Optimal Formula about Ordering the Random Variables Problem

Ivan S. Latif [1]

[1]Department of Mathematics, College of Education, Salahaddin University, Erbil, Iraq
Correspondence: Ivan S. Latif, Salahaddin University, Erbil, Iraq.
Email: ivan.latif@su.edu.krd.

Received: October 3, 2018 Accepted: November 20, 2018 Online Published: December 1, 2019

doi: 10.23918/eajse.v4i2p10

Abstract: In this paper, we propose an algorithm about ordering the random variables into order statistics and find the simple formula of multi order statistics joint probability distribution and endeavor to prove it mathematically. The basic idea is to use the mathematical induction to find the joint probability order statistic distribution. The study found that the new method can be employed in Mathematics.

Key words: Random Variable, Oriented Algorithms, Order Statistics

1. Introduction

Order statistics and their properties have been studied rather extensively since the early part of the last century. Yet, most of these studies focused only on the case when order statistics are from independent and identically distributed (IID) random variables. Motivated by robustness issues, studies of order statistics from outlier models began in early 70s (Barry, Balakrishnan, & Nagaraja, 2008); Larson & Marx, 2012). Though much of the early work in this direction concentrated only on the case when there is one outlier in the sample (single-outlier model), there has been a lot of work during the past fifteen years or so on multiple-outlier models and more generally on order statistics from independent and non-identically distributed (INID) random variables. These results have also enabled useful and interesting discussions on the robustness of different estimators of parameters of a wide range of distributions (Larson & Marx, 2012; Mood, Graybill, & Boes, 1974).

A form of the joint distribution of n ordered random variables is presented that enables a unified approach to a variety of models of ordered random variables, e.g. order statistics and record values (Lindgren, 1976). Several other models are shown. In particular, sequential order statistics are introduced as a modification of order statistics which is naturally suggested by a statistical application in reliability theory. In the distribution theoretical sense, all of these models of ordered random variables are contained in the proposed concept of generalized order statistics. Numerous related results on distribution of order statistics and record values are found in the literature which is deduced separately (Miller & Freund, 1965; Wani, 1971). Generalized order statistics, however, provide a suitable approach to explain and to generalize related results. Through integration of known properties the structure of the embedded models becomes clearer. On the other hand, the validity of these properties and their generalizations for generalized order statistics, and hence for different models of ordered random variables (Goldstein, David, & Martha, 2001) have been obtained. In the present paper we develop the distribution theory for generalized order statistics. Representations for the one-, two- and higher-dimensional marginal densities and a form of the one-dimensional marginal distribution function is given as well as recurrence relations for marginal densities and distribution functions.

2. Order Statistics

The order statistics of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted by $Y_1, \ldots, Y_n$.

The order statistics are random variables that satisfy $Y_1 \leq \cdots \leq Y_n$. In particular,

$Y_1 = \min X_i$, $1 \leq i \leq n$

$Y_2 =$second smallest X_i, $1 \leq i \leq n$

$\vdots$

$Y_n =\max X_i$, $1 \leq i \leq n$

Theorem (2-1):-Let $Y_1, \ldots, Y_n$ denote the order statistics of a random sample, $X_1, \cdots, X_n$, from a continuous population with cdf $F(y)$ and pdf $f(y)$[3], [5]. Then the pdf of Y_i is

$$g(y_i) = \frac{n!}{(i-1)!(n-i)!} f(y_i)[F(y_i)]^{i-1}[1 - F(y_i)]^{n-i} \tag{1}$$

