Distributions of Order Statistics

M. Güngör¹, Y. Bulut² and S. Çalık³

¹,² Department of Mathematics, University of Firat, 23119 Elazığ, Turkey
³ Department of Statistics, University of Firat, 23119 Elazığ, Turkey

Abstract

In this study, we examined distributions of order statistics corresponding to identically, non-identically, dependent, independent, discrete, continuous and discontinuous random variables and the relations among these distributions.

Mathematics Subject Classification: 62G30; 62E15

Keywords: Order statistics; Probability function; Probability density function; Distribution function; Permanent

1. Introduction

Balasubramanian et al.[1] established identities satisfied by distributions of order statistics from non-independent non-identical variables through operator methods based on difference and differential operators.

Childs and Balakrishnan[2] obtained, using multinomial arguments, the probability density function (p.d.f.) of \( X_{r:n+1} \) \((1 \leq r \leq n+1)\) if another independent random variable with distribution function (d.f.) \( F_i \) and p.d.f. \( f_i \) \((i = 1, 2, \ldots, n)\) is added to the original \( n \) variables \( X_1, X_2, \ldots, X_n \).


¹ Corresponding Author.
E-mail addresses: mgungor@firat.edu.tr (M. Güngör), ybulut79@hotmail.com (Y. Bulut), scalik@firat.edu.tr (S. Çalık).
Gan and Bain[4] obtained the joint probability function (p.f.) of any k order statistics and also conditional distributions of discrete order statistics from a general discrete parent by “tie-runs”.

Guilbaud[5] expressed probability of the functions of independent and not necessarily identically distributed (i.n.i.d) random vectors as a linear combination of probabilities of the functions of independent and identically distributed (i.i.d.) random vectors and thus also for order statistics of random variables.

Khatri[6] examined the p.f. and d.f. of a single order statistics, the joint p.f. and d.f. of any two order statistics and joint d.f. of any three order statistics of i.i.d. random variables from a discrete parent.

Reiss[7] considered the joint p.d.f., marginal p.d.f. and d.f. of any k order statistics of i.i.d. random variables under a continuous d.f. and discontinuous d.f. He also considered p.d.f. of bivariate order statistics by marginal ordering of bivariate i.i.d. random vectors with a continuous d.f. by means of multinomial probabilities of appropriate “cell frequency vectors”, defining multivariate order statistics by marginal ordering of i.i.d. random vectors with a continuous d.f.


Let $X_1, X_2, ..., X_n$ be random variables with $X_i$ having p.d.f. $f_i$ and d.f. $F_i$ for $i = 1, 2, ..., n$. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the corresponding order statistics and $F_i$ be discontinuous at $x$. We consider random variable

$$F_i(x^-) + \eta_i(F_i(x) - F_i(x^-))$$

which is uniformly distributed on the interval $(F_i(x^-), F_i(x))$ where $\eta_i$ is uniformly distributed on $(0,1)$. Here, $F_i(x^-)$ denotes the left-hand limit of $F_i$ at $x$. Moreover, $X_i$ and $\eta_i$ are assumed to be independent [7]. If $y_i$ is a realization of $\eta_i$, instead of $F_i(x)$ we may take the random variable

$$H_i(y, x) = F_i(x^-) + y_i(F_i(x) - F_i(x^-)).$$

Note that the discrete case occurs, if the number of discontinuous points is infinite.

If $a_1, a_2, ...$ are column vectors, then

$$[a_1 \quad a_2 \quad ...]_{i_1 \quad i_2 \quad ...}$$

will denote the matrix obtained by taking $i_1$ copies of $a_1$, $i_2$ copies of $a_2$ and so on. Finally, per$A$, where $A = (a_{ij})$ is $n \times n$ square matrix, denotes the permanent of $A$, i.e.,
where $S_n$ is the set of permutations of $\{1,2,...,n\}$. Thus the definition of permanent is equivalent to the determinant except that all signs in the expansion are positive[8].

