On the dependence structure of order statistics

Jean Avérous,[a] Christian Genest,[b,*] and Subhash C. Kochar[c]

[a] Laboratoire de statistique et de probabilités, Université Paul-Sabatier, 118, route de Narbonne, Toulouse 31062, France

[b] Département de mathématiques et de statistique, Université Laval, Québec, Canada G1K 7P4

[c] Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi 110016, India

Received 16 June 2003

Abstract

Given a random sample from a continuous variable, it is observed that the copula linking any pair of order statistics is independent of the parent distribution. To compare the degree of association between two such pairs of ordered random variables, a notion of relative monotone regression dependence (or stochastic increasingness) is considered. Using this concept, it is proved that for $i<j$, the dependence of the jth order statistic on the ith order statistic decreases as i and j draw apart. This extends earlier results of Tukey (Ann. Math. Statist. 29 (1958) 588) and Kim and David (J. Statist. Plann. Inference 24 (1990) 363). The effect of the sample size on this type of dependence is also investigated, and an explicit expression is given for the population value of Kendall's coefficient of concordance between two arbitrary order statistics of a random sample.

Keywords: Concordance ordering; Dispersive ordering; Exponential distribution; Kendall's tau; Monotone regression dependence; Spearman's rho; Stochastic increasingness

1. Introduction

Let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ be the order statistics associated with the first $n \geqslant 2$ observations in a sequence $X_1, X_2, \ldots,$ of continuous random variables. Motivated in part by applications in reliability theory, various authors have investigated the

*Corresponding author. Fax: +1-418-656-2817.

E-mail address: genest@mat.ulaval.ca (C. Genest).

nature of the dependence that may exist between $X_{i:n}$ and $X_{j:n}$ for $1 \leqslant i < j \leqslant n$ under different distributional scenarios. When the X_k are mutually independent and identically distributed, it has been known since the work of Bickel [2] that

$$\text{cov}(X_{i:n}, X_{j:n}) \geqslant 0,$$

but much stronger statements can be made to qualify the association between $X_{i:n}$ and $X_{j:n}$, even when the X_k are from different distributions. For details, refer to the paper by Boland et al. [4] and references therein.

In contrast, very little seems to be known about the relative degree of dependence that may exist between two arbitrary pairs of order statistics, say $(X_{i:n}, X_{j:n})$ and $(X_{i':n}, X_{j':n})$. The only contributions appear to be those of Tukey [21] and Kim and David [12], both of which pertain to the case where the X_k are mutually independent and identically distributed. When the parent distribution has an increasing hazard rate and a decreasing reverse hazard rate, Tukey [21] showed that

$$\text{cov}(X_{i':n}, X_{j':n}) \leqslant \text{cov}(X_{i:n}, X_{j:n}) \tag{1}$$

must hold when

$$\text{either } i = i' \text{ and } j \leqslant j', \quad \text{or } j = j' \text{ and } i' \leqslant i. \tag{2}$$

As for Kim and David [12], they proved that if both the hazard and the reverse hazard rates of the X_k are increasing, then inequality (1) remains valid when $i = i'$ and $j \leqslant j'$, but goes the other way when $j = j'$ and $i' \leqslant i$.

While these results are certainly not contradictory, it may be puzzling at first that different conditions on the common distribution of the X_k could cause the covariance between $X_{i:n}$ and $X_{j:n}$ to increase or to decrease as i and j pull apart. The key to the resolution, of course, is in the fact that the traditional notion of covariance is *not* an appropriate measure of dependence when the pairs being compared do not have the same marginal distributions, as is clearly the case here.

The purpose of this paper is to shed additional light into the dependence structure of pairs of order statistics by showing that for any integers $1 \leqslant i < j \leqslant n$ and $1 \leqslant i' < j' \leqslant n'$ such that

$$i' \leqslant i, \quad j - i \leqslant j' - i', \quad n - i \leqslant n' - i', \quad n' - j' \leqslant n - j, \tag{3}$$

the pair $(X_{i:n}, X_{j:n})$ is more dependent than the pair $(X_{i':n'}, X_{j':n'})$ according to the bivariate monotone regression dependence (or stochastically increasing) ordering. This result, which is independent of the choice of the parent distribution for the X_k, implies in particular that under condition (3), and hence under condition (2) when $n = n'$, one has

$$\kappa(X_{i':n}, X_{j':n}) \leqslant \kappa(X_{i:n}, X_{j:n}),$$

where $\kappa(S, T)$ stands for any measure of concordance between S and T in the sense of Scarsini [16], e.g., Spearman's rho, Kendall's tau, or Gini's coefficient of association. This conclusion is in accordance with the intuition that as order statistics $X_{i:n}$ and $X_{j:n}$ draw apart, they tend to be less dependent.

