JIRSS (2007)
Vol. 6, No. 1, pp 31-46

Ordered Random Variables from Discontinuous Distributions

N. Balakrishnan[1], **A. Dembińska**[2]

[1]Department of Mathematics and Statistics, McMaster University, Canada. (bala@mcmaster.ca)
[2]Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland. (dembinsk@mini.pw.edu.pl)

Abstract. In the absolutely continuous case, order statistics, record values and several other models of ordered random variables can be viewed as special cases of generalized order statistics, which enables a unified treatment of their theory. This paper deals with discontinuous generalized order statistics, continuing on the recent work of Tran (2006). Specifically, we show that in general neither records nor weak records are submodels of discrete generalized order statistics. Next, we show that progressively Type-II right censored order statistics from an arbitrary distribution can be embedded in the model of generalized order statistics and then use this fact to establish some distributional properties of progressively Type-II right censored order statistics. Finally, we present some characterizations of the geometric distribution based on progressively Type-II right censored order statistics.

Key words and phrases: Characterizations, discrete distributions, generalized order statistics, geometric distribution, Markov property, ordered random variables, progressively Type-II right censored order statistics, records, weak records.

1 Introduction

Various models of ordered random variables having different interpretations have been introduced and studied extensively in the literature. The most popular of these are *order statistics*, $X_{1:n} \leq \cdots \leq X_{n:n}$, obtained by arranging n random variables (rv's) $X_1, \cdots, X_n$ in nondecreasing order of magnitude. Order statistics appear in many areas of statistical theory and applications including quality control, robustness, outlier detection, and reliability analysis. If we assume that the underlying rv's $X_1, \cdots, X_n$ are independent and identically distributed (iid) and have probability density function (pdf) f, then the random vector $(X_{1:n}, \cdots, X_{n:n})$ has the joint pdf as

$$f_{X_{1:n},X_{2:n},\cdots,X_{n:n}}(x_1, x_2, \cdots, x_n) = \begin{cases} n! \prod_{i=1}^{n} f(x_i), & x_1 \leq \cdots \leq x_n, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For more details on order statistics and their properties, interested readers may refer to the books by Arnold, Balakrishnan and Nagaraja (1992), Balakrishnan and Rao (1998a,b), Nevzorov (2001), and David and Nagaraja (2003).

Censored order statistics are generalizations of order statistics which arise naturally in life-testing experiments when the experimenter does not observe failure times of all units placed on the test. Various types of censoring schemes have been considered and discussed in the literature; see, for example, Cohen (1991) and Balakrishnan and Cohen (1991). One of them is progressive Type-II right censoring which arises as follows. Let n identical units with lifetimes $X_1, \cdots, X_n$ be placed on a life-test and at the time of the ith failure, one failed item and additional R_i remaining units are randomly withdrawn (or censored) from the experiment, $i = 1, \cdots, m$. Then, the $m = n - \sum_{i=1}^{m} R_i$ failure times so observed are referred to as *progressively Type-II right censored order statistics* with censoring scheme $(R_1, \cdots, R_m)$, and denoted by $X_{1:m:n}^{(R_1,\cdots,R_m)}, \cdots, X_{m:m:n}^{(R_1,\cdots,R_m)}$. If the underlying variables $X_1, \cdots, X_n$ are iid rv's with absolutely continuous cumulative distribution function (cdf) F and pdf f, then the joint pdf of all m progressively Type-II right censored order statistics is given by

$$f_{X_{1:m:n}^{(R_1,\cdots,R_m)},\cdots,X_{m:m:n}^{(R_1,\cdots,R_m)}}(x_1, \cdots, x_m) = \begin{cases} c \prod_{i=1}^{m} f(x_i)\{1 - F(x_i)\}^{R_i}, & x_1 \leq \cdots \leq x_m, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $c = n(n - R_1 - 1) \cdots (n - R_1 - R_2 - \cdots - R_{m-1} - m + 1)$; see Balakrishnan and Aggarwala (2000, p. 8) and the recent review article on this topic by Balakrishnan (2007).

Record values, closely related to extremal order statistics, have several applications including in the statistical study of floods, droughts and earthquakes and also in problems concerning fatigue failure in reliability theory. Suppose $X_1, X_2, \cdots$ is a sequence of iid rv's. Then, *record times* $(T_n, n \geq 1)$ are defined recurrently as

$$T_1 = 1, \quad T_{n+1} = \min\{j : j > T_n,\, X_j > X_{T_n}\}, \quad n \geq 1,$$

and *record values* (or simply *records*) $(R_n, n \geq 1)$ as $R_n = X_{T_n},\ n \geq 1$. It is well-known that if the underlying rv's $X_1, X_2, \cdots$ have absolutely continuous cdf F with pdf f, then the joint pdf of the first n records has the form

$$f_{R_1,\cdots,R_n}(x_1, \cdots, x_n) = \begin{cases} f(x_n) \prod_{i=1}^{n-1} f(x_i)\{1 - F(x_i)\}^{-1}, & x_1 < \cdots < x_n, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

For further details concerning records, one may refer to the books by Arnold, Balakrishnan and Nagaraja (1998) and Nevzorov (2001).

