On the moments of order statistics from discrete distributions

Article *in* Pakistan Journal of Statistics · April 2010

Sinan Calik
Firat University

Pak. J. Statist.
2010 Vol. 26(2), 417-426

ON THE MOMENTS OF ORDER STATISTICS FROM DISCRETE DISTRIBUTIONS

Sinan Çalik[1£], Mehmet Güngör[2€] and Cemil Colak[1§]

[1]Department of Statistics, [2]Department of Mathematics
University of Firat, 23119 Elazığ, Turkey.
Email: [£]scalik@firat.edu.tr; [€]mgungor@firat.edu.tr; [§]cemilcolak@yahoo.com

ABSTRACT

In this paper, the mth raw moments of order statistics from discrete distributions are obtained. By using the mth raw moments, a relation between the moments of sample maximum of order statistics from a discrete uniform distribution and the sum $S(N-1,n)$ is obtained. It is shown that with the help of this relation, one can obtain all the moments for sample maximum of order statistics from a discrete uniform distribution. By using MATLAB, we compute the means and variances of the sample maximum order statistics for sample size N = 10(10)50(50)100 and n = 1(1)10. Further studies may focus on a software program computing the means and variances of sample maximum of order statistics from any discrete distributions.

KEYWORDS

Order statistics; Discrete distribution; Moment; Sum; Uniform distribution.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a discrete distributions with probability mass function $p(x)$ $(x = 0,1,2,\ldots)$ and cumulative distribution function $P(x)$. Let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the order statistics obtained from random sample given above by arranging the observations in increasing order of magnitude. Let $E\left(X_{r:n}^m\right)$ denote by $\mu_{r:n}^{(m)}$ $(1 \le r \le n, m \ge 1)$. For convenience, $\mu_{r:n}$ for $\mu_{r:n}^{(1)}$ and $\sigma_{r:n}^2$ for variance of $X_{r:n}$ will also be used.

The first two moments of order statistics from discrete distributions were obtained by Khatri [6]. The same moments were also obtained by Arnold et al. [1] using a different method.

In this study, we prove the mth raw moments of order statistics from discrete distributions by using the same technique as with by Arnold et al. [1]. A relation between the moments of sample maximum of order statistics from a discrete uniform distribution and the sum $S(N-1,n)$ as given in (5.1) is obtained by using Theorem 4.1 based on

mth raw moments. It is shown that by using this relation, one can obtain all the moments for sample maximum of order statistics from a discrete uniform distribution. We give a numerical application of the moments of order statistics from discrete distributions and show these numerical results in Table 1.

2. MARGINAL DISTRIBUTIONS OF ORDER STATISTICS

Let $F_{r:n}(x)\,(r=1,2,\ldots,n)$ denote the cumulative distribution function (cdf) of $X_{r:n}$. Then it is easy to see that

$$\begin{aligned} F_{r:n}(x) &= P\{X_{r:n} \le x\} \\ &= P\{\text{at least } r \text{ of } X_1, X_2, \ldots, X_n \text{ are at most } x\} \\ &= \sum_{i=r}^{n} P\{\text{exactly } i \text{ of } X_1, X_2, \ldots, X_n \text{ are at most } x\} \\ &= \sum_{i=r}^{n} \binom{n}{i} \left[P(x)\right]^i \left[1-P(x)\right]^{n-i} \\ &= \int_0^{P(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1}(1-t)^{n-r}\, dt \end{aligned} \tag{2.1}$$

for $0 < P(x) < 1$.

For discrete population, the probability mass function (pmf) of $X_{r:n}$ may be obtained from (2.1) by differencing as:

$$f_{r:n}(x) = F_{r:n}(x) - F_{r:n}(x-1) \tag{2.2}$$

$$= \frac{n!}{(r-1)!(n-r)!} \int_{P(x-1)}^{P(x)} t^{r-1}(1-t)^{n-r}\, dt, \tag{2.3}$$

In detailed information, see [1,2,4].