The definition of the monotone regression dependence ordering is recalled in Section 2, where a precise statement of the main result appears as Proposition 2. Auxiliary technical material needed to carry out its proof is collected in Section 3, including a result of possibly independent interest concerning the dispersive properties of generalized spacings from an exponential sample. The argument leading to Proposition 2 appears in Section 4, where some special cases are also discussed. Section 5 contains a closed-form formula for $\tau(X_{i:n}, X_{j:n})$ which extends that just reported by Schmitz [17] in the special case $i = 1$ and $j = n$. Some directions for future work are outlined in Section 6.

2. Preliminaries

For $i = 1, 2$, let (S_i, T_i) be a pair of continuous random variables with joint cumulative distribution function H_i and marginals F_i and G_i. As summarized in the books by Joe [10], Nelsen [15] or Drouet-Mari and Kotz [8], 30 years of research into concepts and measures of association have shown that the proper way of comparing the relative degree of dependence between (S_1, T_1) and (S_2, T_2) is in terms of their associated copulas, implicitly defined in a unique fashion by the relation

$$H_i(s, t) = C_i\{F_i(s), G_i(t)\},$$

valid for all $s, t \in \mathbb{R}$. Thus (S_2, T_2) is said to be more concordant (or more positive quadrant dependent) than (S_1, T_1), denoted by $(S_1, T_1) \prec_{\text{PQD}} (S_2, T_2)$, if and only if, for all $u, v \in (0, 1)$,

$$C_1(u, v) \leqslant C_2(u, v). \tag{4}$$

As shown, e.g., by Tchen [20], condition (4) implies that

$$\kappa(S_1, T_1) \leqslant \kappa(S_2, T_2), \tag{5}$$

where $\kappa(S, T)$ represents Spearman's rho, Kendall's tau, Gini's coefficient, or indeed any other copula-based measure of concordance satisfying the axioms of Scarsini [16]. In the special case where $F_1 = F_2$ and $G_1 = G_2$, it also follows from (4) that the pairs (S_1, T_1) and (S_2, T_2) are ordered by Pearson's correlation coefficient, namely

$$\text{corr}(S_1, T_1) \leqslant \text{corr}(S_2, T_2).$$

In his survey, Joe [10] mentions a number of bivariate stochastic ordering relations $\prec$ that strengthen $\prec_{\text{PQD}}$ and hence imply (5) as well. One such notion that will be pursued here is that of greater monotone regression dependence, originally considered by Yanagimoto and Okamoto [22] and later extended and further investigated by Schriever [18], Capéraà and Genest [5], Block et al. [3], as well as Fang and Joe [9]. Although this ordering, as all other dependence orderings, involves a comparison of the underlying copulas, an equivalent formulation of it will be given in Definition 1 below in terms of the original distributions of (S_1, T_1) and (S_2, T_2). The latter will prove more convenient when time comes to compare pairs of order statistics, in Section 4.

First, recall that according to Lehmann [13], a variable T is said to be stochastically increasing in another variable S if and only if, for all $s, s', t \in \mathbb{R}$,

$$s \leqslant s' \Rightarrow \mathrm{P}(T \leqslant t | S = s') \leqslant \mathrm{P}(T \leqslant t | S = s). \tag{6}$$

If H denotes the joint distribution of the pair (S, T), write $H_{[s]}$ for the distribution function of the conditional distribution of T given $S = s$. The above implication may then be expressed in the alternate form

$$s \leqslant s' \Rightarrow H_{[s']} \circ H_{[s]}^{-1}(u) \leqslant u,$$

where $u \in (0,1)$. For convenience, it will be assumed henceforth that $H_{[s]}$ is continuous and strictly increasing for every $s \in \mathbb{R}$, but obvious adaptations are possible when $H_{[s]}$ has plateaus or jumps, and when the domain of S is restricted to an interval.

Note that property (6) is not symmetric in S and T, but that in case these variables are independent, $H_{[s']} \circ H_{[s]}^{-1}(u) \equiv u$ for all $u \in (0,1)$ and for all $s, s' \in \mathbb{R}$. Observe also that if $\xi_p = F^{-1}(p)$ denotes the pth quantile of the marginal distribution of S, then (6) is equivalent to the condition

$$0 < p \leqslant q < 1 \Rightarrow H_{[\xi_q]} \circ H_{[\xi_p]}^{-1}(u) \leqslant u$$

holding true for all $u \in (0,1)$.