For absolutely continuous cdf F with pdf f, the above three models of ordered rv's are all contained in the model of *generalized order statistics* introduced by Kamps (1995a,b) in the distributional sense.

Definition 1.1. *Let $n \geq 2$ be a given integer and $\tilde{m} = (m_1, m_2, \cdots, m_{n-1}) \in \mathbb{R}^{n-1}$ and $k \in (0, \infty)$ be parameters such that*

$$\gamma_i = k + n - i + \sum_{j=i}^{n-1} m_j > 0 \quad for \quad 1 \leq i \leq n - 1.$$

Then, the rv's $U(1, n, \tilde{m}, k), \cdots, U(n, n, \tilde{m}, k)$ are called generalized uniform order statistics if their joint pdf has the form

$$f_{U(1,n,\tilde{m},k),\cdots,U(n,n,\tilde{m},k)}(u_1, \cdots, u_n) = \begin{cases} k\left(\prod_{j=1}^{n-1} \gamma_j\right)(1 - u_n)^{k-1} \prod_{i=1}^{n-1} (1 - u_i)^{m_i}, & 0 \leq u_1 \leq \cdots \leq u_n \leq 1 \\ 0 & \textit{otherwise.} \end{cases} \tag{4}$$

For an arbitrary F, the rv's

$$X(i, n, \tilde{m}, k) = F^{\leftarrow}(U(i, n, \tilde{m}, k)), \quad i = 1, 2, \cdots, n,$$

are called generalized order statistics (from F), where $F^{\leftarrow}(u) = \inf\{x : F(x) \geq u\}$, $u \in (0, 1]$ is the generalized inverse function.

Remark 1.1. *If the cdf F is absolutely continuous with pdf f, then the rv's $X(1, n, \tilde{m}, k)$, $\cdots$, $X(n, n, \tilde{m}, k)$ can be shown to have the joint pdf [from (4)] as*

$$f_{X(1,n,\tilde{m},k),\cdots,X(n,n,\tilde{m},k)}(x_1, \cdots, x_n) = \begin{cases} k\left(\prod_{j=1}^{n-1}\gamma_j\right) f(x_n)\{1-F(x_n)\}^{k-1}\prod_{i=1}^{n-1} f(x_i)\{1-F(x_i)\}^{m_i}, & x_1 \leq \cdots \leq x_n \\ 0 & \textit{otherwise.} \end{cases} \quad (5)$$

For $\tilde{m} = (0, 0, \cdots, 0)$ and $k = 1$, the pdf in (5) simply reduces to the joint pdf of order statistics in (1). For $n = m$, $\tilde{m} = (R_1, R_2, \cdots, R_{m-1})$ and $k = R_m + 1$, the pdf in (5) reduces to the joint pdf of m progressively Type-II right censored order statistics in (2). The choice of $\tilde{m} = (-1, -1, \cdots, -1)$ and $k = 1$ leads to the pdf of records in (3). In addition to these three models, some other models of ordered rv's such as the kth records, order statistics with non-integral sample size, appropriately restricted versions of sequential order statistics, Pfeifer records, and k_n-records from non-identical distributions are all particular cases of generalized order statistics; see Kamps (1995a,b). For some additional results on generalized order statistics, one may refer to Keseling (1999), Kamps and Cramer (2001), Cramer and Kamps (2003), and Cramer (2003).

If the ordered rv's instead arise from discrete distributions, the distributions become much more complicated simply due to the positive chance of ties among observations. Despite this, a large body of results on discrete order statistics and records exist in the literature; see, for example, the extensive review articles by Nagaraja (1992) and Dembińska (2007a,b). The distribution theory for other models of ordered rv's mentioned above, however, had not been considered in the literature until recently by Balakrishnan and Dembińska (2007) who presented some distributional properties of progressively Type-II right censored order statistics from discrete distributions. Tran (2006) independently considered discrete generalized order statistics, but as she aptly observed, this concept does not cover as wide a class of models as the continuous case does. In fact, she pointed out only one model that can be viewed as a submodel of discrete generalized order statistics, and that is the model of discrete order statistics.

In this paper, we first show in Section 2 that records can be viewed as a submodel of generalized order statistics corresponding to $\tilde{m} = (-1, -1, \cdots, -1)$ and $k = 1$ if and only if (iff) the underlying cdf is continuous, and that weak records as a submodel iff the underlying cdf is degenerate or continuous or has only one atom at the right endpoint of its support. Next, we show in Section 3 that the case of progressively Type-II right censored order statistics discussed recently by Balakrishnan and Dembińska (2007) is indeed a special case of these generalized order statistics in the distributional sense. This allows us to apply the distribution theory developed by Tran (2006) to establish the Markovian property of the bivariate sequence $(X_{i:m:n}^{(R_1,\cdots,R_m)}, V_i)$, where V_i is the number of $X_{j:m:n}^{(R_1,\cdots,R_m)}$ for $j \leq i$ that are tied with $X_{i:m:n}^{(R_1,\cdots,R_m)}$, and also prove some characterizations of the geometric distribution via properties of progressively Type-II right censored order statistics.

