

Characterizations of Geometric and Discrete Pareto Distributions Based on the Conditional Distribution of k th Records

Sevgi Yurt Oncel*

*Department of Statistics, Faculty of Arts and Sciences, Kırıkkale University,
Kırıkkale, Turkey,
syoncel@gmail.com*

Fazil Aliev

*College of Humanities and Sciences, Virginia Commonwealth University,
Richmond, VA, USA
Department of Actuarial and Risk Management, Faculty of Business, Karabuk University,
Karabuk, Turkey
faliev@vcu.edu*

Received 5 November 2015

Accepted 17 September 2016

Abstract

The problem of characterizing of discrete probability distributions is an important problem. Recently many new results are obtained in characterization of distributions using k th records. Based on the distributional properties of k th weak and ordinary records some characterizations of geometric and discrete pareto distributions are given.

Keywords: Characterization; Conditional Distribution; Discrete Pareto Distributions; Geometric Distributions; k th Records.

2000 Mathematics Subject Classification: 62E10, 62G99

1. Introduction

Let X_1, X_2, \dots be a sequence of independent and identically distributed (i.i.d) random variables taking values $0, 1, \dots$ with probability function $p_j = P(X_1 = j)$ and $q_j = P(X_1 \geq j), j \geq 0$. Denote by $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ the order statistics of random sample (X_1, X_2, \dots, X_n) . First let us define weak k th record times as $U_0^{(k)} = k, U_{n+1}^{(k)} = \min \left\{ j > U_n^{(k)} : X_j \geq X_{U_n^{(k)} - k + 1: U_n^{(k)}} \right\}, n = 0, 1, \dots$ and then the n -th weak k th record as $W_n^{(k)} = X_{U_n^{(k)} - k + 1: U_n^{(k)}}, n = 0, 1, \dots$

Note that for $k=1$ the above definition of k th weak records amounts to the definition of weak records. Moreover, if in the definition of k th weak record times we replace the inequality " \leq " with the sharp

* Corresponding author

inequality “<” we will get the definition of k th (ordinary) records $(R_n^{(k)}, n \geq 0)$ introduced by Dziubdziela and Kopocinski (1976).

To deal with k th weak records it is convenient to introduce random variables $\xi_0^{(k)}, \xi_1^{(k)}, \dots$ as follows:

$$\xi_m^{(k)} = i \text{ if there are exactly } i \text{ } k\text{th weak records that are equal to } m.$$

and $\{W_n^{(k)} > m\} = \{\xi_0^{(k)} + \xi_1^{(k)} + \dots + \xi_m^{(k)} \leq n\}$, for $m \geq 0, n \geq 0, k \geq 1$.

The aim of a characterization study is the defining of properties which characterize a specific probability distribution. So, we may summarize distribution properties under some conditions. In literature, many characterization studies are based on k th record values and most of them are based on conditional expectation.

The properties and characterizations for k th record values from continuous and discrete distribution have been widely studied in books as Ahsanullah (1995). Many interesting characterizations for record values can be found also in , refer to, Dembinska and Lopez-Blazquez (2005a, 2005b), Dembinska, (2007, 2008), Lopez-Blazquez et al.(2005), Ahsanullah and Aliev (2011) and Hashorva and Stepanov (2012).

Characterization studies based on k th record values of geometric distribution introduced by Danielak and Dembinska (2007), Dembinska and Lopez-Blazquez (2005a), Dembinska (2008).

We will use the following proposition from Dembinska & Blazquez (2005):

Proposition 1. For any $l > m \geq 0$,

$$P(W_1^{(k)} \geq l | W_0^{(k)} = m) = \frac{k p_m q_l^k}{q_m (q_m^k - q_{m+1}^k)}.$$

2. Main Results

In this section we present characterizations of geometric and discrete Pareto distributions via conditional survival function of k th weak and ordinary record values. We refer the interested readers to the reference Ahsanullah and Hamedani (2012) for characterizations of continuous distributions via conditional survival function of generalized order statistics.

Theorem 1. Let X be a discrete random variable with *cdf* F and g be a non-negative function on $[0, \infty)$. Then

$$P(W_1^{(k)} \geq l | W_0^{(k)} = m) = g(l - m), \quad l \geq m$$

if and only if X has a Geometric distribution with $F(x) = 1 - q^x, x = 0, 1, \dots$.

