Dependence among order statistics - a review and some recent results

Subhash Kochar
Department of Mathematics and Statistics
Portland State University, Portland, OR, USA
Email:kochar@pdx.edu

Maochao Xu
Department of Mathematics
Illinois State University, Normal, IL, USA
Email:mxu2@ilstu.edu

November 1, 2011

Abstract

This paper first reviews some recent developments on dependence among order statistics. Some new results on order statistics from bivariate models are discussed as well.

Key Words Dependence; Exchangeable; Order Statistics; Proportional hazard rates; Relative dependence.

1 Introduction

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of random variables $X_1, X_2, \cdots, X_n$. In the reliability theory context, $X_{n-k+1:n}$ denotes the lifetime of a k-out-of-n system. In particular, the parallel and series systems are 1-out-of-n and n-out-of-n systems. Order statistics have received a tremendous amount of attention from many researchers since they play an important role in reliability, data analysis, goodness-of-fit tests, statistical inference and other applied probability areas. Please refer to *David and Nagaraja* (2003) and *Balakrishnan and Rao* (1998a, 1998b) for more details.

Due to the wide applications, the nature of dependence that may exist between order statistics has received much attention. If $X_1, \ldots, X_n$ are independent and identical random variables, *Bickel* (1967)

This article is dedicated to Professor Jayant Deshpande on his $70th$ birthday. Many of the ideas in this paper are directly or indirectly inspired by a number of conversations with him. We sincerely wish him many more long years of happy and active life.

first showed that

$$\mathrm{Cov}(X_{i:n}, X_{j:n}) \geq 0.$$

This topic has been followed and developed by many researchers including *Karlin and Rinott* (1980), *Kim and David* (1990), *Boland, et al.* (1996), *Avérous* (2005), *Hu and Xie* (2006) and *Dubhashi and Häggström* (2008), among others.

In this paper, we first review some recent developments on dependence among order statistics. Then, we present some new results on dependence among order statistics based on dependent random variables. The rest part of this paper is organized as follows. In Section 2, we recall some dependence notions used in the paper. In Section 3, we review recent results on dependence among order statistics. In Section 4, two common measures of dependence for order statistics are discussed. In Section 5, the topic of relative degree of dependence among order statistics is reviewed, and some new results on dependence among order statistics based on dependent random variables are also developed. In the last section, we mention some open problems in this area.

2 Preliminaries

In this section, we review some dependence notions, which will be used in the sequel.

The following definitions can be found in Chapter 5 of *Barlow and Proschan* (1981).

Definition 2.1 Given two random variables X and Y, we say the following:

(a) Y is stochastically increasing in X, denoted by $\mathrm{SI}(Y|X)$, if $P(Y > y \mid X = x)$ is increasing in x for all y; or equivalently,

$$P(Y \leq y|X = x) \geq P(Y \leq y|X = x^*), \quad x \leq x^*. \tag{2.1}$$

(b) Y is right tail increasing in X, denoted by $\mathrm{RTI}(Y|X)$, if $P(Y > y \mid X > x)$ is increasing in x for all y.

(c) Y is left tail decreasing in X, denoted by $\mathrm{LTD}(Y|X)$, if $P(Y \leq y \mid X \leq x)$ is decreasing in x for all y.

For more dependence concepts, please refer to *Joe* (1997) and *Nelsen* (2006) for a comprehensive discussion.

Observing that when X and Y are continuous, inequality (2.1) can be written as

$$H_{[\xi_q]} \circ H^{-1}_{[\xi_p]}(u) \leq u$$

where $\xi_p = F^{-1}(p)$ stands for the pth quantile of the marginal distribution of X, and $H_{[s]}$ denotes the conditional distribution of Y given $X = s$. *Avérous, et al.* (2005) proposed the following definition to measure the relative degree of monotone dependence between two pairs of bivariate random variables (X_1, Y_1) and (X_2, Y_2).

Definition 2.2 Y_1 is said to be less *stochastic increasing* in X_1 than Y_2 is in X_2, denoted by $(Y_1|X_1) \preceq_{\mathrm{SI}} (Y_2|X_2)$, if and only if, for $0 \leq u \leq 1$, and $0 < p \leq q < 1$,

$$H_{2[\xi_{2q}]} \circ H^{-1}_{2[\xi_{2p}]}(u) \leq H_{1[\xi_{1q}]} \circ H^{-1}_{1[\xi_{1p}]}(u),$$

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

Dolati, et al. (2008) proposed the following weaker dependence order based on RTI criteria, called more RTI order.

Definition 2.3 Y_1 is said to be less *right-tail increasing* (RTI) in X_1 than Y_2 is in X_2, denoted by $(Y_1|X_1) \preceq_{\rm RTI} (Y_2|X_2)$, if and only if, for $0 \le u \le 1$, and $0 < p \le q < 1$,

$$H^*_{2[\xi_{2q}]} \circ {H^*}^{-1}_{2[\xi_{2p}]}(u) \le H^*_{1[\xi_{1q}]} \circ {H^*}^{-1}_{1[\xi_{1p}]}(u),$$

where $\xi_{ip} = F_i^{-1}(p)$ stands for the pth quantile of the marginal distribution of X_i, and $H^*_{i[s]}$ denotes the conditional distribution of Y_i given $X_i > s$, for $i = 1, 2$.

