

## The Probability and the Asymptotic Distribution Functions of Increasing Ordered Statistical Data

A.A. El-Faheem and Eman A. Mahmoud  
 Department of Mathematics, Faculty of Science,  
 South Valley University, Aswan, Egypt

**Abstract: Problem statement:** We considered two types of sequences of ordered statistical discrete data  $i_1, i_2, \dots, i_n, 1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq n$  and  $i_j = 1, 2, \dots, n, j = 1, 2, \dots, n$ . The first type is the order statistics of a sample of size  $n$  taken from the uniform discrete distribution on the set  $S = \{1, 2, \dots, n\}, n \in \mathbb{N}^+$ . The other type is the increasing ordered word of length  $n$  taken from the set  $S = \{1, 2, \dots, n\}, n \in \mathbb{N}^+$ . We studied a common property of both types of samples namely the number of elements  $i_j = j, i_j = 1, 2, \dots, n, j = 1, 2, \dots, n$ . i.e., we find the number of integers in each sequence  $i_1, i_2, \dots, i_n$  satisfying the condition C:  $i_j = j$ , for both types. We obtain the probability distribution functions as well as the asymptotic distributions of the random variables representing the number of integers satisfying condition C. **Approach:** We employed combinatorial tools to obtain the number of samples having  $j$  elements satisfying condition C in both types,  $j = 1, 2, \dots, n$ . **Results:** For large  $n \in \mathbb{N}^+$ , we found that the expected value of the number of samples of the second type, satisfying condition C, is much larger than that of the first type. **Conclusion:** The result can be used to distinguish between these two data.

**Key words:** Discrete order statistics, increasing ordered words, probability distributions, limiting distributions

### INTRODUCTION

We consider two types of sequences of ordered statistical discrete data  $i_1, i_2, \dots, i_n$ , with  $1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq n$  and  $i_j = 1, 2, \dots, n, j = 1, 2, \dots, n$ . The first type is the order statistics of a sample of size  $n$  taken from the uniform discrete distribution on the set  $S = \{1, 2, \dots, n\}, n \in \mathbb{N}^+$ . The number of all possible ordered statistical discrete data is  $n^n$ . The second type of such ordered statistical data is the ordered increasing word,  $i_1 i_2 \dots i_n$  with  $1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq n$ , of length  $n$  taken from the set  $S = \{1, 2, \dots, n\}, n \in \mathbb{N}^+$ . The number of all such ordered increasing words is  $\binom{2n-1}{n}$ , (Tomescu and Rodenau, 1975). Both types of sequences can be considered as ordered statistical discrete data. If one claims that both sequences are the ordered statistics of a sample of size  $n$  from the discrete uniform distribution on the set  $S$ , then we have to develop a method or a test to distinguish between the real set of ordered statistics and the ordered word.

To achieve this, we are going to study a common property of both sequences described in what follows:

For both types of ordered data each we find the number of integers in the sequence satisfying the condition C:  $i_j = j, i_j, j = 1, 2, \dots, n$ . Let  $M^0$  and  $M^1$  denote the random variables representing the number of integers satisfying condition C in the sequences of the first type and the sequences of the second type respectively,  $m^0, m^1 = 1, 2, \dots, n$ . The probability distributions as well as the asymptotic distributions of  $M^0$  and  $M^1$  will be given by the following two theorems.

**Theorem 1:** For  $M^0$  as above, we have:

$$P(M^0 = k) = \frac{n!}{n^n} \binom{k}{n} \frac{n^{n-k}}{(n-k)!}, \quad k = 1, 2, \dots, n$$

**Corollary 1:** Let  $Y^0 = M^0/n^{1/2}$ , then as:

$$n \rightarrow \infty$$

$$P(Y^0 > t) = e^{-(1/2)t^2}, \quad t > 0$$

This is the same limiting distribution for Rayleigh distribution, (Johnson *et al.*, 1994).

