CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Dependence comparisons of order statistics in the proportional hazards model

Subhash Kochar [ORCID]

Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA.
E-mail: kochar@pdx.edu.

Abstract

Let $X_1, \ldots, X_n$ be mutually independent exponential random variables with distinct hazard rates $\lambda_1, \ldots, \lambda_n$ and let $Y_1, \ldots, Y_n$ be a random sample from the exponential distribution with hazard rate $\bar{\lambda} = \sum_{i=1}^{n} \lambda_i / n$. Also let $X_{1:n} < \cdots < X_{n:n}$ and $Y_{1:n} < \cdots < Y_{n:n}$ be their associated order statistics. It is proved that for $1 \leq i < j \leq n$, the generalized spacing $X_{j:n} - X_{i:n}$ is more dispersed than $Y_{j:n} - Y_{i:n}$ according to dispersive ordering and for $2 \leq i \leq n$, the dependence of $X_{i:n}$ on $X_{1:n}$ is less than that of $Y_{i:n}$ on $Y_{1:n}$, in the sense of the more stochastically increasing ordering. This dependence result is also extended to the proportional hazard rates (PHR) model. This extends the earlier work of Genest *et al.* [(2009)]. On the range of heterogeneous samples. *Journal of Multivariate Analysis* 100: 1587–1592] who proved this result for $i = n$.

1. Introduction

Several notions of monotone dependence have been introduced and studied in the literature. Researchers have also developed the corresponding dependence (partial) orderings which compare the degree of (monotone) dependence within the components of different random vectors of the same length. For details, see the pioneering paper of Lehmann [16] and Chapter 5 of Barlow and Proschan [2] for different notions of positive dependence, and that of Kimeldorf and Sampson [12] for a unified treatment of families, orderings and measures of monotone dependence. For more detailed discussion of these concepts, see Chapter 2 of Joe [9] and Chapter 5 of Nelsen [17].

Let $\{X_1, \ldots, X_n\}$ be a set of continuous random variables. Many authors have investigated the nature of the dependence that may exist between two order statistics $X_{i:n}$ and $X_{j:n}$ for $1 \leq i < j \leq n$ under different distributional scenarios. It seems natural to expect some degree of positive dependence between them. It is well known that the order statistics based on independent (but not necessarily identically distributed) random variables are associated (cf., [2, p. 32]). This yields many useful product inequalities for order statistics of independent random variables, and in particular, $\text{Cov}(X_{i:n}, X_{j:n}) \geq 0$ for all i and j, which was initially proved by Bickel [3] when the X's are independent and identically distributed (i.i.d.).

Boland *et al.* [4] studied the problem of dependence among order statistics in detail and discussed different types of dependence that hold between them. It is shown in that paper that in the case of i.i.d. observations, any pair of order statistics is TP_2 dependent (also called monotone likelihood ratio dependent) which is the strongest type of dependence in the hierarchy of various dependence criteria as described in [2]). But this is not the case in the non-i.i.d. case. However, as is shown in Boland *et al.* [4], in general, whereas $X_{j:n}$ may not be stochastically increasing in $X_{i:n}$ for $i < j$, $X_{j:n}$ is always right tail increasing (RTI) in $X_{i:n}$. The reader is referred to Chapter 5 of Barlow and Proschan [2] for basic definitions and relations among various types of dependence.

It is also of interest to compare the strength of dependence that may exist between two pairs of order statistics. When the parent distribution from which the random sample is drawn has an increasing hazard rate and a decreasing reverse hazard rate, Tukey [22] showed that

$$\text{Cov}(X_{i':n}, X_{j':n}) \leq \text{Cov}(X_{i:n}, X_{j:n}) \tag{1.1}$$

for either $i = i'$ and $j \leq j'$; or $j = j'$ and $i' \leq i$. Kim and David [11] proved that if both the hazard and the reverse hazard rates of the X_i's are increasing, then inequality (1.1) remains valid when $i = i'$ and $j \leq j'$; However, the inequality (1.1) is reversed when $j = j'$ and $i' \leq i$.