It is easy to see that both more SI order and more RTI order are *copula-based* orders, and more SI order implies more RTI order which in turn implies more concordance ordering (that is, the two copulas are ordered). For the concept of *copula*, please refer to *Nelsen* (2006) for more details.

As observed in *Avérous, et al.* (2005) and *Genest et al.* (2009), there is a close connection between the above concepts of more dependence and the notion of dispersive ordering.

We shall also use the following notion of multivariate stochastic ordering.

Definition 2.4 The random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is said to be smaller than another random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ (denoted by $\mathbf{X} \preceq_{\rm st} \mathbf{Y}$) according to the multivariate stochastic ordering if

$$\mathrm{E}[\phi(\mathbf{X})] \le \mathrm{E}[\phi(\mathbf{Y})]$$

for all increasing functions ϕ. It is known that multivariate stochastic order implies component-wise stochastic order. For more details on the multivariate stochastic orders, see *Shaked and Shanthikumar* (2007) and *Müller and Stoyan* (2002).

3 Order statistics based on independent observations

Before we review the main results on dependence among order statistics, we first recall the proportional hazard rates (PHR) model.

Independent random variables $X_1, X_2, \ldots, X_n$ are said to follow the PHR model if for $i = 1, 2, \ldots, n$, the survival function of X_i can be expressed as,

$$\overline{F}_i(x) = [\overline{F}(x)]^{\lambda_i}, \text{ for } \lambda_i > 0, \tag{3.1}$$

where $\overline{F}(x)$ is the survival function of some random variable X. If $r(t)$ denotes the hazard rate corresponding to the base line distribution F, then the hazard rate of X_i is $\lambda_i r(t)$, $i = 1, 2, \ldots, n$. We can equivalently express (3.1) as

$$\overline{F}_i(x) = e^{-\lambda_i R(x)}, \; i = 1, 2, \ldots, n$$

where $R(x) = \int_0^x r(t)dt$, is the cumulative hazard rate of X. Many well-known models are special cases of the PHR model, such as Weibull, Pareto and Lomax.

Boland et al. (1996) studied in detail the dependence properties of order statistics. They proved the following dependence result for the PHR model.

Theorem 3.1 Let $X_1, \dots, X_n$ be independent random variables with differentiable densities and follow the PHR model on an interval. Then $X_{i:n}$ is SI in $X_{1:n}$.

They also gave a counterexample to illustrate that, in general, $X_{i:n}$ is not SI in $X_{1:n}$. However, they showed that for $1 \le i < j \le n$, $X_{j:n}$ is RTI in $X_{i:n}$.

Theorem 3.2 Let $X_1, \dots, X_n$ be independent random variables. Then for any $i \le j$, $\mathrm{RTI}(X_{j:n}|X_{i:n})$ and $\mathrm{LTD}(X_{i:n}|X_{j:n})$.

They showed with the help of a counter example that, in general, the relation $\mathrm{RTI}(X_{i:n}|X_{j:n})$ may not hold for $i < j$.

This topic has been further developed by *Hu and Xie* (2006), where they exploited the negative dependence of occupancy numbers in the balls and bins experiment. They proved the following result.

Theorem 3.3 Let $X_1, \dots, X_n$ be independent random variables. For $1 \le i \le j_1 \le j_2 \le \dots \le j_r \le n$, and $(x_1, \dots, x_n) \in \mathbb{R}^n$,

(a) if $A_{i,n,y} = \{X_{i:n} > y\}$, then

$$P(X_{j_1} > x_1, \dots, X_{j_r} > x_r \mid A_{i,n,y}) \tag{3.2}$$

is increasing in y;

(b) if event $A_{i,n,y}$ is either $\{X_{i:n} > y\}$ or $\{X_{i:n} \le y\}$, then the conditional probability in (3.2) is decreasing in i for each y; and

(c) if each X_k has a continuous distribution function, and if $A_{i,n,y}$ is either $\{X_{i:n} = y\}$ or $\{X_{i-1:n} < y < X_{i:n}\}$, then (3.2) is decreasing in i for each y, where $X_{0:n} \equiv -\infty$.

Dubhashi and Häggström (2008) further extended the above result to the multivariate stochastic comparisons.

Theorem 3.4 Let $X_1, \dots, X_n$ be independent random variables. Then

$$[(X_{i:n}, \dots, X_{n:n})|X_{i:n} > y] \preceq_{\mathrm{st}} [(X_{i:n}, \dots, X_{n:n})|X_{i:n} > y'], \quad y \le y'.$$

Subsequently, Theorem 3.3 was further extended by *Hu and Cheng* (2008) as follows.

Theorem 3.5 Let $X_1, \dots, X_n$ be independent random variables.

