

Research Article

On Concordance Measures for Discrete Data and Dependence Properties of Poisson Model

**Taoufik Bouezmarni,¹ Mhamed Mesfioui,²
and Abdelouahid Tajar³**

¹ Department of Economics, McGill University, Leacock Building, 855 Sherbrooke Street West, C.P. 6128, succursale Centre-ville Montreal, QC, Canada H3A 2T7

² Département de Mathématiques et d'Informatique, Université du Québec à Trois-Rivières, Pavillon Ringuet, local 3060, C.P. 500, Trois-Rivières, QC, Canada G9A 5H7

³ ARC Epidemiology Unit, The University of Manchester, Oxford Road, Manchester M13 9PT, UK

Correspondence should be addressed to Mhamed Mesfioui, mhamed.mesfioui@uqtr.ca

Received 2 April 2009; Revised 10 November 2009; Accepted 15 December 2009

Recommended by Chunsheng Ma

We study Kendall's tau and Spearman's rho concordance measures for discrete variables. We mainly provide their best bounds using positive dependence properties. These bounds are difficult to write down explicitly in general. Here, we give the explicit formula of the best bounds in a particular Fréchet space in order to understand the behavior of the ranges of these measures. Also, based on the empirical copula which is viewed as a discrete distribution, we propose a new estimator of the copula function. Finally, we give useful dependence properties of the bivariate Poisson distribution and show the relationship between parameters of the Poisson distribution and both tau and rho.

Copyright © 2009 Taoufik Bouezmarni et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The best known dependence property is "lack of dependence," or what is known as stochastic independence. In many applications, independence between two random variables is assumed; this can be a strong assumption in the undertaken analysis. Taking into account the dependence structure between the variables leads to appropriate modeling approaches and correct conclusions. To study stochastic dependence, concordance concept and positive dependence are well used tools. This is because many dependence properties can be described by means of the joint distribution of the variables and these measures and properties are often margins free. In this paper we study two concordance measures, Kendall's tau Kruskal [1] and Spearman's rho Lehmann [2]. These measures have several

properties known as Rényi's axioms; for more details see Rényi [3]. Among these axioms, we focus on the range of the association measure.

Many researches have been concerned with the study of tau and rho in the case of continuous variables. Schweizer and Wolff [4], in one seminal paper, show that the study of concordance measures for continuous random variables can be characterized as the study of copulas [5]. However, for noncontinuous variables, this interrelationship generally does not hold. There are few papers concerning the discrete version of Kendall's tau and Spearman's rho. Conti [6] gives definitions of two approaches of indifference and links them to concordance and discordance properties of the data. Tajar et al. [7] propose a copula-type representation for random couples with binary margins. They show that appropriate measures of association for binary random variables do not depend on the marginal distribution of the variables under study. Mesfioui and Tajar [8] and Denuit and Lambert [9] have shown independently that the range of tau and rho in the discrete case is not the unit interval as in the continuous case. Nešlehová [10] considers an alternative transformation of an arbitrary random variable to a uniform distribution variable in order to study the rank measures for noncontinuous random variables.

In this paper, we focus on the range of the concordance measures. Aside from identifying the best bounds of tau and rho in the case of discrete random variables, we present some dependence properties of the bivariate Poisson model and discuss their relationship with the concordance measures tau and rho. The paper is organized as follows. The next section provides a method of constructing the ranges of tau and rho for discrete data. Section 3 develops explicit expressions for the best bounds of tau and rho in the discrete Fréchet space with the same marginal. Section 4 provides a new estimator of the copulas based on the so-called empirical copulas. Section 5 discusses some dependence properties of the bivariate Poisson model.

2. Definitions and Properties

Following Hoeffding [11], Kruskal [1], and Lehmann [2], Schweizer and Wolff [4] express Kendall's tau and Spearman's rho for continuous random vector (X, Y) in terms of the joint distribution $H(x, y)$ of (X, Y) and the margins $F(x)$ for X and $G(y)$ for Y . A general representation for each of τ and ρ has been first proposed by Kowalczyk and Niewiadomska-Bugaj [12]; namely

$$\tau = E_H[H(X, Y)] + E_H[H(X^-, Y^-)] + E_H[H(X^-, Y)] + E_H[H(X, Y^-)] - 1, \quad (2.1)$$

$$\rho = 3\{E_\Pi[H(X^-, Y)] + E_\Pi[H(X, Y^-)] + E_\Pi[H(X^-, Y^-)] + E_\Pi[H(X, Y)] - 1\}, \quad (2.2)$$

where $H(x^-, y^-) = P[X < x, Y < y]$, $H(x, y^-) = P[X \leq x, Y < y]$, $H(x^-, y) = P[X < x, Y \leq y]$, and $\Pi(x, y) = F(x)G(y)$.

Several results in this paper are based on the monotonicity property of Kendall's τ and Spearman's ρ . This property has first been proposed for continuous variables by Yanagimoto and Okamoto [13] (see also [14]). Tchen [15] obtained similar monotonicity property for τ and ρ when the supports of the joint distributions consist in a finite number of atoms. Mesfioui and Tajar [8] extend various dependence relationships between Kendall's τ and Spearman's ρ in Capéraà and Genest [16] and Nelsen [5], to the discrete case. One key result

of their paper is the generalization to any kind of random variables for continuous and/or discrete variables.

For the remainder of the paper, we recall the property of concordance orderings, defined as follows.