2 Records and weak records

Tran (2006, p. 31) showed that, in contrast to the continuous case, generalized order statistics based on a discrete cdf F with support $\{0, 1, 2, \cdots\}$ and parameters $\tilde{m} = (-1, -1, \cdots, -1)$ and $k = 1$ can not be viewed as records based on F. Following Nevzorov (2001, p. 66), we can generalize this result as follows to any cdf F having at least one atom.

Lemma 2.1. *Let $(R_n, n \geq 1)$ be records corresponding to a sequence of iid rv's with cdf F such that*

$$\exists\, x_0 \quad \text{for which} \quad F(x_0^-) < F(x_0),$$

and let $(RU_n, n \geq 1)$ denote records from uniform distribution on $[0, 1]$. Then, for any $n = 2, 3, \cdots$,

$$(R_1, R_2, \cdots, R_n) \overset{d}{\neq} (F^{\leftarrow}(RU_1), F^{\leftarrow}(RU_2), \cdots, F^{\leftarrow}(RU_n)), \tag{6}$$

where $\overset{d}{=}$ stands for equality in distribution.

Proof. Suppose the statement in (6) is false. Then, we must have

$$(R_1, R_2) \overset{d}{=} (F^{\leftarrow}(RU_1), F^{\leftarrow}(RU_2)),$$

which by (3) implies that

$$\begin{aligned} P(R_1 = x_0, R_2 = x_0) & \\ &= P(F^{\leftarrow}(RU_1) = x_0, F^{\leftarrow}(RU_2) = x_0) \\ &= P(F(x_0^-) \leq RU_1 \leq F(x_0), F(x_0^-) \leq RU_2 \leq F(x_0)) \\ &= \int_{\mathcal{A}} \frac{1}{1-u_1} du_1 du_2, \end{aligned} \tag{7}$$

where $\mathcal{A} = \{(u_1, u_2) : F(x_0^-) \leq u_1 \leq u_2 \leq F(x_0)\}$. Clearly, the right-hand side of (7) is positive, which contradicts the fact that $R_1 < R_2$. $\square$

Note that if x_0 in Lemma 2.1 is the right endpoint of the support of F, then records $(R_n, n \geq 1)$ are not well-defined, while the right-hand side of (7) is meaningful.

It is well-known that records based on a continuous cdf F admit the representation

$$(R_1, R_2, \cdots, R_n) \stackrel{d}{=} (F^{\leftarrow}(RU_1), F^{\leftarrow}(RU_2), \cdots, F^{\leftarrow}(RU_n)), \quad n = 1, 2, \cdots; \tag{8}$$

see, for example, Nevzorov (2001, p. 66). Upon combining this result, Lemma 2.1 and Eq. (3), we obtain the following theorem.

Theorem 2.1. *The model of generalized order statistics based on a cdf F for $k = 1$ and $\tilde{m} = (-1, -1, \cdots, -1)$ reduces in the distributional sense to the model of records related to F iff F is a continuous cdf.*

In the literature, one can find an alternative model of records called *weak records*. In this model, a repetition of a record is also counted as a record. Weak record times are thereby defined as

$$U_1 = 1, \quad U_{n+1} = \min\{j > U_n : X_j \geq X_{U_n}\} \quad \text{for } n \geq 1,$$

and weak records as

$$W_n = X_{U_n}, \quad n \geq 1.$$

If the parent cdf F is continuous, then the probability of ties among $X_1, X_2, \cdots$ is equal to zero and consequently the sequence of weak records coincides almost surely with the sequence of records. Therefore, in the continuous case, the model of weak records in the distributional sense is a submodel of generalized order statistics corresponding to the parameters $\tilde{m} = (-1, -1, \cdots, -1)$ and $k = 1$. A question that arises is whether this property holds also for discontinuous cdf's F. The following lemma shows that in general the answer to this question is negative.

Lemma 2.2. *Let $(W_n, n \geq 1)$ be weak records arising from a sequence of iid rv's $X_1, X_2, \cdots$ with cdf F such that*

$$\exists\, x_0 \quad \textit{for which} \quad F(x_0^-) < F(x_0) < 1, \tag{9}$$

and let $(WU_n, n \geq 1)$ *denote weak records from uniform distribution on* $[0,1]$*. Then, for any* $n = 2, 3, \cdots$,

$$(W_1, W_2, \cdots, W_n) \overset{d}{\neq} (F^{\leftarrow}(WU_1), F^{\leftarrow}(WU_2), \cdots, F^{\leftarrow}(WU_n)). \quad (10)$$

Proof. Let us once again suppose that the statement in (10) is false. Then, we must have

$$(W_1, W_2) \overset{d}{=} (F^{\leftarrow}(WU_1), F^{\leftarrow}(WU_2)). \quad (11)$$