(a) If $j - i \ge \max\{n - m, 0\}$, then

$$P(X_{j:n} > x_1, X_{j+1:n} > x_2, \dots, X_{n:n} > x_{n-j+1} \mid X_{i:m} > y)$$

is increasing in y for all $(x_1, \dots, x_{n-j+1}) \in \mathbb{R}^{n-j+1}$;

(b) If $j - i \le \min\{n - m, 0\}$, then

$$P(X_{1:n} \le x_1, X_{2:n} \le x_2, \dots, X_{j:n} \le x_{n-j+1} \mid X_{i:m} \le y)$$

is decreasing in y for all $(x_1, \dots, x_j) \in \mathbb{R}^j$.

Recently, *Zhuang, et al.* (2010) discussed the dependence among order statistics in the sense of multivariate stochastic comparisons, which extends the results in *Hu and Cheng* (2008) and *Dubhashi and Häggström* (2008).

Theorem 3.6 Let $X_1, \ldots, X_n$ be independent random variables.

(a) If $j - i \geq \max\{n-m, 0\}$, then

$$[(X_{j:n}, \ldots, X_{n:n})|X_{i:m} > y] \preceq_{\mathrm{st}} [(X_{j:n}, \ldots, X_{n:n})|X_{i:m} > y'], \quad y \leq y'.$$

(b) If $j - i \leq \min\{n-m, 0\}$, then

$$[(X_{1:n}, \ldots, X_{j:n})|X_{i:m} \leq y] \preceq_{\mathrm{st}} [(X_{1:n}, \ldots, X_{j:n})|X_{i:n} \leq y'], \quad y \leq y'.$$

4 Kendall's τ and Spearman's ρ for order statistics

Two popular nonparametric measures of association for bivariate random variables are Kendall's τ and Spearman's ρ, which measure different aspects of the dependence structure. In terms of dependence properties, Spearman's ρ is a measure of average quadrant dependence, while Kendall's τ is a measure of average likelihood ratio dependence (*Nelsen*, 1992 and *Fredricks and Nelsen*, 2007).

Avérous, et al. (2005) made an important observation that in the case of a random sample from a continuous distribution with cdf F, the copula of a pair of order statistics is independent of the parent distribution F. As a result the value of any copula based measure of dependence like Kendall's tau or Spearman's coefficient for any pair of order statistics $(X_{i:n}, X_{j:n})$ will be the same for all continuous distributions F. *Schmitz* (2004) derived the following formulas:

$$\tau(X_{1:n}, X_{n:n}) = \frac{1}{2n-1}, \tag{4.1}$$

and

$$\rho(X_{1:n}, X_{n:n}) = 3 - \frac{12n}{\binom{2n}{n}} \sum_{k=0}^{n} \frac{(-1)^k}{2n-k} \binom{2n}{n+k} + 12\frac{(n!)^3}{(3n)!}(-1)^n. \tag{4.2}$$

Li and Li (2007) proved a conjecture in *Schmitz* (2004) that

$$\frac{\rho(X_{1:n}, X_{n:n})}{\tau(X_{1:n}, X_{n:n})} \longrightarrow \frac{3}{2}, \quad n \longrightarrow \infty. \tag{4.3}$$

Avérous, et al. (2005) used a combinatorial approach to prove the following formula of Kendall's τ for any pair of order statistics from the same continuous distribution:

$$\tau(X_{i:n}, X_{j:n}) \quad = \quad 1 - \frac{2(n-1)}{2n-1}\binom{n-2}{i-1}\binom{n-i-1}{j-i-1}\sum_{s=0}^{n-j}\sum_{r=0}^{i-1}\binom{n}{r}\binom{n-r}{s}\Big/\binom{2n-2}{n-j+s, r+i-1}.$$

Subsequently, *Chen* (2007) developed three new formulas to compute $\rho(X_{1:n}, X_{n:n})$:

$$\rho(X_{1:n}, X_{n:n}) = 3(1 - 4a_n),$$

where a_n can be computed by any one of the following formulas.

(a) Formula 1:

$$a_n = n(n-1)\int_0^1 \int_0^t (1-s)^n t^n (t-s)^{n-2} ds dt;$$

(b) Formula 2:

$$a_n = \sum_{k=0}^{n} (-1)^k \frac{\binom{n}{k}}{\binom{n+k}{k}} \frac{n+k}{2n+k};$$

(c) Formula 3:

$$a_n = \frac{n(n-1)}{(3n)!} \sum_{j=0}^{n} \frac{n!}{j!} \sum_{k=0}^{n} \frac{n!}{k!} (n-2+j+k)!.$$

He further showed the following compound inequality:

$$\frac{3(2n-1)(14n^2+15n+3)}{56n^3+86n^2+43n+7} \le \frac{\rho(X_{1:n}, X_{n:n})}{\tau(X_{1:n}, X_{n:n})} \le \frac{3(2n-1)(14n^2-13n+2)}{56n^3-82n^2+39n-6},$$

from which, Eq. (4.3) follows immediately.

Recently, *Navarro and Balakrishnan* (2010) have also studied this problem and have obtained alternate expressions for computing these measures of dependence.