This leads to the following definition of what it means for a bivariate distribution to be more stochastically increasing (or monotone regression dependent) than another one.

Definition 1. T_2 is said to be more stochastically increasing in S_2 than T_1 is in S_1, denoted by $(T_1 \mid S_1) \prec_{\mathrm{SI}} (T_2 \mid S_2)$ or $H_1 \prec_{\mathrm{SI}} H_2$, if and only if

$$0 < p \leqslant q < 1 \Rightarrow H_{2[\xi_{2q}]} \circ H_{2[\xi_{2p}]}^{-1}(u) \leqslant H_{1[\xi_{1q}]} \circ H_{1[\xi_{1p}]}^{-1}(u) \tag{7}$$

for all $u \in (0,1)$, where for $i = 1, 2$, $H_{i[s]}$ denotes the conditional distribution of T_i given $S_i = s$, and $\xi_{ip} = F_i^{-1}(p)$ stands for the pth quantile of the marginal distribution of S_i.

Obviously, (7) implies that T_2 is stochastically increasing in S_2 if S_1 and T_1 are independent. It also implies that if T_1 is stochastically increasing in S_1, then so is T_2 in S_2; and conversely, if T_2 is stochastically decreasing in S_2, then so is T_1 in S_1.

The bivariate normal family provides a simple illustration of a system of distributions that is ordered by $\prec_{\mathrm{SI}}$; in this case, one has $N_\kappa(\mu, \Sigma) \prec_{\mathrm{SI}} N_{\kappa'}(\mu', \Sigma') \Leftrightarrow \kappa \leqslant \kappa'$, where κ is either one of Pearson's, Spearman's or Kendall's coefficient. Numerous additional examples of bivariate distributions that are ordered in this fashion are given by Yanagimoto and Okamoto [22], Schriever [18], Capéraà and Genest [5,6], Fang and Joe [9], as well as Joe [10, Chapters 2 and 5]. The above definition coincides with theirs when the pairs (S_1, T_1) and (S_2, T_2) have the same margins, i.e., when $F_1 = F_2$ and $G_1 = G_2$. When the margins are different, Definition 1 is then equivalent to that given by these authors, as applied to the underlying copulas C_1 and C_2.

The main result to be proved in this paper may now be stated as follows.

Proposition 2. *Let* $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ *and* $X_{1:n'} \leqslant \cdots \leqslant X_{n':n'}$ *be the order statistics associated with two independent random samples of sizes* n *and* n' *from the same continuous distribution. Under conditions* (3), *one has*

$$(X_{j':n'} \mid X_{i':n'}) \prec_{\mathrm{SI}} (X_{j:n} \mid X_{i:n}).$$

3. Auxiliary material

The proof of Proposition 2 to be given in Section 4 relies heavily on the notion of dispersive ordering between two random variables X and Y, and properties thereof. For completeness, the definition of this concept is recalled below.

Definition 3. A random variable X with distribution function F is said to be less dispersed than another variable Y with distribution G, written as $X \prec_{\mathrm{DISP}} Y$ or $F \prec_{\mathrm{DISP}} G$, if and only if

$$F^{-1}(\beta) - F^{-1}(\alpha) \leqslant G^{-1}(\beta) - G^{-1}(\alpha)$$

for all $0 < \alpha \leqslant \beta < 1$. Equivalently, one must have $F\{F^{-1}(u) - c\} \leqslant G\{G^{-1}(u) - c\}$ for every $c \geqslant 0$ and $u \in (0, 1)$.

For general information about the dispersive ordering and its properties, refer to Section 2.B of Shaked and Shanthikumar [19]. Of immediate relevance here is the following observation, which derives from a connection originally made by Lewis and Thompson [14] between dispersive random variables and strongly unimodal distributions (see, e.g., [11]).

Lemma 4. *Let* X_1, X_2, Y_1, Y_2 *be mutually independent random variables that are strongly unimodal, i.e., whose densities are log-concave. Then*

$$X_1 \prec_{\mathrm{DISP}} X_2 \text{ and } Y_1 \prec_{\mathrm{DISP}} Y_2 \Rightarrow X_1 + Y_1 \prec_{\mathrm{DISP}} X_2 + Y_2.$$

The proof of Proposition 2 will also make use of the following result concerning the dispersive ordering between generalized spacings associated with two random samples of possibly different sample sizes from an exponential distribution. This result may be of independent interest.