Since for continuous underlying distributions the sequence of weak records coincides with the sequence of records with probability 1, Eq. (11) can be rewritten as

$$(W_1, W_2) \overset{d}{=} (F^{\leftarrow}(RU_1), F^{\leftarrow}(RU_2)), \quad (12)$$

where $(RU_n, n \geq 1)$ denotes the sequence of records from uniform distribution on $[0,1]$. Hence, for x_0 satisfying (9),

$$\begin{aligned} P(W_1 = x_0, W_2 = x_0) \\ &= P(F^{\leftarrow}(RU_1) = x_0, F^{\leftarrow}(RU_2) = x_0) \\ &= P(F(x_0^-) \leq RU_1 \leq F(x_0), F(x_0^-) \leq RU_2 \leq F(x_0)) \\ &= \int_{F(x_0^-)}^{F(x_0)} \left[\int_{u_1}^{F(x_0)} \frac{1}{1-u_1} du_2 \right] du_1, \end{aligned}$$

with the last equality being a consequence of (3). An easy computation gives

$$P(W_1 = x_0, W_2 = x_0) = P(X_1 = x_0) + P(X_1 > x_0) \ln \frac{P(X_1 > x_0)}{P(X_1 \geq x_0)}. \quad (13)$$

On the other hand,

$$\begin{aligned} P(W_1 = x_0, W_2 = x_0) \\ &= P(X_1 = x_0, X_2 = x_0) \\ &\quad + \sum_{i=1}^{\infty} P(X_1 = x_0, X_2 < x_0, \cdots, X_{1+i} < x_0, X_{2+i} = x_0) \\ &= \frac{[P(X_1 = x_0)]^2}{P(X_1 \geq x_0)}. \end{aligned} \quad (14)$$

Upon combining Eqs. (13) and (14), we obtain

$$\frac{P(X_1 = x_0)}{P(X_1 \geq x_0)} = 1 + \frac{P(X_1 > x_0)}{P(X_1 = x_0)} \ln \frac{P(X_1 > x_0)}{P(X_1 \geq x_0)},$$

which yields

$$\ln a + 1 - a = 0, \quad \text{where} \quad a = \frac{P(X_1 > x_0)}{P(X_1 \geq x_0)} \in (0, 1).$$

Since $\ln a + 1 - a < 0$ for all $a \in (0, 1)$, we obtain the desired contradiction. □

Lemma 2.2 readily implies the following result.

Corollary 2.1. *Let $(W_n, n \geq 1)$ be weak records corresponding to a sequence of iid rv's with discrete cdf F having at least two points in its support. Then, for any $n = 2, 3, \cdots$,*

$$(W_1, W_2, \cdots, W_n) \overset{d}{\neq} (F^{\leftarrow}(U(1, n, \tilde{m}, 1)), \cdots, F^{\leftarrow}(U(1, n, \tilde{m}, 1))),$$

where $\tilde{m} = (-1, -1, \cdots, -1)$. Therefore, in contrast to the continuous case, weak records based on a discrete non-degenerate cdf can not be viewed as generalized order statistics with $k = 1$ and $\tilde{m} = (-1, -1, \cdots, -1)$.

The only case remaining concerns the behaviour of weak records based on a cdf F with only one atom x_0 such that $F(x_0) = 1$.

Lemma 2.3. *Let $(W_n, n \geq 1)$ be weak records corresponding to a sequence of iid rv's $X_1, X_2, \cdots$ with cdf F having only one atom x_0 such that $F(x_0) = 1$, and let $(WU_n, n \geq 1)$ denote weak records from uniform distribution on $[0, 1]$. Then, for any $n = 1, 2, 3, \cdots$,*

$$(W_1, W_2, \cdots, W_n) \overset{d}{=} (F^{\leftarrow}(WU_1), F^{\leftarrow}(WU_2), \cdots, F^{\leftarrow}(WU_n)). \quad (15)$$

Proof. It is well-known that for any cdf F

$$y \leq F(x) \iff F^{\leftarrow}(y) \leq x \quad (16)$$

and

$$(X_1, \cdots, X_n) \overset{d}{=} (F^{\leftarrow}(U_1), \cdots, F^{\leftarrow}(U_n)), \quad (17)$$

where $U_1, U_2, \cdots$ is a sequence of iid rv's uniformly distributed on $[0, 1]$; see, for example, David and Nagaraja (2003, p. 15). Moreover, for a cdf having only one atom x_0 at the right endpoint of its support, we have

$$F^{\leftarrow}(y) = x_0 \;\Leftrightarrow\; F(x_0^-) \leq y \leq 1 \tag{18}$$

and

$$y_2 < F(x_0^-) \;\Rightarrow\; [F^{\leftarrow}(y_1) < F^{\leftarrow}(y_2) \;\Leftrightarrow\; y_1 < y_2]. \tag{19}$$

In order not to complicate the notation, we will prove (15) only for $n = 3$. From (17) and (18), we see that

$$\begin{aligned} & P(W_1 = W_2 = W_3 = x_0) \\ &\quad = P(X_1 = x_0) = P(F^{\leftarrow}(U_1) = x_0) = P(F^{\leftarrow}(WU_1) = x_0) \\ &\quad = P(F^{\leftarrow}(WU_1) = F^{\leftarrow}(WU_2) = F^{\leftarrow}(WU_3) = x_0). \end{aligned}$$