5 Relative dependence between pairs of order statistics

5.1 Order statistics based on independent observations

Assume that $X_1, \ldots, X_n$ are independent and identically random variables. When the parent distribution has an increasing hazard rate and a decreasing reverse hazard rate, *Tukey* (1958) showed that

$$\mathrm{Cov}(X_{i':n}, X_{j':n}) \le \mathrm{Cov}(X_{i:n}, X_{j:n}) \tag{5.1}$$

for either $i = i'$ and $j \le j'$; or $j = j'$ and $i' \le i$. It is interesting to mention that *Kim and David* (1990) proved that if both the hazard and the reverse hazard rates of the X_i's are increasing, then inequality (5.1) remains valid when $i = i'$ and $j \le j'$; However, the inequality (5.1) is reversed when $j = j'$ and $i' \le i$. *Avérous, et al.* (2005) used the more SI concept to study the relative degree of dependence between two pairs of random variables. They proved the following result.

Theorem 5.1 Let $X_{1:n} \le \ldots X_{n:n}$ and $X_{1:n'} \le \ldots X_{n':n'}$ be the order statistics associated with two independent random samples of sizes n and n' from the same continuous distribution. Then, for $1 \le i \le j \le n$, $1 \le i' \le j' \le n'$, and $i' \le i$, $j - i \le j' - i'$, $n - i \le n' - i'$, $n' - j' \le n - j$, it holds that

$$(X_{j':n'}|X_{i':n'}) \preceq_{\mathrm{SI}} (X_{j:n}|X_{i:n}).$$

As a direct consequence, we have the following result.

Corollary 5.2 Let $X_{1:n} \le \ldots X_{n:n}$ be order statistics from the same continuous distribution. Then,

(a) $(X_{k:n}|X_{i:n}) \preceq_{\mathrm{SI}} (X_{j:n}|X_{i:n})$ for $1 \le i < j < k \le n$;

(b) $(X_{j:n}|X_{i:n}) \preceq_{\mathrm{SI}} (X_{j+1:n+1}|X_{i+1:n+1})$ for $1 \le i < j \le n$;

(c) $(X_{n+1:n+1}|X_{1:n+1}) \preceq_{\mathrm{SI}} (X_{n:n}|X_{1:n})$ for $n \ge 2$.

It can be seen from the above result that the dependence between the components of a pair $(X_{j:n}, X_{i:n})$ of order statistics, decreases in the sense of SI ordering as i and j get further apart. An outline of the proof of Theorem 5.1 is given in the Appendix. This proof is shorter and more elegant than the original proof of *Avérous, et al.* (2005).

Remark 5.1 Since the copula of a pair of order statistics of a random sample is independent of the parent distribution and since the concept of more SI is copula based, it follows that in Theorem 5.1 and Corollary 5.2, the two samples could be from different distributions.

As explained in *Avérous, et al.* (2005), the following result follows immediately from Theorem 5.1.

Corollary 5.3 Under the assumptions of Theorem 5.1, we have

$$\kappa(X_{j':n'}, X_{i':n'}) \leq \kappa(X_{j:n}, X_{i:n}) \tag{5.2}$$

where $\kappa(X, Y)$ represents Spearman's rho, Kendall's tau, Gini's coefficient, or indeed any other copula-based measure of concordance satisfying the axioms of *Scarsini* (1984).

The case of independent but nonidentically distributed random variables

Bapat and Beg (1989) studied the distribution theory of order statistics when the parent observations are independent but nonidentically distributed. *Sathe* (1988) proved that if $X_1, \ldots, X_n$ are independent exponential random variables with distinct parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, then for any $k = 2, \ldots, n$, the Peasrson coefficient of correlation between $X_{k:n}$ and $X_{1:n}$ is maximum when the λ_i's are equal. The natural question is to see if we can extend this result to positive dependence orderings? *Dolati, et al.* (2008) further studied this topic for order statistics from heterogeneous samples. They proved the following result for extreme order statistics.

Theorem 5.4 Let $X_1, \ldots, X_n$ be independent continuous random variables following the PHR model. Let $Y_1, \ldots, Y_n$ be i.i.d. continuous random variables, then

$$(X_{n:n}|X_{1:n}) \prec_{\text{RTI}} (Y_{n:n}|Y_{1:n}).$$

They also wondered whether this result can be strengthened to more SI ordering. *Genest, et al.* (2009) gave a positive answer to this question by showing the following result.

Theorem 5.5 Let $X_1, \ldots, X_n$ be independent continuous random variables following the PHR model. Let $Y_1, \ldots, Y_n$ be i.i.d. continuous random variables, then

$$(X_{n:n}|X_{1:n}) \prec_{\text{SI}} (Y_{n:n}|Y_{1:n}).$$

5.2 Order statistics from exchangeable random variables

Most of the work in the literature on dependence of order statistics has been based on the assumption that the underlying random variables are independent. However, in practice, the underlying random variables are often dependent. For example, in reliability engineering, in a two engine aircraft, the lifetimes of the two engines can not be assumed to be independent since if one of the two engines fails, it affects the working of the other engine. Similarly in a desk top computer with both a central processing unit (CPU) and a co-processor, the components are dependent, as sudden power surge could effect both the components simultaneously (*Nelson*, 1999, p. 46). One of the most famous exponential models, *Marshall-Olkin* model (cf. Barlow and Proschan, 1981), describes this situation perfectly. One quick way to check the dependence of order statistics is though Kendall's tau τ. According to (4.1), for order statistics from independent and identical random variables, it holds that

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

However, for order statistics from dependent random variables, the Kendall's tau depends on the structure of the random variables. There is little work on the dependence of order statistics from dependent samples. *Boland, et al* (1996) initiated some work in their paper. Recently, *Navarro and Balakrishnan* (2010) derived some expressions for Pearson's correlation coefficient between two order statistics. They also mentioned to use a Monte Carlo procedure for Kendall's tau in the case of exchangeable dependent components.

