Journal of Mathematics and Statistics 1(3): 180-183, 2005
ISSN 1549-3644

On the Variances of Distribution of the Sample Range of Order Statistics from a Discrete Uniform Distribution

Sinan Calik
Department of Mathematics, Firat University, 23119, Elazığ, Turkey

Abstract: The purpose of this study was to obtain algebraic expressions for n up to 20 for the variances of distribution of the sample range of order statistics from a discrete uniform distribution.

Key words: Order statistics, expected value, sum, discrete uniform distribution, variance, moment, sample range

INTRODUCTION

Let $X_1, X_2, \ldots X_n$ be a random sample of size n from a discrete distributions with probability mass function (pmf) F(x) (x=0,1,2,...) and cumulative distribution function F(x). Let $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ be the order statistics obtained from above random sample by arranging the observations in increasing order of magnitude. Let us denote the spacing $W_{i,j:n} = X_{j;n} - X_{i;n}$ When i = 1 and j = n, that is, in the case of the sample range W_n. We then have $W_n = X_{n:n} - X_{1:n}$. Let us denote the *m*th moments of distribution of the sample range $E\left(W_n^{(m)}\right)$ by $\mu_{W_n}^{(m)}$ $(n\ge 2, m\ge 1)$. For convenience, μ_{W_n} for $\mu_{W_n}^{(1)}$ and $\sigma_{W_n}^2$ for variance of W_n will also be used.

The distribution of the sample range from a discrete order statistics are given by Arnold *et al.*[1]. For n up to 20, algebraic expressions for the expected values of distribution of the sample range of order statistics from a discrete uniform distribution were obtained by Calik[2]. For more details on discrete order statistics can be found in the works of Nagaraja (1992), Balakrishnan and Rao (1998). In this study, for n up to 20, algebraic expressions for the variances of distribution of the sample range of order statistics from a discrete uniform distribution are obtained.

MARGINAL DISTRIBUTION 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

$$F_{r:n}(x) = P\{X_{r:n} \le x\}$$

$$= P\{\text{at least r of } X_1, X_2, \ldots, X_n \text{ are at most x}\}$$

$$= \sum_{i=r}^{n} P\{\text{exactly i of } X_1, X_2, \ldots, X_n \text{ are at most x}\}$$

$$= \sum_{i=r}^{n} \binom{n}{i} \left[F(x)\right]^i \left[1-F(x)\right]^{n-i}$$

$$= \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r}\, dt \quad (1)$$

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

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

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

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

Arnold *et al.*(1992).

ORDER STATISTICS FROM A DISCRETE UNIFORM DISTRIBUTION

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

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

JOINT DISTRIBUTION OF ORDER STATISTICS

The joint distribution of order statistics can be similarly derived and will naturally look a lot more complicated. 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:

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

$$= \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!}\left\{F\left(x_i\right)\right\}^r \times \left\{F\left(x_j\right) - F\left(x_i\right)\right\}^{s-r}\left\{1 - F\left(x_j\right)\right\}^{n-s} \text{ for } x_i < x_j. \quad (2)$$

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

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

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

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

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

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

And D is k-dimensional space given by:

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

Arnold *et al*.(1992), Balakrishnan and Rao (1998).

THE DISTRIBUTION OF THE SAMPLE RANGE

Let us start with the pmf of the spacing $W_{i,j:n} = X_{j:n} - X_{i:n}$ On using Theorem 1, we can write

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

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. We then have

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

.

Thus, the pmf of W_n is given by:

$$P(W_n = 0) = n(n-1) \sum_{x \in D} \int_{F(x-1)}^{F(x)} \int_{F(x-1)}^{F(x)} (u_n - u_1)^{n-2} du_n du_1 = \sum_{x \in D} \left\{F(x) - F(x-1)\right\}^n = \sum_{x \in D} \left\{f(x)\right\}^n \quad (4)$$

and, for w > 0,

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

$$= \sum_{x \in D} \left\{\begin{matrix}\left[F(x+w) - F(x-1)\right]^n - \left[F(x+w) - F(x)\right]^n \\ -\left[F(x+w-1) - F(x-1)\right]^n \\ +\left[F(x+w-1) - F(x)\right]^n\end{matrix}\right\} \quad (5)$$

Arnold *et al*.(1992).

Expressions (4) and (5) can also be obtained without using the integral expression from Theorem1. One can also use a multinomial argument to obtain an alternative expression for the pmf of W_n.