For $x_1 < x_0$, Eqs. (16) – (19) yield

$$\begin{aligned} & P(W_1 \leq x_1, W_2 = W_3 = x_0) \\ &\quad = P(W_1 \leq x_1, W_2 = x_0) \\ &\quad = P(X_1 \leq x_1, X_2 = x_0) \\ &\qquad + \sum_{j=3}^{\infty} P(X_1 \leq x_1, X_2 < X_1, \cdots, X_{j-1} < X_1, X_j = x_0) \\ &\quad = P(F^{\leftarrow}(U_1) \leq x_1, F^{\leftarrow}(U_2) = x_0) \\ &\qquad + \sum_{j=3}^{\infty} P(F^{\leftarrow}(U_1) \leq x_1, F^{\leftarrow}(U_2) < F^{\leftarrow}(U_1), \cdots, \\ &\qquad\qquad F^{\leftarrow}(U_{j-1}) < F^{\leftarrow}(U_1), F^{\leftarrow}(U_j) = x_0) \\ &\quad = P(U_1 \leq F(x_1), F(x_0^-) \leq U_2 \leq 1) \\ &\qquad + \sum_{j=3}^{\infty} P(U_1 \leq F(x_1), U_2 < U_1, \cdots, U_{j-1} < U_1, \\ &\qquad\qquad F(x_0^-) \leq U_j \leq 1) \\ &\quad = P(WU_1 \leq F(x_1), F(x_0^-) \leq WU_2 \leq 1) \\ &\quad = P(F^{\leftarrow}(WU_1) \leq x_1, F^{\leftarrow}(WU_2) = x_0) \\ &\quad = P(F^{\leftarrow}(WU_1) \leq x_1, F^{\leftarrow}(WU_2) = F^{\leftarrow}(WU_3) = x_0). \end{aligned}$$

Similarly, for $x_1 < x_2 < x_0$, we get

$$
\begin{aligned}
&P(W_1 \le x_1, W_2 \le x_2, W_3 = x_0)\\
&= P(X_1 \le x_1, X_1 \le X_2 \le x_2, X_3 = x_0)\\
&+ \sum_{j=4}^{\infty} P(X_1 \le x_1, X_1 \le X_2 \le x_2, X_3 < X_2, \cdots, X_{j-1} < X_2,\\
&\qquad X_j = x_0)\\
&+ \sum_{j=3}^{\infty} P(X_1 \le x_1, X_2 < X_1, \cdots, X_{j-1} < X_1, X_1 \le X_j \le x_2,\\
&\qquad X_{j+1} = x_0)\\
&+ \sum_{j_1=3}^{\infty} \sum_{j_2=j_1+2}^{\infty} P(X_1 \le x_1, X_2 < X_1, \cdots, X_{j_1-1} < X_1, X_1 \le X_{j_1} \le x_2,\\
&\qquad X_{j_1+1} < X_{j_1}, \cdots, X_{j_2-1} < X_{j_1}, X_{j_2} = x_0)\\
&= P(F^{\leftarrow}(U_1) \le x_1, F^{\leftarrow}(U_1) \le F^{\leftarrow}(U_2) \le x_2, F^{\leftarrow}(U_3) = x_0)\\
&+ \sum_{j=4}^{\infty} P(F^{\leftarrow}(U_1) \le x_1, F^{\leftarrow}(U_1) \le F^{\leftarrow}(U_2) \le x_2, F^{\leftarrow}(U_3) < F^{\leftarrow}(U_2),\\
&\qquad \cdots, F^{\leftarrow}(U_{j-1}) < F^{\leftarrow}(U_2), F^{\leftarrow}(U_j) = x_0)\\
&+ \sum_{j=3}^{\infty} P(F^{\leftarrow}(U_1) \le x_1, F^{\leftarrow}(U_2) < F^{\leftarrow}(U_1), \cdots, F^{\leftarrow}(U_{j-1}) < F^{\leftarrow}(U_1),\\
&\qquad F^{\leftarrow}(U_1) \le F^{\leftarrow}(U_j) \le x_2, F^{\leftarrow}(U_{j+1}) = x_0)\\
&+ \sum_{j_1=3}^{\infty} \sum_{j_2=j_1+2}^{\infty} P(F^{\leftarrow}(U_1) \le x_1, F^{\leftarrow}(U_2) < F^{\leftarrow}(U_1), \cdots,\\
&\qquad F^{\leftarrow}(U_{j_1-1}) < F^{\leftarrow}(U_1), F^{\leftarrow}(U_1) \le F^{\leftarrow}(U_{j_1}) \le x_2,\\
&\qquad F^{\leftarrow}(U_{j_1+1}) < F^{\leftarrow}(U_{j_1}), \cdots, F^{\leftarrow}(U_{j_2-1}) < F^{\leftarrow}(U_{j_1}),\\
&\qquad F^{\leftarrow}(U_{j_2}) = x_0)\\
&= P(U_1 \le F(x_1), U_1 \le U_2 \le F(x_2), F(x_0^-) \le U_3 \le 1)\\
&+ \sum_{j=4}^{\infty} P(U_1 \le F(x_1), U_1 \le U_2 \le F(x_2), U_3 < U_2, \cdots,\\
&\qquad U_{j-1} < U_2, F(x_0^-) \le U_j \le 1)\\
&+ \sum_{j=3}^{\infty} P(U_1 \le F(x_1), U_2 < U_1, \cdots, U_{j-1} < U_1, U_1 \le U_j \le F(x_2),\\
&\qquad F(x_0^-) \le U_{j+1} \le 1)\\
&+ \sum_{j_1=3}^{\infty} \sum_{j_2=j_1+2}^{\infty} P(U_1 \le F(x_1), U_2 < U_1, \cdots, U_{j_1-1} < U_1, U_1 \le U_{j_1} \le F(x_2),\\
&\qquad U_{j_1+1} < U_{j_1}, \cdots, U_{j_2-1} < U_{j_1}, F(x_0^-) \le U_{j_2} \le 1)\\
&= P(WU_1 \le F(x_1), WU_2 \le F(x_2), F(x_0^-) \le WU_3 \le 1)\\
&= P(F^{\leftarrow}(WU_1) \le x_1, F^{\leftarrow}(WU_2) \le x_2, F^{\leftarrow}(WU_3) = x_0).
\end{aligned}
$$