Let (X_1, Y_1) and (X_2, Y_2) be random vectors with identical marginals and respective *cdf's* H_1 and H_2 . The random couple (X_2, Y_2) is said to be more concordant than (X_1, Y_1) , denoted by $(X_1, Y_1) \leq_c (X_2, Y_2)$, if $H_1(x_1, x_2) \leq H_2(x_1, x_2)$ holds for all $x_1, x_2 \in \mathbb{R}$.

In the following proposition, we propose a flexible method to establish the monotonicity property given in Mesfioui and Tajar [8] for purely discrete random vectors. The proof is direct and easy to understand and extends the result to the general random vectors.

Proposition 2.1. *Let (X_1, Y_1) and (X_2, Y_2) be two random couples with respective distribution function H_1 and H_2 in $\Gamma(F, G)$, the Fréchet space of all distribution functions with fixed marginals F and G . Then,*

$$(X_1, Y_1) \leq_c (X_2, Y_2) \implies \tau_{H_1} \leq \tau_{H_2}, \quad (2.3)$$

$$(X_1, Y_1) \leq_c (X_2, Y_2) \implies \rho_{H_1} \leq \rho_{H_2}. \quad (2.4)$$

Proof. Using Fubini's theorem, we note that

$$\begin{aligned} E_{H_1}[H_2(X, Y)] &= E_{H_2}[\overline{H}_1(X^-, Y^-)], \\ E_{H_1}[H_2(X^-, Y)] &= E_{H_2}[\overline{H}_1(X, Y^-)], \\ E_{H_1}[H_2(X, Y^-)] &= E_{H_2}[\overline{H}_1(X^-, Y)], \\ E_{H_1}[H_2(X^-, Y^-)] &= E_{H_2}[\overline{H}_1(X, Y)], \end{aligned} \quad (2.5)$$

where \overline{H}_i denotes the survival functions associated to H_i , $i = 1, 2$.

Now without loss of generality if we assume that $H_1 \leq H_2$, which is equivalent to $\overline{H}_1 \leq \overline{H}_2$, we then get

$$\begin{aligned} E_{H_1}[H_1(X, Y)] &\leq E_{H_1}[H_2(X, Y)] \\ &= E_{H_2}[\overline{H}_1(X^-, Y^-)] \\ &\leq E_{H_2}[\overline{H}_2(X^-, Y^-)] \\ &= E_{H_2}[H_2(X, Y)]. \end{aligned} \quad (2.6)$$

Similarly, we obtain

$$\begin{aligned} E_{H_1} [H_1(X^-, Y)] &\leq E_{H_2} [H_2(X^-, Y)], \\ E_{H_1} [H_1(X, Y^-)] &\leq E_{H_2} [H_2(X, Y^-)], \\ E_{H_1} [H_1(X^-, Y^-)] &\leq E_{H_2} [H_2(X^-, Y^-)]. \end{aligned} \quad (2.7)$$

Combining the later inequalities with (2.1), we then obtain (2.3). It is easy seen that (2.4) is immediate from (2.2). \square

For any bivariate distribution function H with univariate marginals F and G , one has

$$\max[0, F(x) + G(y) - 1] \leq H(x, y) \leq \min[F(x), G(y)]. \quad (2.8)$$

The extreme distributions $H_{\min}(x, y) = \max[0, F(x) + G(y) - 1]$ and $H_{\max}(x, y) = \min[F(x), G(y)]$ are often refereed as Fréchet bounds (see [17]). These bounds play a central role to construct optimal ranges of τ and ρ as stated in the following corollary.

Corollary 2.2. *Let (X, Y) be a random couple with distribution function H in $\Gamma(F, G)$. Then,*

$$\begin{aligned} \tau_{\min} &\leq \tau_H \leq \tau_{\max}, \\ \rho_{\min} &\leq \rho_H \leq \rho_{\max}, \end{aligned} \quad (2.9)$$

where τ_{\min} , ρ_{\min} and τ_{\max} , ρ_{\max} denote the values of Kendall's τ and Spearman's ρ corresponding to the Fréchet lower and upper bounds in $\Gamma(F, G)$, respectively.

As stated earlier, the main objective in this paper is to examine the bounds of τ and ρ in the Fréchet space $\Gamma(F, G)$ when F and G are discrete. To do that, let (X, Y) be a discrete random couple with *cdf* $H \in \Gamma(F, G)$. Since Kendall's τ and Spearman's ρ are scale invariants, they remain unchanged under strictly increasing transformations of the marginal distributions. We can then suppose, without any loss of generality, that X and Y are valued in \mathbb{Z} , the set of all integers. Therefore, we can see from (2.1) and (2.2) that τ and ρ can be written as

$$\tau = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} T_{ij} S_{ij} - 1, \quad (2.10)$$

$$\rho = 3 \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} T_{ij} [F(i) - F(i-1)][G(j) - G(j-1)] - 3, \quad (2.11)$$

where

$$T_{ij} = H(i, j) + H(i-1, j-1) + H(i, j-1) + H(i-1, j), \quad (2.12)$$

$$S_{ij} = H(i, j) - H(i, j-1) - H(i-1, j) + H(i-1, j-1). \quad (2.13)$$

In order to obtain the best bounds τ_{\min} , ρ_{\min} and τ_{\max} , ρ_{\max} , the minimum and maximum values corresponding to lower and upper bound of τ and ρ , respectively, we replace H in (2.10) and (2.11) by the Fréchet bounds $H_{\min}(i, j) = \max[F(i) + G(j) - 1, 0]$ and $H_{\max}(i, j) = \min[F(i), G(j)]$, respectively.

