

## A NOTE ON CONDITIONING AND STOCHASTIC DOMINATION FOR ORDER STATISTICS

DEVDATT DUBHASHI\* AND

OLLE HÄGGSTRÖM,\*\* *Chalmers University of Technology*

### Abstract

For an order statistic  $(X_{1:n}, \dots, X_{n:n})$  of a collection of independent but not necessarily identically distributed random variables, and any  $i \in \{1, \dots, n\}$ , the conditional distribution of  $(X_{i+1:n}, \dots, X_{n:n})$  given  $X_{i:n} > s$  is shown to be stochastically increasing in  $s$ . This answers a question by Hu and Xie (2006).

*Keywords:* Order statistics; coupling; stochastic domination

2000 Mathematics Subject Classification: Primary 62G30; 60E15

### 1. Introduction

Let  $X_1, \dots, X_n$  be independent but not necessarily identically distributed random variables, and let

$$(X_{1:n}, \dots, X_{n:n})$$

be their order statistic. In other words, if  $X_1, \dots, X_n$  is given then  $(X_{1:n}, \dots, X_{n:n})$  is the unique rearrangement of their values such that

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}.$$

It is reasonable to expect the order statistic to exhibit various positive dependence properties. For instance, it turns out that they satisfy what is known in the probability literature as positive associations and as the FKG (Fortuin–Kasteleyn–Ginibre) inequality; a collection  $(W_1, \dots, W_n)$  of random variables is said to have this property if, whenever  $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$  are two bounded and increasing (in the coordinatewise partial order) functions, the correlation inequality

$$E[f(W_1, \dots, W_n)g(W_1, \dots, W_n)] \geq E[f(W_1, \dots, W_n)]E[g(W_1, \dots, W_n)] \quad (1)$$

holds (here and henceforth, ‘increasing’ is taken to mean ‘nondecreasing’). The well-known Harris inequality (see [4] or, for instance, [3]) states that any collection of independent random variables is positively associated. Positive associations for  $(X_{1:n}, \dots, X_{n:n})$  follow from this in conjunction with the observations that each  $X_{i:n}$  is an increasing function of  $(X_1, \dots, X_n)$  and that increasing functions of increasing functions are again increasing.

Received 9 October 2007; revision received 4 February 2008.

\* Postal address: Department of Computing Science, Chalmers University of Technology, S-412 96 Göteborg, Sweden.

\*\* Postal address: Department of Mathematical Statistics, Chalmers University of Technology, S-412 96 Göteborg, Sweden. Email address: olleh@math.chalmers.se

Supported by the Swedish Research Council and by the Göran Gustafsson Foundation for Research in the Natural Sciences and Medicine.

We will be interested in further such properties of the order statistic. To this end, recall the usual notion of stochastic ordering between  $n$ -dimensional random vectors: for two such vectors,  $\mathbf{Y} = (Y_1, \dots, Y_n)$  and  $\mathbf{Z} = (Z_1, \dots, Z_n)$ , we say that  $\mathbf{Y}$  is stochastically dominated by  $\mathbf{Z}$ , writing  $\mathbf{Y} \leq_{\text{st}} \mathbf{Z}$ , if

$$E[f(Y_1, \dots, Y_n)] \leq E[f(Z_1, \dots, Z_n)]$$

whenever  $f$  is bounded and increasing. By Strassen’s theorem (see [6] or [8]),  $\mathbf{Y} \leq_{\text{st}} \mathbf{Z}$  is equivalent to the existence of a coupling of  $\mathbf{Y}$  and  $\mathbf{Z}$  such that  $P(Y_i \leq Z_i \text{ for each } i) = 1$ .

Conditioning on  $X_{i:n}$  being large is good news for the other components of the order statistic, in the sense that

$$(X_{1:n}, \dots, X_{n:n}) \leq_{\text{st}} [(X_{1:n}, \dots, X_{n:n}) \mid X_{i:n} > s];$$

this follows from (1) applied with  $(W_1, \dots, W_n) = (X_{1:n}, \dots, X_{n:n})$  and  $g$  equal to the indicator  $\mathbf{1}_{\{X_{i:n} > s\}}$ . But is it better news the larger  $s$  is? In other words, we may wonder whether

$$[(X_{1:n}, \dots, X_{n:n}) \mid X_{i:n} > s] \leq_{\text{st}} [(X_{1:n}, \dots, X_{n:n}) \mid X_{i:n} > s'] \tag{2}$$

whenever  $s < s'$ . This, however, turns out to be too much to ask for, as the following simple example (similar to one given in [1]) shows. Let  $n = 2$ , let  $X_1 = 1$  or  $3$  with probability  $\frac{1}{2}$  each, and let  $X_2 = 2$  or  $4$  with probability  $\frac{1}{2}$  each. A direct calculation shows that

$$P(X_{1:2} > 1 \mid X_{2:2} > 2) = \frac{2}{3} > \frac{1}{2} = P(X_{1:2} > 1 \mid X_{2:2} > 3),$$

disproving (2). Not all is lost, however, and we will prove the following weaker version of (2) obtained by considering only order statistics larger than  $X_{i:n}$ .

