

# On Distributions of Order Statistics from Nonidentical Discrete Variables

M. Güngör  
Department of Econometrics  
University of Inonu, Malatya, Turkey  
mgungor44@gmail.com

## Abstract

In this study, the distributions of  $r$ th order statistic of *innid* discrete random variables are obtained. In addition, the distributions are also expressed in the form of an integral. Then, the results related to *pdf* and *df* of minimum and maximum order statistics of *innid* discrete random variables are given.

**Keywords:** Order statistics, Discrete random variable, Probability function, Distribution function, Permanent.

**MSC 2010:** 62G30, 62E15.

## 1. Introduction

The joint probability density function(*pdf*) and marginal *pdf* of order statistics of independent but not necessarily identically distributed(*innid*) random variables was derived by Vaughan and Venables (1972) by means of permanents. In addition, Balakrishnan (2007), and Bapat and Beg (1989) obtained the joint *pdf* and distribution function(*df*) of order statistics of *innid* random variables by means of permanents. In the first of two papers, Balasubramanian et al. (1991) obtained the distribution of single order statistic in terms of distribution functions of the minimum and maximum order statistics of some subsets of  $\{X_1, X_2, \dots, X_n\}$  where  $X_i$ 's are *innid* random variables. Later, Balasubramanian et al. (1996) generalized their previous results (1991) to the case of the joint distribution function of several order statistics. Recurrence relationships among the distribution functions of order statistics arising from *innid* random variables were obtained by Cao and West (1997). Using multinomial arguments, the *pdf* of  $X_{r:n+1}$  ( $1 \leq r \leq n+1$ ) was obtained by Childs and Balakrishnan (2006) by adding another independent random variable to the original  $n$  variables  $X_1, X_2, \dots, X_n$ . Also, Balasubramanian et al. (1994) established the identities satisfied by distributions of order statistics from non-independent non-identical variables through operator methods based on the difference and differential operators. In a paper published in 1991, Beg obtained several recurrence relations and identities for product moments of order statistics of *innid* random variables using permanents. Recently, Cramer et al. (2009) derived the expressions for the distribution and density functions by Ryser's method and the distribution of maxima and minima based on permanents.

A multivariate generalization of classical order statistics for random samples from a continuous multivariate distribution was defined by Corley (1984). Guilbaud (1982) expressed the probability of the functions of *innid* random vectors as a linear combination of probabilities of the functions of independent and identically distributed(*iid*) random vectors and thus also for order statistics of random variables. Expressions for generalized joint densities of order statistics of *iid* random variables in terms of Radon-Nikodym derivatives with respect to product measures based on *df* were derived by Goldie and Maller (1999).

Several identities and recurrence relations for *pdf* and *df* of order statistics of *iid* random variables were established by numerous authors including Arnold et al. (1992), Balasubramanian and Beg (2003), David (1981), and Reiss (1989). Furthermore, Arnold et al. (1992), David (1981), Gan and Bain (1995), and Khatri (1962) obtained the probability function(*pf*) and *df* of order statistics of *iid* random variables from a discrete parent. Balakrishnan (1986) showed that several relations and identities that have been derived for order statistics from continuous distributions also hold for the discrete case. In a paper published in 1986, Nagaraja explored the behavior of higher order conditional probabilities of order statistics in a attempt to understand the structure of discrete order statistics. Later, Nagaraja (1992) considered some results on order statistics of a random sample taken from a discrete population.

In general, the distribution theory for order statistics is complicated when the discrete random variables are *innid*. In this study, the distributions of order statistics of *innid* discrete random variables are obtained.

As far as we know, these approaches have not been considered in the framework of order statistics from *innid* discrete random variables.

From now on, the subscripts and superscripts are defined in the first place in which they are used and these definitions will be valid unless they are redefined.

If  $a_1, a_2, \dots$  are defined as column vectors, then the matrix obtained by taking  $m_1$  copies of  $a_1$ ,  $m_2$  copies of  $a_2, \dots$  can be denoted as

$$\begin{bmatrix} a_1 & a_2 & \dots \\ m_1 & m_2 & \end{bmatrix}$$

and *per*A denotes the permanent of a square matrix A, which is defined as similar to determinant except that all terms in the expansion have a positive sign.

