Journal of Modern Applied Statistical Methods

Volume 8 | Issue 1 Article 16

5-1-2009

On The Expected Values of Distribution of the Sample Range of Order Statistics from the Geometric Distribution

Sinan Calik
Firat University, Turkey, scalik@firat.edu.tr

Cemil Colak
Inonu University, Turkey, cemilcolak@yahoo.com

Ayse Turan
Firat University, Turkey, ayseturan23@hotmail.com

Journal of Modern Applied Statistical Methods
May 2009, Vol. 8, No. 1, 194-199

1538 – 9472/09/$95.00

On The Expected Values of Distribution of the Sample Range of Order Statistics from the Geometric Distribution

Sinan Calik
Firat University, Turkey

Cemil Colak
Inonu University, Turkey

Ayse Turan
Firat University, Turkey

The expected values of the distribution of the sample range of order statistics from the geometric distribution are presented. For n up to 10, algebraic expressions for the expected values are obtained. Using the algebraic expressions, expected values based on the p and n values can be easily computed.

Key words: Order statistics, expected value, moment, sample range, geometric distribution.

Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a discrete distribution with a probability mass function (pmf) $f(x)$ $(x = 0,1,2,\ldots)$ and a cumulative distribution function $F(x)$. Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be the order statistics obtained from the above random sample by arranging the observations in increasing order of magnitude. When spacing is denoted as $W_{i,j:n} = X_{j:n} - X_{i:n}$, and $i = 1$ and $j = n$, that is, in the case of the sample range W_n, then $W_n = X_{n:n} - X_{1:n}$. Denote the expected values of distribution of the sample range $E(W_n)$ by $\mu_{W_n}^{(k)}$ $(n \geq 2)$. For convenience, denote $\mu_{W_n}^{(1)}$ simply by μ_{W_n}.

Order statistics from the geometric distribution have been studied by many authors, for example, see Abdel-Aty (1954) and Morgolin and Winokur (1967). In particular, characterizations of the geometric distribution using order statistics have received great attention; for example, see Uppuliri (1964), Ferguson (1965, 1967), Crawford (1966), Srivastava (1974), Galambos (1975), El-Neweihi and Govindarajulu (1979), and Govindarajulu (1980). Expressions for the first two single moments of order statistics have been obtained by Morgolin and Winokur (1967).

The calculation of the exact sampling distribution of ranges from a discrete population was obtained by Burr (1955). The distribution of the sample range from a discrete order statistics were given by Arnold, et al. (1992). Additional details on discrete order statistics can be found in the works of Khatri (1962), David (1981), Nagaraja (1992), and Balakrishnan and Rao (1998). In this study, for n up to 10, algebraic expressions for the expected values of the distribution of the sample range of order statistics from the geometric distribution are obtained.

Cemil Colak is Assistant Professor in the Department of Biostatistics. Email: cemilcolak@yahoo.com. Sinan Calik is Assistant Professor of Statistics. Email: scalik@firat.edu.tr. Ayse Turan is research assistant in the Department of Statistics. Email: ayseturan23@hotmail.com.

Methodology

Marginal Distribution of Order Statistics

If $F_{r:n}(x)\,(r = 1,2,\ldots,n)$ denotes the cumulative distribution function (cdf) of $X_{r:n}$, then the following results:

$$F_{r:n}(x) = P\{X_{r:n} \le x\}$$
$$= P\{\textit{at least } r \textit{ of } X_1, X_2, \ldots, X_n \textit{ are at most } x\}$$
$$= \sum_{i=r}^{n} P\{\textit{exactly } i \textit{ of } X_1, X_2, \ldots, X_n \textit{ are at most } x\}$$
$$= \sum_{i=r}^{n} \binom{n}{i} [F(x)]^i [1-F(x)]^{n-i}$$
$$= \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r}\, dt \tag{1}$$

for $-\infty < x < \infty$.

For a discrete population, the probability mass function (pmf) of $X_{r:n}$ may be obtained from (1) by differencing as

$$f_{r:n}(x) = F_{r:n}(\mathrm{x}) - F_{r:n}(x-1)$$
$$= \frac{n!}{(r-1)!(n-r)!} \int_{F(x-1)}^{F(x)} t^{r-1}(1-t)^{n-r}\, dt$$

(Arnold, et al., 1992; Balakrishnan, 1986).