In particular, we also have

$$f_{1:n}(x) = \left[1-P(x-1)\right]^n - \left[1-P(x)\right]^n$$

and

$$f_{n:n}(x) = \left[P(x)\right]^n - \left[1-P(x-1)\right]^n.$$

3. THE DISCRETE UNIFORM DISTRIBUTION

Let the random variable X be discrete uniform with support $B=\{1,2,\ldots,N\}$. We then write, X is discrete uniform $[1\,,N]$. Note that its *pmf* is given by $p(x)=\dfrac{1}{N}$ and its *cdf* is $P(x)=\dfrac{x}{N}$ for $x\in B$. Consequently, the *cdf* of the rth order statistics is given by

$$F_{r:n}(x)=\sum_{i=r}^{n}\binom{n}{i}\left(\frac{x}{N}\right)^{i}\left(1-\frac{x}{N}\right)^{n-i},\quad x\in B\,.$$

4. MOMENTS OF $X_{r:n}$

The mth moments of $X_{r:n}$ can be written as:

$$\mu_{r:n}^{(m)}=\sum_{x=0}^{\infty}x^{m}f_{r:n}(x) \tag{4.1}$$

where $f_{r:n}(x)$ is as given in (2.3).

The associated quantile function (or inverse distribution function, if you wish) is defined by

$$F^{-1}(y)=\sup\{x:F(x)\le y\}.$$

Now it is well known that if U is a Uniform (0,1) random variable, then $F^{-1}(U)$ has distribution function F. Moreover if we envision $U_1,\ldots,U_n$ as being i.i.d. Uniform (0,1) random variables with common distribution F, then

$$X_{r:n}\overset{d}{=}F^{-1}(U_{r:n})$$

where $\overset{d}{=}$ is to be read as "has the same distribution as". We can use the transformation $X_{r:n}\overset{d}{=}F^{-1}(U_{r:n})$ to be express the moments of $X_{r:n}$. For example, we can express the mean of $X_{r:n}$ as

$$\mu_{r:n}=\frac{n!}{(r-1)!(n-r)!}\int_{0}^{1}F^{-1}(u)u^{r-1}(1-u)^{n-r}\,du\,,$$

see [1]. However, since $F^{-1}(u)$ does not have an explicit form for most of the discrete (as well as absolutely continuous) distributions, this approach is often impractical. When the support B is a subset of nonnegative integers which is the case with several standard discrete distributions, one can use the *cdf* $F_{r:n}(x)$ directly to obtain the mth raw moments of $X_{r:n}$.

Theorem 4.1: *Let B, the support of the any discrete distribution, be a subset of nonnegative integers. Then*

$$\mu_{r:n}^{(m)} = \sum_{x=0}^{\infty}\left[(x+1)^m - x^m\right]\left[1-F_{r:n}(x)\right] \tag{4.2}$$

whenever the moment on the left-hand side is assumed to exist.

Proof. Let us note that if $\mu_{r:n}^{(m)}$ exists, $k^m P(X_{r:n} > k) \to 0$ as $k \to \infty$. Now consider

$$\sum_{x=0}^{\infty} x^m f_{r:n}(x) = \sum_{x=0}^{k} x^m \left[P(X_{r:n} > x-1) - P(X_{r:n} > x)\right]$$

$$= \sum_{x=0}^{k-1}\left[(x+1)^m - x^m\right] P(X_{r:n} > x) - k^m P(X_{r:n} > k).$$

On letting $k \to \infty$, we obtain

$$\mu_{r:n}^{(m)} = \lim_{k\to\infty}\sum_{x=0}^{k-1}\left[(x+1)^m - x^m\right] P(X_{r:n} > x) - \lim_{k\to\infty} k^m P(X_{r:n} > k)$$

$$= \sum_{x=0}^{\infty}\left[(x+1)^m - x^m\right]\left[1-F_{r:n}(x)\right]$$

which establishes (4.2).

