Sankhyā: The Indian Journal of Statistics
1997, Volume 59, Series A, Pt. 1, pp. 102-116

STOCHASTIC PROPERTIES OF ORDER STATISTICS FROM FINITE POPULATIONS

By SUBHASH C. KOCHAR
Indian Statistical Institute, New Delhi
and
RAMESH KORWAR*
University of Massachusetts, Amherst

SUMMARY. We study stochastic orders and dependence relations between order statistics from a linearly ordered finite population when using either simple random sampling without replacement (SRSWOR) or Midzuno sampling schemes. It is shown that when there are no multiplicities in the population, the denisty functions of order statistics, in the cases of SRSWOR and a special case of Midzuno sampling, are logconcave and hence they have *increasing failure rate* (IFR) distributions. Also in this case the successive order statistics are likelihood ratio ordered. It is also seen that whereas any pair of order statistics is positively quadrant dependent under Midzuno sampling, it may not satisfy many of the stronger notions of positive dependence like positive regression dependence and TP_2 dependence etc. However, we are able to prove that $X_{(j)}$ is *right tail increasing* in $X_{(1)}$ for any $j > 1$. We further discuss some unresolved problems in this area.

1. Introduction

There are many ways in which one can say that a random variable X is smaller than another random variable Y. In the *usual* stochastic ordering, one says that X is stochastically smaller than Y (and write as $X \leq_{st} Y$) if $F_X(t) \geq F_Y(t)$ for all t. That is, $X \leq_{st} Y$ if the distribution function F_Y of Y is dominated by that of X at all points. A very useful characterizing property of stochastic ordering is that $X \leq_{st} Y \Leftrightarrow E[g(X)] \leq E[g(Y)]$ for all nondecreasing functions g whenever the expectations exist.

Paper received. August 1995; revised October 1996.
AMS (1991) *subject classifications.* 62G30, 60E15, 62D05.
Key words and phrases. Simple random sampling without replacement, Midzuno sampling, log-concave density, likelihood ratio ordering, failure rate ordering, TP_2 dependence, RTI property, positive quadrant dependence.
* Work done while the second author visiting ISI, New Delhi. He wishes to sincerely thank ISI for extending an invitation and making available all its facilities.

In some cases, a pair of distributions may satisfy a stronger condition called *likelihood ratio ordering*. If distributions F and G possess densities (or probability mass functions) f and g, respectively, then the condition required for likelihood ratio ordering is given by

$$f(x)/g(x) \quad \text{is nonincreasing in } x. \qquad \ldots(1.1)$$

This ordering is denoted by $X \leq_{l.r.} Y$ and has the interpretation that (1.1) holds if and only if for every $a < b$, the conditional distribution of X given $X \in [a,b]$ is stochastically smaller than that of Y given $Y \in [a,b]$. It is known that $X \leq_{l.r} Y$ implies $\overline{F}(x)/\overline{G}(x)$ is nonincreasing in x, where $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$ denote the survival functions of X and Y, respectively. This latter condition defines *hazard rate ordering* and this, in turn, implies stochastic ordering. It is shown in Shanthikumar and Yao (1991) that $X \leq_{l.r} Y$ if and only if $E_g(X,Y) \geq E_g(Y,X)$, $\forall g \in \mathcal{G}_{l.r.} := \{g(x,y) : g(x,y) \geq g(y,x),\ \forall x \leq y\}$.

There are also many notions of positive dependence between random variables. Perhaps the strongest of them is what is called TP_2 dependence. Two random variables X and Y are TP_2 (totally positive of order 2) dependent if their joint density or mass function $f(x,y)$ is TP_2 or more precisely, if

$$\begin{vmatrix} f(x_1,y_1) & f(x_1,y_2) \\ f(x_2,y_1) & f(x_2,y_2) \end{vmatrix} \geq 0, \qquad \ldots(1.2)$$

for any $x_1 \leq x_2, y_1 \leq y_2$. Two random variables X and Y are *right corner set increasing* (RCSI) if for any fixed x and y, $P[X > x, Y > y | X > x', Y > y']$ is increasing in x' and y'. One says that Y is *stochastically increasing* in X if $P[Y > y | X = x]$ is increasing in x for all y, and write $SI(Y|X)$. Lehmann (1966) uses the term *positively regression dependent* to describe SI. We say that Y is *right tail increasing* in X if $P[Y > y | X > x]$ is increasing in x for all y, and write $RTI(Y|X)$. The random variables X and Y are *associated* (written as $A(X,Y)$) if $Cov[\Gamma(X,Y), \Delta(X,Y)] \geq 0$ for all pairs of increasing binary functions Γ and Δ. Finally we say that X and Y are *positively quadrant dependent* if

$$P[X \leq x, Y \leq y] \geq P[X \leq x] P[Y \leq y].$$

for all x, y and write $PQD(X,Y)$. The various implications between these notions of dependence are summarized in the following figure (cf. Barlow and Proschan (1981) and Shaked (1977)).