Let $X_1, \ldots, X_n$ be mutually independent exponentials with distinct hazard rates $\lambda_1, \ldots, \lambda_n$ and let $Y_1, \ldots, Y_n$ form a random sample from the exponential distribution with hazard rate $\bar{\lambda} = (\sum_{i=1}^{n} X_i)/n$. Sathe [18] proved that for any $i \in \{2, \ldots, n\}$

$$\text{corr}(X_{1:n}, X_{i:n}) \leq \text{corr}(Y_{1:n}, Y_{i:n}). \tag{1.2}$$

Although this observation is interesting, it merely compares the relative degree of *linear* association within the two pairs. It is now widely recognized, however, that margin-free measures of association are more appropriate than Pearson's correlation, because they are based on the unique underlying copula which governs the dependence between the components of a continuous random pair.

In this paper, a long-standing open problem that for $1 \leq i < j \leq n$, the generalized spacing $X_{j:n} - X_{i:n}$ is more dispersed than $Y_{j:n} - Y_{i:n}$ according to dispersive ordering, has been solved (cf. [23]). This result is used to solve a related open problem that for $2 \leq i \leq n$, the dependence of $X_{i:n}$ on $X_{1:n}$ is less than that of $Y_{i:n}$ on $Y_{1:n}$, in the sense of the more stochastically increasing ordering. This dependence result is also extended to the proportional hazard rates (PHR) model. This extends the earlier work of Genest *et al.* [8] who proved this result for $i = n$. Section 2 is on some preliminaries where various definitions and notations are given. The main results of this paper are given in Section 3.

2. Preliminaries

For $i = 1, 2$, let (X_i, Y_i) be a pair of continuous random variables with joint cumulative distribution function H_i and margins F_i, G_i. Let

$$C_i(u, v) = H_i\{F_i^{-1}(u), G_i^{-1}(v)\}, \quad u, v \in (0, 1)$$

be the unique copula associated with H_i. In other words, C_i is the distribution of the pair $(U_i, V_i) \equiv (F_i(X_i), G_i(Y_i))$ whose margins are uniform on the interval $(0, 1)$. See, for example, Chapter 1 of Nelsen [17] for details.

The most well-understood partial order to compare the strength of dependence within two random vectors is that of *positive quadrant dependence* (PQD) as defined below.

Definition 2.1. *A copula C_1 is said to be less dependent than copula C_2 in the positive quadrant dependence (PQD) ordering, denoted $(X_1, Y_1) \prec_{\text{PQD}} (X_2, Y_2)$, if and only if*

$$C_1(u, v) \leq C_2(u, v), \quad u, v \in (0, 1).$$

This condition implies that $\kappa(S_1, T_1) \leq \kappa(S_2, T_2)$ for all concordance measures meeting the axioms of Scarsini [19] like Kendall's τ and Spearman's ρ. See Tchen [21] for details.

Lehmann [16] in his seminal work introduced the notion of *monotone regression dependence (MRD)* which is also known as *stochastic increasingness* (SI) in the literature.

Definition 2.2. *Let (X, Y) be a bivariate random vector with joint distribution function H. The variable Y is said to be stochastically increasing (SI) in X if for all $(x_1, x_2) \in \mathbf{R}^2$,*

$$x_1 < x_2 \Rightarrow P(Y \leq y \mid X = x_2) \leq P(Y \leq y \mid X = x_1), \quad \textit{for all } y \in \mathbf{R}. \tag{2.1}$$

If we denote by H_x the distribution function of the conditional distribution of Y given $X = x$, then (2.1) can be rewritten as

$$x_1 < x_2 \Rightarrow H_{x_2} \circ H_{x_1}^{-1}(u) \leq u, \quad \text{for } 0 \leq u \leq 1. \tag{2.2}$$

Note that in case X and Y are independent, $H_{x_2} \circ H_{x_1}^{-1}(u) = u$, for $0 \leq u \leq 1$ and for all (x_1, x_2). The SI property is not symmetric in X and Y. It is a very strong notion of positive dependence and many of the other notions of positive dependence follow from it.

Denoting by $\xi_p = F^{-1}(p)$, the pth quantile of the marginal distribution of X, we see that (2.2) will hold if and only if for all $0 \leq u \leq 1$,

$$0 \leq p < q \leq 1 \Rightarrow H_{\xi_q} \circ H_{\xi_p}^{-1}(u) \leq u.$$

In his book, Joe [9] mentions a number of bivariate stochastic ordering relations. One such notion is that of greater monotone regression (or more SI) dependence, originally considered by Yanagimoto and Okamoto [24] and later extended and further investigated by Schriever [20], Capéraà and Genest [5], and Avérous *et al.* [1], among others.