For discrete data, the ranges of τ and ρ are different from the usual unit interval $[-1, 1]$. This is a violation of the monotone dependence properties of concordance measures, as stated in Nelsen [5]. To correct this problem, we propose the following corrections:

$$\tau_c = \begin{cases} \frac{\tau}{\tau_{\max}} & \text{if } \tau \geq 0, \\ -\frac{\tau}{\tau_{\min}} & \text{if } \tau < 0, \end{cases} \quad (2.14)$$

$$\rho_c = \begin{cases} \frac{\rho}{\rho_{\max}} & \text{if } \rho \geq 0, \\ -\frac{\rho}{\rho_{\min}} & \text{if } \rho < 0. \end{cases}$$

The main importance of these corrections is that they allow to interpret the levels of the new measures, τ_c and ρ_c , as percentages. Illustrations of these transformations are proposed in Section 5 with the bivariate Poisson distribution.

3. Explicit Bounds of Discrete τ and ρ in $\Gamma(F, F)$

The aim of this section is to study the effect of the marginal distributions on the range of τ and ρ for discrete data. Note that it is difficult to obtain explicit expressions of the extreme values of τ and ρ in $\Gamma(F, G)$ for noncontinuous distribution F and G . This problem is very complicated and requires several assumptions on F and G . In order to analyze the behavior of these bounds, we consider the particular space $\Gamma(F, F)$, where F is a discrete distribution function. To this end, consider the integer function defined by

$$\phi(i) = \min\{j \in \mathbb{Z} : F(i) + F(j) > 1\}, \quad i \in \mathbb{Z}. \quad (3.1)$$

This function plays an important role to explicit lower bounds of τ and ρ in the space $\Gamma(F, F)$. The next proposition presents explicit optimal bounds of Spearman's ρ .

Proposition 3.1. *The best bounds for ρ in the space $\Gamma(F, F)$ are given by*

$$\rho_{\max} = 3E\left[1 - F^2(X) - F^2(X - 1)\right], \quad (3.2)$$

$$\rho_{\min} = 3E[\psi(X) + \psi(X - 1) - 1], \quad (3.3)$$

where

$$\psi(i) = \sum_{j=\phi(i)}^{\infty} [F(i) + F(j) - 1][F(j + 1) - F(j - 1)]. \quad (3.4)$$

Proof. Let $H(i, j) = \min[F(i), F(j)]$. From (2.12), we observe that

$$T_{ij} = [F(i) + 3F(i-1)]\mathbb{I}_{[i=j]} + 2[F(i) + F(i-1)]\mathbb{I}_{[i<j]} + 2[F(j) + F(j-1)]\mathbb{I}_{[i>j]}, \quad (3.5)$$

and writing $F(i) - F(i-1) = p_i$, we get from (2.11) that

$$\begin{aligned} \rho_{\max} &= 3 \sum_{i=-\infty}^{\infty} [F(i) + 3F(i-1)][F(i) - F(i-1)]p_i \\ &\quad + 6 \sum_{i=-\infty}^{\infty} [F(i) + F(i-1)]p_i \sum_{j=i+1}^{\infty} [F(j) - F(j-1)] \\ &\quad + 6 \sum_{i=-\infty}^{\infty} p_i \sum_{j=-\infty}^{i-1} [F(j) - F(j-1)][F(j) + F(j-1)] - 3, \end{aligned} \quad (3.6)$$

which may be simplified as

$$\begin{aligned} \rho_{\max} &= 3E\{[F(X) + 3F(X-1)][F(X) - F(X-1)]\} \\ &\quad + 6E\{[F(X) + F(X-1)][1 - F(X)]\} \\ &\quad + 6E[F^2(X-1)] - 3. \end{aligned} \quad (3.7)$$

The result then follows from the fact that $E[F(X) + F(X-1)] = 1$. Now, choose $H(i, j) = \sup[F(i) + F(j) - 1, 0]$ and put $H^+(i, j) = F(i) + F(j) - 1$. From (2.11), we see that

$$\begin{aligned} \rho_{\min} &= 3 \sum_{i=-\infty}^{\infty} \sum_{j=\phi(i)}^{\infty} H^+(i, j)p_i p_j + 3 \sum_{i=-\infty}^{\infty} \sum_{j=\phi(i)}^{\infty} H^+(i, j)p_{i+1} p_{j+1} \\ &\quad + 3 \sum_{i=-\infty}^{\infty} \sum_{j=\phi(i)}^{\infty} H^+(i, j)p_i p_{j+1} + 3 \sum_{i=-\infty}^{\infty} \sum_{j=\phi(i)}^{\infty} H^+(i, j)p_{i+1} p_j - 3. \end{aligned} \quad (3.8)$$

It follows that

$$\rho_{\min} = 3 \sum_{i=-\infty}^{\infty} (p_i + p_{i+1}) \sum_{j=\phi(i)}^{\infty} [F(i) + F(j) - 1](p_j + p_{j+1}) - 3, \quad (3.9)$$

which may be rewritten as

$$\rho_{\min} = 3 \sum_{i=-\infty}^{\infty} [\psi(i) + \psi(i-1) - 1]p_i, \quad (3.10)$$

where

$$\psi(i) = \sum_{j=\phi(i)}^{\infty} [F(i) + F(j) - 1] [F(j+1) - F(j-1)]. \quad (3.11)$$

The result is therefore obtained from (3.11) and (3.10). \square

Using (2.10) with $H(i, j) = \min[F(i), F(j)]$, we notice that the upper bound of Kendall's τ in the space $\Gamma(F, F)$ can be expressed as

$$\tau_{\max} = 2E[F(X-1)]. \quad (3.12)$$

Note that the sharp upper bound given in Denuit and Lambert [9] coincides with (3.12) in $\Gamma(F, F)$. However, the behavior of Kendall's tau lower bound in terms of the distribution F is not evident. The following proposition gives an explicit form of this bound in $\Gamma(F, F)$.