In particular, we also have

$$\mu_{r:n}^{(1)} = \sum_{x=0}^{\infty}\left[1-F_{r:n}(x)\right]$$

and

$$\mu_{r:n}^{(2)} = 2\sum_{x=0}^{\infty} x\left[1-F_{r:n}(x)\right] + \mu_{r:n}.$$

The first two moments of order statistics from discrete distributions were obtained by Khatri [6] and Arnold et al. [1].

5. SPECIAL SUMS

In the theory of nonparametric statistics, particularly when we deal with rank sums, we often need for the sums of powers of the first n positive integers; namely, expression for

$$S(N-1, n) = 1^n + 2^n + \cdots + (N-1)^n = \sum_{x=1}^{N-1} x^n \tag{5.1}$$

for $n = 0,1,2,\ldots$. In the following theorem, we provide a convenient way of obtaining these sums.

Lemma 5.1:

$$\sum_{n=0}^{k-1}\binom{k}{n}S(N-1,n) = N^k - 1$$

for any positive N *and* k (see, [5]).

A disadvantage of this theorem is that we have to find the sums $S(N-1,n)$ one at a time, first for *n*=0, then *n*=1, then *n*=2 and so forth. For instance, for *k*=1, we get

$$\binom{1}{0}S(N-1,0) = N-1$$

and, hence, $S(N-1,0) = 1^0 + 2^0 + \cdots + (N-1)^0 = N-1$. Similarly, for *k*=2, we get

$$\binom{2}{0}S(N-1,0) + \binom{2}{1}S(N-1,1) = N^2 - 1$$

$$N - 1 + 2S(N-1,1) = N^2 - 1$$

and, hence, $S(N-1,1) = 1^1 + 2^1 + \cdots + (N-1)^1 = \frac{1}{2}(N-1)N$. Using the same technique, we can find the sums

$$S(N-1,2) = \frac{1}{6}(N-1)N(2N-1);\ S(N-1,3) = \frac{1}{4}(N-1)^2 N^2 \text{ and so on.}$$

Theorem 5.1: *For a discrete uniform distribution and* $m \ge 1$

$$\mu_{n:n}^{(m)} = N^m - \frac{\sum_{k=1}^{m}\binom{m}{k}S(N-1,\ n+m-k)}{N^n}. \tag{5.2}$$

Proof of Theorem 5.1: Let us consider the expression of $\mu_{r:n}^{(m)}$ in (4.2). By noting that $F_{n:n}(x) = \left[P(x)\right]^n = \left(\frac{x}{N}\right)^n$ for the discrete uniform distribution, we may write for $r = n$ and $m \ge 1$

$$\mu_{n:n}^{(m)} = \sum_{x=0}^{N}\left[(x+1)^m - x^m\right]\left[1 - \left(\frac{x}{N}\right)^n\right]$$

$$= \sum_{x=0}^{N-1}\left[(x+1)^m - x^m\right]\left[1-\left(\frac{x}{N}\right)^n\right]$$

$$= \sum_{x=0}^{N-1}(x+1)^m\left[1-\left(\frac{x}{N}\right)^n\right] - \sum_{x=0}^{N-1}x^m\left[1-\left(\frac{x}{N}\right)^n\right]$$

$$= N^m - \frac{1}{N^n}\left[\sum_{x=0}^{N-1}\left[(x+1)^m - x^m\right]x^n\right]. \tag{5.3}$$

Upon using the sum on the right-hand side of (5.3) in (5.1), we also derive the identity

$$\sum_{x=0}^{N-1}\left[(x+1)^m - x^m\right]x^n = \sum_{k=1}^{m}\binom{m}{k}S(N-1,\, n+m-k). \tag{5.4}$$

Thus, we obtain (5.2).