Let  $X_1, X_2, \dots, X_n$  be *innid* discrete random variables and  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  be the order statistics obtained by arranging the  $n$   $X_i$ 's ( $i = 1, 2, \dots, n$ ) in increasing order of magnitude. Let  $F_i$  and  $f_i$  be *df* and *pdf* of  $X_i$ , respectively.

The paper is organized as follows. In section 2, we give the theorems concerning  $pf$  and  $df$  of order statistics of *innid* discrete random variables. In the last section, some results related to  $pf$  and  $df$  will be given.

## 2. Probability and distribution functions

In this section, the theorems related to  $pf$  and  $df$  of  $X_{r:n}$  will be given. We will now express the following theorem for the  $pf$  of  $r$ th order statistic of *innid* discrete random variables.

### Theorem 2.1.

$$f_{r:n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(k+1+m)!(n-m-r)!} \text{per} \begin{bmatrix} F(x-) & f(x) & 1-F(x) \\ r-1-k & k+1+m & n-m-r \end{bmatrix}, \quad (1)$$

where  $F(x-) = (F_1(x-), F_2(x-), \dots, F_n(x-))'$ ,  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))'$ ,  $1-F(x) = (1-F_1(x), 1-F_2(x), \dots, 1-F_n(x))'$  are column vectors,  $F_i(x-) = P(X_i < x)$  and  $x = 0, 1, 2, \dots$ .

**Proof.** Consider the event

$$\{X_{r:n} = x\}, \quad r = 1, 2, \dots, n.$$

The above event can be realized mutually exclusive ways as follows:  $r-1-k$  observations are less than  $x$ ,  $k+1+m$  observations are equal to  $x$  and  $n-m-r$  observations are exceed  $x$  with respective probabilities  $F_i(x-)$ ,  $f_i(x)$  and  $1-F_i(x)$  ( $k = 0, 1, \dots, r-1$  and  $m = 0, 1, \dots, n-r$ ). Therefore, the  $pf$  of the above event can be written as

$$f_{r:n}(x) = P\{X_{r:n} = x\}. \quad (2)$$

(2) can be expressed as (1). Thus, the proof is completed.

In Theorem 2.1., the  $pf$  of  $r$ th order statistic from *innid* discrete random variables is given by using permanent.

Consider (1). Using expansion of permanent in (1), the following theorem can be expressed.

### Theorem 2.2.

$$f_{r:n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(k+1+m)!(n-m-r)!} \sum_P \left( \prod_{l_1=1}^{r-1-k} F_{i_{l_1}}(x-) \right) \left( \prod_{l_2=r-k}^{r+m} f_{i_{l_2}}(x) \right) \cdot \prod_{l_3=r+m+1}^n [1-F_{i_{l_3}}(x)], \quad (3)$$

where  $\sum_P$  denotes the sum over all  $n!$  permutations of  $(i_1, i_2, \dots, i_n)$  of  $(1, 2, \dots, n)$ .

**Proof.** (1) can also be expressed as

$$f_{r,n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(k+1+m)!(n-m-r)!} \sum_{n_{s_1}, n_{s_2}}^{r-1-k} \text{per}[F(x-)]_{[s_1/\cdot]} \\ \cdot \text{per}[f(x)]_{[s_2/\cdot]} \text{per}[1-F(x)]_{[s_3/\cdot]},$$

where  $\sum_{n_{s_1}, n_{s_2}}$  denotes the sum over  $\bigcup_{l=1}^2 s_l$  for which  $s_\nu \cap s_\nu = \emptyset$  for  $\nu \neq \nu, s = \bigcup_{l=1}^3 s_l$ ,

$s_1 = \{s_1^{(1)}, s_1^{(2)}, \dots, s_1^{(r-1-k)}\}$ ,  $s_2 = \{s_2^{(1)}, s_2^{(2)}, \dots, s_2^{(k+1+m)}\}$ ,  $s_3 = \{s_3^{(1)}, s_3^{(2)}, \dots, s_3^{(n-m-r)}\}$ ,  $s = \{1, 2, \dots, n\}$ ,  $s_l \neq \emptyset$  and  $n_{s_l}$  is the cardinality of  $s_l$ . Furthermore,  $A_{[s_l/\cdot]}$  is the matrix obtained from  $A$  by taking rows whose indices are in  $s_l \subset s$ . Using expansion of permanent in the above identity, we get (3). Thus, the proof is completed.