Proof: First we prove the *Necessity* part: Assume that

$$P(W_1^{(k)} \geq l | W_0^{(k)} = m) = g(l - m), \quad l \geq m.$$

For all $l > m \geq 0$ by Proposition 1 we get

$$\frac{kp_m q_l^k}{q_m (q_m^k - q_{m+1}^k)} = g(l - m) \tag{1}$$

or with $m = 0$,

$$\frac{kp_0 q_l^k}{(1 - q_1^k)} = g(l).$$

By changing l to $l - m$ we obtain

$$g(l - m) = \frac{kp_0 q_{l-m}^k}{(1 - q_1^k)}.$$

Now combining with (1) and taking into account that $p_m = q_m - q_{m+1}$

$$\frac{q_l^k}{q_{l-m}^k} = \frac{p_0}{(1 - q_1^k)} \frac{q_m (q_m^k - q_{m+1}^k)}{(q_m - q_{m+1})}$$

Let's put $t = \frac{q_{m+1}}{q_m}$, $\lambda = \frac{p_0}{(1 - q_1^k)}$ and $l = m + 1$ in the last expression and rewrite it as $t^k = \lambda q_1^k \frac{1 - t^k}{1 - t}$ which is same (for $t \neq 1$) as

$$t^{k+1} - (1 + \lambda q_1^k) t^k + \lambda q_1^k = 0. \tag{2}$$

The left side of (2) is a polynomial function and is positive when $t = 0$ and is zero at $t = 1$. Equating the derivative of (5) to 0 in $0 < t < 1$,

$$(k + 1)t^k - k(1 + \lambda q_1^k)t^{k-1} = 0, \quad t^{k-1} [(k + 1)t - k(1 + \lambda q_1^k)] = 0.$$

The only positive root of the last equation is $t_0 = \frac{k(1 + \lambda q_1^k)}{(k + 1)}$, so the polynomial on the left side of (2) is strictly decreasing on the left side of t_0 and strictly increasing on the right side of t_0 . It means that the polynomial can cross x axis at most twice on the interval $t > 0$ and only once on $0 < t < t_0$. Using the inequality of arithmetic and geometric means $a_1 + a_2 + \dots + a_k > k \sqrt[k]{a_1 \cdot a_2 \cdot \dots \cdot a_n}$ ($a_i \geq 0$) we can write

$$1 + q_1 + \dots + q_1^{k-1} > k \sqrt[k]{1 \cdot q_1 \cdot \dots \cdot q_1^{k-1}} = k \sqrt[k]{q_1^{\frac{(k-1)k}{2}}} = k q_1^{\frac{(k-1)}{2}}.$$

Then

$$t_0 = \frac{k}{(k + 1)} \left(1 + \frac{1 - q_1}{1 - q_1^k} q_1^k \right) = \frac{k}{(k + 1)} \frac{1 - q_1^k + q_1^k - q_1^{k+1}}{1 - q_1^k} =$$

$$\begin{aligned}
 &= \frac{k}{(k+1)} \frac{1 - q_1^{k+1}}{1 - q_1^k} = \frac{k}{(k+1)} \frac{(1 - q_1)(1 + q_1 + \dots + q_1^k)}{(1 - q_1)(1 + q_1 + \dots + q_1^{k-1})} = \\
 &= \frac{k}{(k+1)} \left(1 + \frac{q_1^k}{1 + q_1 + \dots + q_1^{k-1}} \right) = \frac{k}{(k+1)} \left(1 + \frac{q_1^k}{k q_1^{\frac{k-1}{2}}} \right) = \frac{k}{(k+1)} \left(1 + \frac{q_1^{\frac{k+1}{2}}}{k} \right) < \\
 &< \frac{k}{(k+1)} \left(1 + \frac{1}{k} \right) = 1.
 \end{aligned}$$

Now by simple checking we can show that the solution on $0 < t < t_0$ is $t = q_1$;
 for $t = q_1$ we can write

$$\begin{aligned}
 t^{k+1} - (1 + \lambda q_1^k) t^k + \lambda q_1^k &= q_1^{k+1} - \left(1 + \frac{1 - q_1}{1 - q_1^k} q_1^k \right) q_1^k + \frac{1 - q_1}{1 - q_1^k} q_1^k = \\
 &= q_1^{k+1} - q_1^k - \frac{1 - q_1}{1 - q_1^k} [q_1^{2k} - q_1^k] = q_1^k (q_1 - 1) + \frac{1 - q_1}{1 - q_1^k} q_1^k [1 - q_1^k] = 0.
 \end{aligned}$$

It is also trivial to check that $t = 1$ satisfies (2).