Lemma 5. *Let* $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ *be the order statistics associated with a random sample of size* n *from an exponential distribution, and for* $0 \leqslant i < j \leqslant n$, *let*

$$D_{ij}^{(n)} = X_{j:n} - X_{i:n}$$

stand for the (i,j)*th generalized spacing, with* $X_{0:n} \equiv 0$*. Then for* $j - i \leqslant j' - i'$ *and* $n' - j' \leqslant n - j$*, one has* $D_{ij}^{(n)} \prec_{\mathrm{DISP}} D_{i'j'}^{(n')}$.

Proof. Let $X_1, \ldots, X_n$ and $X'_1, \ldots, X'_{n'}$ be two independent random samples from an exponential distribution with hazard rate λ. Then $D_{ij}^{(n)}$ may be expressed as a convolution of $j-i$ consecutive spacings, namely

$$D_{ij}^{(n)} = (X_{j:n} - X_{j-1:n}) + \cdots + (X_{i+1:n} - X_{i:n}) \equiv \sum_{k=1}^{j-i} E_{n-j+k},$$

where the E_ℓ are mutually independent exponential random variables, the hazard rate of E_ℓ being $\ell\lambda$. Similarly,

$$D_{i'j'}^{(n')} \equiv \sum_{k=1}^{j'-i'} E'_{n'-j'+k}$$

for some mutually independent exponential random variables E'_ℓ with hazard rate $\ell\lambda$.

Now it is easy to see that for $k = 1, \ldots, j-i$ and $n' - j' \leqslant n - j$, one has

$$E_{n-j+k} \prec_{\mathrm{DISP}} E'_{n'-j'+k}.$$

Since the class of distributions with log-concave densities is closed under convolutions of independent random variables (see [7, p. 17]), it thus follows from repeated applications of Lemma 4 that

$$\sum_{k=1}^{j-i} E_{n-j+k} \prec_{\mathrm{DISP}} \sum_{k=1}^{j-i} E'_{n'-j'+k}.$$

A further application of Lemma 4 implies that

$$\sum_{k=1}^{j-i} E'_{n'-j'+k} \prec_{\mathrm{DISP}} \sum_{k=1}^{j-i} E'_{n'-j'+k} + \sum_{k=j-i+1}^{j'-i'} E'_{n'-j'+k},$$

since the two summands on the right-hand side are sums of mutually independent exponential random variables, and hence are independent and have log-concave densities. This concludes the proof. □

Note in passing that if $i = i' = 0$ in Lemma 5, then one has

$$j \leqslant j' \text{ and } n' - j' \leqslant n - j \Rightarrow X_{j:n} \prec_{\mathrm{DISP}} X_{j':n'}, \tag{8}$$

a fact that was already established by Khaledi and Kochar [11].

Finally, the following lemma formalizes the observation that the copula associated with a pair of order statistics does not depend on the parent distribution.

Lemma 6. *Let* $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ *be the order statistics associated with a random sample of size n from a continuous distribution F. The pairs* $(X_{i:n}, X_{j:n})$ *and* $(U_{i:n}, U_{j:n}) = (F(X_{i:n}), F(X_{j:n}))$ *then share the same copula, whatever the choices of* $1 \leqslant i < j \leqslant n$.

Proof. Let F_i and G_i denote the marginal distributions of $X_{i:n}$ and $U_{i:n}$, respectively. Then $F_i = G_i \circ F$, since the probability integral transformation $U = F(X)$ is order preserving, and thus converts the ith order statistic of F into the ith order statistic of a uniform random variable on $(0, 1)$. Thus

$$\mathrm{P}\{X_{i:n} \leqslant F_j^{-1}(u), X_{j:n} \leqslant F_j^{-1}(v)\} = \mathrm{P}\{U_{i:n} \leqslant G_j^{-1}(u), U_{j:n} \leqslant G_j^{-1}(v)\}$$

for all $u, v \in (0, 1)$, which establishes the coincidence of the copulas. □

4. Proof of Proposition 2

In view of Lemma 6, it may be assumed without loss of generality that the parent distribution of the X_k is exponential. Now under this assumption, the consecutive spacings are mutually independent. Therefore,

$$H_{2[x]}(y) \equiv \mathrm{P}(X_{j:n} \leqslant y \mid X_{i:n} = x) = \mathrm{P}(X_{j:n} - X_{i:n} \leqslant y - x \mid X_{i:n} = x)$$
$$= \mathrm{P}(X_{j:n} - X_{i:n} \leqslant y - x) = L_{ijn}(y - x) \text{ (say)},$$

namely the distribution function of $D_{ij}^{(n)}$ at $y - x$.