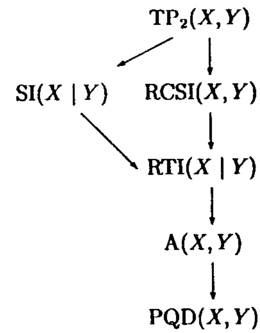

Figure 1. Implications among notions of positive dependence

Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the order statistics of the random variables $X_1, X_2, \ldots, X_n$. There is an extensive literature on order statistics when $X_1, \ldots, X_n$ is a random sample from an absolutely continuous distribution. There are excellent books by David (1980) and Arnold, Balakrishnan and Nagaraja (1992) on the distribution theory of order statistics from parametric families of distributions.

However, it would be more interesting to study the general stochastic properties of order statistics. Developments on this topic up to 1988 are reviewed in the expository article by Kim, Proschan and Sethuraman (1988). The recent book of Shaked and Shanthikumar (1994) contains many results on stochastic orders between order statistics based on identically as well as non-identically distributed *independent* random variables. Also see the survey article by Boland, Shaked and Shanthikumar (1995) on this topic. In particular, it is shown in Kim, Proschan and Sethuraman (1988) that in case the order statistics are based on a random sample from a continuous distribution, then, in general, $X_{(i)} \leq_{l.r.} X_{(j)}$ for $i < j$. However, this relation may not hold in case the original observations are *not* identically distributed. Bapat and Kochar (1994) have proved that this is true if the independent random variables are themselves ordered according to likelihood ratio ordering. It is also easy to prove that in the indepenendent, identically distributed (i.i.d.) case $X_{(i)}$ and $X_{(j)}$ are TP_2 dependent for any $i \neq j$. Boland *et al.* (1996) have shown with the help of a counter example that in the case of the independent (but not necessarily identically distributed) random variables $X_1, X_2, \ldots, X_n$, it is not necessarily true that $X_{(i)}$ and $X_{(j)}$ are TP_2 dependent. They have proved, however, that $RTI(X_{(j)}|X_{(i)})$ does hold in this case for any $i < j$.

As mentioned above, most of the results on order statistics are available only in the case of independent random variables of continuous type. It should be interesting to examine to what extent the above results can be generalized to the case when the original observations do not constitute a random sample from an absolutely continuous distribution.

Chapter 4 of Arnold, Balakrishnan and Nagaraja (1992) contains some distribution theory of order statistics from discrete distributions. Boland et al.

(1996) discuss the dependence properties of order statistics from simple random sampling without replacement (SRSWOR) from a linearly ordered finite population. Note that in the case of SRSWOR, the original observations are identically distributed but are *not* independent. A variation on and in some respects the closest to SRSWOR is Midzuno sampling (see, Gabler (1987)), also known as sampling with probability proportional to aggregate size. In this sampling scheme, introduced by Midzuno (1950), n units from a finite population of size N are drawn one by one without replacement as in SRSWOR, but a unit k has probability p_k (compare $p_k = 1/N$ for SRSWOR), $k = 1, \ldots, N, \sum_{i=1}^{N} p_k = 1$, of being chosen at the first draw and the remaining $n-1$ units in the sample forming a SRSWOR of size $(n-1)$ from the remaining $(N-1)$ units (after the first draw) in the population. Thus an *unordered* sample $s = \{i_1, \ldots, i_n\}$ has a probability $p(s) = \sum_{j=1}^{n} p_{i_j} / \binom{N-1}{n-1}$ of being chosen by the sampling scheme. In this case, the observations are neither independent nor identically distributed.

In this paper we will study some stochastic properties of order statistics when sampling from a linealy ordered finite population using either the SRSWOR or the Midzuno sampling scheme. In the next section we prove that when there are no multiplicites in the population, the density function of $X_{(i)}$ is logconcave for any i. We also establish likelihood ratio ordering between the successive order statistics in thise case. The third section is devoted to the study of dependence properties of a pair of order statistics. The paper ends with Section 4 on concluding remarks and some open problems in this area.

2. Stochastic orderings between order statistics

In this section we will investigate the stochastic properties of the *marginal* distributions of order statistics when sampling from a linearly ordered finite population using the SRSWOR and the Midzuno schemes.

A finite population which is linearly ordered and which has no replications can be represented as $\{1, \ldots, N\}$. Let $X_1, \ldots, X_n$ be the observations in a random sample of size n drawn using either SRSWOR or Midzuno sampling and let $X_{(1)}, \ldots, X_{(n)}$ be the corresponding order statistics. If X_i represents the value obtained on the ith draw ($i = 1, \ldots, n$), then X_i's are identically but not independently distributed in SRSWOR, and neither identically nor independently distributed in Midzuno sampling.