This proves that $u_m = q_{m+1} / q_m = q_1 < 1$ for all $m \geq 0$ which corresponds to geometric distributin with $p_j = p(1 - p)^j, j \geq 0, p > 0$. \square

To prove the *Sufficiency*: When the parent distribution is geometric $Ge(p)$, i.e., $p_j = pq^j, p > 0, q = 1 - p, j \geq 0$ and $q_j = q^j$, we obtain for any $m \geq 0$

$$\begin{aligned}
 P(W_1^{(k)} \geq l | W_0^{(k)} = m) &= \frac{kp_m q_1^k}{q_m (q_m^k - q_{m+1}^k)} = \frac{kp q^m q^{lk}}{q^m (q^{mk} - q^{(m+1)k})} = \frac{kp q^{(l-m)k}}{1 - q^k} = g(l - m). \\
 P(W_1^{(k)} \geq l | W_0^{(k)} = m) &= \frac{kp_m q_1^k}{q_m (q_m^k - q_{m+1}^k)} = \frac{kp q^m q^{lk}}{q^m (q^{mk} - q^{(m+1)k})} \\
 &= \frac{kp q^{lk}}{(q^{mk} - q^{mk} q^k)} = \frac{kp q^{lk}}{q^{mk} (1 - q^k)} \\
 &= \frac{kp}{(1 - q^k)} q^{lk} q^{-mk} = \frac{kp}{(1 - q^k)} q^{k(l-m)}. \quad \square
 \end{aligned}$$

Remark 1. By the similar way as in Theorem 2.1 we can prove the following:

Let X be a discrete random variable with *cdf* F and g be a non-negative function on $[0, \infty)$. Then

$$P(R_1^{(k)} \geq l | R_0^{(k)} = m) = g(l - m), \quad l \geq m$$

if and only if X has a Geometric distribution with $F(x) = 1 - q^x, x = 0, 1, \dots$

The proof is based on the next equality

$$P\left(R_1^{(k)} \geq l \mid R_0^{(k)} = m\right) = \frac{k p_m q_l^k}{q_{m+1} (q_m^k - q_{m+1}^k)}, \text{ for any } l > m \geq 0,$$

proof of which is similar to Proposition 1.

Remark 2. The equality $P\left(W_i^{(k)} \geq l \mid W_0^{(k)} = m\right) = g(l - m), l \geq m$ does not hold for geometric distribution for any $i > 1$.

Theorem 2. Let $k = 1, X$ be a discrete random variable having values on $1, 2, \dots$ with *cdf* F and g be a discrete non-negative function on $(0, \infty)$. Then

$$P\left(W_1^{(1)} \geq l \mid W_0^{(1)} = m\right) = g(l / m), l \geq m$$

if and only if X has a Pareto distribution with $F(x) = 1 - x^{-\theta}, x = 1, 2, \dots$

Proof: First we prove the *Necessity* part: For all $l > m \geq 1$ by Proposition 1 and the condition $P\left(W_1^{(1)} \geq l \mid W_0^{(1)} = m\right) = g(l / m)$ we get

$$\frac{q_l}{q_m} = g(l / m). \tag{3}$$

For $m = 1$

$$q_l = g(l). \tag{4}$$

Denote $t = l / m$ and assume that t is integer. By (3) one can write $\frac{q_{lm}}{q_m} = g(t)$, combining with (4) we have

$q_{lm} = q_l q_m, m \geq 1, l \geq m, q_m \geq 0, q_i$ is decreasing and $q_1 = 1$ and q_2 is positive. Assume that $q_2 = 2^{-\theta}$. Then $q_4 = q_2 q_2 = 2^{-2\theta}$, similarly $q_{2^x} = 2^{-x\theta}, q_{3^t} = (q_3)^t$. Take any prime number $z > 2$. We will show that $q_z = z^{-\theta}$. For any given integer s there exists an integer d such that $2^{d-1} < z^s \leq 2^d$. By taking logarithm $(d-1) / s < \log_2 z \leq d / s$. Taking a limit when $s \rightarrow \infty$ we get $\lim_{s \rightarrow \infty} [(d-1) / s] \leq \log_2 z \leq \lim_{s \rightarrow \infty} (d / s)$. Since left and right hand side limits are equal