Theorem (2-2):-Let $Y_1, \ldots, Y_n$ denote the order statistics of a random sample, $X_1, \cdots, X_n$, from a continuous population with cdf $F(y)$ and pdf $f(y)$[3], [5]. Then the joint pdf of Y_i and Y_j, $1 \leq i \leq j \leq n$, is

$$g(y_i, y_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \{F(y_i)\}^{i-1} \{F(y_j) - F(y_i)\}^{j-i-1} \{1 - F(y_j)\}^{n-j} f(y_i) f(y_j) \tag{2}$$

$$-\infty < y_i < y_j < \infty.$$

Theorem (2-3):- Let $Y_1, \cdots, Y_n$ denote the order statistics of a random sample, $X_1, \cdots, X_n$, from a continuous population with cdf $F(y)$ and pdf $f(y)$[5]. Then the joint pdf of Y_i, Y_j, and Y_k ($1 \leq i \leq j \leq k \leq n$) is

$$g(y_i, y_j, y_k) = \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(n-k)!} \times \{F(y_i)\}^{i-1} \{F(y_j) - F(y_i)\}^{j-i-1}$$

$$\times \{F(y_k) - F(y_j)\}^{k-j-1} \{1 - F(y_k)\}^{n-k} f(y_i) f(y_j) f(y_k), \tag{3}$$
$$-\infty < y_i < y_j < y_k < \infty$$

In this section we propose new method for find general forms about joint probability density function of k-random variables.

Theorem(3):- Let $Y_1, \cdots, Y_n$ denote the order statistics of a random sample, $X_1, \cdots, X_n$, from a continuous population with cdf $F(y)$ and pdf $f(y)$. Then the joint pdf of Y_i, Y_j, Y_k, and Y_l ($1 \leq i \leq j \leq k \leq \cdots \leq l \leq n$) is

$$f_{i,j,k,l:n}(y_i, y_j, y_k, y_l, y_n) = \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(l-k-1)!\,(n-l)!} \times \{F(y_i)\}^{i-1}\{F(y_j) - F(y_i)\}^{j-i-1}$$

$$\times \{F(y_k) - F(y_j)\}^{k-j-1}\{F(y_l) - F(y_k)\}^{l-k-1}\{1 - F(y_l)\}^{n-l} f(y_i) f(y_j) f(y_k) f(y_l), \quad (4)$$

$$-\infty < y_i < y_j < y_k < y_l < \infty$$

Proof: By the same way we can prove this theorem using mathematical induction for one and two, three joint probability distribution and assumed for $k - order\ statistics$and proved for more than $k - order\ statistics$,that is

Integrating out the variables $(Y_{1:n}, \cdots, Y_{i-1:n})$, $(Y_{i+1:n}, \cdots, Y_{j-1:n})$, $(Y_{j+1:n}, \cdots, Y_{k-1:n})$, $(Y_{k+1:n}, \cdots, Y_{l-1:n})$, and $(Y_{l+1:n}, \cdots, Y_{n:n})$, then

$$f_{i,j,k,l:n}(y_i, y_j, y_k, y_l) = n!\, f(y_i) f(y_j) f(y_k) f(y_l) \left\{\int_{-\infty}^{y_i} \cdots \int_{-\infty}^{y_2} f(y_1) \cdots f(y_{i-1}) dy_1 \cdots dy_{i-1}\right\}$$

$$\times \left\{\int_{y_i}^{y_j} \cdots \int_{y_i}^{y_{i+2}} f(y_{i+1}) \cdots f(y_{j-1}) dy_{i+1} \cdots dy_{j-1}\right\}$$

$$\times \left\{\int_{y_j}^{y_k} \cdots \int_{y_j}^{y_{j+2}} f(y_{j+1}) \cdots f(y_{k-1}) dy_{j+1} \cdots dy_{k-1}\right\}$$

$$\times \left\{\int_{y_k}^{y_l} \cdots \int_{y_k}^{y_{k+2}} f(y_{k+1}) \cdots f(y_{l-1}) dy_{k+1} \cdots dy_{l-1}\right\}$$

$$\times \left\{\int_{y_l}^{\infty} \cdots \int_{y_l}^{y_{l+2}} f(y_{l+1}) \cdots f(y_n) dy_{l+1} \cdots dy_n\right\}. \quad (5)$$