In this section, we study the dependence properties of order statistics from *Marshall-Olkin* bivariate and *Pareto* bivariate models. It is shown that order statistics from those two models have more dependence than those from independent cases according to *more SI order*.

5.2.1 Bivariate Marshall-Olkin model

An exchangeable random vector (X_1, X_2) is said to follow Marshall-Olkin bivariate exponential distribution with parameters (λ, λ_{12}), if the survival function can be expressed as

$$P(X_1 > x, X_2 > y) = \exp\left\{-\lambda x - \lambda y - \lambda_{12} \max\{x, y\}\right\}, \quad \lambda > 0,\, \lambda_{12} > 0.$$

It is seen that

$$P(X_1 > x, X_2 > y) \geq P(X_1 > x)P(X_2 > y),$$

which indicates that (X_1, X_2) have positive dependence. This property is actually called positively quadrant dependence (see Chapter 9 of Shaked and Shanthikumar, 2007).

A natural question would be see how the parameters affect the dependence between order statistics. The following result gives some insight into this problem.

Theorem 5.6 Assume that exchangeable random vector (X_1, X_2) follows a bivariate Marshall-Olkin model with parameters (λ, λ_{12}), and exchangeable random vector (Y_1, Y_2) follows a bivariate Marshall-Olkin model with parameters $(\lambda^*, \lambda_{12}^*)$. Then

$$\frac{\lambda_{12}^*}{\lambda^*} \leq \frac{\lambda_{12}}{\lambda} \Longrightarrow (X_{2:2}^* | X_{1:2}^*) \prec_{\text{SI}} (X_{2:2} | X_{1:2}).$$

Proof: a) From Theorem 1.4 in *Barlow and Proschan* (1981, p. 131), we have

$$\begin{aligned} H_{2[s]}(x) &= P(X_{2:2} \leq x | X_{1:2} = s) \\ &= P(X_{2:2} - X_{1:2} \leq x - s | X_{1:2} = s) \\ &= P(X_{2:2} - X_{1:2} \leq x - s) \\ &= 1 - \frac{2\lambda \exp\{-(\lambda + \lambda_{12})(x - s)\}}{\gamma}, \end{aligned}$$

where $\gamma = 2\lambda + \lambda_{12}$. Since

$$P(X_{1:2} > x) = e^{-\gamma x},$$

the pth quantile of $X_{1:2}$ can be written as

$$\xi_p = -\frac{1}{\gamma}\log(1-p), \quad 0 < p < 1.$$

Hence, for $0 < p \leq q < 1$, it holds that

$$H_{2[\xi_q]} \circ H^{-1}_{2[\xi_p]}(u) = 1 - (1-u) \times \left(\frac{1-p}{1-q}\right)^{\frac{\lambda+\lambda_{12}}{\gamma}}, \tag{5.3}$$

from which, the required result follows. ■

The other interesting question is that if we ignore the dependence between (X_1, X_2), how would this affect the dependence of ($X_{2:2}$ on $X_{1:2}$)? Observing that SI order only depends on the copula structure, without loss of generality, setting $\lambda^*_{12} = 0$ in Eq. (5.3), we immediately have the following result.

Corollary 5.7 Assume that exchangeable random vector (X_1, X_2) follows a bivariate Marshall-Olkin model, and X^*_1 and X^*_2 are independent and identical random variables with arbitrary continuous distributions. Then

$$(X^*_{2:2}|X^*_{1:2}) \prec_{\rm SI} (X_{2:2}|X_{1:2}).$$

Remark 5.2: From Eq. (5.3), it is easy to see that the relative dependence of $(X_{2:2}, X_{1:2})$ is increasing in λ_{12}, but decreasing in λ. Corollary 5.7 also gives a lower bound for $\tau(X_{2:2}, X_{1:2})$ according to (4.1),

$$\tau(X_{2:2}, X_{1:2}) \geq \tau(X^*_{1:2}, X^*_{2:2}) = \frac{1}{3}.$$

Remark 5.3: Let us compute Kendall's tau of $(X_{1:2}, X_{2:2})$ to further understand the implications of the above result. First, note that,

$$\begin{aligned}
& P\left(X_{1:2} > x, X_{2:2} < y\right) \\
= & \int_x^y \int_x^v f_{1:2,2:2}(u,v)dudv \\
= & \int_x^y \int_x^v h_{2,1}(v-u) f_{1:2}(u)dudv \\
= & 2\lambda(\lambda+\lambda_{12}) \int_x^y \int_x^v e^{-(\lambda+\lambda_{12})(v-u)} e^{-\gamma u} dudv \\
= & 2\lambda(\lambda+\lambda_{12}) \int_x^y e^{-(\lambda+\lambda_{12})v} \int_x^v e^{-\lambda u} dudv \\
= & 2(\lambda+\lambda_{12}) \int_x^y e^{-(\lambda+\lambda_{12})v} \left(e^{-\lambda x} - e^{-\lambda v}\right) dv \\
= & 2e^{-\lambda x}\left[e^{-(\lambda+\lambda_{12})x} - e^{-(\lambda+\lambda_{12})y}\right] - \frac{2(\lambda+\lambda_{12})}{\gamma}\left(e^{-\gamma x} - e^{-\gamma y}\right).
\end{aligned}$$