The probability mass function (p.m.f.) $f_{(i)}$ of $X_{(i)}, i = 1, \ldots, n$ is given by
(a) for SRSWOR

$$f_{(i)}(x) = \begin{cases} \dfrac{\binom{s-1}{i-1}\binom{N-s}{n-i}}{\binom{N}{n}}, & s = i, \ldots, N-n+i; \\ 0, & \text{otherwise}; \end{cases} \qquad \ldots (2.1)$$

(b) for Midzuno sampling

$$f_{(i)}(s) = \begin{cases} \left[\binom{N-1}{n-1}\right]^{-1}\left[\left(\sum_{l=1}^{s-1} p_l\right)\binom{s-2}{i-2}\binom{N-s}{n-i} \right. \\ \left. +p_s\binom{s-1}{i-1}\binom{N-s}{n-i} + \left(\sum_{l=s+1}^{N} p_l\right)\binom{s-1}{i-1}\binom{N-s-1}{n-1-i}\right] & s = i, \ldots N-n+i; \\ 0, & \text{otherwise.} \end{cases} \quad \ldots(2.2)$$

We show in this section that in the case of SRSWOR, the probability mass function of $X_{(i)}$ is logconcave for any i. However this result holds in the case of Midzuno sampling only for the special case $p_k \propto k, k = 1, \ldots, N$. Recall that an integer valued random variable with probability mass function g is said to be logconcave if $g(s)/g(s-1)$ is nonincreasing in s in its support. It is also known that a random variable with logconcave density (or probability mass function) has *increasing failure rate* (IFR) distribution. See Chapter 1 of Shaked and Shanthikumar (1994) for further properties of distributions with logconcave densities.

THEOREM 2.1. *For $i = 1, \ldots, n, f_{(i)}(s)$ is logconcave in s,*
(a) for SRSWOR,
(b) for Midzuno sampling when $p_k \propto k, k = 1, \ldots, N$.

PROOF. (a) From (2.1) it follows that

$$\begin{aligned} \frac{f_{(i)}(s)}{f_{(i)}(s-1)} &= \left[\binom{s-1}{i-1}\binom{N-s}{n-i}\right] / \left[\binom{N}{n}\right] / \left[\binom{s-2}{i-1}\binom{N-s+1}{n-i} / \binom{N}{n}\right] \\ &= (s-1)(N-s+1-n+i)/\{(N-s+1)(s-i)\} \\ &= \left(1+\frac{i-1}{s-i}\right)\left(1-\frac{n-i}{N-s+1}\right), \end{aligned}$$

which is a decreasing function of s, since both the factors of the last equality are positive and decreasing in s. Thus $f^2_{(i)}(s) \geq f_{(i)}(s-1)f_{(i)}(s+1)$. That is, $f_{(i)}(s)$ is logconcave.

(b) From (2.2) and $p_k = ck, k = 1, \ldots, N, c^{-1} = N(N+1)/2$, we have, after a little bit of rearranging that

$$f_{(i)}(s) = \frac{c}{2}\{(n+1)s + (n-i)(N+1)\}\frac{N}{n}P(Y'_{(i)} = s), \quad \ldots(2.3)$$

where $Y'_{(i)}$'s are the order statistics from a SRSWOR of size n from the linearly ordered finite population $\{1, \ldots, N\}$. Using (2.3) we obtain that

$$\begin{aligned}\frac{f_{(i)}(s)}{f_{(i)}(s-1)} &= \frac{\{(n+1)s+(n-i)(N+1)\}P(Y'_{(i)}=s)}{\{(n+1)(s-1)+(n-i)(N+1)\}P(Y'_{(i)}=s-1)} \\ &= \left\{1+\frac{n+1}{(n+1)(s-1)+(N+1)(n-i)}\right\}\frac{P(Y'_{(i)}=s)}{P(Y'_{(i)}=s-1)}\end{aligned}$$

which is decreasing function in s, since both the factors of the last equality are positive and decreasing in s, the monotonicity of the second factor being a consequence of (a) above. □

Our next two examples show that logconcavity holds (1) neither when multiplicities in the population are present and SRSWOR is used (2) nor in the general Midzuno scheme.

EXAMPLE 2.1. Let the value 1 occur in the population 3 times and the values 2,3,4 once each. That is, let there be $m_1 = 3$ units with value 1 each, $m_i = 1$ unit with value $i = 2, 3, 4$ and $N = m_1 + m_2 + m_3 + m_4 = 6$. Let $n = 2$ be the sample size of a SRSWOR. We then compute $P(X_{(1)} = 1) = 12/15, P(X_{(1)} = 2) = 2/15$ and $P(X_{(1)} = 3) = 1/15$. Thus $f^2_{(1)}(2) = 4/225 < 12/15 \times 1/15 = f_{(1)}(1)f_{(1)}(3)$. That is, $f_{(1)}(s)$ is not logconcave.