Theorem 2.2. can also be written in form of an integral as follows.

**Theorem 2.3.**

$$f_{r,n}(x) = \frac{1}{(r-1)!(n-r)!} \sum_P \int_{F_{i_r}(x-)}^{F_{i_r}(x)} \left( \prod_{l_1=1}^{r-1} v_{i_{l_1}} \right) \left( \prod_{l_2=r+1}^n [1-v_{i_{l_2}}] \right) dv_{i_r}, \quad (4)$$

where  $v_{i_j} = [v_{i_r} - F_{i_r}(x-)] \frac{f_{i_j}(x)}{f_{i_r}(x)} + F_{i_j}(x-)$ ,  $j = 1, 2, \dots, n$ .

**Proof.** Consider (3). It can be written as

$$f_{r,n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(k+1+m)!(n-m-r)!} \sum_P \left( \prod_{j_1=1}^{r-1-k} F_{i_{j_1}}(x-) \right) \left( \prod_{j_2=r-k}^{r-1} f_{i_{j_2}}(x) \right) \\ \cdot f_{i_r}(x) \left( \prod_{j_3=r+1}^{r+m} f_{i_{j_3}}(x) \right) \left( \prod_{j_4=r+m+1}^n [1-F_{i_{j_4}}(x)] \right) \int_0^1 \frac{(k+1+m)!}{k!m!} y^k (1-y)^m dy. \quad (5)$$

Also, (5) can be clearly written as

$$f_{r,n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(n-m-r)!k!m!} \sum_P \int_0^1 \left( \prod_{j_1=1}^{r-1-k} F_{i_{j_1}}(x-) \right) \left( \prod_{j_2=r-k}^{r-1} y f_{i_{j_2}}(x) \right) \\ \cdot f_{i_r}(x) \left( \prod_{j_3=r+1}^{r+m} (1-y) f_{i_{j_3}}(x) \right) \left( \prod_{j_4=r+m+1}^n [1-F_{i_{j_4}}(x)] \right) dy. \quad (6)$$

In (6), if  $v_{i_j} = y f_{i_j}(x) + F_{i_j}(x-)$ , the following identity is obtained

$$f_{r:n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!k!m!(n-m-r)!} \sum_P \int_{F_{i_r}(x-)}^{F_{i_r}(x)} \left( \prod_{j_1=1}^{r-1-k} F_{i_{j_1}}(x-) \right) \cdot \left( \prod_{j_2=r-k}^{r-1} [v_{i_{j_2}} - F_{i_{j_2}}(x-)] \right) \left( \prod_{j_3=r+1}^{r+m} [F_{i_{j_3}}(x) - v_{i_{j_3}}] \right) \left( \prod_{j_4=r+m+1}^n [1 - F_{i_{j_4}}(x)] \right) dv_{i_r}. \quad (7)$$

Now consider the following identity

$$\sum_{k=0}^{r-1} \frac{1}{(r-1-k)!k!} \sum_P \left( \prod_{j_1=1}^{r-1-k} A_{i_{j_1}}(x) \right) \prod_{j_2=r-k}^{r-1} B_{i_{j_2}}(x) = \frac{1}{(r-1)!} \sum_P \prod_{l=1}^{r-1} [A_{i_l}(x) + B_{i_l}(x)]. \quad (8)$$

Using (8) in (7), we get (4). Thus, the proof is completed.

We will now express the following theorem to obtain the  $df$  of  $r$ th order statistic from *innid* discrete random variables.

**Theorem 2.4.**

$$F_{r:n}(x) = \sum_{z=0}^x \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(k+1+m)!(n-m-r)!} \text{per} \begin{bmatrix} F(z-) & f(z) & 1-F(z) \\ r-1-k & k+1+m & n-m-r \end{bmatrix}. \quad (9)$$

**Proof.** It can be written

$$F_{r:n}(x) = \sum_{z=0}^x f_{r:n}(z) \quad (10)$$

and using (1) in (10), (9) is obtained.

Using expansion of permanent in the above theorem, we will establish the following theorem.