$$\lim_{s \rightarrow \infty} (d / s) = \log_2 z. \tag{5}$$

Series $\{q_m\}_{m \geq 1}$ is decreasing and $q_{2^{d-1}} < q_{z^s} \leq q_{2^d}$. It means that $2^{-\theta(d-1)} < (q_z)^s \leq 2^{-\theta d}$ or, equivalently, $2^{-\theta \frac{(d-1)}{s}} < q_z \leq 2^{-\theta \frac{d}{s}}$. Taking limits on both sides and using (5) $2^{-\theta \log_2 z} < q_z \leq 2^{-\theta \log_2 z}$. Hence $q_z = 2^{\log_2 z^{-\theta}}$ or $q_z = z^{-\theta}$.

We showed that $q_z = z^{-\theta}$ for all $z > 2$, therefore $F(z) = 1 - z^{-\theta}, z > 1$. It proves the Theorem. \square

To prove the *Sufficiency*: When the parent distribution is discrete Pareto distribution we get $p_j = j^{-\theta} - (j+1)^{-\theta}$ and $q_j = j^{-\theta}$, $j = 1, 2, \dots$, we obtain for $m \geq 1$

$$P\left(W_1^{(k)} \geq l \mid W_0^{(k)} = m\right) = \frac{kp_m q_l^k}{q_m(q_m^k - q_{m+1}^k)} = (l/m)^{-\theta} = g(l/m). \quad \square$$

3. Concluding Remarks

Characterizations of a probability distribution plays an important role in statistic. A probability distribution can be characterized through various methods. This paper extends the result by Ahsanullah and Hamedani (2012) to the discrete case using k-records. There are many papers characterizing continuous distributions with order statistics and records. In the literature, most of work is devoted to the case where the underlying distribution is continuous, because the absence of ties makes possible a relatively straightforward analysis of continuous record sequences. For discrete case due to positive probability of ties among observations, the distribution theory is much more complicated.

Most of continuous kth records properties can be obtained as trivial extensions of corresponding properties for records. It is not the case for discrete kth records. So, the number of papers on discrete kth records is considerably smaller than on continuous kth record statistics. In the literature, most characterization theorems through kth records were about geometric distribution.

Acknowledgements

The authors are grateful to an anonymous referee for valuable comments and suggestions.

References

- [1] M. Ahsanullah, *Record Statistics*. (Nova Science Publishers, Commack, New York,1995).
- [2] M. Ahsanullah and F. Aliev, A characterization of geometric distribution based on weak records, *Stat Papers* **52** (2011) 651–655.
- [3] M. Ahsanullah and G. G. Hamedani, Characterizations of certain continuous univariate distributions based on the conditional distribution of generalized order statistics, *Pak. J. Statis.* **28**(2) (2012) 253-258.
- [4] K. Danielak, and A. Dembinska, Some characterizations of discrete distributions based on weak records. *Stat. Papers* **48**(3) (2007) 479-489.
- [5] A. Dembinska and F. Lopez-Blazquez, A characterization of geometric distribution through kth weak records, *Comm. Stat. Theort and Methods* **34**(12) (2005a). 2345-2351.
- [6] A. Dembinska and F. Lopez-Blazquez, kth records from discrete distributions, *Statistics and Probability Letters*, **71**(3) (2005b) 203-214.
- [7] A. Dembinska, Records from discrete distributions, in recent development in ordered random variables, in *Recent Development in Ordered Random Variables*, eds M. Ahsanullah and M. Raqab, (Nova Science Publishers, Inc. 2007), pp.77-95.
- [8] A. Dembinska, kth records from geometric distribution, *Statistics and Probability Letters* **78**(12) (2008) 1662–1670.
- [9] W. Dziubdziela and B. Kopocinski, Limiting properties of the kth record values, *Appl. Math.* **15** (1976). 187-190.
- [10] E. Hashorva. and A. Stepanov,. Limit theorems for the spacings of weak records, *Metrika* **75**(2) (2012) 163-180.
- [11] F.Lopez-Blazquez, B. S. Mino and A. Dembinska, A note on the distribution of kth records from discrete distributions, *Statistics and Probability Letters* **75**(4) (2005) 325–330.