Now, assume that $(Y_{1:2}, Y_{2:2})$ also follows a bivariate Marshall-Olkin model, then

$$\tau(X_{1:2}, X_{2:2}) = 1 - 4p,$$

where,

$$p = P(X_{1:2} > Y_{1:2}, X_{2:2} < Y_{2:2}).$$

Hence,

$$
\begin{aligned}
& P(X_{1:2} > Y_{1:2}, X_{2:2} < Y_{2:2}) \\
= & \int_0^\infty \int_0^y P(X_{1:2} > x, X_{2:2} < y) f_{1:2,2:2}(x,y)\,dxdy \\
= & \; 2\lambda(\lambda+\lambda_{12}) \int_0^\infty \int_0^y P(X_{1:2} > x, X_{2:2} < y) e^{-(\lambda+\lambda_{12})y} e^{-\lambda x}\,dxdy.
\end{aligned}
$$

After some calculation, the above equality reduces to,

$$P(X_{1:2} > Y_{1:2}, X_{2:2} < Y_{2:2}) = \frac{\lambda(\lambda^2 + 4\lambda^2 - 2\gamma\lambda - \lambda_{12}^2)}{2\gamma^2(\gamma + \lambda + \lambda_{12})}.$$

Thus,

$$\tau(X_{1:2}, X_{2:2}) = 1 - \frac{2\lambda(\lambda^2 + 4\lambda^2 - 2\gamma\lambda - \lambda_{12}^2)}{\gamma^2(\gamma + \lambda + \lambda_{12})}.$$

In Table 1, we present some numerical evidences. It is seen that $\tau(X_{1:2}, X_{2:2})$ is increasing in λ_{12}, decreasing in λ, which agrees with Eq. (5.3).

$\tau(X_{1:2}, X_{2:2})$					
$\lambda \diagdown \lambda_{12}$	0.5	1	1.5	2	2.5
0.5	0.733	0.859	0.911	0.939	0.956
1	0.6	0.733	0.809	0.857	0.889
1.5	0.532	0.654	0.733	0.788	0.828
2	0.492	0.6	0.677	0.733	0.766
2.5	0.465	0.561	0.634	0.690	0.733

Table 1: Kendall's tau of $(X_{1:2}, X_{2:2})$.

5.2.2 Bivariate Pareto distribution

Assume that (X_1, X_2) is an exchangeable random vector with joint Pareto survival function

$$\bar{F}(x,y) = P(X_1 > x, X_2 > y) = (1 + ax + ay)^{-c},$$

where $x, y \geq 0$ and $a > 0$ and $c > 0$ are called scale parameter and shape parameter, respectively. It should be noted that (X_1, X_2) is also positively quadrant dependent.

The following result reveals that the dependence of $(X_{2:2}, X_{1:2})$ only relies on the shape parameter c.

Theorem 5.8 Assume that (X_1, X_2) is an exchangeable Pareto random vector with shape parameter c, and (X_1^*, X_2^*) be other exchangeable Pareto random vector with shape parameter c^*. Then

$$c^* \leq c \Longrightarrow (X_{2:2}^* | X_{1:2}^*) \prec_{\rm SI} (X_{2:2} | X_{1:2}).$$

Proof: Note that the conditional density function of $X_{2:2}$ given $X_{1:2}$ is

$$f(y|x) = \frac{a(1+c)(1+2ax)^{c+1}}{(1+ax+ay)^{c+2}}.$$

Hence,

$$\begin{aligned}
H_{2[s]} &= P(X_{2:2} \leq x | X_{1:2} = s) \\
&= \int_s^x f(u|s) du \\
&= (1+2as)^{c+1}\left[(1+2as)^{-c-1} - (1+as+ax)^{-c-1}\right] \\
&= 1-(1+2as)^{c+1}(1+as+ax)^{-c-1}.
\end{aligned}$$

Since

$$P(X_{1:2} \leq x) = 1-(1+2ax)^{-c},$$

it holds that

$$\begin{aligned}
H_{2[\xi_p]}(x) &= P(X_{2:2} \leq x | X_{1:2} = F_{1:2}^{-1}(p)) \\
&= 1-(1-p)^{-1-1/c}\left[\frac{1+(1-p)^{-1/c}}{2} + ax\right]^{-c-1}.
\end{aligned}$$

Hence, we have for $0 < p \leq q < 1$,

$$H_{2[\xi_q]} \circ H_{2[\xi_p]}^{-1}(u) \quad = \quad 1-(1-q)^{-1-1/c}\left[\frac{(1-q)^{-\frac{1}{c}} - (1-p)^{-\frac{1}{c}}}{2} + (1-u)^{\frac{-1}{c+1}}(1-p)^{-\frac{1}{c}}\right]^{-(c+1)} \tag{5.4}$$

which can be verified to be increasing in $c > 0$.