This example also serves to illustrate that logconcavity of $X_{(i)}$ does not hold for Midzuno sampling when multiplicities are present. This is so since SRSWOR is a special case of Midzuno sampling.

EXAMPLE 2.2. To illustrate the logconcavity does not hold in the general Midzuno sampling, consider a finite population $U = \{1, 2, 3, 4\}$ with $p_1 = .97, p_2 = p_3 = p_4 = .01$. Take $n = 2$. Then

$$f_{(1)}(1) = (3p_1 + p_2 + p_3 + p_4)/3 = .98, f_{(1)}(2) = (2p_2 + p_3 + p_4)/3 = .0133$$

and $f_{(1)}(3) = (p_3 + p_4)/3 = .0067$. Hence

$$f^2_{(1)}(2) = (.0133)^2 < .98 \times .0067 = f_{(1)}(1)f_{(1)}(3),$$

showing that logconcavity of $X_{(i)}$ does not hold even for $i = 1$.

The next restult concerns itself with likelihood ratio ordering or $X_{(i)}$'s.

THEOREM 2.2. *Assume that there are no multiple values in the linearly ordered finite population. Then for* $i < j$; $j = 2, \ldots, n$, *(a) for SRSWOR* $X_{(i)} \leq_{l.r} X_{(j)}$ *and (b) for Midzuno sampling with* $p_k \propto k, X_{(i)} \leq_{l.r.} X_{(j)}$.

PROOF. (a) From (2.1) we get

$$\frac{f_{(i+1)}(s)}{f_{(i)}(s)} = \frac{\binom{s-1}{i}\binom{N-i}{n-i-1}/\binom{N}{n}}{\binom{s-1}{i-1}\binom{N-s}{n-i}/\binom{N}{n}} = \left(\frac{n-i}{i}\right)\cdot\frac{s-i}{N-s-n+i+1},$$

which is an increasing function of s, showing that $X_{(i)} \leq_{l.r.} X_{(i+1)}, i = 1, \ldots, n-1$. The result follows from this.

(b) From (2.3), we get for $i < j$

$$\begin{aligned} \frac{f_{(j)}(s)}{f_{(i)}(s)} &= \frac{\{(n+1)s + (n-j)(N+1)\}}{\{(n+1)s + (n-i)(N+1)\}} \frac{P(Y'_{(j)} = s)}{P(Y'_{(i)} = s)} \\ &= \left\{1 - \frac{(j-i)(N+1)}{(n+1)s + (n-i)(N+1)}\right\} \frac{P(Y'_{(j)} = s)}{P(Y'_{(i)} = s)}. \end{aligned}$$

Now the second factor in the last equality is positive and by (a), increasing in s, while the first factor is positive and increasing in s. Hence $f_{(j)}(s)/f_{(i)}(s)$ is increasing in s and the result follows. □

It is an open question whether $X_{(i)} \leq_{l.r.} X_{(j)}, i < j;\ j = 2, \ldots, n$ holds either in SRSWOR when multiplicities are present or in the general Midzuno sampling. We close this section by showing that the variances of $X_{(i)}$'s are not ordered. That is, $Var(X_{(i)}) \leq Var(X_{(j)}), i < j;\ j = 2, \ldots, n$ does not obtain. This we do by actually computing the variance of $X_{(i)}$ in general for SRSWOR and by a numerical example in the Midzuno case. The result (2.4) below was first proved by Arnold, Balakrishnan and Nagaraja (1992) by using exchangeable random variables and their proof is rather long and complicated. Our proof is simple, short and elementary. A similar proof can yield moment of any order of $X_{(i)}$ in a closed form.

THEOREM 2.3. *For SRSWOR from a linearly ordered finite population without multiplicities*

$$Var(X_{(i)}) = [(N+1)(N-n)/\{(n+1)^2(n+2)\}]\{(\frac{n+1}{2})^2 - (i - \frac{n+1}{2})^2\} \quad \ldots (2.4)$$

and hence for $i < j$,

$$Var(X_{(i)}) < Var(X_{(j)}),\ j \leq [(n+1)/2]$$

and

$$Var(X_{(i)}) > Var(X_{(j)}),\ i \geq [(n+1)/2]$$

where $[x]$ represents the greatest integer in x less than or equal to x.