It is known that distributions of order statistics is conveniently expressed in terms of permanents.

2. Distributions of order statistics

Now, the p.d.f. (or p.f.) and d.f. of order statistics in any case will be given. Theorem 2.1. and Theorem 2.2. can be proved using the same techniques in David [3], therefore we omitted.

**Theorem 2.1.** Let $F_i$ be discontinuous at $x$. The d.f. of $r$th order statistic of dependent random variables can be expressed as

$$H_{r:n}(y, x) = P\{X_{r:n} \leq x\} = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} H[\{i_1, i_2, \ldots, i_m\}]((y, x), (y, x), ..., (y, x)) \quad (2.1)$$

where $H[\{i_1, i_2, \ldots, i_m\}]$ is the joint d.f. of $X_{i_1}, X_{i_2}, \ldots, X_{i_m}$.

**Result 2.1.** If the random variables in (2.1) are continuous, the d.f. of the $r$th order statistic is given by

$$F_{r:n}(x) = P\{X_{r:n} \leq x\} = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} F[\{i_1, i_2, \ldots, i_m\}](x, x, ..., x) \quad (2.2)$$

where $F[\{i_1, i_2, \ldots, i_m\}]$ is the joint d.f. of $X_{i_1}, X_{i_2}, \ldots, X_{i_m}$.

**Result 2.2.** If the random variables in (2.1) are i.n.n.i.d., the d.f. of the $r$th order statistic is given by

$$H_{r:n}(y, x) = P\{X_{r:n} \leq x\} = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} H_{i_1}(y, x)H_{i_2}(y, x) \cdots H_{i_m}(y, x)$$
\[ \sum_{m=r}^{n} \frac{1}{m! (n-m)!} \text{per} \left[ \binom{H(y,x)}{m} \binom{1 - H(y,x)}{n-m} \right] \quad (2.3) \]

where \( H(y,x) = (H_1(y,x), H_2(y,x), ..., H_n(y,x))^\prime \) and \( 1 - H(y,x) = (1 - H_1(y,x), 1 - H_2(y,x), ..., 1 - H_n(y,x))^\prime \) are column vectors.

**Result 2.3.** If the random variables in (2.3) are identically distributed, the d.f. of \( r \)th order statistic is given by

\[ H_{r:n}(y,x) = P\{X_{r:n} \leq x\} \]

\[ = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \binom{n}{m} H^m(y,x) \]

\[ = \sum_{m=r}^{n} \binom{n}{m} H^m(y,x) (1 - H(y,x))^{n-m}. \quad (2.4) \]

**Result 2.4.** If the random variables in (2.2) are i.n.n.i.d., the d.f. of the \( r \)th order statistic is given by

\[ F_{r:n}(x) = P\{X_{r:n} \leq x\} \]

\[ = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} F_{i_1}(x) F_{i_2}(x) \cdots F_{i_m}(x) \]

\[ = \sum_{m=r}^{n} \frac{1}{m! (n-m)!} \text{per} \left[ \binom{F(x)}{m} \binom{1 - F(x)}{n-m} \right] \quad (2.5) \]

where \( F(x) = (F_1(x), F_2(x), ..., F_n(x))^\prime \) and \( 1 - F(x) = (1 - F_1(x), 1 - F_2(x), ..., 1 - F_n(x))^\prime \) are column vectors.

**Result 2.5.** If the random variables in (2.5) are identically distributed, the d.f. of \( r \)th order statistic is given by

\[ F_{r:n}(x) = P\{X_{r:n} \leq x\} \]

\[ = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \binom{n}{m} F^m(x) \]

\[ = \sum_{m=r}^{n} \binom{n}{m} F^m(x) (1 - F(x))^{n-m}. \quad (2.6) \]
Theorem 2.2. Let $P_i$ be d.f. of the discrete random variable $X_i$. The d.f. of the $r$th order statistic of dependent random variables can be expressed as

$$P_{r:n}(x) = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \sum_{1 \leq i_2 < \cdots < i_m \leq n} p[i_1, i_2, \ldots, i_m](x, x, \ldots, x)$$

(2.7)

where $p[i_1, i_2, \ldots, i_m]$ is the joint d.f. of $X_{i_1}, X_{i_2}, \ldots, X_{i_m}$.