Let ξ_{2p} and ξ_{2q} denote the pth and qth quantiles of $X_{i:n}$, respectively. Then for $0 < p \leqslant q < 1$,

$$H_{2[\xi_{2q}]} \circ H_{2[\xi_{2p}]}^{-1}(v) = L_{ijn}\{L_{ijn}^{-1}(v) - (\xi_{2q} - \xi_{2p})\}, \tag{9}$$

for arbitrary $v \in (0, 1)$. Similarly, for the order statistics $X_{j':n'}$ and $X_{i':n'}$, one has

$$H_{1[\xi_{1q}]} \circ H_{1[\xi_{1p}]}^{-1}(v) = L_{i'j'n'}\{L_{i'j'n'}^{-1}(v) - (\xi_{1q} - \xi_{1p})\},$$

for all $v \in (0, 1)$, where ξ_{1p} and ξ_{1q}, respectively, denote the pth and qth quantiles of the distribution of $X_{i':n'}$.

In order to prove Proposition 2, therefore, one needs only show that under conditions (3), one has

$$0 < p \leqslant q < 1 \Rightarrow H_{2[\xi_{2q}]} \circ H_{2[\xi_{2p}]}^{-1}(v) \leqslant H_{1[\xi_{1q}]} \circ H_{1[\xi_{1p}]}^{-1}(v),$$

i.e.,

$$L_{ijn}\{L_{ijn}^{-1}(v) - (\xi_{2q} - \xi_{2p})\} \leqslant L_{i'j'n'}\{L_{i'j'n'}^{-1}(v) - (\xi_{1q} - \xi_{1p})\} \tag{10}$$

for all $v \in (0, 1)$.

Now under the assumed condition that $i' \leqslant i$ and $n - i \leqslant n' - i'$, it follows from (8) that $X_{i':n'} \prec_{\mathrm{DISP}} X_{i:n}$, so that $0 \leqslant \xi_{1q} - \xi_{1p} \leqslant \xi_{2q} - \xi_{2p}$ for $0 < p \leqslant q < 1$. Thus for fixed $v \in (0, 1)$, it follows that

$$L_{ijn}\{L_{ijn}^{-1}(v) - (\xi_{2q} - \xi_{2p})\} \leqslant L_{ijn}\{L_{ijn}^{-1}(v) - (\xi_{1q} - \xi_{1p})\}. \tag{11}$$

At the same time, however, Lemma 5 implies that $D_{ij}^{(n)} \prec_{\mathrm{DISP}} D_{i'j'}^{(n')}$, so that

$$L_{ijn}\{L_{ijn}^{-1}(v) - c\} \leqslant L_{i'j'n'}\{L_{i'j'n'}^{-1}(v) - c\} \tag{12}$$

for every $c \geqslant 0$ and hence in particular when $c = \xi_{1q} - \xi_{1p}$. The conjunction of (11) and (12) yields (10), so the proof is complete. □

The following set of immediate consequences of Proposition 2 is of special interest.

Corollary 7. *Let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ be the order statistics associated with a random sample $X_1, \ldots, X_n$ from some continuous distribution. Then*

(a) $(X_{k:n} \mid X_{i:n}) \prec_{\mathrm{SI}} (X_{j:n} \mid X_{i:n})$ *for all* $1 \leqslant i < j < k \leqslant n$;
(b) $(X_{j:n} \mid X_{i:n}) \prec_{\mathrm{SI}} (X_{j+1:n+1} \mid X_{i+1:n+1})$ *for all* $1 \leqslant i < j \leqslant n$;
(c) $(X_{n+1:n+1} \mid X_{1:n+1}) \prec_{\mathrm{SI}} (X_{n:n} \mid X_{1:n})$ *for every integer* $n \geqslant 2$.

It is clear from the above result that for fixed n, the association between the components of a pair $(X_{i:n}, X_{j:n})$ of order statistics, as measured by the $\prec_{\mathrm{SI}}$ ordering, decreases as i and j get further apart. This finding generalizes those of Tukey [21] and Kim and David [12]. It may also be seen from the above that the dependence of the largest order statistic on the smallest one decreases as sample size increases.