So, the required result follows. ■

The following result discusses order statistics from bivariate Pareto distributions.

Theorem 5.9 Assume that (X_1, X_2) is an exchangeable Pareto random vector with shape parameter c, and X_1^* and X_2^* be independent and identically distributed random variables with arbitrary distributions. Then

$$(X_{2:2}^*|X_{1:2}^*) \prec_{\text{SI}} (X_{2:2}|X_{1:2}).$$

Proof: Note that, for $0 < p \leq q < 1$,

$$H_{1[\xi_q^*]}^* \circ {H^*}_{1[\xi_p^*]}^{-1}(u) = 1-(1-u) \times \left(\frac{1-p}{1-q}\right)^{1/2}.$$

Now, from Eq. (5.4), we have

$$\begin{aligned}
H_{2[\xi_q]} \circ H_{2[\xi_p]}^{-1}(u) &\leq 1-(1-u) \times \left(\frac{1-p}{1-q}\right)^{1+1/c} \\
&\leq 1-(1-u) \times \left(\frac{1-p}{1-q}\right)^{1/2} \\
&= H_{2[\xi_q^*]}^* \circ {H^*}_{2[\xi_p^*]}^{-1}(u).
\end{aligned}$$

6 Concluding Remarks

It will be of interest to know whether Theorem 5.5 can be extended to other order statistics, that is, for $2 \leq j \leq n-1$,

$$(X_{j:n}|X_{1:n}) \preceq_{\text{SI}} (Y_{j:n}|Y_{1:n}).$$

It is also true for $j = 2$. The proof follows on the same lines by noting that

$$X_{2:n} - X_{1:n} \leq_{\text{DISP}} Y_{2:n} - Y_{1:n}$$

as proved in Kochar and Korwar (2005) under the given conditions on the parameters of the exponential distributions. To prove our conjecture one needs to prove that

$$X_{j:n} - X_{1:n} \leq_{\text{DISP}} Y_{j:n} - Y_{1:n}$$

whose proof is still elusive for $3 \leq j \leq n-1$.

It is also worth noting that *Dolati, et al.* (2008) got a nice bound for Kendall's tau of $(X_{n:n}, X_{1:n})$ by using Theorem 5.4,

$$\tau(X_{n:n}, X_{1:n}) \leq \frac{1}{2n-1}.$$

Khaledi and Kochar (2005) has extended this work to **Generalized Order Statistics** (which contain order statistics and record values as special cases besides many other models of ordered random variables). It will also be of interest to extend the bivariate exchangeable cases in Section 5 to the general exchangeable cases.

Appendix

The proof of Theorem 5.1 depends 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 1 A random variable X with distribution function F is said to be less dispersed than another variable Y with distribution G, written as $X \prec_{\text{DISP}} Y$ or $F \prec_{\text{DISP}} G$, if and only if

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

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

For general information about the dispersive ordering and its properties, refer to Section 3.B of *Shaked and Shanthikumar* (2007). A related earlier paper on this topic is by *Deshpandé and Kochar* (1983).

The proof of Theorem 5.1 will also makes 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.

Lemma 2 (*Avérous, et al. 2005*) Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics associated with a random sample of size n from an exponential distribution, and for $0 \leq i < j \leq n$, 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 $j - i \leq j' - i'$ and $n' - j' \leq n - j$, one has

$$D_{ij}^{(n)} \prec_{\text{DISP}} D_{i'j'}^{(n')},$$

and

$$i' \leq i \ \text{ and } \ n - i \leq n' - i' \Longrightarrow X_{i':n'} \prec_{\text{DISP}} X_{i:n}, \tag{.1}$$

There is an intimate connections between the concepts of dispersive ordering and more SI ordering as shown in the next Lemma.

Lemma 3 (*Khaledi and Kochar, 2005*) Let X_i and Y_i be independent random variables with distribution functions F_i and G_i, respectively for $i = 1, 2$. Then

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

$$(X_2 + Y_2)|X_2 \prec_{\text{SI}} (X_1 + Y_1)|X_1$$

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 4 (*Avérous, et al. 2005*) Let $X_{1:n} \leq \cdots \leq X_{n:n}$ 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 $1 \leq i < j \leq n$.

Proof of Theorem 5.1

In view of Lemma 4, it may be assumed without loss of generality that the parent distribution of the X_i's is exponential. Now under this assumption, the consecutive spacings are mutually independent. We can write

$$X_{j:n} = X_{i:n} + D_{ij}^{(n)} \tag{.2}$$

and

$$X_{j':n'} = X_{i':n'} + D_{i'j'}^{(n')}, \tag{.3}$$

Note that $X_{i:n}$ and $D_{i,j:n}$ are independent as well as $X_{i':n'}$ and $D_{i'j'}^{(n')}$ are independent. It follows from Lemma 2 that,

$$i' \leq i \quad \text{and} \quad n - i \leq n' - i' \Longrightarrow X_{i':n'} \prec_{\text{DISP}} X_{i:n}, \tag{.4}$$

and

$$D_{ij}^{(n)} \prec_{\text{DISP}} D_{i'j'}^{(n')}$$

by Lamma 3, it follows that

$$(X_{j':n'}|X_{i':n'}) \preceq_{\text{SI}} (X_{j:n}|X_{i:n}),$$

thus proving Theorem 5.1.