Alternative Proof of Theorem 5.1: Let us consider the expression of $\mu_{r:n}^{(m)}$ in (4.1). By noting that for the discrete uniform distribution, from (2.2)

$$f_{n:n}(x) = \left(\frac{x}{N}\right)^n - \left(\frac{x-1}{N}\right)^n$$

From (4.1), for $r = n, m \geq 1$, we can write

$$\mu_{n:n}^{(m)} = \sum_x x^m f_{n:n}(x)$$

$$= \sum_{x=1}^{N} x^m\left[\left(\frac{x}{N}\right)^n - \left(\frac{x-1}{N}\right)^n\right]$$

$$= \frac{1}{N^n}\left[\sum_{x=1}^{N} x^m[x^n - (x-1)^n]\right]$$

$$= \frac{1}{N^n}\left[\sum_{x=1}^{N} x^m x^n - \sum_{x=1}^{N} x^m (x-1)^n\right]$$

$$= \frac{1}{N^n}\left[\sum_{x=1}^{N-1} x^m x^n + N^m N^n - \sum_{x=1}^{N} x^m (x-1)^n\right]$$

$$= N^m - \frac{1}{N^n}\left[\sum_{x=1}^{N} x^m (x-1)^n - \sum_{x=1}^{N-1} x^m x^n\right]$$

$$= N^m - \frac{1}{N^n}\left[\sum_{x=1}^{N-1} x^m (x-1)^n + N^m (N-1)^n - \sum_{x=1}^{N-1} x^m x^n\right]$$

$$= N^m - \frac{1}{N^n}\left[\sum_{x=1}^{N-1} \{x^m (x-1)^n - x^m x^n\} + N^m (N-1)^n\right]$$

$$= N^m - \frac{1}{N^n}\left[\sum_{x=1}^{N-1} x^m [(x-1)^n - x^n] + N^m (N-1)^n\right]. \tag{5.5}$$

On the other hand, we get following equation

$$\sum_{x=1}^{N-1} x^m [(x-1)^n - x^n] + N^m (N-1)^n = \sum_{k=1}^{m} \binom{m}{k} S(N-1, n+m-k)$$

If this last expression is written in (5.5), (5.4) is obtained. Therefore, the proof is complete.

These moments can be evaluated easily. Using this relation, the moments of the sample maximum of order statistics from a discrete uniform distribution are also more practical.

6. THE CALCULATION OF THE MOMENTS FOR THE DISCRETE UNIFORM DISTRIBUTION

Using Relation 5.1, we can conclude, for example, that

$$\mu_{2:2} = \frac{(N+1)(4N-1)}{6N} \text{ and } \mu_{2:2}^{(2)} = \frac{(N+1)(3N^2+N+1)}{6N}.$$

When $n = 2$, using the values of $\mu_{2:2}$ and $\mu_{2:2}^{(2)}$, we obtain

$$\sigma_{2:2}^2 = \mu_{2:2}^{(2)} - \mu_{2:2}^2 = \frac{(2N^2+1)(N^2-1)}{36N^2}. \tag{6.1}$$

The same results on the right-hand side of (6.1) were also obtained by Arnold et al. [1] in a different manner. For *n* up to 15, algebraic expressions for the expected values of the sample maximum of order statistics from a discrete uniform distribution were obtained by Çalik and Güngör [3].

By using Relation 5.1., the values $\mu_{n:n}$ and $\sigma_{n:n}^2$ are presented in Table 1, for example, for n = 1(1)10 and N = 10(10)50(50)100.

Table 1
Means and Variances of Sample Maximum of Order Statistics for the Discrete Uniform Population