By direct integration we get

$$\int_{-\infty}^{y_2} f(y_1) dy_1 = F(y_1) \Big|_{-\infty}^{y_2}$$

$$= F(y_2) - F(-\infty)$$

$$= F(y_2) - 0$$

$$= F(y_2),$$

$$\int_{-\infty}^{y_3} F(y_2) f(y_2) dy_2 = \frac{[F(y_2)]^2}{2} \Big|_{-\infty}^{y_3}$$

$$= \frac{[F(y_3)]^2 - [F(-\infty)]^2}{2}$$

$$= \frac{[F(y_3)]^2 - [0]^2}{2}$$

$$= \frac{[F(y_3)]^2}{2!},$$

$$\int_{-\infty}^{y_4} \frac{[F(y_3)]^2}{2!} f(y_3) dy_3 = \frac{[F(y_3)]^3}{2! \times 3} \Big|_{-\infty}^{y_4}$$

$$= \frac{[F(y_4)]^3 - [F(-\infty)]^3}{3!}$$

$$= \frac{[F(y_4)]^3}{3!},$$

$$\vdots$$

$$\int_{-\infty}^{y_i} \frac{[F(y_{i-1})]^{i-2}}{(i-2)!} f(y_{i-1}) dy_{i-1} = \frac{[F(y_{i-1})]^{i-2+1}}{(i-2)!\,(i-2+1)}$$

$$= \frac{[F(y_i)]^{i-1} - [F(-\infty)]^{i-1}}{(i-2)!\,(i-1)}$$

$$= \frac{[F(y_i)]^{i-1} - [0]^{i-1}}{(i-1)!}$$

$$= \frac{[F(y_i)]^{i-1}}{(i-1)!},$$

$$\therefore \int_{-\infty}^{y_i} \cdots \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} f(y_1) f(y_2) \cdots f(y_{i-1}) dy_1 dy_2 \cdots dy_{i-1} = \frac{\{F(y_i)\}^{i-1}}{(i-1)!},$$

$$\int_{y_i}^{y_{i+2}} f(y_{i+1})\ dy_{i+1} = F(y_{i+1}) \Big|_{y_i}^{y_{i+2}}$$

$$= F(y_{i+2}) - F(y_i),$$

$$\int_{y_i}^{y_{i+3}} [F(y_{i+2}) - F(y_i)] f(y_{i+2})\ dy_{i+2} = \frac{[F(y_{i+2}) - F(y_i)]^2}{2} \Big|_{y_i}^{y_{i+3}}$$

$$= \frac{[F(y_{i+3}) - F(y_i)]^2 - [F(y_i) - F(y_i)]^2}{2}$$

$$= \frac{[F(y_{i+3}) - F(y_i)]^2}{2!},$$

$$\int_{y_i}^{y_{i+4}} \frac{[F(y_{i+3}) - F(y_i)]^2}{2!} f(y_{i+3})\ dy_{i+3} = \frac{[F(y_{i+3}) - F(y_i)]^{2+1}}{2! \times 3} \Bigg|_{y_i}^{y_{i+4}}$$

$$= \frac{[F(y_{i+4}) - F(y_i)]^3 - [F(y_i) - F(y_i)]^3}{6}$$

$$= \frac{[F(y_{i+4}) - F(y_i)]^3}{3!}\ ,$$

$$\vdots$$

$$\int_{y_i}^{y_j} \frac{[F(y_{j-1}) - F(y_i)]^{j-i-2}}{(j-i-2)!} f(y_{j-1})\ dy_{j-1} = \frac{[F(y_{j-1}) - F(y_i)]^{j-i-2+1}}{(j-i-2)! \times (j-i-2+1)} \Bigg|_{y_i}^{y_j}$$

$$= \frac{[F(y_j) - F(y_i)]^{j-i-1} - [F(y_i) - F(y_i)]^{j-i-1}}{(j-i-1)!}$$

$$= \frac{[F(y_j) - F(y_i)]^{j-i-1}}{(j-i-1)!},$$

$$\int_{y_i}^{y_j} \cdots \int_{y_i}^{y_{i+3}} \int_{y_i}^{y_{i+2}} f(y_{i+1}) f(y_{i+2}) \cdots f(y_{j-1}) dy_{i+1} dy_{i+2} \cdots dy_{j-1} = \frac{\{F(y_j) - F(y_i)\}^{j-i-1}}{(j-i-1)!},$$