Order Statistics from the Geometric Distribution

To explore the properties of the geometric distribution order statistics, begin by stating that X is a Geometric (p) random variable. Note that it's pmf is given by $f(x) = pq^{x-1}$, and it's cdf is $F(x) = 1 - q^x$, for $x = 1,2,\ldots,$. Consequently the cdf of the r^{th} order statistic is given by

$$F_{r:n}(x) = \sum_{i=r}^{n} \binom{n}{i} (1-q^x)^i (q^x)^{n-i},\ x = 1,2,\ldots,.$$

Joint Distribution of Order Statistics

The joint distribution of order statistics can be similarly derived. For example, the joint cumulative distribution function of $X_{i:n}$ and $X_{j:n}$ ($1 \le i \le j \le n$) can be shown to be

$$F_{i,j:n}(x_i, x_j) = F_{j:n}(x_j),\ \text{for } x_i \ge x_j.$$

For $x_i < x_j$,

$$F_{i,j:n}(x_i, x_j) =$$
$$= \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!} \{F(x_i)\}^r$$
$$\{F(x_j) - F(x_i)\}^{s-r} \{1 - F(x_j)\}^{n-s} \tag{2}$$

This expression holds for any arbitrary population whether continuous or discrete.

For discrete populations, the joint probability mass function of $X_{i:n}$ and $X_{j:n}$ $(1 \le i \le j \le n)$ may be obtained from (2) by differencing as:

$$f_{i,j:n}(x_i, x_j) = P(X_{i.n} = x_i,\ X_{j:n} = x_j)$$
$$= F_{i,j:n}(x_i, x_j) - F_{i,j:n}(x_i - 1, x_j) -$$
$$F_{i,j:n}(x_i, x_j - 1) + F_{i,j:n}(x_i - 1, x_j - 1)$$

Theorem 1. For $1 \le i_1 \le i_2 \le \ldots \le i_k \le n$, the joint pmf of $X_{i_1:n}, X_{i_2:n}, \ldots, X_{i_k:n}$ is given by

$$f_{i_1,i_2,\ldots,i_k:n}(x_{i_1:n}, x_{i_2:n}, \ldots, x_{i_k:n})$$
$$= [\mathrm{C}(i_1, i_2, \ldots, i_k : n) \times$$
$$\int_D \left\{ \prod_{r=1}^{k} (u_{i_r} - u_{i_{r-1}})^{i_r - i_{r-1} - 1} \right\} (1 - u_{i_k})^{n-i_k}\, du_{i_1} \ldots du_{i_k}],$$

where $i_0 = 0,\ u_0 = 0,$

$$\mathrm{C}(i_1, i_2, \ldots, i_k : n) =$$
$$\frac{n!}{\left\{ (n-i_k)! \prod_{r=1}^{k} (i_r - i_{r-1} - 1)! \right\}},$$

and D is k-dimensional space given by

$$D = \begin{Bmatrix} (u_{i_1}, \ldots, u_{i_k}) : u_{i_1} \le u_{i_2} \le \ldots \le u_{i_k}, \\ F(x_r - 1) \le u_r \le F(x_r), \\ r = i_1, i_2, \ldots, i_k \end{Bmatrix}$$

(Nagaraja, 1986; Arnold, et al., 1992; Balakrishnan & Rao, 1988). Khatri (1962) presented this result for $k \le 3$, but only proved it for $k \le 2$ for the case of no ties.

Distribution of the Sample Range

Starting with the *pmf* of the spacing $W_{i,j:n} = X_{j:n} - X_{i:n}$, and using Theorem 1, results in

$$P(W_{i,j:n} = w) = \sum_{x \in D} \int_{F(x-1)}^{F(x)} \int_{\substack{F(x+w-1) \\ u_i < u_j}}^{F(x+w)} u_i^{i-1}(u_j - u_i)^{j-i-1}(1-u_j)^{n-j}\, du_j du_i \tag{3}$$

Substantial simplification of the expression in (3) is possible when $i = 1$ and $j = n$, that is, in the case of the sample range W_n, this results in:

$$P(W_n = w) = C(1, n:n) \sum_{x \in D} \int_{F(x-1)}^{F(x)} \int_{F(x+w-1)}^{F(x+w)} (u_n - u_1)^{n-2}\, du_n du_1$$

Thus, the *pmf* of W_n is given by

$$\begin{aligned} P(W_n = 0) &= n(n-1) \sum_{x \in D} \int_{F(x-1)}^{F(x)} \int_{F(x-1)}^{F(x)} (u_n - u_1)^{n-2}\, du_n du_1 \\ &= \sum_{x \in D} \{F(x) - F(x-1)\}^n \\ &= \sum_{x \in D} \{f(x)\}^n \end{aligned} \tag{4}$$

and, for $w > 0$, from Arnold, et al., 1992,

$$\begin{aligned} P(W_n = w) &= n(n-1) \sum_{x \in D} \int_{F(x-1)}^{F(x)} \int_{F(x+w-1)}^{F(x+w)} (u_n - u_1)^{n-2}\, du_n du_1 \\ &= \sum_{x \in D} \begin{Bmatrix} [F(x+w) - F(x-1)]^n \\ -[F(x+w) - F(x)]^n \\ -[F(x+w-1) - F(x-1)]^n \\ +[F(x+w-1) - F(x)]^n \end{Bmatrix} \end{aligned} \tag{5}$$

Expressions (4) and (5) can also be obtained without using the integral expression from Theorem 1, and a multinomial argument can also be used to obtain an alternative expression for the *pmf* of W_n.