PROOF. From (2.1) we have

$$\begin{aligned} E(X_{(i)}) &= \sum_{s=i}^{N-n+i} s \binom{s-1}{i-1} \binom{N-s}{n-i} / \binom{N}{n} \\ &= i \binom{N+1}{n+1} / \binom{N}{n} \\ &= i(N+1)/(n+1) \end{aligned}$$

and similarly

$$E\{(X_{(i)}+1)X_{(i)}\} = i(i+1)(N+1)(N+2)/\{(n+1)(n+2)\}$$

From these we get

$$\begin{aligned} Var(X_{(i)}) &= E\{X_{(i)}(X_{(i)}+1)\} - E(X_{(i)}) - E^2\{X_{(i)}\} \\ &= \frac{(N+1)(N-n)}{(n+1)^2(n+2)}\left\{\left(\frac{n+1}{2}\right)^2 - \left(i-\frac{n+1}{2}\right)^2\right\} \end{aligned}$$

The second statement follows from this expression for $Var(X_{(i)})$. □

The next example shows that the variances of order statistics in case of Midzuno sampling are not ordered even in the case when $p_k \propto k, k = 1, \ldots, N$.

EXAMPLE 2.3. Let $N = 5, n - 3, p_k \propto k, k = 1, \ldots, N$. Then the p.m.f's of $X_{(i)}$'s are given by

s	1	2	3	4	5
$f_{(1)}(s)$	$\frac{16}{30}$	$\frac{10}{30}$	$\frac{4}{30}$		
$f_{(2)}(s)$		$\frac{7}{30}$	$\frac{12}{30}$	$\frac{11}{30}$	
$f_{(3)}(s)$			$\frac{2}{30}$	$\frac{8}{30}$	$\frac{20}{30}$

From these we compute

$$Var(X_{(1)}) = \frac{38}{75},\ Var(X_{(2)}) = \frac{131}{225} \text{ and } Var(X_{(3)}) = \frac{28}{75}$$

and note that

$$Var(X_{(1)}) < Var(X_{(2)}) > Var(X_{(3)}).$$

3. Bivariate dependence relations

In this section we explore dependence relations between order statistics of observations from either of the two sampling procedures. Boland *et al.* (1996) have studied such relations for SRSWOR from linearly ordered populations with/without multiplicities. They have shown that the joint p.m.f. of order statistics $X_{(i)}$ and $X_{(j)}$ from a linearly ordered finite population with/without replications is, for all i, j, TP_2. From this strong property follow weaker properties such as $SI(X_{(j)}|X_{(i)})$ and $RTI(X_{(j)}|X_{(i)})$.

A natural question to ask is : How far these and other bivariate dependence notions among order statistics continue to hold when the sampling scheme is changed even slightly as in Midzuno sampling ? We can supply only a partial answer to this question. We first show with the help of a counter example that even $RCSI(X_{(i)}, X_{(j)})$ does not hold in this case.

EXAMPLE 3.1. Consider the linearly ordered population $U = \{1,2,3,4,5\}$ with no multiplicities, and take a sample of $n = 2$ using Midzuno sampling with $p_1 = p_2 = p_4 = .001, p_3 = .008$ and $p_5 = .989$. We then compute

$$\begin{aligned} P(X_{(2)} &= 3|X_{(1)} > 1) = .0030, P(X_{(2)} = 4|X_{(1)} > 1) = .0037 \\ P(X_{(2)} &= 5|X_{(1)} > 1) = .9933; \\ P(X_{(2)} = 4|X_{(1)} > 2) &= .0045, P(X_{(2)} = 5|X_{(1)} > 2) = .9955. \end{aligned}$$

The hazard rates $h(t|s)$ of these conditional distribution are

t	3	4	5
$h(t\|1)$	.0030	.0.0037	1
$h(t\|2)$		0.0045	1

Thus $h(4|1) < h(4|2)$. Now, it can be shown (see Shaked (1977) and Kochar and Deshpande (1986)) that $RCSI(S,T)$ holds iff the hazard rate $h(t|S > s)$ of the conditional distribution of T given $S > s$ is decreasing in s for every fixed t. Since in our case $h(4|1) < h(4|2)$ we conclude that $RCSI(X_{(1)}, X_{(2)})$ does not hold.

Our next example shows that $SI(X_{(j)}|X_{(i)})$ does not hold for Midzuno sampling.

EXAMPLE 3.2. Let $U = \{1,2,3,4,5\}$ with no multiplicities and consider taking a Midzuno sample of size $n = 2$ with probabilities $p_1 = p_3 = .47, p_2 = p_4 = p_5 = .02$. We find that

$$P(X_{(2)} > 3|X_{(1)} = 1) = 0.4066 \text{ and } P(X_{(2)}) > 3|X_{(1)} = 2) = 0.1403.$$

Thus $P(X_{(2)}) > 3|X_{(1)} = 2) < P(X_{(2)}) > 3|X_{(1)} = 1)$ and the property $SI(X_{(j)}|X_{(i)})$ does not hold even for $i = 1$.