**Theorem 1.** *Let  $X_1, \dots, X_n$  be independent random variables, and let  $(X_{1:n}, \dots, X_{n:n})$  be their order statistic. For any  $i \in \{1, \dots, n\}$  and any  $s, s' \in \mathbb{R}$  such that  $s < s'$ , we have*

$$[(X_{i:n}, \dots, X_{n:n}) \mid X_{i:n} > s] \leq_{\text{st}} [(X_{i:n}, \dots, X_{n:n}) \mid X_{i:n} > s'] \tag{3}$$

This answers a question by Hu and Xie [5], who established the weaker result that

$$P(X_{i+1:n} > x_{i+1}, X_{i+2:n} > x_{i+2}, \dots, X_{n:n} > x_n \mid X_{i:n} > s) \tag{4}$$

is increasing as a function of  $s$  for any  $(x_{i+1}, \dots, x_n) \in \mathbb{R}^{n-i}$ .

(To see that the orthant property, (4), is not enough to immediately deduce the stochastic domination property, (3), consider two pairs of  $\{0, 1\}$ -valued random variables  $(W_1, W_2)$  and  $(W'_1, W'_2)$  such that  $(W_1, W_2)$  equals  $(0, 0), (0, 1), (1, 0)$ , or  $(1, 1)$  with probability  $\frac{1}{4}$  each, while  $(W'_1, W'_2)$  equals  $(0, 0)$  or  $(1, 1)$  with probability  $\frac{1}{2}$  each. Then  $P(W_1 < w_1, W_2 < w_2) \leq P(W'_1 < w_1, W'_2 < w_2)$  for all  $w_1, w_2 \in \mathbb{R}$ , while, on the other hand,  $(W_1, W_2) \not\leq_{\text{st}} (W'_1, W'_2)$ , since  $P(W_1 + W_2 \geq 1) = \frac{3}{4} > \frac{1}{2} = P(W'_1 + W'_2 \geq 1)$ .)

We will prove Theorem 1 in Section 2. Before that, let us comment on two easy extensions. First, (3) still holds if we replace the event  $X_{i:n} > s$  by  $X_{i:n} \geq s$  and/or replace the event  $X_{i:n} > s'$  by  $X_{i:n} \geq s'$ ; this follows from trivial changes to our proof. Second, by substituting  $-X_1, \dots, -X_n$  for  $X_1, \dots, X_n$  we see that Theorem 1 implies that (under the same conditions)

$$[(X_{1:n}, \dots, X_{i:n}) \mid X_{i:n} < s] \leq_{\text{st}} [(X_{1:n}, \dots, X_{i:n}) \mid X_{i:n} < s'].$$

### 2. Proof of main result

We will make use of the following known result.

**Lemma 1.** *Let  $X_1, \dots, X_n$  be independent, but not necessarily identically distributed,  $\{0, 1\}$ -valued random variables, and let  $T = \sum_{i=1}^n X_i$  be their sum. We then have, for any  $t, t' \in \{1, \dots, n\}$  such that  $t < t'$ ,*

$$(X_1, \dots, X_n \mid T = t) \preceq_{st} (X_1, \dots, X_n \mid T = t').$$

The seemingly weaker result that, for each fixed  $i$ ,

$$P(X_i = 1 \mid S = t) \leq P(X_i = 1 \mid S = t')$$

is stated in [3, Proposition 1]—a result which, by the way, goes back all the way to Newton [7]—but the proof in [3] of that result contains a coupling of  $(X_1, \dots, X_n \mid T = t)$  and  $(X_1, \dots, X_n \mid T = t')$  from which Lemma 1 can immediately be deduced. Efron [2], who proved a similar result for a certain class of continuous distributions, mentioned an unpublished proof of Lemma 1 due to Proschan and Barlow.

*Proof of Theorem 1.* The core of the proof consists of establishing the desired stochastic domination under the additional assumption that

$$\text{the supports of the distributions of } X_1, \dots, X_n \text{ are finite and disjoint.} \tag{5}$$

Once that is done, we will complete the proof using a couple of standard limiting arguments.