Proposition 3.2. *The best lower bounds of τ in $\Gamma(F, F)$ is*

$$\tau_{\min} = 2E[F[\phi(X-1)]] - 2 \sum_{k=-\infty}^{\infty} \xi(k) - 2, \quad (3.13)$$

where

$$\xi(k) = [F(k-1) + F[\phi(k-1)] - 1] [F(k) + F[\phi(k-1) - 1] - 1] \mathbb{I}_{[\phi(k) < \phi(k-1)]}. \quad (3.14)$$

Proof. From (2.12) and (2.13), we observe that

$$\begin{aligned} S_{ij}T_{ij} &= H^2(i, j) + H^2(i-1, j-1) - H^2(i-1, j) - H^2(i, j-1) \\ &\quad + 2H(i, j)H(i-1, j-1) - 2H(i-1, j)H(i, j-1) \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} [H^2(i, j) + H^2(i-1, j-1) - H^2(i-1, j) - H^2(i, j-1)] = 1. \end{aligned} \quad (3.15)$$

Consider now $H(i, j) = \sup[F(i) + F(j) - 1, 0]$ and write $H^+(i, j) = F(i) + F(j) - 1$. From (2.10), we get

$$\begin{aligned} \tau_{\min} &= 2 \sum_{i=-\infty}^{\infty} \sum_{j=\phi(i-1)+1}^{\infty} H^+(i, j)H^+(i-1, j-1) \\ &\quad - 2 \sum_{i=-\infty}^{\infty} \sum_{j=\max[\phi(i-1), \phi(i)+1]}^{\infty} H^+(i-1, j)H^+(i, j-1). \end{aligned} \quad (3.16)$$

Using the fact that

$$H^+(i, j)H^+(i-1, j-1) - H^+(i-1, j)H^+(i, j-1) = -p_i p_j, \quad (3.17)$$

we have

$$\begin{aligned} \tau_{\min} &= -2 \sum_{i=-\infty}^{\infty} p_i \left[\sum_{j=\phi(i-1)+1}^{\infty} p_j \right] \mathbb{I}_{[\phi(i)=\phi(i-1)]} - 2 \sum_{i=-\infty}^{\infty} p_i \left[\sum_{j=\phi(i-1)+1}^{\infty} p_j \right] \mathbb{I}_{[\phi(i)<\phi(i-1)]} \\ &\quad - 2 \sum_{i=-\infty}^{\infty} [F(i-1) + F(\phi(i-1)) - 1] [F(i) + F(\phi(i-1) - 1) - 1] \mathbb{I}_{[\phi(i)<\phi(i-1)]}, \end{aligned} \quad (3.18)$$

which is equivalent to

$$\tau_{\min} = -2 \sum_{i=-\infty}^{\infty} p_i [1 - F[\phi(i-1)]] - 2 \sum_{i=-\infty}^{\infty} r(i) \quad (3.19)$$

with

$$r(i) = [F(i-1) + F(\phi(i-1)) - 1] [F(i) + F(\phi(i-1) - 1) - 1] \mathbb{I}_{[\phi(i)<\phi(i-1)]}, \quad (3.20)$$

which completes the proof. \square

Remark 3.3. Let $F_{n,p}$ be a binomial distribution with parameters n and p , and denote the extreme values of τ and ρ in $\Gamma(F_{n,p}, F_{n,p})$ by $\tau_{\max}(n, p)$ and $\rho_{\max}(n, p)$. One can show the following symmetry properties, namely:

$$\tau_{\max}(n, p) = \tau_{\max}(n, 1-p), \quad \rho_{\max}(n, p) = \rho_{\max}(n, 1-p). \quad (3.21)$$

Indeed, since $F_{n,p}(k) = F_{n,1-p}(n-k)$, then from (3.7), we have

$$\begin{aligned} \tau_{\max}(n, p) &= 2 \sum_{k=0}^n F_{n,p}(k-1) [F_{n,p}(k) - F_{n,p}(k-1)] \\ &= 2 \sum_{k=0}^n F_{n,1-p}(k-1) [F_{n,1-p}(k) - F_{n,1-p}(k-1)] \\ &= \tau_{\max}(n, 1-p). \end{aligned} \quad (3.22)$$

Similar arguments provide $\rho_{\max}(n, p) = \rho_{\max}(n, 1-p)$.

In this section, we examine the symmetry of the ranges of τ and ρ associated to discrete data. In continuous case, it is well known that the ranges of these parameters are symmetric, that is, $\tau_{\max} = -\tau_{\min}$ and $\rho_{\max} = -\rho_{\min}$. This conclusion is of course invalid for noncontinuous data. In order to clarify this question, we consider again the space $\Gamma(F, F)$ with discrete distribution F . We present below a situation which ensures that $\rho_{\max} = -\rho_{\min}$ and $\tau_{\max} = -\tau_{\min}$. As consequence of Propositions 3.1 and 3.2 and (3.12), one can establish the following results.

