. *Canad. J. Statist.*, **18**, 315-326.

CARLEY, H. (2002) Maximum and minimum extensions of finite subcopulas. *Comm. Statist. Theory Methods*, **31**, 2151-2166.

CHOULAKIAN, V. and DE TIBEIRO, J. (2000) Copules archimédiennes et tableaux de contingence à variables qualitatives ordinales. *Rev. Statist. Appl.*, **48**, 83-96.

CONTI, P.L. (1993) On some descriptive aspects of measures of monotone dependence. *Metron*, **51**, 43-60 (1994).

DENUIT, M. and LAMBERT, P. (2005) Constraints on concordance measures in bivariate discrete data. *J. Multivariate Anal.*, **93**, 40-57.

EMBRECHTS, P., McNEIL, A.J. and STRAUMANN, D. (2002) Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond (Cambridge, 1998)*, pages 176-223. Cambridge Univ. Press, Cambridge.

ESARY, J.D. and PROSCHAN, F. (1972) Relationships among some concepts of bivariate dependence. *Ann. Math. Statist.*, **43**, 651-655.

FANG, Z. and JOE, H. (1992) Further developments on some dependence orderings for continuous bivariate distributions. *Ann. Inst. Statist. Math.*, **44**, 501-517.

FREES, E.W. and VALDEZ, E.A. (1998) Understanding relationships using copulas. *N. Amer. Act. J.*, **2**, 1-25.

GENEST, C. and FAVRE, A.-C. (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrologic Eng.*, **12**, 347-368.

GENEST, C., GHOUDI, K. and RIVEST, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543-552.

GENEST, C., MARCEAU, É. and MESFIOUI, M. (2003) Compound Poisson approximations for individual models with dependent risks. *Insurance Math. Econom.*, **32**, 73-91.

GENEST, C. and RÉMILLARD, B. (2004) Tests of independence and randomness based on the empirical copula process. *Test*, **13**, 335-370.

GENEST, C., RÉMILLARD, B. and BEAUDOIN, D. (2007) Omnibus goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.*, **42**, in press.

GOODMAN, L.A. and KRUSKAL, W.H. (1954) Measures of association for cross classifications. *J. Amer. Statist. Assoc.*, **49**, 732-764.

HOEFFDING, W. (1940) Maßstabinvariante Korrelationstheorie für diskontinuierliche Verteilungen. *Arch. Math. Wirt. Sozialforsch.*, **7**, 4-70.

JOE, H. (1993) Multivariate dependence measures and data analysis. *Comput. Statist. Data Anal.*, 16, 279-297.

JOE, H. (1997) *Multivariate Models and Dependence Concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

JOE, H. (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.*, **94**, 401-419.

KENDALL, M.G. (1945) The treatment of ties in ranking problems. *Biometrika*, **33**, 239-251.

KIM, G., SILVAPULLE, M.J. and SILVAPULLE, P. (2007) Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.*, **51**, 2836-2850.

KIMELDORF, G. and SAMPSON, A.R. (1987) Positive dependence orderings. *Ann. Inst. Statist. Math.*, **39**, 113-128.

KOWALCZYK, T. and NIEWIADOMSKA-BUGAJ, M. (2001) An algorithm for maximizing Kendall's τ. *Comput. Statist. Data Anal.*, **37**, 181-193.

LEHMANN, E.L. (1966) Some concepts of dependence. *Ann. Math. Statist.*, **37**, 1137-1153.

MARSHALL, A.W. (1996) Copulas, marginals, and joint distributions. In *Distributions with Fixed Marginals and Related Topics (Seattle, WA, 1993)*, volume 28 of *IMS Lecture Notes Monogr. Ser.*, pages 213-222. Inst. Math. Statist., Hayward, CA.

MCNEIL, A.J., FREY, R. and EMBRECHTS, P. (2005) *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, Princeton, NJ.

MEESTER, S.G. and MACKAY, R.J. (1994) A parametric model for cluster correlated categorical data. *Biometrics*, **50**, 954-963.

MESFIOUI, M. and TAJAR, A. (2005) On the properties of some nonparametric concordance measures in the discrete case. *J. Nonparametr. Stat.*, **17**, 541-554.

MIKUSIŃSKI, P., SHERWOOD, H. and TAYLOR, M.D. (1992) Shuffles of min. *Stochastica*, **13**, 61-74.

NELSEN, R.B. (1987) Discrete bivariate distributions with given marginals and correlation. *Comm. Statist. B–Simulation Comput.*, **16**, 199-208.

NELSEN, R.B. (1999) *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer, New York.

NEŠLEHOVÁ, J. (2004) *Dependence of Non-Continuous Random Variables*. Doctoral dissertation, Universität Oldenburg, Oldenburg, Germany.

NEŠLEHOVÁ, J. (2007) On rank correlation measures for non-continuous random variables. *J. Multivariate Anal.*, **98**, 544-567.

OAKES, D. (1982) A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B*, **44**, 414-422.

PFEIFER, D. and NEŠLEHOVÁ, J. (2004) Modeling and generating dependent risk processes for IRM and DFA. *Astin Bull.*, **34**, 333-360.

SCARSINI, M. (1984) On measures of concordance. *Stochastica*, **8**, 201-218.

SCHWEIZER, B. and SKLAR, A. (1974) Operations on distribution functions not derivable from operations on random variables. *Studia Math.*, **52**, 43-52.

SHIH, J.H. and LOUIS, T.A. (1995) Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 1384-1399.

SKLAR, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229-231.

TCHEN, A.H. (1980) Inequalities for distributions with given marginals. *Ann. Probab.*, **8**, 814-827.

TRÉGOUËT, D.-A., DUCIMETIÈRE, P., BOCQUET, V., VISVIKIS, S., SOUBRIER, F. and TIRET, L. (2004) A parametric copula model for analysis of familial binary data. *Am. J. Hum. Genet.*, **64**, 886-893.

VANDENHENDE, F. and LAMBERT, P. (2003) Improved rank-based dependence measures for categorical data. *Statist. Probab. Lett.*, **63**, 157-163.

WHITT, W. (1976) Bivariate distributions with given marginals. *Ann. Statist.*, **4**, 1280-1289.

YAARI, M.E. (1987) The dual theory of choice under risk. *Econometrica*, **55**, 95-115.

YANAGIMOTO, T. and OKAMOTO, M. (1969) Partial orderings of permutations and monotonicity of a rank correlation statistic. *Ann. Inst. Statist. Math.*, **21**, 489-506.

CHRISTIAN GENEST
Département de mathématiques et de statistique
Université Laval, 1045, avenue de la Médecine
Québec, Canada G1V 0A6

JOHANNA NEŠLEHOVÁ
Department of Mathematics
ETH Zurich
CH-8092 Zurich, Switzerland