We next prove that in case there are no repeated values in the population, $X_{(i)}$ is right tail increasing in $X_{(1)}$, the first order statistic for $i = 2, \ldots, n$.

THEOREM 3.1. *Assume that there are no repeated values in the population. Then, for Midzuno sampling* $RTI(X_{(i)}|X_{(1)}), i = 2, \ldots, n$ *holds*

PROOF. We have

$$P(X_{(1)} > s) = \left(\sum_{k=s+1}^{N} p_k\right) \binom{N-s-1}{n-1} / \binom{N-1}{n-1}$$

and

$$P(X_{(i)} > t, X_{(1)} > s) = \left[\sum_{k=0}^{i-1} \binom{t-s}{k} \binom{N-t-1}{n-k-1} (\sum_{k=s+1}^{t} p_k)\right.$$

$$+(\sum_{k=t+1}^{N} p_k)\sum_{k=0}^{i-2}\binom{t-s-1}{k}\binom{N-t}{n-k-1}\Big] / \binom{N-1}{n-1}.$$

From these we get

$$\begin{aligned} g(t|s) &= P(X_{(i)} > t | X_{(1)} > s) \\ &= p_s^* P(X \le i-2) + (1-p_s^*) P(X' \le i-1) \end{aligned}$$

where

$$X' \sim \text{ Hypergeometric } (t-s, N-t-1, n-1),$$

$$X \sim \text{ Hypergeometric } (t-s-1, N-t, n-1),$$

and

$$p_s^* = \sum_{k=s+1}^{t} p_k / \sum_{k=s+1}^{N} p_k.$$

Now consider a finite population of size $N-s-1$ with three categories of sizes $t-s-1, 1$ and $N-t-1$. Then in a simple random sample without replacement of size $n-1$ from the above population the random vector $(X, Y, n-1-X-Y)$ of the number of units coming from the three respective categories has a multivariate hypergeometric distribution $MH(t-s-1, 1, N-t-1, n-1)$. Thus X' has the same distribution as $X+Y$ (written $X' \stackrel{d}{=} X+Y$). Hence

$$\begin{aligned} g(t|s) &= P(X' \le i-1) - p_s^*\{P(X' \le i-1) - P(X \le i-2)\} \\ &= P(X' \le i-1) - p_s^*\{P(X+Y \le i-1, Y=0) \\ &\quad +P(X+Y \le i-1, Y=1) - P(X \le i-2, Y=0) \\ &\quad -P(X \le i-2, Y=1)\} \\ &= P(X' \le i-1) - p_s^*\{P(X \le i-1, Y=0) \\ &\quad -P(X \le i-2, Y=0)\} \\ &= P(X' \le i-1) - p_s^* P(X = i-1, Y=0) \\ &= P(X' \le i-1) - p_s^* \binom{t-s-1}{i-1}\binom{N-t-1}{n-i} / \binom{N-s-1}{n-1}. \end{aligned}$$

Now

$$g(t|s+1) - g(t|s) = P(X'' \leq i-1) - P(X' \leq i-1)$$
$$+p_s^* \binom{t-s-1}{i-1} \binom{N-t-1}{n-i} \Big/ \binom{N-s-1}{n-1}$$
$$-p_{s+1}^* \binom{t-s-2}{i-1} \binom{N-t-1}{n-i} \Big/ \binom{N-s-2}{n-1}$$

where

$$X'' \sim \text{ Hypergeometric } (t-s-1, N-t-1, n-1).$$

Now we will show that

$$P(X'' \leq i-1) - p(X' \leq i-1)$$
$$(n-i) \binom{t-s-1}{i-1} \binom{N-s-1}{n-i} /\{\binom{N-s-1}{n-1} (N-s-n)\} \qquad \ldots (3.2)$$

In proving (3.2) and simplifying the right side of (3.1), we repeatedly make use of the identity

$$(m+1)\binom{m}{r} = \binom{m+1}{r}(m-r+1) \qquad \ldots (3.3)$$

If (3.2) is granted, then using (3.2) and (3.3) on the right side of (3.1), we get

$$\begin{aligned} g(t|s+1) - g(t|s) &= c\{(n-i)(t-s-1) + p_s^*(N-s-n)(t-s-1) \\ &\quad -p_{s+1}^*(N-s-1)(t-s-i)\} \\ &\geq c(N-t)(i-1)p_s^* \\ &\geq 0, \end{aligned}$$

since $1 \geq p_s^* \geq p_{s+1}^*$. Here

$$c = \binom{t-s-1}{i-1} \binom{N-t-1}{n-i} /\{(N-s-n)(t-s-1)\binom{N-s-1}{n-1}\}$$