Result 2.6. If the random variables in (2.7) are i.n.n.i.d., the d.f. of the $r$th order statistic is given by

$$P_{r:n}(x) = \sum_{x=0}^{n} \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{n!}{(r-1-k)! (k+m+1)! (n-r-m)!} \cdot \text{per}[P(x-1) p(x) 1 - P(x)] \quad \frac{1}{r-1-k} \frac{1}{k+m+1} \frac{1}{n-r-m}$$

(2.8)

where $P(x-1) = (P_1(x-1), P_2(x-1), \ldots, P_n(x-1))^\prime$, $p(x) = (p_1(x), p_2(x), \ldots, p_n(x))^\prime$ and $1 - P(x) = (1 - P_1(x), 1 - P_2(x), \ldots, 1 - P_n(x))^\prime$ are column vectors. Moreover, $P_i(x-1) = 0$ for $x = 0$.

Result 2.7. If the random variables in (2.8) are identically distributed, the d.f. of the $r$th order statistic is given by

$$P_{r:n}(x) = \sum_{x=0}^{n} \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{n!}{(r-1-k)! (k+m+1)! (n-r-m)!} \cdot \text{per}\{P(x-1) p(x) 1 - P(x)\} \quad \frac{1}{r-1-k} \frac{1}{k+m+1} \frac{1}{n-r-m}$$

(2.9)

Result 2.8. If the number of the random variables in Result 2.6 are increased for continuity and $k = m = 0$, the d.f. of the $r$th order statistic is obtained as in (2.5).

Result 2.9. If the random variables in Result 2.8 are identically distributed, the d.f. of the $r$th order statistic is obtained as in (2.6).

Result 2.10. If we find the derivative of (2.3) with respect to the argument $x$, the p.d.f. of the $r$th order statistic is given by

$$h_{r:n}(y, x) = \frac{1}{(r-1)! (n-r)!} \text{per}[H(y, x) h(y, x) 1 - H(y, x)]$$

(2.10)
where $\mathbf{h}(y, x) = (h_1(y, x), h_2(y, x), ..., h_n(y, x))'$ is column vector and $h_i(y, x) = f_i(x^-) + y_i(f_i(x) - f_i(x^-))$.

**Result 2.11.** If the random variables in (2.10) are identically distributed, the p.d.f. of the $r$th order statistic is given by

$$h_{r:n}(y, x) = \frac{n!}{(r-1)! (n-r)!} H^{r-1}(y, x)[1 - H(y, x)]^{n-r}h(y, x).$$  \hspace{1cm} (2.11)

**Result 2.12.** If we find the derivative of (2.5) with respect to the argument $x$, the p.d.f. of the $r$th order statistic is given by

$$f_{r:n}(x) = \frac{1}{(r-1)! (n-r)!} \text{per}[F(x)] \frac{f(x)}{r-1} \frac{1 - F(x)}{n-r}$$  \hspace{1cm} (2.12)

where $f(x) = (f_1(x), f_2(x), ..., f_n(x))'$ is column vector.

**Result 2.13.** If the random variables in (2.12) are identically distributed, the p.d.f. of the $r$th order statistic is given by

$$f_{r:n}(x) = \frac{n!}{(r-1)! (n-r)!} F^{r-1}(x)[1 - F(x)]^{n-r}f(x).$$

**Result 2.14.** From (2.8), the p.f. of the $r$th order statistic is given by

$$p_{r:n}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{1}{(r-1-k)! (k+m+1)! (n-r-m)!} \text{per}[p(x-1)] p(x) \frac{1 - p(x)}{n-r-m}.$$  \hspace{1cm} (2.13)