As discussed in the books by Joe [9] and Nelsen [17], a reasonable way to compare the relative degree of dependence between two random vectors is through their copulas. We will discuss here the concept of more SI, a partial order to compare the strength of dependence that may exist between two bivariate random vectors in the sense of monotone regression dependence (stochastic increasingness).

Suppose we have two pairs of continuous random variables (X_1, Y_1) and (X_2, Y_2) with joint cumulative distribution functions H_i and marginals F_i and G_i for $i = 1, 2$.

Definition 2.3. *Y_2 is said to be more stochastically increasing in X_2 than Y_1 is in X_1, denoted by $(Y_1 \mid X_1) \prec_{\mathrm{SI}} (Y_2 \mid X_2)$ or $H_1 \prec_{\mathrm{SI}} H_2$, if*

$$0 < p \leq q < 1 \Longrightarrow H_{2, \xi_{2q}} \circ H_{2, \xi_{2p}}^{-1}(u) \leq H_{1, \xi_{1q}} \circ H_{1, \xi_{1p}}^{-1}(u), \tag{2.3}$$

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

Obviously, (2.3) implies that Y_2 is stochastically increasing in X_2 if X_1 and Y_1 are independent. It also implies that if Y_1 is stochastically increasing in X_1, then so is Y_2 in X_2; and conversely, if Y_2 is stochastically decreasing in X_2, then so is Y_1 in X_1. The above definition of more SI is copula based and

$$H_1 \prec_{\mathrm{SI}} H_2 \Rightarrow H_1 \prec_{\mathrm{PQD}} H_2.$$

Avérous *et al.* [1] have shown in their paper that in the case of i.i.d. observations, the copula of any pair of order statistics is independent of the distribution of the parent observations as long as they are continuous. In a sense, their copula has a distribution-free property. But this is not the case if the observations are not i.i.d.

The natural question is to see if we can extend the result given in (1.2) to a copula-based positive dependence ordering. Genest *et al.* [8] proved that under the given conditions,

$$(X_{n:n} \mid X_{1:n}) \prec_{\mathrm{SI}} (Y_{n:n} \mid Y_{1:n}).$$

It has been an open problem to see whether this result can be generalized from the the largest order statistic to other order statistics. We prove in this paper that for $2 \leq i \leq n$,

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

This implies in particular that

$$\kappa(X_{i:n}, X_{1:n}) \leq \kappa(Y_{i:n}, Y_{1:n}),$$

where $\kappa(S,T)$ represents any concordance measure between random variables S and T in the sense of Scarsini [19], for example, Spearman's rho or Kendall's tau. A related work to this problem is Dolati *et al.* [7].

3. Main results

The proof of our main result 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 3.1. *A random variable X with distribution function F is said to be less dispersed than another variable Y with distribution G, written as $X \leq_{\text{disp}} Y$ or $F \leq_{\text{disp}} G$, if and only if*

$$F^{-1}(v) - F^{-1}(u) \leq G^{-1}(u) - G^{-1}(v)$$

for all $0 < u \leq v < 1$.

The dispersive order is closely related to the star-ordering which is a partial order to compare the relative aging or skewness and is defined as below.

Definition 3.2. *A random variable X with distribution function F is said to be star ordered with respect to another random variable Y with distribution G, written as $X \leq_* Y$ or $F \leq_* G$, if and only if*

$$\frac{G^{-1} \circ F(x)}{x} \quad \text{is increasing in } x.$$

For nonnegative random variables, the star order is related to the dispersive order by

$$X \leq_* Y \Leftrightarrow \log X \leq_{\text{disp}} \log Y.$$

The proof of the next lemma, which is a refined version of a result by Deshpande and Kochar [6], on relation between the star order and the dispersive order, can be found in Kochar and Xu [15].

Lemma 3.1. *Let X and Y be two random variables. If $X \leq_* Y$ and $X \leq_{\text{st}} Y$, then $X \leq_{\text{disp}} Y$.*

The following result of Khaledi and Kochar [10] will be used to prove our main theorem.

Lemma 3.2 ([10]). *Let X_i and Y_i be independent random variables with distribution functions F_i and G_i, respectively for $i = 1, 2$. Then*

$$X_2 \leq_{\text{disp}} X_1 \text{ and } Y_1 \leq_{\text{disp}} Y_2 \Rightarrow (X_2 + Y_2)|X_2 \prec_{\text{SI}} (X_1 + Y_1)|X_1.$$

The next theorem on dispersive ordering between general spacings which is also of independent interest, will be used to prove our main result.