**Corresponding Author:** A.A. El-Faheem, Department of Mathematics, Faculty of Science, South Valley University, Aswan, Egypt

**Theorem 2:** For  $M^1$  as above, we have:

$$\binom{2n-1}{n} P(M^1 = k) = \binom{2n}{n-k} (k/n), k = 1, 2, \dots, n$$

**Corollary 2:** Let  $Y^1 = M^1/n^{1/2}$ , then as:

$$n \rightarrow \infty$$

$$P(Y^1 > t) = e^{-t^2}, \quad t > 0$$

This is the same limiting distribution for Weibull distribution with shape parameter 2 (Johnson *et al.*, 1994).

### MATERIALS AND METHODS

To prove Theorem 1, let  $I_1, I_2, \dots, I_n$  be a random sample of size  $n$  from the discrete uniform distribution on the set  $S = \{1, 2, \dots, n\}$  and  $i_1, i_2, \dots, i_n$ , with  $1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq n$ , be its order statistics. Let  $E_{ij}$  be the event that  $i_j = j$  and  $P(n, k)$  be the probability (inclusive) of  $k$  order statistics being such that  $i_{j_s} = j_s, s = 1, 2, \dots, k, k = 1, 2, \dots, n$ . Then:

$$P(n, k) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(E_{i_1} E_{i_2} \dots E_{i_k}) k = 1, 2, 3, \dots, n$$

Using the joint distribution of  $k$  discrete order statistics (David, 1970), then

$$P(n, k) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \frac{n!}{(i_1 - 1)!(i_2 - i_1 - 1)! \dots (n - i_k)!} \int_{I(i_1)} \dots \int_{I(i_k)} U_{(i_1)}^{i_1-1} [U_{(i_2)} - U_{(i_1)}]^{i_2-i_1-1} \dots [1 - U_{(i_k)}]^{n-i_k} \dots dU_{(i_k)} \dots dU_{(i_2)} dU_{(i_1)} \quad (1)$$

where,  $\int_{I(i_s)} = \int_{(i_s-1)/n}^{i_s/n}$ , for example we have:

$$P(n, 1) = \sum_{i=1}^n \frac{n!}{(i-1)!(n-i)!} \int_{\frac{i-1}{n}}^{\frac{i}{n}} U^{i-1} (1-U)^{n-i} dU$$

which after integrating by parts and summing over  $i$ , we get:

$$P(n, 1) = \frac{n!}{n^n} \sum_{i=1}^n \frac{(i)^i [n-i]^{n-i-1}}{i!(n-i-1)!}$$

In general the sum (1) is found to be:

$$\frac{n^n}{n!} P(n, k) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \frac{i_1^{i_1} (i_2 - i_1)^{i_2 - i_1} \dots (n - i_k)^{n - i_k}}{i_1! (i_2 - i_1)! \dots (n - i_k)!} \quad (2)$$

By using the well known Abel extension of the binomial theorem (Riordan, 1979), we get:

$$\frac{n^n}{n!} P(n, k) = \sum_{j=0}^{n-k} \binom{j+k-1}{j} \frac{n^j}{j!} \quad (3)$$

Employing the inclusion-exclusion formula (Riordan, 1978), we get:

$$\frac{n^n}{n!} P(M^0 \geq r) = \sum_{s=r}^n (-1)^{s-r} \binom{j}{r} \sum_{j=0}^{n-s} \binom{j+s-1}{j} \frac{n^j}{j!} = 1, 2, \dots, n \quad (4)$$

Simple manipulation after interchanging the summation signs in (4), we get:

$$P(M^0 \geq r) = \frac{n! n^{n-r}}{n^n (n-r)!} r = 1, 2, \dots, n \quad (5)$$

From (5) we get:

$$P(M^0 = k) = \frac{n! k n^{n-k}}{n^n n(n-k)!} \quad (6)$$

and Theorem 1 is proved.