**Result 2.15.** If the random variables in (2.13) are identically distributed, the p.f. of the $r$th order statistic is given by

$$p_{r:m}(x) = \sum_{k=0}^{r-1} \sum_{m=0}^{n-r} \frac{n!}{(r-1-k)! (k+m+1)! (n-r-m)!} p^{r-1-k}(x-1)p^{k+m+1}(x)[1 - p(x)]^{n-r-m}$$

$$= \frac{n!}{(r-1)! (n-r)!} \int_{p(x-1)}^{p(x)} w^{r-1} (1 - w)^{n-r} dw.$$  \hspace{1cm} (2.16)

**Result 2.16.** If the number of the random variables in Result 2.14 are increased for continuity and $k = m = 0$, the p.d.f. of the $r$th order statistic is obtained as in (2.12).
Result 2.17. If the random variables in Result 2.16 are identically distributed, the p.d.f. of the \( r \)th order statistic is obtained as in (2.13).

Result 2.18. If we integrate (2.10) from \(-\infty\) to \(x\), the d.f. of the \( r \)th order statistic is given by

\[
H_r:n(y, x) = \int_{-\infty}^{x} h_{r:n}(y, t) dt
\]

\[
= \int_{-\infty}^{x} \frac{1}{(r-1)! (n-r)!} p_{\text{per}}[H(y, t) \frac{h(y, t)}{r-1} 1 \frac{1-H(y, t)}{n-r}] dt
\]

\[
= \int_{-\infty}^{x} \frac{1}{(r-1)! (n-r)!} p_{\text{per}}[H(y, t) \frac{dH(y, t)}{r-1} 1 \frac{1-H(y, t)}{n-r}] dt
\]

\[
= \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_m \leq n} H_{i_1}(y, x)H_{i_2}(y, x) \ldots H_{i_m}(y, x)
\]

\[
= \sum_{m=r}^{n} \frac{1}{m! (n-m)!} p_{\text{per}}[H(y, x) \frac{1-H(y, x)}{m} 1 \frac{n-m}{n-r}]
\]

where \( dH(y, t) = h(y, t) dt \) is column vector.

Result 2.19. If we integrate (2.11) from \(-\infty\) to \(x\), the d.f. of the \( r \)th order statistic is given by

\[
H_r:n(y, x) = \int_{-\infty}^{x} h_{r:n}(y, t) dt
\]

\[
= \int_{-\infty}^{x} \frac{n!}{(r-1)! (n-r)!} H^{r-1}(y, t)[1-H(y, t)]^{n-r} h(y, t) dt. \quad (2.14)
\]

Then, taking \( H(y, t) = u \) and writing its binom expansion instead of \((1-u)^{n-r}\) in (2.14), the d.f. can be expressed as

\[
H_r:n(y, x) = \sum_{j=0}^{n-r} \int_{0}^{H(y,x)} \binom{n-r}{j} \frac{n!}{(r-1)! (n-r)!} u^{r-1+j} (-1)^j du.
\]

Then, taking \( r + j = m \), the d.f. of the \( r \)th order statistic, as in (2.4), is obtained as
\[ H_{r:n}(y, x) = \sum_{m=r}^{n} (-1)^{m-r} \binom{m-1}{r-1} \binom{n}{m} H^m(y, x) \]

\[ = \sum_{m=r}^{n} \binom{n}{m} H^m(y, x) (1 - H(y, x))^{n-m}. \]

**Result 2.20.** If we integrate (2.12) from \(-\infty\) to \(x\), the d.f. of the \(r\)th order statistic is obtained as in (2.5).

**Result 2.21.** If the random variables in Result 2.20 are identically distributed, the d.f. of the \(r\)th order statistic is obtained as in (2.6).

**Result 2.22.** If we sum (2.13) from zero to \(x\), the d.f. of the \(r\)th order statistic is obtained as in (2.8).

**Result 2.23.** If the random variables in Result 2.22 are identically distributed, the d.f. of the \(r\)th order statistic is obtained as in (2.9).

**References**


Received: September, 2008