In a similar manner, we conclude that for $x_1 < x_2 < x_3 < x_0$

$$P(W_1 \le x_1, W_2 \le x_2, W_3 \le x_3) = \\ P(F^{\leftarrow}(WU_1) \le x_1, F^{\leftarrow}(WU_2) \le x_2, F^{\leftarrow}(WU_3) \le x_3),$$

and the proof for $n = 3$ is thus complete. Now, for an arbitrary n, representation (15) can be established following exactly the same steps. $\square$

Combining all these, we have proved the following theorem.

Theorem 2.2. *The model of generalized order statistics based on a cdf F for $k = 1$ and $\tilde{m} = (-1, -1, \cdots, -1)$ reduces in the distributional sense to the model of weak records related to F iff F is degenerate or continuous or has only one atom at the right endpoint of its support.*

Remark 2.1. *Note that a distribution that has only one atom at the right endpoint of its support is a special mixture of a continuous distribution and a degenerate distribution.*

3 Progressively Type-II right censored order statistics

It is well-known that order statistics $X_{1:n} \le \cdots \le X_{n:n}$ arising from iid rv's $X_1, \cdots, X_n$ from an arbitrary cdf F admit the representation

$$(X_{1:n}, \cdots, X_{n:n}) \stackrel{d}{=} (F^{\leftarrow}(U_{1:n}), \cdots, F^{\leftarrow}(U_{n:n})),$$

where $U_{1:n} \le \cdots \le U_{n:n}$ are order statistics from a random sample from the uniform distribution on $[0, 1]$; see, for example, Nagaraja (1992) and David and Nagaraja (2003, p. 15). Balakrishnan and Dembińska (2007) recently established that a similar result holds for progressively Type-II right censored order statistics. More precisely, if $X_{1:m:n}^{(R_1,\cdots,R_m)} \le \cdots \le X_{m:m:n}^{(R_1,\cdots,R_m)}$ are progressively Type-II right censored order statistics from an arbitrary cdf F and $U_{1:m:n}^{(R_1,\cdots,R_m)} \le \cdots \le U_{m:m:n}^{(R_1,\cdots,R_m)}$ denote progressively Type-II right censored order statistics from the uniform distribution on $[0, 1]$, then

$$(X_{1:m:n}^{(R_1,\cdots,R_m)}, \cdots, X_{m:m:n}^{(R_1,\cdots,R_m)}) \stackrel{d}{=} \\ (F^{\leftarrow}(U_{1:m:n}^{(R_1,\cdots,R_m)}), \cdots, F^{\leftarrow}(U_{m:m:n}^{(R_1,\cdots,R_m)})).$$

It then follows that, for an arbitrary cdf F, both order statistics and progressively Type-II right censored order statistics are special cases (in the distributional sense) of generalized order statistics. In fact, the model of order statistics is a submodel of generalized order statistics corresponding to $\tilde{m} = (0, 0, \cdots, 0)$ and $k = 1$, while the model of progressively Type-II right censored order statistics corresponds to generalized order statistics with $n = m$, $\tilde{m} = (R_1, R_2, \cdots, R_{m-1})$ and $k = R_m + 1$. Consequently, any property that holds for generalized order statistics is also valid for these two submodels. In the following two subsections, we will apply the theory of generalized order statistics to study dependence structure of progressively Type-II right censored order statistics from an arbitrary cdf and to obtain some characterizations of the geometric and modified geometric type distributions.

3.1 Dependence structure

In contrast to the continuous case, order statistics from a discrete cdf with at least three points in its support do not form a Markov chain; see Nagaraja (1982) and Arnold, Balakrishnan and Nagaraja (1992, Section 3.4). Balakrishnan and Dembińska (2007) recently showed that progressively Type-II right censored order statistics also similarly do not possess Markov property.