References

[1] Avérous, J., Genest, C. and Kochar, S. C. (2005). On dependence structure of order statistics. *Journal of Multivariate Analysis* **94**, 159-171.

[2] Balakrishnan, N. and Rao, C. R. (1998a) *Handbook of Statistics 16-Order Statistics: Theory and Methods.* Elsevier, New York.

[3] Balakrishnan, N. and Rao, C. R. (1998b) *Handbook of Statistics 17-Order Statistics: Applications.* Elsevier, New York.

[4] Bapat, R. B. and Beg, M. I (1989). Order statistics for nonidentically distributed variables and permanents. *Sankhya Ser. A* **51**, 79-93.

[5] Barlow, R. E. and Proschan, F. (1981) *Statistical Theory of Reliability and Life Testing*. To Begin with, Silver Spring, Maryland.

[6] Bickel, P.J.. *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.

[7] Boland,P.J ., Hollander, M., Joag-Dev, K. and Kochar, S. C. (1996). Bivariate dependence properties of order statistics. *Journal of Multivariate Analysis* **56**, 75-89.

[8] Chen,Y. P. (2007). A note on the relationship between Spearman's ρ and Kendall's τ for extreme order statistics. *Journal of Statistical Planning and Inference* **137**, 2165-2171.

[9] David, H. A. and Nagaraja, H.N. (2003) *Order statistics*. 3rd ed. Wiley, New York.

[10] Deshpandé, J.V. and Kochar, S.C. Dispersive ordering is the same as tail ordering. *Adv. Appl. Prob.* **15** (1983), 686-87.

[11] Dolati, A., Genest, C. and Kochar, S. C. (2008). On the dependence between the extreme order statistics in the proportional hazards model. *Journal of Multivariate Analysis* **99**, 777-786.

[12] Dubhashi, D. and Häggström, O. (2008). A note on conditioning and stochastic domination for order statistics. *Journal of Applied Probability* **45**, 575-579.

[13] Fredricks, G. A. and Nelsen, R. B. (2007). On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference* **137**, 2143-2150.

[14] Genest, C., Kochar, S. and Xu, M. (2009) On the range of heterogeneous samples. *Journal of Multivariate Analysis* **100**, 1587-1592.

[15] Hu, T. and Chen, H. (2008). Dependence properties of order statistics. *Journal of Statistical Planning and Inference* **138**, 2214-2222.

[16] Hu, T. and Xie, C. (2006). Negative dependence in the balls and bins experiment with applications to order statistics. *Journal of Multivariate Analysis* **97**, 1342-1354.

[17] Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman & Hall.

[18] Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis* **10**, 467-498.

[19] Kochar, S. C. and Khaledi, B. (2005). Dependence orderings for generalized order statistics. *Statistics and Probability Letters* **73**, 357-367.

[20] Kim, S. H. and David, H. A. (1990). On the dependence structure of order statistics and concomitants of order statistics. *Journal of Statistical Planning and Inference* **24**, 363-368.

[21] Kochar, S. C. and Korwar, R. (1996). Stochastic orders for spacings of heterogeneous exponential random variables. *Journal of Multivariate Analysis* **57**, 69-83.

[22] Li, X. and Li, Z. (2007). Proof of a conjecture on Spearman's rho and Kendall's tau for sample minimum and maximum. *Journal of Statistical Planning and Inference* **137**, 359-361.

[23] Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks.* John Wiley & Sons, New York.

[24] Navarro, J. and Balakrishnan, N. (2010). Study of some measures of dependence between order statistics and systems. *Journal of Multivariate Analysis* **101**, 52-67.

[25] Nelsen, R.B. (1992). On measures of association as measures of positive dependence. *Statistics and Probability Letters* **14**, 269-274.

[26] Nelsen, R. B. (2006). *An Introduction to Copulas.* Lecture Notes in Statistics. 2nd ediiton. Springer, New York.

[27] Sathe, Y. S. (1988). On the correlation coefficient between the first and the r-th smallest order statistics based on n independent exponential random variables. *Comm. Statist. Theory Meth.* **17** 3295–3299.

[28] Scarsini, M. (1984). On measures of concordance. *Stochastica* **8**, 201-218.

[29] Schmitz, V. (2004). Revealing the dependence structure between $X_{(1)}$ and $X_{(n)}$. *Journal of Statistical Planning and Inference* **123**, 41-47.

[30] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders and Their Applications.* Springer, New York.

[31] Tukey, J. W. (1958). A problem of Berkson, and minimum variance orderly estimators, *Ann. Math. Statist.* **29**, 588-592.

[32] Zhuang, W., Yao, C. and Hu, et al. (2010). Conditional ordering of order statistics. *Journal of Multivariate Analysis* **101**, 640-644.