Corollary 3.4. *In space, $\Gamma(F, F)$, if $E[F^2(X)] = E[\psi(X)]$ and $E[F^2(X - 1)] = E[\psi(X - 1)]$, then $\rho_{\max} = -\rho_{\min}$.*

Corollary 3.5. *In space, $\Gamma(F, F)$, if $\phi(i) = \phi(i - 1)$, $i \in \mathbb{Z}$, and $E[F(X)] = E[F(\phi(X))]$, then $\tau_{\max} = -\tau_{\min}$.*

4. Empirical Copulas Viewed as a Discrete Distribution

It is well recognized that copula provides a flexible approach to model the joint behavior of random variables. In fact, this method allows to represent a bivariate distribution as function of its univariate marginals through a linking function called a copula. Specifically, if H is a distribution function of a bivariate random vector (X, Y) with continuous marginals, then Sklar [18] ensures that there exists a unique copula $C : [0, 1]^2 \rightarrow [0, 1]$ such that for all $(x, y) \in \mathbb{R}^2$,

$$H(x, y) = C[F(x), G(y)]. \quad (4.1)$$

Hence, C is a bivariate distribution function with uniform marginals on $[0, 1]$ that captures all the information about the dependence among the components of (X, Y) . For a comprehensive introduction to a copula, the reader is referred to monographs by Nelsen [5].

Suppose that the random sample $(X_1, Y_1), \dots, (X_n, Y_n)$ is given from some pair (X, Y) of continuous variable with copula $C(u, v)$. To estimate the copula C , Deheuvels [19] proposes the so-called empirical copula defined by

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(F_n(X_i) \leq u, G_n(Y_i) \leq v), \quad (4.2)$$

where F_n and G_n are the empirical distribution functions of X and Y based on the sample X_1, \dots, X_n and Y_1, \dots, Y_n given by

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq t), \quad G_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(Y_i \leq t). \quad (4.3)$$

Let R_i be the rank of X_i among the sample X_1, \dots, X_n and T_i stands the rank of Y_i among the sample Y_1, \dots, Y_n . Observe that C_n is a function of ranks $(R_1, T_1), \dots, (R_n, T_n)$, because $F_n(X_i) = R_i/n$ and $G_n(Y_i) = T_i/n$, $i = 1, \dots, n$, namely,

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\left(\frac{R_i}{n} \leq u, \frac{T_i}{n} \leq v\right). \quad (4.4)$$

From this representation, one can consider $C_n(u, v)$ as a discrete bivariate distribution with uniform marginals taking values in the set $\{1/n, 2/n, \dots, 1\}$. Observe that

$$C_n(u, v) = C_n\left(\frac{i}{n}, \frac{j}{n}\right) \quad \text{for } (u, v) \in \left[\frac{i}{n}, \frac{i+1}{n}\right] \times \left[\frac{j}{n}, \frac{j+1}{n}\right]. \quad (4.5)$$

Now, one can observe that the C_n is not copula. Indeed, $C_n(u, 1) = \lfloor nu \rfloor / n \neq u$, where $\lfloor nu \rfloor$ denotes the integer part of nu .

Our goal in this section is to transform the empirical copula in order to obtain a new estimator C_n^* which is a copula. To this end, let (Z_n, W_n) be a discrete random vector with distribution function C_n which is defined in (4.2). The idea is to transform the uniform discrete random variables Z_n and W_n into a continuous variables Z_n^* and W_n^* by defining

$$Z_n^* = Z_n - U_n, \quad W_n^* = W_n - V_n, \quad (4.6)$$

where U_n and V_n are independents and uniformly distributed in $[0, 1/n]$. We also suppose that the random vectors Z_n and U_n (resp, W_n and V_n) are independents. The next result shows that the distribution function of the continuous version (Z_n^*, W_n^*) is a copula.

Proposition 4.1. *The distribution function C_n^* of the random vector (Z_n^*, W_n^*) is a copula which may be expressed in terms of the empirical copula as follows:*

$$\begin{aligned} C_n^*(u, v) &= [1 - nu + \lfloor nu \rfloor][1 - nv + \lfloor nv \rfloor]C_n\left(\frac{\lfloor nu \rfloor}{n}, \frac{\lfloor nv \rfloor}{n}\right) \\ &\quad + [nu - \lfloor nu \rfloor][1 - nv + \lfloor nv \rfloor]C_n\left(\frac{\lfloor nu \rfloor + 1}{n}, \frac{\lfloor nv \rfloor}{n}\right) \\ &\quad + [1 - nu + \lfloor nu \rfloor][nv - \lfloor nv \rfloor]C_n\left(\frac{\lfloor nu \rfloor}{n}, \frac{\lfloor nv \rfloor + 1}{n}\right) \\ &\quad + [nu - \lfloor nu \rfloor][nv - \lfloor nv \rfloor]C_n\left(\frac{\lfloor nu \rfloor + 1}{n}, \frac{\lfloor nv \rfloor + 1}{n}\right), \quad u, v \in [0, 1], \end{aligned} \quad (4.7)$$

where $\lfloor x \rfloor$ is the integer part of x .