$$\int_{y_j}^{y_{j+2}} f(y_{j+1})\ dy_{j+1} = F(y_{j+1}) \Bigg|_{y_j}^{y_{i+2}}$$

$$= F(y_{j+2}) - F(y_j),$$

$$\int_{y_j}^{y_{j+3}} [F(y_{j+2}) - F(y_j)] f(y_{j+2})\ dy_{j+2} = \frac{[F(y_{j+2}) - F(y_j)]^2}{2} \Bigg|_{y_j}^{y_{j+3}}$$

$$= \frac{[F(y_{j+3}) - F(y_j)]^2 - [F(y_j) - F(y_j)]^2}{2}$$

$$= \frac{[F(y_{j+3}) - F(y_j)]^2}{2!}\ ,$$

$$\int_{y_j}^{y_{j+4}} \frac{[F(y_{j+3}) - F(y_j)]^2}{2!} f(y_{j+3})\ dy_{j+3} = \frac{[F(y_{j+3}) - F(y_j)]^{2+1}}{2! \times 3} \Bigg|_{y_j}^{y_{j+4}}$$

$$= \frac{[F(y_{j+4}) - F(y_j)]^3 - [F(y_j) - F(y_j)]^3}{6}$$

$$= \frac{[F(y_{j+4}) - F(y_j)]^3}{3!} \ ,$$

$$\vdots$$

$$\int_{y_j}^{y_k} \frac{[F(y_{k-1}) - F(y_j)]^{j-i-2}}{(k-j-2)!} f(y_{k-1}) \ dy_{k-1} = \frac{[F(y_{k-1}) - F(y_j)]^{k-j-2+1}}{(k-j-2)! \times (k-j-2+1)} \Big|_{y_j}^{y_k}$$

$$= \frac{[F(y_k) - F(y_j)]^{k-j-1} - [F(y_j) - F(y_j)]^{k-j-1}}{(k-j-1)!}$$

$$= \frac{[F(y_k) - F(y_j)]^{k-j-1}}{(k-j-1)!},$$

$$\int_{y_j}^{y_k} \cdots \int_{y_j}^{y_{j+3}} \int_{y_j}^{y_{j+2}} f(y_{j+1}) f(y_{j+2}) \cdots f(y_{k-1}) dy_{j+1} dy_{j+2} \cdots dy_{k-1}$$

$$= \frac{\{F(y_k) - F(y_j)\}^{k-j-1}}{(k-j-1)!},$$

$$\int_{y_k}^{y_{k+2}} f(y_{k+1}) \ dy_{k+1} = F(y_{k+1}) \Big|_{y_k}^{y_{k+2}}$$

$$= F(y_{k+2}) - F(y_k),$$

$$\int_{y_k}^{y_{k+3}} [F(y_{k+2}) - F(y_k)] f(y_{k+2}) \ dy_{k+2} = \frac{[F(y_{k+2}) - F(y_k)]^2}{2} \Big|_{y_k}^{y_{k+3}}$$

$$= \frac{[F(y_{k+3}) - F(y_k)]^2 - [F(y_k) - F(y_k)]^2}{2}$$

$$= \frac{[F(y_{k+3}) - F(y_k)]^2}{2!} \ ,$$

$$\int_{y_k}^{y_{k+4}} \frac{[F(y_{k+3}) - F(y_k)]^2}{2!} f(y_{k+3}) \ dy_{k+3} = \frac{[F(y_{k+3}) - F(y_k)]^{2+1}}{2! \times 3} \Big|_{y_k}^{y_{k+4}}$$