N	n	$\mu_{n:n}$	$\sigma^2_{n:n}$	N	n	$\mu_{n:n}$	$\sigma^2_{n:n}$
10	1	5.5000	8.2500	40	1	20.5000	133.2500
	2	7.1500	5.5275		2	27.1625	88.8611
	3	7.9750	3.7084		3	30.4937	59.9583
	4	8.4667	2.6167		4	32.4917	42.6167
	5	8.7918	1.9288		5	33.8229	31.6905
	6	9.0216	1.4716		6	34.7732	24.4303
	7	9.1920	1.1536		7	35.4854	19.3820
	8	9.3227	0.9241		8	36.0389	15.7377
	9	9.4257	0.7533		9	36.4813	13.0244
	10	9.5086	0.6229		10	36.8428	10.9513
20	1	10.5000	33.2500	50	1	25.5000	208.2500
	2	13.8250	22.1944		2	33.8300	138.8611
	3	15.4875	14.9583		3	37.9950	93.7083
	4	16.4833	10.6167		4	40.4933	66.6167
	5	17.1458	7.8810		5	42.1583	49.5476
	6	17.6179	6.0630		6	43.3471	38.2058
	7	17.9709	4.7988		7	44.2383	30.3195
	8	18.2445	3.8861		8	44.9311	24.6266
	9	18.4626	3.2065		9	45.4850	20.3879
	10	18.6403	2.6872		10	45.9379	17.1495
30	1	15.5000	74.9167	100	1	50.5000	833.2500
	2	20.4944	49.9722		2	67.1650	555.5278
	3	22.9917	33.7083		3	75.4975	374.9583
	4	24.4889	23.9500		4	80.4967	266.6167
	5	25.4861	17.8016		5	83.8292	198.3571
	6	26.1976	13.7160		6	86.2093	153.0017
	7	26.7306	10.8751		7	87.9942	121.4653
	8	27.1445	8.8242		8	89.3822	98.7006
	9	27.4750	7.2972		9	90.4925	81.7515
	10	27.7450	6.1304		10	91.4008	68.8024

7. DISCUSSION

The current study presents the obtained *m*th raw moments of order statistics from discrete distributions. From the statistical point of view, the moments of order statistics carry great importance to estimate means and variances for the discrete distributions. For the purpose of giving an example for discrete distributions, we showed the relationship between the moments of sample maximum of order statistics from a discrete uniform distribution and the sum $S(N-1,n)$ in this study. Then, an illustrative example was given on the means and variances of sample maximum of order statistics for the discrete uniform population.

During the last decades, computer technologies have been considerably developed in relation to statistical analyses and computations. Also, software programs such as artificial neural networks, several algorithms etc. have enormous performance for carrying out the statistical problems. In a study, Evans et al. [7] presented an algorithm for computing the probability density function of order statistics drawn from discrete parent populations, and used exact bootstrapping analysis which illustrates the utility of the presented algorithm. Computer-aided algorithms give good results on the computations related to order statistics. Adatia [8] derived an explicit expression for the expected value of the product of two order statistics from the geometric distribution and discussed a method of computation of the expected values and covariances of order statistics. The studies and others reported have been focused on computer-aided computations or algorithms by some software programs. In parallel with the developments in computer-based technology, in the next phase of the study we want to create a program computing the means and variances of sample maximum of order statistics for the discrete distributions.

In conclusion, using the obtained relationship previously described, all the moments for sample maximum of order statistics from the discrete uniform distribution can be achieved. It is recommended that this relationship may be applied to other discrete distributions with illustrative examples. Further studies may focus on a software program computing the means and variances of sample maximum of order statistics from any discrete distributions.

REFERENCES

1. Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. (1992). *A First Course in Order Statistics*, John Wiley and Sons, New York.
2. Balakrishnan, N. and Rao, C.R. (1998). *Handbook of Statistics 16-Order Statistics*: Theory and Methods, Elsevier, New York.
3. Çalik, S. and Güngör, M. (2004). On the expected values of the sample maximum of order statistics from a discrete uniform distribution, *App. Math. and Compu.*, 157, 695-700.
4. David, H.A. (1981). *Order Statistics*, John Wiley and Sons, New York.
5. Freund, J.E. and Walpole, R.E. (1962). *Mathematical Statistics*, Prentice-Hall International, Inc., London.
6. Khatri, C.G. (1962). Distribution of order statistics for discrete case. *Ann. Inst. Statist. Math.*, 14, 167-171.
7. Evans, D.L., Leemis, L.M., Drew, J.H. (2006). The Distribution of Order Statistics for Discrete Random Variables with Applications to Bootstrapping. *INFORMS Journal on Computing*, 18(1), 19-30.
8. Adatia, A. (1991). Computation of variances and covariances of order statistics from the geometric distribution. *J. Statist. Compu. and Simul.*, 39, 91-94.