Theorem 3.1.1. [Balakrishnan and Dembińska, 2007] *Let F be a discrete cdf with at least three points in its support. Then, for an arbitrary censoring scheme $(R_1, \cdots, R_m)$, the progressively Type-II right censored order statistics $X_{1:m:n}^{(R_1,\cdots,R_m)}, \cdots, X_{m:m:n}^{(R_1,\cdots,R_m)}$ arising from F do not form a Markov chain.*

Even though order statistics from cdf with at least three points in its support do not exhibit the Markov property, they constitute a component of a bivariate Markov chain. Rüschendorf (1985) showed that for a discrete underlying distribution, the sequence $(X_{i:n}, V_i)$ forms a Markov chain, where V_i is the number of $X_{k:n}$'s for $k \leq i$ that are tied with $X_{i:n}$. It is of interest to examine whether an analogous property holds for discrete progressively Type-II right censored order statistics. To answer this question, we will need the following result of Tran (2006) concerning generalized order statistics.

Theorem 3.1.2. [Tran, 2006] *Let* $X(1, n, \tilde{m}, k), \cdots, X(n, n, \tilde{m}, k)$

be generalized order statistics based on an arbitrary cdf F, and let

$$V_i = \#\{j \leq i : X(j,n,\tilde{m},k) = X(i,n,\tilde{m},k)\},\ i = 1,2,\cdots n.$$

Then, the bivariate sequence $(X(i,n,\tilde{m},k), V_i)$ forms a Markov chain.

From Theorem 3.1.2, we readily obtain the following result.

Theorem 3.1.3. *Let $X_{1:m:n}^{(R_1,\cdots,R_m)}, \cdots, X_{m:m:n}^{(R_1,\cdots,R_m)}$ be progressively Type-II right censored order statistics arising from an arbitrary cdf, and let*

$$V_i = \#\{j \leq i : X_{j:m:n}^{(R_1,\cdots,R_m)} = X_{i:m:n}^{(R_1,\cdots,R_m)}\}, \qquad i = 1,2,\cdots,m.$$

Then, the bivariate sequence $(X_{i:m:n}^{(R_1,\cdots,R_m)}, V_i)$ is a Markov chain.

Proof. Since progressively Type-II right censored order statistics from an arbitrary cdf are a special case of generalized order statistics in the distributional sense, the assertion follows directly from Theorem 3.1.2. □

3.2 Characterizations of the geometric distribution

There are many characterizations of the geometric distribution based on order statistics; see Nagaraja (1992) and Rao and Shanbhag (1998). Tran (2006) extended some of these characterization results via properties of generalized order statistics. She proved, in particular, that in the geometric case the rv's $X(1,n,\tilde{m},k)$ and $X(i,n,\tilde{m},k)-X(1,n,\tilde{m},k)$ are independent for any $2 \leq i \leq n$. Moreover, to obtain a characterization of the geometric distribution, the complete independence between $X(1,n,\tilde{m},k)$ and $X(i,n,\tilde{m},k) - X(1,n,\tilde{m},k)$ does not seem to be necessary as shown in the following theorem.

Theorem 3.2.1. [Tran, 2006] *Let $X(1,n,\tilde{m},k), \cdots, X(n,n,\tilde{m},k)$ be generalized order statistics based on a discrete cdf F with support $\{0,1,\cdots\}$. Then, the rv $X(1,n,\tilde{m},k)$ and the event $\{X(i,n,\tilde{m},k) = X(1,n,\tilde{m},k)\}$ are independent for some positive integer $2 \leq i \leq n$ iff F is a geometric cdf with $F(j) = 1-(1-p)^j$, $j = 0,1,\cdots$, $p \in (0,1)$.*

Theorem 3.2.1 will not hold if we replace $X(1,n,\tilde{m},k)$ by $X(j,n,\tilde{m},k)$, $2 \leq j < i \leq n$. In the case of order statistics, Deheuvels (1984) showed that the only discrete distributions for which $X_{j:n}$ and $X_{i:n} - X_{j:n}$ are independent for some $2 \leq j < i \leq n$ are degenerate cdf's. But among discrete cdf's, the conditional independence of $X_{j:n}$ and

$\{X_{i:n} = X_{j:n}\}$ for some $2 \leq j < i \leq n$, given the event $\{X_{j:n} > X_{j-1:n}\}$, implies that F is a modified geometric type cdf; see Nagaraja and Srivastava (1987). Tran (2006) proved that this result also holds more generally for generalized order statistics.

Theorem 3.2.2. [Tran, 2006] *Let $X(1, n, \tilde{m}, k), \cdots, X(n, n, \tilde{m}, k)$ be generalized order statistics based on a discrete cdf F with at least three values in its support S. Then, if for some $2 \leq j < i \leq n$, the rv $X(j, n, \tilde{m}, k)$ and the event $\{X(i, n, \tilde{m}, k) = X(j, n, \tilde{m}, k)\}$ are conditionally independent, given the event $\{X(j, n, \tilde{m}, k) > X(j-1, n, \tilde{m}, k)\}$, the underlying distribution is of modified geometric type; that is, $S = \{a_0, a_1, \cdots,$ where $a_i < a_j$ for $i < j\}$, $F(a_0) = \theta$ and $F(a_j) = 1 - (1-\theta)(1-p)^j$, $j = 1, 2, \cdots$, for some $\theta \in (0, 1)$ and $p \in (0, 1)$.*

Since the model of progressively Type-II right censored order statistics is a submodel of generalized order statistics in the distributional sense, Theorems 3.2.1 and 3.2.2 can readily be stated in terms of progressively Type-II right censored order statistics as follows.