Now to prove (3.2), we write the left side, using the definitions of X' and X'', as

$$\sum_{k=0}^{i-1} c' \left\{ \binom{t-s-1}{k} \binom{N-t-1}{n-1-k} (N-s-1) - \binom{t-s}{k} \binom{N-t-1}{n-1-k} (N-s-n) \right\}$$

$$= \sum_{k=0}^{i-1} c' \left\{ \binom{t-s}{k} (t-s-k) \binom{N-t-1}{n-1-k} + (N-t-1) \binom{t-s-1}{k} \binom{N-t-1}{n-1-k} \right\}$$

$$-\sum_{k=0}^{i-1} c' \left\{ \binom{t-s}{k} \binom{N-t-1}{n-1-k} (t-s-k) + (N-t-1) \binom{t-s}{k} \binom{N-t-2}{n-1-k} \right\}$$

$$= c'(N-t-1) \binom{N-s-2}{n-1} \sum_{k=0}^{i-1} \left\{ \binom{t-s-1}{k} \binom{N-t-1}{n-1-k} / \binom{N-s-2}{n-1} \right.$$

$$\left. - \binom{t-s}{k} \binom{N-t-2}{n-1-k} / \binom{N-s-2}{n-1} \right\}$$

$$= c'(N-t-1) \binom{N-s-2}{n-1} \{P(X''' \leq i-1) - P(X'''' \leq i-1)\} \quad \ldots(3.4)$$

where

$$X''' \sim \text{Hypergeometric } (t-s-1, N-t-1, n-1),$$

$$X'''' \sim \text{Hypergeometric } (t-s, N-t-2, n-1) \text{ and}$$

$$c' = 1 / \left\{ \binom{N-s-1}{n-1} (N-s-n) \right\} \quad \ldots(3.5)$$

To complete the proof, we use by now a familiar argument. Consider a finite population of size $N-s-2$ with three categories of sizes $t-s-1, 1$ and $N-t-2$. In a simple random sample without replacement of size $n-1$ the random vector $(X''', Y', n-1-X'''-Y')$ of the units coming from the above respective categories has a $MH(t-s-1, 1, N-t-2, n-1)$ distribution. Hence

$$X'''' \stackrel{d}{=} X''' + Y'$$

and

$$P(X''' \leq i-1) - P(X'''' \leq i-1)$$

$$= P(X''' \leq i-1) - P(X''' + Y' \leq i-1)$$

$$= P(X''' \leq i-1, Y' = 0) + P(X''' \leq i-1, Y' = 1)$$

$$-P(X''' + Y' \leq i-1, Y' = 0) - P(X''' + Y' \leq i-1, Y' = 1)$$

$$= P(X''' \leq i-1, Y' = 1) - P(X''' \leq i-2, Y' = 1)$$

$$= P(X''' = i-1, Y' = 1)$$

$$= \binom{t-s-1}{i-1} \binom{N-t-2}{n-i-1} / \binom{N-s-2}{n-1}. \quad \ldots(3.6)$$

Now (3.2) follows from (3.4) - (3.6), completing the proof. □

Our final result concerns itself with the weakest of the notions of bivariate dependency.

THEOREM 3.2. *In Midzuno sampling from a linearly ordered population without multiplicites, $X_{(i)}$ and $X_{(j)}$ are PQD, $i \neq j, i, j = 1, \ldots, n$.*

PROOF. The survival function of $X_{(i)}$ and the joint survival function of $X_{(i)}$ and $X_{(j)}$ are, respectively, given by

$$\begin{aligned} P(X_{(i)} > s) &= \left[\binom{N-1}{n-1}\right]^{-1} \left[\left(\sum_{l=1}^{s} p_l\right) \sum_{k=0}^{i-2} \binom{s-1}{k} \binom{N-s}{n-1-k} \right. \\ &\quad \left. + \left(\sum_{l=s+1}^{N} p_l\right) \sum_{k=0}^{i-1} \binom{s}{k} \binom{N-s-1}{n-1-k}\right] \\ &= P(s)P(Y_{(i-1)} > s-1) + (1-P(s))P(Y_{(i)} > s) \qquad \ldots (3.7) \end{aligned}$$

and

$$\begin{aligned} P(X_{(j)} > t, X_{(i)} > s) &= P(s)P(Y_{(i-1)} > s-1, Y_{(j-1)} > t-1) \\ &\quad (P(t) - P(s))P(Y_{(i)} > s, Y_{(j-1)} > t-1) \qquad \ldots (3.8) \\ &\quad +(1-P(s))P(Y_{(i)} > s, Y_{(j)} > t), \end{aligned}$$

where $P(l) = \sum_{k=1}^{l} p_k, l = 1, \ldots, N$ and $Y_{(i)}$'s are order statistics in a SRSWOR of size $(n-1)$ from $U = \{1, \ldots, N-1\}$ (without multiplicities). From (3.7) and (3.8), we have