**Theorem 2.5.**

$$F_{r:n}(x) = \sum_{z=0}^x \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)!(k+1+m)!(n-m-r)!} \sum_P \left( \prod_{l_1=1}^{r-1-k} F_{i_{l_1}}(z-) \right) \left( \prod_{l_2=r-k}^{r+m} f_{i_{l_2}}(z) \right) \cdot \prod_{l_3=r+m+1}^n [1 - F_{i_{l_3}}(z)]. \quad (11)$$

**Proof.** Using (3) in (10), (11) is obtained.

(9) can also be written in form of an integral as follows.

**Theorem 2.6.**

$$F_{r:n}(x) = \frac{1}{(r-1)!(n-r)!} \sum_P \int_0^{F_{i_r}(x)} \left( \prod_{l_1=1}^{r-1} v_{i_{l_1}} \right) \left( \prod_{l_2=r+1}^n [1-v_{i_{l_2}}] \right) dv_{i_r}. \quad (12)$$

**Proof.** Using (4) in (10), (12) is obtained.

**3. Results for probability and distribution functions**

In this section, the results related to  $pf$  and  $df$  of  $X_{r:n}$  will be given. We will now express the following result for  $pf$  of minimum order statistic of *innid* discrete random variables.

**Result 3.1.**

$$\begin{aligned} f_{1:n}(x) &= \sum_{m=0}^{n-1} \frac{1}{(1+m)!(n-m-1)!} per[f(x)_{1+m} \quad 1-F(x)_{n-m-1}] \\ &= \sum_{m=0}^{n-1} \frac{1}{(1+m)!(n-m-1)!} \sum_P \left( \prod_{l_2=1}^{1+m} f_{i_{l_2}}(x) \right) \prod_{l_3=m+2}^n [1-F_{i_{l_3}}(x)] \\ &= \frac{1}{(n-1)!} \sum_P \int_{F_{i_1}(x-)}^{F_{i_1}(x)} \left( \prod_{l_2=2}^n [1-v_{i_{l_2}}] \right) dv_{i_1}. \end{aligned} \quad (13)$$

**Proof.** In (1), (3) and (4), if  $r=1$ , (13) is obtained.

We will express the following result for  $pf$  of maximum order statistic of *innid* discrete random variables.

**Result 3.2.**

$$\begin{aligned} f_{n:n}(x) &= \sum_{k=0}^{n-1} \frac{1}{(n-1-k)!(k+1)!} per[F(x-)_n \quad f(x)_{k+1}] \\ &= \sum_{k=0}^{n-1} \frac{1}{(n-1-k)!(k+1)!} \sum_P \left( \prod_{l_1=1}^{n-1-k} F_{i_{l_1}}(x-) \right) \prod_{l_2=n-k}^n f_{i_{l_2}}(x) \\ &= \frac{1}{(n-1)!} \sum_P \int_{F_{i_n}(x-)}^{F_{i_n}(x)} \left( \prod_{l_1=1}^{n-1} v_{i_{l_1}} \right) dv_{i_n}. \end{aligned} \quad (14)$$

**Proof.** In (1), (3) and (4), if  $r=n$ , (14) is obtained.

In the following result, we will express  $df$  of minimum order statistic of *innid* discrete random variables.

### Result 3.3.

$$\begin{aligned}
 F_{1:n}(x) &= \sum_{z=0}^x \sum_{m=0}^{n-1} \frac{1}{(1+m)!(n-m-1)!} \text{per} \begin{bmatrix} f(z) & 1-F(z) \\ 1+m & n-m-1 \end{bmatrix} \\
 &= \sum_{z=0}^x \sum_{m=0}^{n-1} \frac{1}{(1+m)!(n-m-1)!} \sum_P \left( \prod_{l_2=1}^{1+m} f_{i_{l_2}}(z) \right) \prod_{l_3=m+2}^n [1-F_{i_{l_3}}(z)] \\
 &= \frac{1}{(n-1)!} \sum_P \int_0^{F_{i_1}(x)} \left( \prod_{l_2=2}^n [1-v_{i_{l_2}}] \right) dv_{i_1}.
 \end{aligned} \tag{15}$$

**Proof.** In (9), (11) and (12), if  $r=1$ , (15) is obtained.

We will express the following result for  $df$  of maximum order statistic of *innid* discrete random variables.