To prove Corollary 1, from (6) we write  $P(Y > r-1)$  as:

$$P(M^{0+} > r-1) = \prod_{i=0}^{r-1} (1 - i/n) \quad (7)$$

by taking the logarithm of both sides of (18) and using the well known approximation  $\ln(1+x) \approx x$ , for  $-1 < x < 1$ , we get Corollary 3.

**Corollary 3:** For large  $n$ , the expected value  $E(Y^0)$  is given by:

$$E(Y^0) \approx \frac{\sqrt{n\pi}}{\sqrt{2}}$$

We outline the proof of Theorem 2 as follows, we consider a sequence  $i_1, i_2, \dots, i_n$  and let  $j_1, j_2, \dots, j_k$  be the subscripts of  $k$  of the  $i$ 's satisfying the condition C:

$i_{js} = j_s, s = 1, 2, \dots, k$ . The number of sequences  $H(n, k)$ , with  $k$  or more elements (inclusive) satisfying this condition is given by the following lemma.

**Lemma 1:** (without proof):

$$H(n, k) = \sum_{1 \leq j_1 < j_2 < \dots < j_k \leq n} Q(j_1, j_2, \dots, j_k)$$

Where:

$$Q(j_1, j_2, \dots, j_k) = \binom{2j_1-2}{j_1-1} \binom{2(j_2-j_1)-1}{j_2-j_1} \dots \binom{2(j_k-j_{k-1})-1}{j_k-j_{k-1}} \binom{2n-2j_k}{n-j_k} \quad (8)$$

**Lemma 2:** Let  $H(n, k)$  as in (8), then:

$$H(n, k) = \sum_{j=0}^{n-k} \binom{j+k-1}{j} \binom{2n-1}{n-j} \quad (9)$$

**Proof of Lemma 2:** We write  $H(n, k)$  as:

$$H(n, k) = 2^{-(k-1)} \sum_{1 \leq j_1 < j_2 < \dots < j_k \leq n} \binom{2j_1}{j_1} \binom{2(j_2-j_1)}{j_2-j_1} \dots \binom{2(j_k-j_{k-1})}{j_k-j_{k-1}} \binom{2n-2j_k}{n-j_k} \quad (10)$$

in (10), we put  $j_1-1 = l_1, j_2-j_1 = 1 = l_2, \dots, j_k-j_{k-1} = l_k, n-j_k = l_{k+1}$ , then we have  $l_j \geq 0, j = 0, 1, \dots, k+1$  and  $\sum_{j=0}^{k+1} l_j = n - k$  accordingly  $H(n, k)$  becomes:

$$H(n, k) = 2^{-(k-1)} \sum_{l_j \geq 0, \sum_{j=0}^{k+1} l_j = n-k} \left( \prod_{j=1}^k \binom{l_j+2}{l_j+1} \right) \binom{2l_{k+1}}{l_{k+1}} \quad (11)$$

$$= 2^{-(k-1)} \sum_{l_j \geq 0, \sum_{j=0}^{k+1} l_j = n-k} \binom{2l_1}{l_1} \left( \prod_{j=2}^k \binom{l_j+2}{l_j+1} \right) \binom{2l_{k+1}}{l_{k+1}}$$

The sum in (11) is the coefficient of  $x^{(n-k)}$  in:

$$g(x) = ((1-4x)^{-1/2})^2 ((1-4x)^{-1/2} - 1/x)^{k-1}$$

Using the substitution  $x = z / (1+z)^2$  and using Lagrange's theorem for implicit function (Goulden and Jackson, 1983), (11) will be equal to the coefficient of  $z^{n-k}$  in the expression  $(1-z)^{-k} (1+z)^{2n-1}$ , Lemma 2 is proved, satisfying this condition is given by the following lemma.

To prove Theorem 2, we employ the inclusion-exclusion formula (Riordan, 1978) to get:

$$\binom{2n-1}{n} P(M^1 \geq k) = \sum_{s=k}^n (-1)^{s-k} \binom{s-1}{k-1} H(n, s) \quad (12)$$

$$= \sum_{s=k}^n (-1)^{s-k} \binom{s-1}{k-1} \sum_{j=0}^{n-s} \binom{j+s-1}{j} \binom{2n-1}{n-j}$$

Interchanging the summation signs in (12), after simple manipulation, we get:

$$\binom{2n-1}{n} P(M^1 \geq k) = \binom{2n-1}{n-k} \quad (13)$$

from which Theorem 2 follows.