Expected Values of the Sample Range

The m^{th} moments of W_n can be written as

$$\mu_{w_n}^{(m)} = E(W_n^m) = \sum_{w=0}^{\infty} w^m P(W_n = w), \tag{6}$$

where $P(W_n = w)$ is as given in (5).

When X is a geometric (p) random variable, as in the case of the expected values of the sample range, (6) yields

$$\begin{aligned} \mu_{w_n}^{(1)} &= E(W_n) \\ &= \sum_{w=0}^{\infty} wP(W_n = w) \\ &= \sum_{w=1}^{\infty} wP(W_n = w), \end{aligned} \tag{7}$$

where $P(W_n = w)$ is as given in (7).

Distribution of the Sample Range from the Geometric Distribution

The distribution of higher order statistics is not as simple for the geometric distribution. For the sample range W_n, from (2),

$$P(W_n = 0) = \sum_{x=1}^{\infty}\left(pq^{x-1}\right)^n = p^n \sum_{x=1}^{\infty}\left(q^n\right)^{x-1} = p^n \Big/ \left(1-q^n\right)$$

and from (3) the following is obtained:

$$P(W_n = w) = \sum_{x=1}^{\infty}\left\{\begin{array}{l}\left(q^{x-1}\left(1-q^{w+1}\right)\right)^n + \left(q^x\left(1-q^{w-1}\right)\right)^n \\ -\left(q^x\left(1-q^w\right)\right)^n - \left(q^{x-1}\left(1-q^w\right)\right)^n\end{array}\right\} = \frac{1}{1-q^n}\left\{\begin{array}{l}\left(1-q^{w+1}\right)^n - \left(1-q^w\right)^n - \\ q^n\left[\left(1-q^w\right)^n - \left(1-q^{w-1}\right)^n\right]\end{array}\right\}, \quad (7)$$

for $w > 0$.

In particular,

$$P(W_2 = w) = \frac{1}{1-q^2}\left\{\begin{array}{l}\left(1-q^{w+1}\right)^2 - \left(1-q^w\right)^2 - \\ q^2\left[\left(1-q^w\right)^2 - \left(1-q^{w-1}\right)^2\right]\end{array}\right\}$$

for $w > 0$, thus

$$P(W_2 = w) = \frac{q^w\left(2-4q+2q^2\right)}{1-q^2}$$

for $w > 0$.

Using the above pmf, the moments of W_n can be determined. For example, when $n = 2$, using the pmf in (7), the following results:

$$E(W_2) = \sum_{w=1}^{\infty} w \frac{q^w\left(2-4q+2q^2\right)}{1-q^2} = \frac{2q}{p(1+q)} = \frac{2q}{1-q^2}$$

For n up to 10, algebraic expressions for the expected values of the distribution of the sample range of order statistics from the geometric distribution are obtained; these are shown in Table 1.

Conclusion

Algebraic expressions are presented for n up to 10 for the expected values of distribution of the sample range of order statistics from the geometric distribution. Using the obtained algebraic expressions, these expected values can be computed. As it is shown in Table 1, different values can be obtained for q and n. For example, for q=0.50, using the value n=2 in Table 1, $\mu_{W_2} \approx 0{,}011765$ is obtained. Further studies may focus on a software program for estimating the expected values found in this study.

References

Abdel-Aty, S. H. (1954). Ordered variables in discontinuous distributions. *Statistica Neerlandica*, *8*, 61-82.

Arnold, B. C., Balakrihnan, N. & Nagaraja, H. N. (1992). *A first course in order statistics.* NY: John Wiley & Sons.

Balakrihnan, N. & Rao, C. R. (1998). *Handbook of statistics 16-order statistics: Theory and methods*. NY: Elsevier.

Balakrihnan, N. (1986). Order Statistics from discrete distribution. *Communications in Statistics: Theory and Methods*, *15(*3*),* 657-675.