Theorem 3.2.3. *Let $X_{1:m:n}^{(R_1,\cdots,R_m)}, \cdots, X_{m:m:n}^{(R_1,\cdots,R_m)}$ be progressively Type-II right censored order statistics from a discrete cdf F with support S.*

1. *If $S = \{0, 1, \cdots\}$, then the rv $X_{1:m:n}^{(R_1,\cdots,R_m)}$ and the event $\{X_{1:m:n}^{(R_1,\cdots,R_m)} = X_{i:m:n}^{(R_1,\cdots,R_m)}\}$ are independent for some $2 \leq i \leq m$ iff F is a geometric cdf;*

2. *If S has at least three elements and for some $2 \leq j < i \leq n$, the rv $X_{j:m:n}^{(R_1,\cdots,R_m)}$ and the event $\{X_{j:m:n}^{(R_1,\cdots,R_m)} = X_{i:m:n}^{(R_1,\cdots,R_m)}\}$ are conditionally independent, given the event $\{X_{j:m:n}^{(R_1,\cdots,R_m)} > X_{j-1:m:n}^{(R_1,\cdots,R_m)}\}$, then F is a modified geometric type cdf.*

Other characterizations of the geometric and modified geometric type distributions established by Tran (2006), based on identical distribution, independence and conditional expectations of some functions of generalized order statistics, can all be stated in terms of progressively Type-II right censored order statistics as well.

References

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992), A First Course in Order Statistics. New York: John Wiley &

Sons.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998), Records. New York: John Wiley & Sons.

Balakrishnan, N. (2007), Progressive censoring methodology: An appraisal (with discussion). Test, **16**, 211–259.

Balakrishnan, N. and Aggarwala, R. (2000), Progressive Censoring: Theory, Methods, and Applications. Boston: Birkhäuser.

Balakrishnan, N. and Cohen, A. C. (1991), Order Statistics and Inference: Estimation Methods. Boston: Academic Press.

Balakrishnan, N. and Dembińska, A. (2007), Progressively Type-II right censored order statistics from discrete distributions. Journal of Statistical Planning and Inference, doi: 10.1016/j.jspi.2007.02.006.

Balakrishnan, N. and Rao, C. R. (Eds.) (1998a), Handbook of Statistics. Vol. 16 – Order Statistics: Theory and Methods, Amsterdam: North-Holland.

Balakrishnan, N. and Rao, C. R. (Eds.) (1998b), Handbook of Statistics. Vol. 17 – Order Statistics: Applications, Amsterdam: North-Holland.

Cohen, A. C. (1991), Truncated and Censored Samples: Theory and Applications. New York: Marcel Dekker.

Cramer, E. (2003), Contributions to Generalized Order Statistics. Habilitationsschrift, University of Oldenburg, Germany.

Cramer, E., Kamps, U. (2003), Marginal distributions of sequential and generalized order statistics. Metrika, **38**, 293–310.

David, H. A. and Nagaraja, H. N. (2003), Order Statistics. Third edition, Hoboken, New Jersey: John Wiley & Sons.

Deheuvels, P. (1984), The characterization of distributions by order statistics and record values – A unified approach. Journal of Applied Probability, **21**, 326–334. Correction: **22**, 997.

Dembińska, A. (2007a), A review of characterizations of discrete distributions based on records and kth records. Communications in Statistics – Theory and Methods, **36**, 1381–1387.

Dembińska, A. (2007b), Records from discrete distributions. In Ordered Random Variables: Theory and Applications, Eds., Ahsanullah, M. and Raqab, M. Z., to appear.

Kamps, U. (1995a), A concept of generalized order statistics. Journal of Statistical Planning and Inference, **48**, 1–23.

Kamps, U. (1995b), A Concept of Generalized Order Statistics. Teubner, Stuttgart, Germany.

Kamps, U. and Cramer, E. (2001), On distributions of generalized order statistics. Statistics, **35**, 269–280.

Keseling, C. (1999), Conditional distributions of generalized order statistics and some characterizations. Metrika, **49**, 27-40.

Nagaraja, H. N. (1982), On the non-Markovian structure of discrete order statistics. Journal of Statistical Planning and Inference, **7**, 29–33.

Nagaraja, H. N. (1992), Order statistics from discrete distributions. Statistics, **23**, 189–216.

Nagaraja, H. N. and Srivastava, R. C. (1987), Some characterizations of geometric type distributions based on order statistics. Journal of Statistical Planning and Inference, **17**, 181–191.

Nevzorov, V. B. (2001), Records: Mathematical Theory, Translations of Mathematical Monographs. Vol. 194, American Mathematical Society, Providence, Rhode Island.

Rao, C. R., Shanbhag, D. N. (1998), Recent approaches to characterizations based on order statistics and record values. In Handbook of Statistics, **16**: Theory and Applications, Eds., Balakrishnan, N. and Rao,C. R., 231–256, North-Holland, Amsterdam.

Rüschendorf, L. (1985), Two remarks on order statistics. Journal of Statistical Planning and Inference, **11**, 71–74.

Tran, T. T. H. (2006), Discrete Generalized Order Statistics. Ph.D. Thesis, University of Oldenburg, Germany.