Proof. For any $u \in [i/n, (i+1)/n]$, $i = 0, \dots, n-1$, one sees from the definition of Z_n^* that

$$P(Z_n^* \leq u) = \frac{1}{n} \sum_{k=1}^n P\left(U_n \geq \frac{k}{n} - u\right), \quad (4.8)$$

and by using the fact that

$$P\left(U_n \geq \frac{k}{n} - u\right) = \mathbb{I}_{(k \leq i)} + (nu - i)\mathbb{I}_{(k=i+1)}, \quad (4.9)$$

it follows that $P(Z_n^* \leq u) = u$, which ensures that Z_n^* is uniformly distributed in $[0, 1]$. Similar arguments imply that W_n^* is also uniformly distributed in $[0, 1]$, so that C_n^* is a copula.

Now, we show the expression of C_n^* given in (4.7). Let (u, v) be in the set $[i/n, (i+1)/n] \times [j/n, (j+1)/n]$, $i, j = 0, \dots, n-1$. In view of relations (4.6), one has

$$\begin{aligned} C_n^*(u, v) &= P(Z_n^* \leq u, W_n^* \leq v) \\ &= \sum_{k=1}^n \sum_{p=1}^n P\left(U_n \geq \frac{k}{n} - u\right) P\left(V_n \geq \frac{p}{n} - v\right) P(Z_n = k, W_n = p) \\ &= \sum_{k=1}^n \sum_{p=1}^n [\mathbb{I}_{(k \leq i)} + (nu - i)\mathbb{I}_{(k=i+1)}] [\mathbb{I}_{(p \leq j)} + (nv - j)\mathbb{I}_{(p=j+1)}] P(Z_n = k, W_n = p). \end{aligned} \quad (4.10)$$

After simplifications, one observes

$$\begin{aligned} C_n^*(u, v) &= C_n\left(\frac{i}{n}, \frac{j}{n}\right) \\ &\quad + (nu - i) \left[C_n\left(\frac{i+1}{n}, \frac{j}{n}\right) - C_n\left(\frac{i}{n}, \frac{j}{n}\right) \right] \\ &\quad + (nv - j) \left[C_n\left(\frac{i}{n}, \frac{j+1}{n}\right) - C_n\left(\frac{i}{n}, \frac{j}{n}\right) \right] \\ &\quad + (nu - i)(nv - j) \left[C_n\left(\frac{i+1}{n}, \frac{j+1}{n}\right) - C_n\left(\frac{i+1}{n}, \frac{j}{n}\right) - C_n\left(\frac{i}{n}, \frac{j+1}{n}\right) + C_n\left(\frac{i}{n}, \frac{j}{n}\right) \right]. \end{aligned} \quad (4.11)$$

which can be rewritten as

$$\begin{aligned} C_n^*(u, v) &= [1 - nu + i][1 - nv + j] C_n\left(\frac{i}{n}, \frac{j}{n}\right) \\ &\quad + [nu - i][1 - nv + j] C_n\left(\frac{i+1}{n}, \frac{j}{n}\right) \\ &\quad + [1 - nu + i][nv - j] C_n\left(\frac{i}{n}, \frac{j+1}{n}\right) \\ &\quad + [nu - i][nv - j] C_n\left(\frac{i+1}{n}, \frac{j+1}{n}\right), \end{aligned} \quad (4.12)$$

and hence the result is obtained, since $i = \lfloor nu \rfloor$ and $j = \lfloor nv \rfloor$. \square

Finally, one concludes that it will be convenient to estimate the theoretical copula C by using the proposal estimator C_n^* instead of the empirical copula. The reason is that C_n^* is a copula which uses all the points $(i/n, j/n)$, $(i/n, (j+1)/n)$, $(i+1/n, j/n)$, and $((i+1)/n, (j+1)/n)$ in order to estimate C in $[i/n, (i+1)/n] \times [j/n, (j+1)/n]$.

5. Understanding Dependence Structure of the Bivariate Poisson Distribution

Our purpose in this section is to study dependence properties of the bivariate Poisson distribution H of a random couple (X, Y) and the relationship between τ and ρ and the parameters of H . Several bivariate Poisson distributions have been proposed in the statistical literature, for example, S. Kocherlakota and K. Kocherlakota [20]. In applied statistics, however, the focus is on the trivariate reduction method described by Johnson et al. [21] who construct the Bivariate Poisson distribution using three independent random variables X_1, X_2 , and Z all distributed as Poisson with parameters λ_1, λ_2 , and α , respectively:

$$X = X_1 + Z, \quad Y = X_2 + Z. \quad (5.1)$$

The cumulative distribution of (X, Y) is given by

$$H_{\alpha, \lambda_1, \lambda_2}(i, j) = \sum_{k=0}^{i \wedge j} F_{\lambda_1}(i-k) F_{\lambda_2}(j-k) \frac{\alpha^k e^{-\alpha}}{k!}, \quad (5.2)$$

where F_{λ_i} denotes the *cdf* of $X_i, i = 1, 2$. We notice that X and Y are Poisson model with means $\lambda_1 + \alpha$ and $\lambda_2 + \alpha$, respectively. Note that the covariance and the correlation between X and Y are expressed by

$$\text{cov}(X, Y) = \alpha, \quad \text{corr}(X, Y) = \frac{\alpha}{\sqrt{(\lambda_1 + \alpha)(\lambda_2 + \alpha)}}, \quad (5.3)$$

which are positive and nondecreasing functions of α .