It is worth emphasizing here that contrary to Tukey [21] and Kim and David [12], Proposition 2 and Corollary 7 do not rely on any specific assumption about the parent distribution of the order statistics. This is in contrast with the results of Avérous and Dortet-Bernadet [1] concerning the ordering of the largest order statistic on the smallest one in the non-copula-based formulation of the more stochastically increasing ordering that they use.

The following corollary makes it clear that under the conditions given in Proposition 2, any measure of concordance satisfying the axioms of Scarsini [16] will agree with the ordering $\prec_{\mathrm{SI}}$, whereas covariance (which is not a margin-free measure of association) may not.

Corollary 8. *Let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ and $X_{1:n'} \leqslant \cdots \leqslant X_{n':n'}$ be the order statistics associated with two independent random samples of sizes n and n' from the same continuous distribution. Under conditions* (3), *one has*

$$\kappa(X_{i':n'}, X_{j':n'}) \leqslant \kappa(X_{i:n}, X_{j:n}),$$

where κ may stand for Spearman's rho, Kendall's tau, Gini's coefficient, or any other measure of concordance in the sense of Scarsini [16].

5. Kendall's tau for a pair of order statistics

In the course of checking the validity of Corollary 7 in specific cases, it came to the authors' attention that a simple closed-form formula could be found for the population value of Kendall's τ coefficient of concordance between any two order statistics associated with a random sample from a continuous distribution. This result, which is given next, may be viewed as an extension of a contemporaneous finding of Schmitz [17], who only considered the case $i = 1$, $j = n$.

Proposition 9. *Let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ be the order statistics associated with a random sample of size n from some continuous distribution. Then for $1 \leqslant i < j \leqslant n$, the population value of Kendall's coefficient of concordance between $X_{i:n}$ and $X_{j:n}$ is given by*

$$\tau(X_{i:n}, X_{j:n}) = 1 - \frac{2(n-1)}{2n-1}\binom{n-2}{i-1}\binom{n-i-1}{j-i-1}$$

$$\times \sum_{s=0}^{n-j}\sum_{r=0}^{i-1}\binom{n}{r}\binom{n-r}{s}\Big/\binom{2n-2}{n-j+s,\, r+i-1}.$$

Proof. Let $Y_1, \dots, Y_n$ be an independent random sample from the same distribution as the X_k, and let $Y_{1:n} \leqslant \cdots \leqslant Y_{n:n}$ be the corresponding order statistics. By definition, $\tau(X_{i:n}, X_{j:n}) = 1 - 4p$, where

$$p = \mathrm{P}(X_{i:n} < Y_{i:n}, X_{j:n} > Y_{j:n}).$$

To compute this probability, it suffices to determine the proportion of the $(2n)!$ equally likely arrangements of the X_k and the Y_ℓ for which the event

$$X_{i:n} < Y_{i:n} < Y_{j:n} < X_{j:n} \tag{13}$$

occurs. To this end, suppose that $X_{i:n} = X_m$ and $X_{j:n} = X_{m'}$ for some fixed $m, m' \in \{1, \dots, n\}$ with $m \neq m'$. In order that (13) holds, the remaining $n-2$ of the X_k and all the Y_ℓ must then be positioned in such a way that, for some $r \in \{0, \dots, i-1\}$ and $s \in \{0, \dots, n-j\}$,

(i) exactly $i-1$ of the X_k and exactly r of the Y_ℓ are less than X_m;
(ii) exactly $n-j$ of the X_k and exactly s of the Y_ℓ are greater than $X_{m'}$;
(iii) the remaining $j-i-1$ values of the X_k and $n-r-s$ values of the Y_ℓ are located in the interval $(X_m, X_{m'})$.

Upon summing over the different possible values of r and s, one finds

$$p = n(n-1)\binom{n-2}{i-1}\binom{n-i-1}{j-i-1}$$

$$\times \sum_{s=0}^{n-j}\sum_{r=0}^{i-1}\binom{n}{r}\binom{n-r}{s}\frac{(n-r-s+j-i-1)!(r+i-1)!(n-j+s)!}{(2n)!},$$

where the factor $n(n-1)$ at the beginning of the formula comes because there are that many ways of choosing X_m and $X_{m'}$, and the fraction inside the sum is obtained through an enumeration of the possible arrangements of the other X_k and Y_ℓ, conditionally on (i)–(iii) and the positions of X_m and $X_{m'}$. A simple algebraic manipulation then yields the final formula for tau. □