$$= \frac{[F(y_{k+4}) - F(y_k)]^3 - [F(y_k) - F(y_k)]^3}{6}$$

$$= \frac{[F(y_{k+4}) - F(y_k)]^3}{3!} \ ,$$

$$\vdots$$

$$\int_{y_k}^{y_l} \frac{[F(y_{l-1}) - F(y_k)]^{l-k-2}}{(l-k-2)!} f(y_{l-1}) \ dy_{l-1} = \frac{[F(y_{l-1}) - F(y_k)]^{l-k-2+1}}{(l-k-2)! \times (l-k-2+1)} \Big|_{y_k}^{y_l}$$

$$= \frac{[F(y_l) - F(y_j)]^{l-j-1} - [F(y_k) - F(y_k)]^{l-k-1}}{(l-k-1)!}$$

$$= \frac{[F(y_l) - F(y_k)]^{l-k-1}}{(l-k-1)!}$$

$$\int_{y_k}^{y_l} \cdots \int_{y_k}^{y_{k+3}} \int_{y_k}^{y_{k+2}} f(y_{k+1}) f(y_{k+2}) \cdots f(y_{l-1}) dy_{k+1} dy_{k+2} \cdots dy_{l-1} = \frac{\{F(y_l) - F(y_k)\}^{l-k-1}}{(l-k-1)!},$$

and

$$\int_{y_l}^{y_{l+2}} f(y_{l+1})\ dy_{l+1} = F(y_{l+1}) \Big|_{y_l}^{y_{l+2}}$$

$$= F(y_{l+2}) - F(y_l),$$

$$\int_{y_l}^{y_{l+3}} [F(y_{l+2}) - F(y_l)] f(y_{l+2})\ dy_{l+2} = \frac{[F(y_{l+2}) - F(y_l)]^2}{2} \Big|_{y_l}^{y_{l+3}}$$

$$= \frac{[F(y_{l+3}) - F(y_l)]^2 - [F(y_l) - F(y_l)]^2}{2}$$

$$= \frac{[F(y_{l+3}) - F(y_l)]^2}{2!},$$

$$\int_{y_l}^{y_{l+4}} \frac{[F(y_{l+3}) - F(y_l)]^2}{2!} f(y_{l+3})\ dy_{l+3} = \frac{[F(y_{l+3}) - F(y_l)]^3}{2! \times 3} \Big|_{y_l}^{y_{l+4}}$$

$$= \frac{[F(y_{l+4}) - F(y_l)]^3 - [F(y_l) - F(y_l)]^3}{3!}$$

$$= \frac{[F(y_{l+4}) - F(y_l)]^3}{3!},$$

$$\vdots$$

$$\int_{y_l}^{y_n} \frac{[F(y_{n-1}) - F(y_l)]^{n-l-2}}{(n-l-2)!} f(y_{n-1})\ dy_{n-1} = \frac{[F(y_{n-1}) - F(y_l)]^{n-l-2+1}}{(n-l-2)!\,(n-l-2+1)} \Big|_{y_l}^{y_n}$$

$$= \frac{[F(y_n) - F(y_l)]^{n-l-1} - [F(y_l) - F(y_l)]^{n-l-1}}{(n-l-1)!}$$

$$= \frac{[F(y_n) - F(y_l)]^{n-l-1}}{(n-l-1)!},$$

$$\int_{y_l}^{\infty} \frac{[F(y_n) - F(y_l)]^{n-l-1}}{(n-l-1)!} f(y_n)\ dy_n = \frac{[F(y_n) - F(y_l)]^{n-l-1+1}}{(n-l-1)!\,(n-l-1+1)} \Big|_{y_l}^{\infty}$$

$$= \frac{[F(\infty) - F(y_l)]^{n-l} - [F(y_l) - F(y_l)]^{n-l}}{(n-l)!}$$

$$= \frac{[1 - F(y_l)]^{n-l}}{(n-l)!},$$

$$\int_{y_l}^{\infty} \cdots \int_{y_l}^{y_{l+3}} \int_{y_l}^{y_{l+2}} f(y_{l+1}) f(y_{l+2}) \cdots f(y_n) dy_{l+1} dy_{l+2} \cdots dy_n = \frac{\{1 - F(y_l)\}^{n-l}}{(n-l)!}$$