Assume that (5) holds, and write  $\{x_1, \dots, x_l\}$  with  $x_1 \leq \dots \leq x_l$  for the union of the supports of the distributions of the  $X_i$ s. To show (3) is then tantamount to showing that

$$[(X_{i:n}, \dots, X_{n:n}) \mid X_{i:n} > x_{m-1}] \preceq_{st} [(X_{i:n}, \dots, X_{n:n}) \mid X_{i:n} > x_m] \tag{6}$$

for  $m = 2, 3, \dots, l - 1$ . We may assume that  $P(X_{i:n} = x_m) > 0$ , because otherwise (6) holds trivially (with equality). Under that assumption, showing (6) is equivalent to showing that

$$[(X_{i:n}, \dots, X_{n:n}) \mid X_{i:n} = x_m] \preceq_{st} [(X_{i:n}, \dots, X_{n:n}) \mid X_{i:n} > x_m] \tag{7}$$

for each such  $m$ . Fix such an  $m$ . We will show (7) using the coupling method: to establish (7), it is enough to jointly construct two random vectors  $(X_1^*, \dots, X_n^*)$  and  $(X_1^{**}, \dots, X_n^{**})$  whose distributions are those of

$$[(X_1, \dots, X_n) \mid X_{i:n} = x_m]$$

and

$$[(X_1, \dots, X_n) \mid X_{i:n} > x_m],$$

respectively, and such that their order statistics satisfy

$$X_{j:n}^* \leq X_{j:n}^{**} \quad \text{for } j = i, i + 1, \dots, n. \tag{8}$$

The construction of  $(X_1^*, \dots, X_n^*)$  and  $(X_1^{**}, \dots, X_n^{**})$  is carried out in several steps. We know that  $X_{i:n}^* = x_m$ . By the disjointness assumption in (5) we can read off the (unique)  $k$  for which  $X_k^* = X_{i:n}^* = x_m$ . We then reveal the value of  $X_k^{**}$ . Next, we define  $L$  as the number of indices  $j \in \{1, \dots, k - 1, k + 1, \dots, n\}$  such that  $X_j^{**} > x_m$ , and pick  $L$  according to its

correct conditional distribution given  $X_k^{**}$ . If  $X_k^{**} \leq x_m$  then  $L \geq n - i + 1$ , and if  $X_k^{**} > x_m$  then  $L \geq n - i$ . In either case

$$L \geq n - i. \tag{9}$$

At this point, the joint distribution of  $X_1^*, \dots, X_{k-1}^*, X_{k+1}^*, \dots, X_n^*$  is precisely the distribution of  $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n$  conditional on exactly  $n - i$  of them taking a value that exceeds  $x_m$ , while the joint distribution of  $X_1^{**}, \dots, X_{k-1}^{**}, X_{k+1}^{**}, \dots, X_n^{**}$  is precisely the distribution of  $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n$  conditional on exactly  $L$  of them taking a value that exceeds  $x_m$ .

Before obtaining full information about the variables  $X_1^*, \dots, X_{k-1}^*, X_{k+1}^*, \dots, X_n^*$  and  $X_1^{**}, \dots, X_{k-1}^{**}, X_{k+1}^{**}, \dots, X_n^{**}$ , we peek only at the indicator variables

$$\mathbf{1}_{\{X_1^* > x_m\}}, \dots, \mathbf{1}_{\{X_{k-1}^* > x_m\}}, \mathbf{1}_{\{X_{k+1}^* > x_m\}}, \dots, \mathbf{1}_{\{X_n^* > x_m\}} \tag{10}$$

and

$$\mathbf{1}_{\{X_1^{**} > x_m\}}, \dots, \mathbf{1}_{\{X_{k-1}^{**} > x_m\}}, \mathbf{1}_{\{X_{k+1}^{**} > x_m\}}, \dots, \mathbf{1}_{\{X_n^{**} > x_m\}}. \tag{11}$$