The above formula for Kendall's tau simplifies as follows in a few special cases:

(a) for $1 \leqslant i < j = n$,

$$\tau(X_{i:n}, X_{n:n}) = -1 + \frac{2(n-1)}{2n-1}\binom{n-2}{i-1}\sum_{r=i}^{n}\binom{n}{r}\Big/\binom{2n-2}{r+i-1};$$

(b) for $1 \leqslant i < j = i+1 \leqslant n$,

$$\tau(X_{i:n}, X_{i+1:n}) = 1 - \binom{n}{i}^2\Big/\binom{2n}{2i};$$

(c) for $i = 1$ and $j = n$, $\tau(X_{1:n}, X_{n:n}) = 1/(2n-1)$, as reported by Schmitz [17].

For illustration purposes, Tables 1 and 2 give the values of $\tau(X_{i:n}, X_{j:n})$ for all choices of $1 \leqslant i < j \leqslant n$ and for $n = 6$ and 7, respectively. The various monotonicity properties stated in Proposition 2 and its corollaries can be readily verified from these tables. In addition, the tables show an obvious diagonal symmetry property that is not immediately clear from Proposition 9. This is a simple consequence of the following result.

Proposition 10. *Let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ be the order statistics associated with a random sample of size n from some continuous distribution. Then for arbitrary $i, j \in \{1, \ldots, n\}$, the pairs $(-X_{i:n}, -X_{j:n})$ and $(X_{n-i+1:n}, X_{n-j+1:n})$ have the same copula. Consequently, one has*

$$\kappa(X_{i:n}, X_{j:n}) = \kappa(X_{n-i+1:n}, X_{n-j+1:n}), \tag{14}$$

where κ is any measure of concordance in the sense of Scarsini [16].

Proof. Since by Lemma 6 the copula of a pair of order statistics has the distribution-free property, it can be assumed without loss of generality that the parent distribution is uniform on the interval $(0, 1)$. Under this assumption, it can be easily verified that the pairs $(X_{n-i+1:n}, X_{n-j+1:n})$ and $(1 - X_{i:n}, 1 - X_{j:n})$ have the same joint

Table 1
The values of $3003 \times \tau(X_{i:6}, X_{j:6})$

i	j				
	2	3	4	5	6
1	1365	910	650	455	273
2		1638	1118	767	455
3			1703	1118	650
4				1638	910
5					1365

Table 2
The values of $3003 \times \tau(X_{i:7}, X_{j:7})$

i	j					
	2	3	4	5	6	7
1	1386	945	700	525	378	231
2		1680	1190	875	623	378
3			1778	1253	875	525
4				1778	1190	700
5					1680	945
6						1386

distribution. Therefore, $(-X_{i:n}, -X_{j:n})$ and $(X_{n-i+1:n}, X_{n-j+1:n})$ have the same copula, and (14) holds by Axiom 5 of Scarsini (see [15, p. 136]). □

It may also be seen from Tables 1 and 2 that $\tau(X_{i:n}, X_{j:n})$ increases with sample size n for fixed $1 \leqslant i < j \leqslant n$, a fact that can be verified readily in the special cases discussed above, as well as when $j = i + 1$. This is possibly true in general. Furthermore, it is easy to check from the special case (b) mentioned above that $\tau(X_{i:n}, X_{i+1:n})$ increases in i for $1 \leqslant i \leqslant \lceil (n-1)/2 \rceil$, where $\lceil x \rceil$ denotes the smallest integer $y \geqslant x$. More generally, it would appear (but remains to be shown) that

$$\tau(X_{i:n}, X_{i+k:n}) \leqslant \tau(X_{i+1:n}, X_{i+k+1:n})$$

for all $1 \leqslant i \leqslant \lceil (n-k)/2 \rceil$.

In his paper, Schmitz [17] gives an explicit formula for the value of Spearman's rho between the smallest and largest order statistics in a random sample of arbitrary size. Unfortunately, it does not seem possible to generalize this expression to any two order statistics, although the coefficient can be computed easily in specific cases using a symbolic calculator such as MAPLE. Still, it may be observed (as Schmitz does in his special case) that

$$\rho(X_{i:n}, X_{j:n}) \geqslant \tau(X_{i:n}, X_{j:n}),$$

since Capéraà and Genest [6] showed that these two measures of dependence are so ordered whenever the pair of variables under consideration is in positive likelihood ratio dependence. That such is the case for a pair $(X_{i:n}, X_{j:n})$ of order statistics from a random sample is a well known fact (see, e.g., [4, p. 78]).