To study further the relationships between α and each of τ and ρ for the bivariate Poisson model, we propose an alternative parametrization which consists in fixing the marginal parameters $\alpha + \lambda_1 = m_1$ and $\alpha + \lambda_2 = m_2$. In this context, the *cdf* (5.2) becomes

$$H_{\alpha}(i, j) = \sum_{k=0}^{i \wedge j} F_{m_1 - \alpha}(i-k) F_{m_2 - \alpha}(j-k) \frac{\alpha^k e^{-\alpha}}{k!}. \quad (5.4)$$

As a consequence of the above representation, we can see $\{H_{\alpha}\}$ as a family of bivariate Poisson models with fixed marginals which are univariate Poisson models with parameters m_1 and m_2 , respectively. This means that the set $\{H_{\alpha}\}, 0 \leq \alpha \leq m_1 \wedge m_2$ is included in the particular Fréchet space $\Gamma(F_{m_1}, F_{m_2})$, where F_{m_i} denotes the *cdf* of a Poisson model with mean $m_i, i = 1, 2$. The advantage of the parametrization (5.4) rather than (5.2) is that the coefficient α may be interpreted as a dependence parameter in the family $\{H_{\alpha}\}$.

Now, let τ_{α} and ρ_{α} be Kendall's τ and Spearman's ρ associated with the distribution H_{α} . The result below provides the monotonicity of τ_{α} and ρ_{α} as functions of α .

Proposition 5.1. *Let H_{α_1} and H_{α_2} be two *cdf* of the set $\{H_{\alpha}\}$. Then,*

$$\alpha_1 \leq \alpha_2 \implies H_{\alpha_1} \leq H_{\alpha_2}, \quad (5.5)$$

and consequently,

$$\alpha_1 \leq \alpha_2 \implies \tau_{\alpha_1} \leq \tau_{\alpha_2}, \quad \rho_{\alpha_1} \leq \rho_{\alpha_2}. \quad (5.6)$$

Proof. From (5.4),

$$\begin{aligned} \frac{\partial H_\alpha(i, j)}{\partial \alpha} &= \sum_{k=0}^{i \wedge j} \frac{\partial F_{m_1-\alpha}(i-k)}{\partial \alpha} F_{m_2-\alpha}(j-k) \frac{\alpha^k e^{-\alpha}}{k!} \\ &+ \sum_{k=0}^{i \wedge j} F_{m_1-\alpha}(i-k) \frac{\partial F_{m_2-\alpha}(j-k)}{\partial \alpha} \frac{\alpha^k e^{-\alpha}}{k!} \\ &+ \sum_{k=0}^{i \wedge j} F_{m_1-\alpha}(i-k) F_{m_2-\alpha}(j-k) \left[\frac{\alpha^{k-1} e^{-\alpha}}{(k-1)!} - \frac{\alpha^k e^{-\alpha}}{k!} \right], \end{aligned} \quad (5.7)$$

and using the fact that

$$\begin{aligned} \frac{\partial F_{m_1-\alpha}(i-k)}{\partial \alpha} &= F_{m_1-\alpha}(i-k) - F_{m_1-\alpha}(i-k-1), \\ \frac{\partial F_{m_2-\alpha}(j-k)}{\partial \alpha} &= F_{m_2-\alpha}(j-k) - F_{m_2-\alpha}(j-k-1), \end{aligned} \quad (5.8)$$

(5.7) becomes, upon simplifications,

$$\frac{\partial H_\alpha(i, j)}{\partial \alpha} = \sum_{k=0}^{i \wedge j} \frac{(m_1 - \alpha)^{i-k} e^{-(m_1-\alpha)}}{(i-k)!} \frac{(m_2 - \alpha)^{j-k} e^{-(m_2-\alpha)}}{(j-k)!} \frac{\alpha^k e^{-\alpha}}{k!} \geq 0. \quad (5.9)$$

Therefore (5.9) together with Proposition 2.1 provides (5.5) and (5.6). \square

Many statistical researches have focused on studying concepts of positive dependence for bivariate distributions, example right tail increasing, and positive quadrant dependence which are widely used in actuarial literature [22]. There are natural relationships between dependence properties and measures of concordance. An interesting property of positive dependence is the concept of positive quadrant dependence (PQD) defined as follows: let (X, Y) be a random couple valued in $\mathbb{R} \times \mathbb{R}$ with joint *cdf* H , and marginals F and G . These random variables are said to be positively quadrant dependent if, and only if, for all $(x, y) \in \mathbb{R}^2$

$$H(x, y) \geq F(x)G(y). \quad (5.10)$$

The following corollary is a direct consequence of the previous result.

Corollary 5.2. *The family $\{H_\alpha\}$ is positively quadrant dependent.*

Table 1: τ_α , ρ_α , $\tau_{\alpha,c}$ and $\rho_{\alpha,c}$ for the Poisson model.

α	τ_α	$\tau_{\alpha,c}$	ρ_α	$\rho_{\alpha,c}$
0.2	0.059	0.075	0.089	0.094
0.4	0.120	0.152	0.180	0.189
0.6	0.183	0.231	0.272	0.286
0.8	0.248	0.313	0.365	0.383
1.0	0.316	0.398	0.459	0.482
1.2	0.388	0.490	0.554	0.582
1.4	0.467	0.589	0.651	0.684
1.6	0.556	0.701	0.749	0.787
1.8	0.660	0.832	0.849	0.892

Proof. Since H_α is a nondecreasing function of α , then $H_0 \leq H_\alpha$ for all $0 \leq \alpha \leq m_1 \wedge m_2$. Now, from (5.4), $H_0(i, j) = F_{m_1}(i)F_{m_2}(j)$ for all i, j . Therefore the family $\{H_\alpha\}$ is PQD. Consequently, $\tau_\alpha \geq 0$, $\rho_\alpha \geq 0$, and $3\tau_\alpha \geq \rho_\alpha$ for all $0 \leq \alpha \leq m_1 \wedge m_2$. \square

Remark 5.3. When $m_1 = m_2 = m$, the upper bound of the family $\{H_\alpha\}$ is given by the *cdf* H_m , and using (5.4), we then obtain that $H_m(i, j) = F_m(i \wedge j) = \min[F_m(i), F_m(j)]$, for all i, j , which is the upper Fréchet bound.