Exactly  $n - i$  of the indicators in (10) take value 1, while exactly  $L$  of those in (11) take value 1. Using (9), Lemma 1 allows us to couple them in such a way that  $\mathbf{1}_{\{X_j^* > x_m\}} \leq \mathbf{1}_{\{X_j^{**} > x_m\}}$  for  $j = 1, \dots, k - 1, k + 1, \dots, n$ . For all  $j \in \{1, \dots, k - 1, k + 1, \dots, n\}$  such that  $\mathbf{1}_{\{X_j^* > x_m\}} < \mathbf{1}_{\{X_j^{**} > x_m\}}$ , we then pick  $X_j^*$  and  $X_j^{**}$  independently according to their respective conditional distributions. For all other  $j \in \{1, \dots, k - 1, k + 1, \dots, n\}$ , we have  $\mathbf{1}_{\{X_j^* > x_m\}} = \mathbf{1}_{\{X_j^{**} > x_m\}}$ , and the conditional distributions of  $X_j^*$  and  $X_j^{**}$  are thus identical (and equal to the conditional distribution  $X_j$  given  $X_j > x_m$  or  $X_j < x_m$  depending on the value of the indicators  $\mathbf{1}_{\{X_j^* > x_m\}}$  and  $\mathbf{1}_{\{X_j^{**} > x_m\}}$ ). We may therefore take  $X_j^* = X_j^{**}$  for each such  $j$ . This defines the coupling. It has the particular property that  $X_j^{**} = X_j^*$  whenever  $X_j^* > x_m$ . It follows that  $X_{j:n}^* \leq X_{j:n}^{**}$  for  $j = i + 1, \dots, n$ . Since the corresponding inequality for  $j = i$  is automatic by the conditioning, we have (8). Thus, Theorem 1 is established in the special case where assumption (5) holds. It remains to remove this assumption.

First drop the disjointness part of (5), again write  $x_1 \leq \dots \leq x_i$  for the union of the supports of the distributions of the  $X_i$ s, and define  $\delta = \min_{i \neq j} |x_i - x_j|$ . For each  $\varepsilon > 0$  and for each  $j \in \{1, \dots, n\}$ , define  $X_j^\varepsilon = X_j - j\varepsilon$ . As soon as  $\varepsilon < \delta/n$ , the distributions of the  $X_j^\varepsilon$  variables have disjoint support, so, for all such  $\varepsilon$ , the desired conclusion, (3), holds with  $(X_1^\varepsilon, \dots, X_n^\varepsilon)$  in place of  $(X_1, \dots, X_n)$ . Sending  $\varepsilon \rightarrow 0$  gives  $X_j^\varepsilon \rightarrow X_j$  for each  $j$ , as well as  $\mathbf{1}_{\{X_{i:n}^\varepsilon > s\}} \rightarrow \mathbf{1}_{\{X_{i:n} > s\}}$  and  $\mathbf{1}_{\{X_{i:n}^\varepsilon > s'\}} \rightarrow \mathbf{1}_{\{X_{i:n} > s'\}}$ . It follows that (3) holds whenever the  $X_i$ s have finite support.

Finally, remove assumption (5) altogether. For  $j \in \{1, \dots, n\}$  and any positive integer  $N$ , define  $X_j^N$  to be  $X_j$  rounded down to the nearest ‘number’ in the set  $\{-\infty\} \cup \{-N, -N + 1/N, -N + 2/N, \dots, N\}$ . Then (3) holds with  $(X_1^N, \dots, X_n^N)$  in place of  $(X_1, \dots, X_n)$ . Sending  $N \rightarrow \infty$  gives  $X_j^N \rightarrow X_j$  for each  $j$ , as well as  $\mathbf{1}_{\{X_{i:n}^N > s\}} \rightarrow \mathbf{1}_{\{X_{i:n} > s\}}$  and  $\mathbf{1}_{\{X_{i:n}^N > s'\}} \rightarrow \mathbf{1}_{\{X_{i:n} > s'\}}$ . So (3) holds with assumption (5) removed, and the proof is complete.

**Remark.** Note that our coupling of  $(X_1^*, \dots, X_n^*)$  and  $(X_1^{**}, \dots, X_n^{**})$  satisfies  $X_j^* \leq X_j^{**}$  for every  $j$  except possibly for  $j = k$ . So in a sense the incorrect statement, (2), is not terribly far from being true.

### Acknowledgement

We thank Taizhong Hu for pointing out an error in an earlier draft and for other helpful comments.

### References

- [1] BOLAND, P. J., HOLLANDER, M., JOAG-DEV, K. AND KOCHAR, S. (1996). Bivariate dependence properties of order statistics. *J. Multivariate Anal.* **56**, 75–89.
- [2] EFRON, B. (1965). Increasing properties of Pólya frequency functions. *Ann. Math. Statist.* **36**, 272–279.
- [3] HÄGGSTRÖM, O. (2007). Problem solving is often a matter of cooking up an appropriate Markov chain. *Scand. J. Statist.* **34**, 768–780.
- [4] HARRIS, T. E. (1960). Lower bound for the critical probability in a certain percolation process. *Proc. Camb. Philos. Soc.* **56**, 13–20.
- [5] HU, T. AND XIE, C. (2006). Negative dependence in the balls and bins experiment with applications to order statistics. *J. Multivariate Anal.* **97**, 1342–1354.
- [6] LINDVALL, T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.
- [7] NEWTON, I. (1707). *Arithmetica Universalis: Sive de Compositione et Resolutione Arithmetica Liber*.
- [8] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–439.