6. Conclusion

This paper has continued the work of Tukey [21] and Kim and David [12] by comparing the degree of association present in two pairs of order statistics from the same continuous distribution. Conditions were found under which the copulas of two such pairs are ordered in the stochastically increasing, or monotone regression dependence, ordering of Yanagimoto and Okamoto [22].

There are several ways in which this investigation might be continued. One possibility would be to seek more restrictive conditions under which stochastic increasingness could be replaced by stronger dependence orderings. Other options would be to consider the case of discrete or non-identically distributed observations. For example, partial results are already available from the authors for an ordering of Capéraà and Genest [5] that strengthens $\prec_{SI}$, and for heterogenous exponential parent distributions. Because of the possibility of ties, however, extensions to the case of discrete parent distributions will probably prove most challenging.

Acknowledgments

This research benefited from the financial support of the Natural Sciences and Engineering Research Council of Canada and from the Fonds québécois de la recherche sur la nature et les technologies. The work was carried out in part while the third author made extended visits to Université Laval in 2002–2003.

References

[1] J. Avérous, J.-L. Dortet-Bernadet, LTD and RTI dependence orderings, Canad. J. Statist. 28 (2000) 151–157.
[2] P.J. Bickel, Some contributions to the theory of order statistics, in: L.M. LeCam, J. Neyman (Eds.), Fifth Berkeley Symposium on Mathematics and Statistics, Vol. 1, University of California Press, Berkeley, CA, 1967, pp. 575–591.
[3] H.W. Block, D. Chhetry, Z. Fang, A.R. Sampson, Partial orders on permutations and dependence orderings on bivariate empirical distributions, Ann. Statist. 18 (1990) 1840–1850.
[4] P.J. Boland, M. Hollander, K. Joag-Dev, S.C. Kochar, Bivariate dependence properties of order statistics, J. Multivariate Anal. 56 (1996) 75–89.
[5] P. Capéraà, C. Genest, Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles, Canad. J. Statist. 18 (1990) 315–326.
[6] P. Capéraà, C. Genest, Spearman's rho is larger than Kendall's tau for positively dependent random variables, J. Nonparametric Statist. 2 (1993) 183–194.
[7] S.W. Dharmadhikari, K. Joag-Dev, Unimodality, Convexity and Applications, Academic Press, San Diego, CA, 1988.
[8] D. Drouet-Mari, S. Kotz, Correlation and Dependence, Imperial College Press, London, 2001.
[9] Z. Fang, H. Joe, Further developments on some dependence orderings for continuous bivariate distributions, Ann. Inst. Statist. Math. 44 (1992) 501–517.
[10] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, London, 1997.
[11] B.-E. Khaledi, S.C. Kochar, On dispersive ordering between order statistics in one-sample and two-sample problems, Statist. Probab. Lett. 46 (2000) 257–261.
[12] S.H. Kim, H.A. David, On the dependence structure of order statistics and concomitants of order statistics, J. Statist. Plann. Inference 24 (1990) 363–368.
[13] E.L. Lehmann, Some concepts of dependence, Ann. Math. Statist. 37 (1966) 1137–1153.
[14] T. Lewis, J.W. Thompson, Dispersive distributions, and the connection between dispersivity and strong unimodality, J. Appl. Probab. 18 (1981) 76–90.

[15] R.B. Nelsen, An Introduction to Copulas, in: Lecture Notes in Statistics, Vol. 139, Springer, New York, 1999.
[16] M. Scarsini, On measures of concordance, Stochastica 8 (1984) 201–218.
[17] V. Schmitz, Revealing the dependence structure between $X_{(1)}$ and $X_{(n)}$, J. Statist. Plann. Inference, to appear.
[18] B.F. Schriever, An ordering for positive dependence, Ann. Statist. 15 (1987) 1208–1214.
[19] M. Shaked, J.G. Shanthikumar, Stochastic Orders and Their Applications, Academic Press, San Diego, CA, 1994.
[20] A.H. Tchen, Inequalities for distributions with given marginals, Ann. Probab. 8 (1980) 814–827.
[21] J.W. Tukey, A problem of Berkson, and minimum variance orderly estimators, Ann. Math. Statist. 29 (1958) 588–592.
[22] T. Yanagimoto, M. Okamoto, Partial orderings of permutations and monotonicity of a rank correlation statistic, Ann. Inst. Statist. Math. 21 (1969) 489–506.