THE VARIANCE OF SAMPLE RANGE

The mth moments of W_n can be immediately written down as:

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

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

When X is a discrete uniform [1,N] random variable, in the case of the sample range, (6) yields

Table 1: The variances of distribution of the sample range of order statistics from discrete uniform distribution

n	$\sigma^2_{w_n}$
2	$(1/(18))N^{-2}(N^2+2)(N^2-1)$
3	$(1/(20))N^{-2}(2N^2-1)$
4	$(1/(450))N^{-6}(3N4-38N^2+18N^6+2)(N^2-1)$
5	$(1/(252))N^{-6}(2N4-N^2-7)(4N^2-1)(N^2-1)$
6	$(1/(1764))N^{-10}(312N^4-52N^2-186N^6-81N^8+45N^{10}+4)(N^2-1)$
7	$(1/(720))N^{-10}(261N^4-120N^2-59N^6-45N^8+15N^{10}+20)(N^2-1)$
8	$(1/(4050))N^{-14}(1082N^4-222N^2-2998N^6+2375N^8-155N^{10}-305N^{12}+70N^{14}+18)(N^2-1)$
9	$(1/(1100))N^{-14}(572N^4-264N^2-127N^6-46N^8+16N^{10}+99)(N^2-N-1)(N+N^2-1)(N^2-1)$
10	$(1/(4356))N^{-18}(5776N^4-1220N^2-14288N^6+22760N^8-17458N^{10}+4553N^{12}+461N^{14}-408N^{16}+54N^{18}+100)(N^2-1)$
11	$(1/(65520))N^{-18}(641095N^4-254800N^2-980525N^6+992041N^8-480339N^{10}+79220N^{12}+13440N^{14}-6580N^{16}+700N^{18}+45500)(N^2-1)$
12	$(1/(3726450))N^{-22}(54684638N^4-11621238N^2-132994282N^6+199319700N^8-208505900N^{10}+137066545N^{12}-45416210N^{14}+4728885N^{16}+1200360N^{18}-395325N^{20}+34650N^{22}+954962)(N^2-1)$
13	$(1/(44100))N^{-22}(6659202N^4-2666569N^2-9968838N^6+10192303N^8-7565227N^{10}+3480384N^{12}-830286N^{14}+51595N^{16}+20135N^{18}-4890N^{20}+360N^{22}+477481)(N^2-1)$
14	$(1/(16200))N^{-26}(5040368N^4-1072680N^2-12213232N^6+18122840N^8-18348920N^{10}+13712320N^{12}-7386980N^{14}+2499030N^{16}-443910N^{18}+13755N^{20}+9855N^{22}-1863N^{24}+117N^{26}+88200)(N^2-1)$
15	$(1/(48960))N^{-26}(208610740N^4-83680800N^2-310871860N^6+313890720N^8-231687560N^{10}+131040293N^{12}-53162187N^{14}+13720165N^{16}-1855915N^{18}+11535N^{20}+38055N^{22}-5805N^{24}+315N^{26}+14994000)(N^2-1)$
16	$(1/(130050))N^{-30}(1494529362N^4-318173022N^2-3618213838N^6+5357087530N^8-5392599990N^{10}+3959794970N^{12}-2236122950N^{14}+978863895N^{16}-308127565N^{18}+62378935N^{20}-6517815N^{22}-97425N^{24}+125275N^{26}-15825N^{28}+750N^{30}+26165378)(N^2-1)$
17	$(1/(1077300))N^{-30}(217\ 748448141N^4-87\ 383287467N^2-324\ 142030359N^6+326\ 404529269N^8-239\ 200266851N^{10}+134\ 181336457N^{12}-59\ 748108933N^{14}+20716099185N^{16}-5186278575N^{18}+841579725N^{20}-68427555N^{22}-2361135N^{24}+1256465N^{26}-134050N^{28}+5600N^{30}+15\ 659978733)(N^2-1)$
18	$(1/(15920100))N^{-34}(10990\ 053373216N^4-2339\ 892977540N^2-26600\ 944716704N^6+39364\ 234902320N^8-39573090389680N^{10}+28961\ 477494880N^{12}-16199\ 947049440N^{14}+7191\ 320385480N^{16}-2578\ 292032050N^{18}+722\ 128583085N^{20}-146\ 742515115N^{22}+19\ 414517619N^{24}-1229271561N^{26}-66651795N^{28}+22069845N^{30}-2019780N^{32}+74970N^{34}+192\ 431368900)(N^2-1)$
19	$(1/(1940400))N^{-34}(29428\ 581476459N^4-11811\ 038847520N^2-43796\ 203580701N^6+44073\ 620107407N^8-32247\ 349060113N^{10}+18015\ 589326999N^{12}-7968\ 541625781N^{14}+2870\ 154101505N^{16}-843\ 381663555N^{18}+194\ 089805325N^{20}-32\ 560612515N^{22}+3560365081N^{24}-174354719N^{26}-13265252N^{28}+3149608N^{30}-250404N^{32}+8316N^{34}+2116\ 745057900)(N^2-1)$
20	$(1/(24012450))N^{-38}(1535756\ 632486358N4-326985\ 026884278N^2-3717035\ 883669442N^6+5499777\ 622562120N^8-5527205\ 631616280N^{10}+4041943\ 140883960N^{12}-2256561\ 161631080N^{14}+996522\ 724403100N^{16}-358014\ 241816260N^{18}+106593\ 367281225N^{20}-26054\ 672796900N^{22}+4999\ 922977995N^{24}-702\ 371816355N^{26}+64\ 168782975N^{28}-2389743675N^{30}-246156075N^{32}+45069750N^{34}-3145725N^{36}+94050N^{38}+26891\ 299165122)(N^2-1)$