Table 1: The expected values of distribution of the sample range of order statistics from the geometric distribution

n	μ_{Wn}
2	$\frac{2q}{1-q^2}$
3	$\frac{3q}{1-q^2}$
4	$\frac{4q^5+2q^4+10q^3+2q^2+4q}{(1-q^4)(1+q+q^2)}$
5	$\frac{5q^5+15q^3+5q}{(1-q^4)(1+q+q^2)}$
6	$\frac{6q^{11}+3q^{10}+29q^9-4q^8+34q^7+13q^6+34q^5-4q^4+29q^3-3q^2+6q}{(1-q^4)(1+q+q^4)(1+q+q^2+q^3+q^4)}$
7	$\frac{7q^{11}-7q^{10}+42q^9-21q^8+49q^7+7q^6+49q^5-21q^4+42q^3-7q^2+7q}{(1-q^4)(1+q+q^4)(1+q+q^2+q^3+q^4)}$
8	$\frac{8q^{21}-49q^{20}+56q^{19}-6q^{18}+92q^{17}+68q^{16}+208q^{15}+94q^{14}+246q^{13}+162q^{12}+306q^{11}+162q^{10}+246q^9+94q^8+208q^7+68q^6+92q^5+6q^4+56q^3-4q^2+8q}{(1-q^8)(1+q^2+q^4)(1+q+q^2+q^3+q^4)(1+q+q^2+q^3+q^4+q^5+q^6)}$
9	$\frac{9q^{21}-9q^{20}+75q^{19}-21q^{18}+120q^{17}+45q^{16}+270q^{15}+42q^{14}+285q^{13}+126q^{12}+399q^{11}+126q^{10}+285q^9+42q^8+270q^7+45q^6+120q^5-21q^4+75q^3-9q^2+9q}{(1-q^8)(1+q+q^2+q^3+q^4+q^5+q^6)}$
10	$\frac{[10q^{31}-25q^{30}+125q^{29}-180q^{23}+337q^{27}-233q^{26}+536q^{25}-337q^{24}+724q^{23}-212q^{22}+991q^{21}-290q^{20}+1153q^{19}-260q^{18}+1381q^{17}-311q^{16}+1381q^{15}-260q^{14}+1153q^{13}-290q^{12}+99q^{11}-212q^{10}+724q^9-337q^8+536q^7-233q^6+337q^5-180q^4+125q^3-25q^2+99]}{\left[\left(1-q^4\right)\left(1+q^2+q^4\right)\left(1+q^3+q^6\right)\left(1+q+q^2+q^3+q^4\right)\right]}$

Burr, I. W. (1955). Calculation of exact sampling distribution of ranges from a discrete population. *Annals of Mathematical Statistics*, *26*, 530-532. *Correction*, *38*, 280.

Crawford, B. G. (1966). Characterization of geometric and exponential distributions. *Annals of Mathematical Statistics*, *37*, 1790-95.

David, H. A. (1981). *Order statistics* (*2nd Ed.*), NY: John Wiley & Sons.

El-Neweihi, E. & Gavindarajulu, Z. (1979). Characterization of geometric and exponential distribution and discrete ifr (dfr) distributions using order statistics. *Journal of Statistical Planning and Inference*, *3*, 85-90.

Ferguson, T. S. (1965.) A characterization of the geometric distribution. *American Mathematical Monthly*, *72*, 256-260.

Ferguson, T. S. (1967). On characterizing distributions by properties of order statistics. *Sankhya Series A, 29*, 265-278.

Galambos, J. (1975). Characterizations of probability distributions by properties of order statistics. In: G. P. Patil, S. Kotz and G. K. Ord, Eds., Statistical distributions in scientific work, *Characterization and Applications*, Reidel Publishing Company, Dordrecht- Holland, 2, 289-101.

Gavindarajulu, Z. (1980). Characterization of the Geometric Distribution using properties of order statistics *Journal of Statistical Planning and Inference, 4*, 237-47.

Khatri, C. G. (1962). Distributions of order statistics for discrete case. *Annals of Institute of Statistical Mathematics*, *14*, 167-171.

Margolin, B. H. & Winokur, H. S., Jr. (1967). exact moments of the order statistics of the geometric distribution and their relation to inverse sampling and reliability of redundant systems. *Journal of the American Statistical Association*, *62*, 915-925.

Nagaraja, H. N. (1986). Structure of discrete order statistics. *Journal of Statistical Planning and Inference*, *13*, 165-177.

Nagaraja, H. N. (1992). Order statistics from discrete distribution (with discussion). *Statistics*, *23,* 189-216.

Srivastava, R. C. (1974). Two characterizations of the geometric distribution. *Journal of the American Statistical Association*, *69*, 267-269.

Uppuluri, V. R. R. (1964). A characterization of the geometric distribution. *Annals of Mathematical Statistics. 4*, 1841.