In order to appreciate the corrections of τ and ρ given by (2.14), we consider the family of Poisson model $\{H_\alpha\}$ with marginal parameters $m_1 = m_2 = 2$. Using (3.2) and (3.12) with F_m instead of F , we obtain that $\rho_{\max} = 0.951$ and $\tau_{\max} = 0.792$. Table 1 provides τ_α and ρ_α with their corrections $\tau_{\alpha,c}$ and $\rho_{\alpha,c}$ for chosen values of α .

From Table 1, we note that the differences $D_{\tau,\alpha} = \tau_{\alpha,c} - \tau_\alpha$ and $D_{\rho,\alpha} = \rho_{\alpha,c} - \rho_\alpha$ are increasing as function of the dependence parameter α . This constatation is true in general because $D_{\tau,\alpha}$ and $D_{\rho,\alpha}$ can be expressed as

$$D_{\tau,\alpha} = \frac{(1 - \tau_{\max})\tau_\alpha}{\tau_{\max}}, \quad D_{\rho,\alpha} = \frac{(1 - \rho_{\max})\rho_\alpha}{\rho_{\max}}, \quad (5.11)$$

which shows that these parameters are in fact increasing with α .

Acknowledgment

The second author acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada.

References

- [1] W. H. Kruskal, "Ordinal measures of association," *Journal of the American Statistical Association*, vol. 53, pp. 814–861, 1958.
- [2] E. L. Lehmann, "Some concepts of dependence," *Annals of Mathematical Statistics*, vol. 37, pp. 1137–1153, 1966.
- [3] A. Rényi, "On measures of dependence," *Acta Mathematica Academiae Scientiarum Hungaricae*, vol. 10, pp. 441–451, 1959.

- [4] B. Schweizer and E. F. Wolff, "On nonparametric measures of dependence for random variables," *The Annals of Statistics*, vol. 9, no. 4, pp. 879–885, 1981.
- [5] R. B. Nelsen, *An Introduction to Copulas*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2006.
- [6] P. L. Conti, "On some descriptive aspects of measures of monotone dependence," *Metron*, vol. 51, no. 3-4, pp. 43–60, 1993.
- [7] A. Tajar, M. Denuit, and Ph. Lambert, "Copula-type representation for random couples with Bernoulli margins," Discussion paper 0118, Institute of Statistics, U.C.L., Leuven, Belgium, 2001.
- [8] M. Mesfioui and A. Tajar, "On the properties of some nonparametric concordance measures in the discrete case," *Journal of Nonparametric Statistics*, vol. 17, no. 5, pp. 541–554, 2005.
- [9] M. Denuit and P. Lambert, "Constraints on concordance measures in bivariate discrete data," *Journal of Multivariate Analysis*, vol. 93, no. 1, pp. 40–57, 2005.
- [10] J. Nešlehová, "On rank correlation measures for non-continuous random variables," *Journal of Multivariate Analysis*, vol. 98, no. 3, pp. 544–567, 2007.
- [11] W. Hoeffding, "Masstabinvariante korrelationstheorie," *Schriften der Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, vol. 5, pp. 179–233, 1940.
- [12] T. Kowalczyk and M. Niewiadomska-Bugaj, "Grade correspondence analysis based on Kendall's tau," in *Proceedings of the Conference of the International Federation of Classification Societies (IFCS '98)*, pp. 182–185, Rome, Italy, July 1998.
- [13] T. Yanagimoto and M. Okamoto, "Partial orderings of permutations and monotonicity of a rank correlation statistic," *Annals of the Institute of Statistical Mathematics*, vol. 21, pp. 489–506, 1969.
- [14] H. Joe, *Multivariate Models and Dependence Concepts*, vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1997.
- [15] A. H. Tchen, "Inequalities for distributions with given marginals," *The Annals of Probability*, vol. 8, no. 4, pp. 814–827, 1980.
- [16] P. Capéraà and C. Genest, "Spearman's ρ is larger than Kendall's τ for positively dependent random variables," *Journal of Nonparametric Statistics*, vol. 2, no. 2, pp. 183–194, 1993.
- [17] M. Fréchet, "Sur les tableaux de corrélation dont les marges sont données," *Annales de l'Université Lyon A*, vol. 14, pp. 53–77, 1951.
- [18] M. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, vol. 8, pp. 229–231, 1959.
- [19] P. Deheuvels, "La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance," *Bulletin de la Classe des Sciences. Académie Royale de Belgique*, vol. 65, no. 6, pp. 274–292, 1979.
- [20] S. Kocherlakota and K. Kocherlakota, *Bivariate Discrete Distributions*, vol. 132 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1992.
- [21] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distributions*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1997.
- [22] J. Dhaene and M. J. Goovaerts, "Dependency of risks and loss orders," *Astin Bulletin*, vol. 26, pp. 201–212, 1996.