By substituting the above results of integrations in the Eq.(5) we get the joint pdf of Y_i, Y_j, Y_k, and Y_l $(1 \leq i \leq j \leq k \leq l \leq n)$, i.e

$$f_{i,j,k,l:n}(y_i, y_j, y_k, y_l)$$
$$= \frac{n!}{(i-1)!\,(j-i-1)!\,(k-j-1)!\,(l-k-1)!\,(n-l)!} \times \{F(y_i)\}^{i-1} \{F(y_j) - F(y_i)\}^{j-i-1}$$

$$\times \{F(y_k) - F(y_j)\}^{k-j-1} \{F(y_l) - F(y_k)\}^{l-k-1} \{1 - F(y_l)\}^{n-l} f(y_i) f(y_j) f(y_k) f(y_l),$$

3. Algorithm to our Technique

This algorithm to obtain the optimal solution formula defined earlier can be summarized as follows:

Step 1: Assign arbitrary value to each of the random variables$X_1, X_2, \dots, X_n$.and its order statistics $Y_1, Y_2, \dots, Y_n$

Step 2: Compute probability distribution function of random variable and its order statistics.

Step 3: Compute the commutative probability distribution function of random variable and its order statistics.

Step 4: Set the size number of order statistics to compute the joint probability distribution of order statistics.

Step 5: Evaluate the joint probability distribution of K −order statistics use the formula (4).

Step 6: Print the joint probability distribution of K −order statistics.

Step 7**:** Stop.

4. Conclusion and Recommendation

We introduce some topics about order statistics which are given by statisticians or mathematicians to construct some results about the probability distribution function. Therefore generally the order statistics reduced random variable X_i ,$i = 1,2, \dots, n$ into ordered or arranged in increasing order, such that the smallest of the X_i s 'is denoted by $Y_1$and the largest denoted by Y_n. In other words, we may change X_i s 'into $Y_1 \leq Y_2 \leq \cdots \leq Y_n$, and we use them to construct a new probability distribution function a bout $Y_i, i = 1,2, \dots, n$. Let $X_i$be identically independent distribution function about Y_i such that $Y_i \sim g(Y_i)$. So in these work we propose an algorithm about ordering the general forms about p.d.f. of random variables Y_i, for example about one variable, two variables, and $K -$ variables and find the simple formula of multi order statistics joint probability distribution and finally we proved mathematically.

In this paper only general forms about joint probability distribution of K −random variables order statistics is considered, it is recommended to extend these proposed procedures to find simple formula and prove for the conditional joint probability distribution and also find special formula for each famous distribution of K −random variables order statistics.

References

Balakrishnan, N., & Nevsorov,V. B. (2006). *A primer on statistical distributions*. New Jersey: Wiley.

Barry, C. A., Balakrishnan, N., & Nagaraja, H. N. (2008). *A first course in order statistics*. Philadelphia: Society for Industrial and Applied Mathematics.

David, H. A., & Nagaraja, H. A. (2003). *Order statistics*. 3-rd Edition. New Jersey: Wiley.

Goldstein, L. J., David, I. S., & Martha, S. (2001). *Finite mathematics and applications* (7th Edition). Prentice-Hall.

Hogg, R., & Graig, A. (1978). *Introduction to mathematical Statistics*. New York: Collier Macmillan 4th Edition.

Larson, R. J., & Marx, L. M. (2012). *An introduction to mathematical statistics and applications*. 5th Edition, New York: Prentice Hall.

Lindgren, B. W. (1976). *Statistical theory*. 3rd Edition. New York: Macmillan.

Miller, I., & Freund, J. E. (1965). *Probability and statistics for engineers*. New Jersey: Prentice-Hall, Englewood Cliffs.

Mood, A, M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. 3rd Edition. New York: Mc Graw-Hill.

Wani, J. (1971). *Probability & statistical inference*. USA: Meredith Corporation.