$$\mu^{(1)}_{w_n} = E(W_n) = \sum_{w=0}^{\infty} wP(W_n = w)$$

$$= \sum_{w=1}^{N} wP(W_n = w) \qquad (7)$$

where $P(W_n = w)$ is as given in (9).

For n up to 20, algebraic expressions for the μ_{w_n} of distribution of the sample range of order statistics from a discrete uniform distribution were obtained by Calik (2005).

Further from (6)

$$\mu^{(2)}_{w_n} = E(W_n^2) = \sum_{w=0}^{\infty} w^2 P(W_n = w) \qquad (8)$$

and hence the variance of the sample range we obtain

$$\sigma^2_{W_n} = \mu^{(2)}_{W_n} - \mu^2_{W_n}.$$

THE DISTRIBUTION OF THE SAMPLE RANGE FROM A DISCRETE UNIFORM DISTRIBUTION

When X is a discrete uniform [1,N] random variable, the expression in (4) and (5) can be further simplified. We then have:

$$P(W_n = 0) = \sum_{x=1}^{N} \left(\frac{1}{N}\right)^n = \frac{1}{N^{n-1}}$$

and

$$P(W_n = w) = \sum_{x=1}^{N-w} \left\{ \left(\frac{x+w}{N} - \frac{x-1}{N}\right)^n - \left(\frac{x+w}{N} - \frac{x}{N}\right)^n - \left(\frac{x+w-1}{N} - \frac{x-1}{N}\right)^n + \left(\frac{x+w-1}{N} - \frac{x}{N}\right)^n \right\}$$

$$= \sum_{x=1}^{N-w} \frac{1}{N^n} \left\{ (w+1)^n - 2w^n + (w-1)^n \right\}$$

$$= \frac{(N-w)}{N^n}\left\{(w+1)^n - 2w^n + (w-1)^n\right\}, \quad (9)$$
$$w = 1, \ldots, N-1$$

In particular, we also have

$$P(W_2 = w) = 2\frac{N-w}{N^2}.$$

Using the above pmf, one can determine the moments of W_n.

Example 1: Using (7) and (8), we can conclude, for example,that

$$\mu_{w_2} = \frac{N^2-1}{3N} \quad \text{and} \quad \mu_{w_2}^{(2)} = \frac{N^2-1}{6}$$

we obtain the values, using the values of μ_{w_2} and $\mu_{w_2}^{(2)}$

$$\sigma_{w_2}^2 = \frac{(N^2+2)(N^2-1)}{18N^2}$$

For n up to 20, algebraic expressions for the variances of distribution of the sample range of order statistics from a discrete uniform distribution are obtained, see Table 1.

CONCLUSIONS

As it is understood from Table 1, different values can be obtained for N and n. For instance, for N=100, using the value n=2 in the Table I, we obtain $\sigma_{w_2}^2$ =555,6111.

REFERENCES

1. Arnold, B.C., N. Balakrishnan and H.N. Nagaraja, 1992. A First Course in Order Statistics. John Wiley and Sons, New York.
2. Calik, S., 2005. On the expected values of distribution of the sample range of order statistics from a discrete uniform distribution. (submitted).
3. Nagaraja, H.N., 1992. Order Statistics from discrete distribution (with discussion). Statistics, 23: 189-216.
4. Balakrishnan, N. and C.R. Rao, 1998. Handbook of statistics 16-Order Statistics: Theory and Methods, Elsevier, New York.