$$\begin{aligned} &P(Y_{(j)} > t, Y_{(i)} > s) - P(Y_{(j)} > t)P(Y_{(i)} > s) \\ &= P(s)P(Y_{(i-1)} > s-1, Y_{(j-1)} > t-1) + (P(t) - P(s)) \\ &\quad \times P(Y_{(i)} > s, Y_{(j-1)} > t-1) + (1-P(s))P(Y_{(i)} > s, Y_{(j)} > t) \\ &\quad -\{P(s)P(Y_{(i-1)} > s-1) + (1-P(s)P(Y_{(i)} > s)\}\{P(t)P(Y_{(j-1)} > t-1) \\ &\quad +(1-P(t)P(Y_{(j)} > t)\}. \qquad \ldots (3.9) \end{aligned}$$

Now in SRSWOR $PQD(Y_{(i)}, Y_{(j)})$ holds (see Boland *et al.* (1996)). Hence, $P(Y_{(i)} > s, Y_{(j)} > t) \geq P(Y_{(i)} > s)P(Y_{(j)} > t)$, $\forall s, t$ (and for all i, j). Hence using the monotonicity of $P(s)$, the right hand side of (3.9)) is

$$\geq P(s)(1-P(t)\{P(Y_{(i)} > s) - P(Y_{(i-1)} > s-1)\}\{P(Y_{(j)} > t) - P(Y_{(j-1)} > t-1)\}$$

and the result follows from the fact that $\{Y_{(l-1)} > u - 1\}$ implies $\{Y_{(l)} > u\}$. $\square$

4. Concluding remarks

In this paper we have tried to extend to two particular sampling situations in which observations are dependent / or nonidentically distributed, results on order statistics that exist for observations which are independent / or identically distributed. As stated before, the observations from a SRSWOR are dependent buit identically distributed, while those from the Midzuno sample are neither independent nor identically distributed. The answers to questions raised here are at best partial and the questions need further investigation. For example, does $RTI(X_{(j)}|X_{(i)})$ hold, in general, for $i < j$ and not just for $i = 1$? Do some of the results presented here for a linearly ordered finite population without replication go through when there are replications ?

References

ARNOLD, B., BALAKRISHNAN, N. and NAGARAJA, H.N. (1992). *A First Course in Order Statistics*, Wiley : New York.

BAPAT, R.B. and KOCHAR, S.C. (1994). On likelihood ratio ordering of order statistics. *Linear Algebra and Its Applications*, **199**, 281-291.

BARLOW, R.E. and PROSCHAN, F. (1981). *Statistical Thoeory of Reliability and Life Testing*, To Begin With : Silver Spring, Maryland.

BOLAND, P.J., HOLLANDER, M., JOAG-DEV, K. and KOCHAR, S. (1996). Bivariate dependence properties of order statistics. *J. Multivariate Anal.*, **56**, 75-89.

BOLAND, P.J., SHAKED, M. and SHANTHIKUMAR, J.G. (1995). Stochastic ordering of order statistics. Preprint.

DAVID, H.A. (1980). *Order Statistics*, Wiley : New York.

GABLER, S. (1987). The nearest proportional to size sampling design. *Commun. Statist. Theor-Meth.*, **16**, 1117-1131.

HÁJEK, J. (1981). *Sampling from a Finite Population*. Marcel Dekker : New York.

KIM, J.S., PROSCHAN, F. and SETHURAMAN, J. (1988). Stochastic comparisons of order statistics, with applications in reliability. *Commun. Statist. Theor-Meth.*, **17**, 2151-2172.

KOCHAR, S.C. and DESHPANDE, J.V. (1986). Nonparametric tests for independence in bivariate reliability models. *Proc. International Symp. On Advances in Multivariate Statistical Analysis*, ed. J.K. Ghosh and S. DasGupta, 203-217.

LEHMANN, E.L. (1966). Some concepts of depedndence. *Ann. Math. Statist.*, **37**, 1137-1153.

MIDZUNO, H. (1950). An outline of the theory of sampling systems. *Ann. Inst. Stat. Math.*, **1**, 149-156.

SHAKED, M. (1977). A family of concepts of dependence for bivariate distributions. *J. Amer. Statist. Assoc.*, **72**, 642-650.

SHAKED, M. and SHANTHIKUMAR, J.G. (1994). *Stochastic Orders and their Applications.* Academic Press : San Diego.

SHANTHIKUMAR, J.G. and YAO, D.D. (1991). Bivariate characterization of some stochastic order relations. *Adv. Appl. Probab.*, **23**, 642-659.

INDIAN STATISTICAL INSTITUTE
7, S.J.S. SANSANWAL MARG
NEW DELHI - 110016
INDIA

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MA - 01003
U.S.A.