Corollary 2, can be easily proved by expanding the binomial coefficients in (13) and take the logarithm of both sides using the approximation  $\ln(1-z)^{-1} \approx z$  for large  $n, -1 < z < 1$ , (Balakrishnan, 1997), we get

**Corollary 4:** Let  $Y^1 = M^1/n^{1/2}$ , then as:

$$n \rightarrow \infty$$

$$P(Y^1 > t) = e^{-t^2}, \quad t > 0$$

## RESULTS

From Corollary 1 and 2, we see that the expected values,  $E(Y^1)$  and  $E(Y^0)$ , of  $Y^1$  and  $Y^0$  are approximated, respectively, by:

$$E(Y^1) \approx \frac{\sqrt{n\pi}}{2}$$

and

$$E(Y^0) \approx \frac{\sqrt{n\pi}}{\sqrt{2}}$$

The difference  $DE = (E(Y^1) - E(Y^0))$  will be approximately equal to  $DE \approx \alpha \sqrt{n\pi} \rightarrow \infty$  as  $n \rightarrow \infty$ , where  $\alpha = 1/\sqrt{2} - 1/2$ .

## DISCUSSION

The sequence  $i_1, i_2, \dots, i_n$ , with  $1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq n$ , whether it is from the first type or the second type, is a kind of weak discrete records. Thus, in this article, we found two distributions of functions defined on these types of records which can be considered an addition to records' literature. A feasible extension to these results

is to consider the distributions of  $M^0$  and  $M^1$  under the condition  $C'$ :

$$i_j = j \text{ or } (j+1), j=1,2,\dots,(n-1)$$

### CONCLUSION

If we consider the condition  $C$ , introduced before, as a concordance condition of the element  $i_j$  with  $j$ ,  $i_j$ ,  $j = 1,2,\dots,n$ , then the number of concordances  $M^0$  is smaller than that of  $M^1$ . This makes sense, since the sequences of the first type represents the reality, while the sequences of the second type (ordered increasing words) can be looked at as fabricated sequences. Thus, obtaining a large number of concordances is an indication of fabrication.

### ACKNOWLEDGEMENT

I express my sincere thanks and appreciations to Professor Dr. N. Balakrishnan for inviting me to visit Mathematics and Statistics Department, McMaster University as a Postdoctoral fellowship funded by the Government of Egypt and also for his constant support and supervision, whose comments substantially improved the presentation of this work.

### REFERENCES

- Balakrishnan, N., 1997. *Advances in Combinatorial Methods and Applications to Probability and Statistics*. Birkhauser, Berlin, ISBN: 3-7643-3908-X.
- David, H.A., 1970. *Order Statistics*. John Wiley, ISBN: 431-19675-4, pp: 12.
- Goulden, I.P. and D.M. Jackson, 1983. *Combinatorial Enumeration. With a foreword by Gian-Carlo Rota*. John Wiley and Sons, Inc., New York, ISBN: 0-471-86654-7, pp: 569.
- Johnson, N.L., S. Kotz and N. Balakrishnan, 1994. *Continuous Univariate Distributions. 2nd Edn.*, John Wiley and Sons, New York, ISBN: 978-0-471-58494-0, pp: 402.
- Riordan, J., 1978. *An Introduction to Combinatorial Analysis*. Princeton University Press, ISBN: 0-691-08262-6, pp: 52.
- Riordan, J., 1979. *Combinatorial Identities*. Robert E Krieger Pub. Co., ISBN: 0-88275-829-2, pp: 22.
- Tomescu, I. and S. Rodenau, 1975. *Introduction to Combinatorics*. Collets, England, ISBN: 0569080576.