Theorem 3.1. *Let* $X_1, \ldots, X_n$ *be mutually independent exponential random variables with distinct hazard rates* $\lambda_1, \ldots, \lambda_n$ *and let* $Y_1, \ldots, Y_n$ *be a random sample from the exponential distribution with hazard rate* $\bar{\lambda}$. *Then for* $1 \leq i < j \leq n$,

$$(Y_{j:n} - Y_{i:n}) \leq_{\text{disp}} (X_{j:n} - X_{i:n}). \tag{3.1}$$

Proof. Yu [25] proved in Corollary 1 of his paper that for $1 \leq i < j \leq n$,

$$(Y_{j:n} - Y_{i:n}) \leq_* (X_{j:n} - X_{i:n}). \tag{3.2}$$

and Kochar and Rojo [14] proved in their Corollary 2.1 that

$$(Y_{j:n} - Y_{i:n}) \leq_{\text{st}} (X_{j:n} - X_{i:n}). \tag{3.3}$$

Using (3.2) and (3.3), we get (3.1) using Lemma 3.1. □

Xu and Balakrishnan [23] proved a special case of the above result when $i = n$.

Now, we give the main result of this paper.

Theorem 3.2. *Let* $X_1, \ldots, X_n$ *be mutually independent exponentials with distinct hazard rates* $\lambda_1, \ldots, \lambda_n$ *and let* $Y_1, \ldots, Y_n$ *form a random sample from the exponential distribution with hazard rate* $\bar{\lambda} = (\sum_{i=1}^{n} \lambda_i)/n$. *Then, for* $i \in \{2, \ldots, n\}$,

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

Proof. It follows from Theorem 3.1 above that for $2 \leq i \leq n$,

$$(Y_{i:n} - Y_{1:n}) \leq_{\text{disp}} (X_{i:n} - X_{1:n}). \tag{3.4}$$

Kochar and Korwar [13] proved that $(X_{i:n} - X_{1:n})$ is independent of $X_{1:n}$ and $X_{1:n} \stackrel{st}{=} Y_{1:n}$. Similarly, $(Y_{i:n} - Y_{1:n})$ is independent of $Y_{1:n}$. Moreover,

$$X_{1:n} \stackrel{\text{st}}{=} Y_{1:n}.$$

Now, we can express $X_{i:n}$ and $Y_{i:n}$ as

$$X_{i:n} = (X_{i:n} - X_{1:n}) + X_{1:n} \text{ and } Y_{i:n} = (Y_{i:n} - Y_{1:n}) + Y_{1:n}.$$

Using (3.4) and the above two equations, it follows from Lemma 3.2 that

$$(X_{i:n} \mid X_{1:n}) \prec_{\text{SI}} (Y_{i:n} \mid Y_{1:n}), \quad \text{for } 2 \leq i \leq n.$$

This proves the required result. □

Remark 3.1. *Since the copula of the order statistics of a random sample has the distribution-free property, it is not required that the common hazard rate of Y's is necessarily* $\bar{\lambda}$ *in the above theorems. In fact, the Y's could be a random sample from any continuous distribution.*

The above theorem can be extended to the PHR model using the technique used in Genest *et al.* [8].

Theorem 3.3. *Let* $X_1, \ldots, X_n$ *be independent continuous random variables following the PHR model with* $\lambda_1, \ldots, \lambda_n$ *as the proportionality parameters. Let* $Y_1, \ldots, Y_n$ *be i.i.d. continuous random variables with common survival function* $\bar{F}^{\bar{\lambda}}$, *then*

$$(X_{i:n} \mid X_{1:n}) \prec_{\text{SI}} (Y_{i:n} \mid Y_{1:n}) \quad \text{for } 2 \leq i \leq n.$$

Proof. Let $R(t) = -\log\{\bar{F}(t)\}$ in the PHR model. Make the monotone transformations $X_i^* = R(X_i)$ and $Y_i^* = -\log\{\bar{F}(Y_i)\}$. Then, the transformed variable X_i^* has exponential distribution with hazard rate λ_i, $i = 1, \ldots, n$ and $Y_1^*, \ldots, Y_i^*$ is a random sample from an exponential distribution with parameter $\bar{\lambda}$. Let $X_{(1)}^* < \cdots < X_{(n)}^*$ and $Y_{(1)}^* < \cdots < Y_{(n)}^*$ be the order statistics corresponding to the new sets of variables.