### Result 3.4.

$$\begin{aligned}
 F_{n:n}(x) &= \sum_{z=0}^x \sum_{k=0}^{n-1} \frac{1}{(n-1-k)!(k+1)!} \text{per} \begin{bmatrix} F(z-) & f(z) \\ n-1-k & k+1 \end{bmatrix} \\
 &= \sum_{z=0}^x \sum_{k=0}^{n-1} \frac{1}{(n-1-k)!(k+1)!} \sum_P \left( \prod_{l_1=1}^{n-1-k} F_{i_{l_1}}(z-) \right) \prod_{l_2=n-k}^n f_{i_{l_2}}(z) \\
 &= \frac{1}{(n-1)!} \sum_P \int_0^{F_{i_n}(x)} \left( \prod_{l_1=1}^{n-1} v_{i_{l_1}} \right) dv_{i_n}.
 \end{aligned} \tag{16}$$

**Proof.** In (9), (11) and (12), if  $r=n$ , (16) is obtained.

### References

1. Arnold, B. C., Balakrishnan, N. & Nagaraja, H. N. (1992). *A first course in order statistics*. John Wiley and Sons Inc., New York.
2. Balakrishnan, N. (1986). Order statistics from discrete distributions. *Communications in Statistics-Theory and Methods* 15, 657-675.
3. Balakrishnan, N. (2007). Permanents, order statistics, outliers and robustness. *Revista Matemática Complutense* 20, 7-107.
4. Balasubramanian, K. & Beg, M. I. (2003). On special linear identities for order statistics. *Statistics* 37, 335-339.
5. Balasubramanian, K., Beg, M. I. & Bapat, R. B. (1991). On families of distributions closed under extrema. *Sankhya Ser. A* 53, 375-388.
6. Balasubramanian, K., Beg, M. I. & Bapat, R. B. (1996). An identity for the joint distribution of order statistics and its applications. *Journal of Statistical Planning and Inference* 55, 13-21.

7. Balasubramanian, K., Balakrishnan, N. & Malik, H. J. (1994). Identities for order statistics from non-independent non- identical variables. *Sankhya* Ser. B 56, 67-75.
8. Bapat, R. B. & Beg, M. I. (1989). Order statistics for nonidentically distributed variables and permanents. *Sankhya* Ser. A 51, 79-93.
9. Beg, M. I. (1991). Recurrence relations and identities for product moments of order statistics corresponding to nonidentically distributed variables. *Sankhya* Ser. A 53, 365-374.
10. Cao, G. & West, M. (1997). Computing distributions of order statistics. *Communications in Statistics-Theory and Methods* 26, 755-764.
11. Childs, A. & Balakrishnan, N. (2006). Relations for order statistics from non-identical logistic random variables and assessment of the effect of multiple outliers on bias of linear estimators. *Journal of Statistical Planning and Inference* 136, 2227-2253.
12. Corley, H. W. (1984). Multivariate order statistics. *Communications in Statistics-Theory and Methods* 13, 1299-1304.
13. Cramer, E., Herle, K. & Balakrishnan, N. (2009). Permanent Expansions and Distributions of Order Statistics in the INID Case. *Communications in Statistics - Theory and Methods* 38, 2078-2088.
14. David, H. A. (1981). *Order statistics*. John Wiley and Sons Inc., New York.
15. Gan, G. & Bain, L. J. (1995). Distribution of order statistics for discrete parents with applications to censored sampling. *Journal of Statistical Planning and Inference* 44, 37-46.
16. Goldie, C. M. & Maller, R. A. (1999). Generalized densities of order statistics. *Statistica Neerlandica* 53, 222-246.
17. Guilbaud, O. (1982). Functions of non-i.i.d. random vectors expressed as functions of i.i.d. random vectors. *Scandinavian Journal of Statistics* 9, 229-233.
18. Khatri, C. G. (1962). Distributions of order statistics for discrete case. *Annals of the Institute of Statistical Mathematics* 14, 167-171.
19. Nagaraja, H. N. (1986). Structure of discrete order statistics. *Journal of Statistical Planning and Inference* 13, 165-177.
20. Nagaraja, H. N. (1992). Order statistics from discrete distributions. *Statistics* 23, 189-216.
21. Reiss, R. D. (1989). *Approximate distributions of order statistics*. Springer-Verlag, New York,.
22. Vaughan, R. J. & Venables, W. N. (1972). Permanent expressions for order statistics densities, *Journal of the Royal Statistical Society* Ser. B 34, 308-310.