In view of their invariance by monotone increasing transformations of the margins, the copulas associated with the pairs $(X_{1:n}, X_{n:n})$ and $(X_{1:n}^*, X_{n:n}^*)$ coincide. Similarly, the pairs $(Y_{1:n}, Y_{n:n})$ and $(Y_{1:n}^*, Y_{i:n}^*)$ have the same copula.

Also, since the more SI dependence ordering is copula-based,

$$(X_{n:n} \mid X_{1:n}) \prec_{\text{SI}} (Y_{n:n} \mid Y_{1:n}) \Leftrightarrow (X_{n:n}^* \mid X_{1:n}^*) \prec_{\text{SI}} (Y_{n:n}^* \mid Y_{1:n}^*).$$

$\square$

Under the conditions of Theorem 3.3, an upper bound on $\kappa(X_{1:n}, X_{i:n})$ is given by $\kappa(Y_{1:n}, Y_{i:n})$. Avérous *et al.* [1] obtained an analytic expression for computing the exact values of the Kendall's τ for any pair of order statistics from a random sample from a continuous distribution which in our case reduces to

$$\tau(Y_{1:n}, Y_{i:n}) = 1 - \frac{2(n-1)}{2n-1} \binom{n-2}{i-2} \sum_{s=0}^{n-i} \binom{n}{s} \bigg/ \binom{2n-2}{n-i+s}.$$

References

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

[2] Barlow, R.E. & Proschan, F. (1981). *Statistical theory of reliability and life testing*. Silver Spring, MD: To Begin with.

[3] Bickel, P.J. (1967). Some contributions to the theory of order statistics. In L.M. LeCam & J. Neyman (eds), *Fifth Berkeley symposium on mathematics and statistics*, vol. 1. Berkeley, CA: University of California Press, pp. 575–591.

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

[5] Capéraà, P. & Genest, C. (1990). Concepts de dépendance et ordres stochastiques pour des lois bidimensionnelles. *Canadian Journal of Statistics* 18: 315–326.

[6] Deshpandé, J.V. & Kochar, S.C. (1983). Dispersive ordering is the same as tail ordering. *Advances in Applied Probability* 15: 686–687.

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

[8] Genest, C., Kochar, S., & Xu, M. (2009). On the range of heterogeneous samples. *Journal of Multivariate Analysis* 100: 1587–1592.

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

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

[11] Kim, S.H. & 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.

[12] Kimeldorf, G. & Sampson, A.R. (1989). A framework for positive dependence. *Annals of Institute of Statistical Mathematics* 41: 31–45.

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

[14] Kochar, S.C. & Rojo, J. (1996). Some new results on stochastic comparisons of spacings from heterogeneous exponential distributions. *Journal of Multivariate Analysis* 59: 272–281.

[15] Kochar, S.C. & Xu, M. (2012). Some unified results on comparing linear combinations of independent gamma random variables. *Probability in the Engineering and Informational Sciences* 26: 393–404.

[16] Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics* 37: 1137–1153.

[17] Nelsen, R.B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics No 139. New York: Springer.

[18] Sathe, Y.S. (1988). On the correlation coefficient between the first and the rth smallest order statistics based on n independent exponential random variables. *Communications in Statistics Theory and Methods* 17: 3295–3299.

[19] Scarsini, M. (1984). On measures of concordance. *Stochastica* 8: 201–218.

[20] Schriever, B.F. (1987). An ordering for positive dependence. *Annals of Statistics* 15: 1208–1214.

[21] Tchen, A.H. (1980). Inequalities for distributions with given marginals. *Annals of Probability* 8: 814–827.

[22] Tukey, J.W. (1958). A problem of Berkson, and minimum variance orderly estimators,. *Annals of Mathematical Statistics* 29: 588–592.

[23] Xu, M. & Balakrishnan, N. (2012). On the sample ranges from heterogeneous exponential variables. *Journal of Multivariate Analysis* 109: 1–9.

[24] Yanagimoto, T. & Okamoto, M. (1969). Partial orderings of permutations and monotonicity of a rank correlation statistic. *Annals of Institute of Statistical Mathematics* 21: 489–506.

[25] Yu, Y. (2021). On stochastic comparisons of order statistics from heterogeneous exponential samples. *Probability in Engineering and Information Sciences*. 35: 532–537.