Bounds for CDFs of Order Statistics Arising from INID Random Variables

Jaber Kazempoor [1]**, Arezou Habibirad** [1]**, and Kheirolah Okhli** [1]

[1] Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran.

Abstract. In recent decades, studying order statistics arising from independent and not necessary identically distributed (INID) random variables has been a remarkable concern for researchers. The cumulative distribution function (CDF) of these random variables ($F_{i:n}$) is a complex manipulating, long time consuming and a software-intensive tool that takes considerable time. Therefore, obtain approximations and bounds for $F_{i:n}$ and other theoretical properties of these variables, such as moments, quantiles, characteristic functions, and some related probabilities, has always been the main challenge. Recently, Bayramoglu (2018), Bayramoglu (2018), has introduced a set of CDFs ($F_{[i]}$), whose definitions are based on a point to point ordering of the original CDFs (F_i), that can be used to approximate the CDF of i-th order statistics ($F_{i:n}$). Here, by using just $F_{[1]}$ and $F_{[n]}$, we provide new upper and lower bounds for the $F_{i:n}$. Furthermore, new approximations for $F_{1:n}$ and $F_{n:n}$, as well as for other cases, are derived. Comparisons with respect to approximations suggested by Bayramoglu Bayramoglu (2018) are also provided.

Keywords. Approximation, Bounds, Cumulative Distribution Function, Independent and not Necessary Identically Distributed, Order Statistics.

Jaber Kazempoor (kazempoor.jaber@mail.um.ac.ir)
Corresponding Author: Arezou Habibirad (ahabibi@um.ac.ir)
Kheirolah Okhli (kh.okhli@mail.um.ac.ir)

MSC: 62E17, 62E99, 62G07.

1   Introduction

Order statistics play a significant role in the mathematical statistics and other disciplines like in issues such as the investigation of natural disasters, the lifespan of a coherent system, extreme values, records, time series, discussion about the range of variables, and so forth. In all of the mentioned cases, we have to use the functions of a variety of ordered statistics. Order statistics arising from independent and identically distributed (IID) random variables have been studied in many sources, such as Ahsanullah et al. (2013), Arnold et al. (1992), and David and Nagaraja (2004), but in some situations, we must assess the behavior of order statistics arising from independent and not necessary identically distributed (INID) random variables. In this case, some theoretical properties, such as relations and formulas quickly became complicated and overwhelming (see David and Nagaraja (2004) and Reiss (2012)). However, many improvements have been made in the theoretical part of these cases, and although many formulas have a closed form in these situations, they are tedious and time consuming to deal with. Hence, these formulas cannot be applied in practice due to the complexity of the calculations and ample time consumption. Because of the mentioned reasons, many scholars have sought to approximate these relationships, build bounds for them, or compute them through the recurrence relationships in INID case as well as in IID (see Reiss (2012), Balakrishnan and Sultan (1998), Arnold and Balakrishnan (2012), and Balakrishnan et al. (1992)).

Now, consider INID random variables $X_1, \ldots, X_n$ with corresponding absolutely continuous CDFs $F_1, \ldots, F_n$. Denote by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ the order statistics constructed from $X_1, \ldots, X_n$. The CDF of rth order statistics $X_{r:n}$ is (see David and Nagaraja (2004))

$$
\begin{aligned}
F_{r:n} &= P(X_{r:n} \leq x) \\
&= \frac{1}{r!(n-r)!} \sum_{i=r}^{n} \sum_{S} \prod_{j=1}^{k} F_{t_j}(x) \prod_{j=k+1}^{n} \bar{F}_{t_j}(x),
\end{aligned}
$$

where the summation S extends over all $n!$ permutations $t_1, \ldots, t_n$ of $1, \ldots, n$.

Given a set of INID variables Bayramoglu Bayramoglu (2018) defined the D-order functions ($F_{[i]}$) based on point to point ordering of given CDFs (F_i). He discussed

some useful properties of these definitions and showed that the corresponding random variables are independent. Finally, he suggested that these new functions could be used as an alternative or an approximation of the order statistics CDFs ($F_{i:n}$).

His new definition of D-order functions is as follows:

First, consider INID random variables $X_1, X_2, \ldots, X_n$ with, respectively, corresponding CDFs $F_1, F_2, \ldots, F_n$, and next suppose that $a_i = inf\{ x : F_i(x) > 0 \}$ and $b_i = sup\{ x : F_i(x) < 1 \}$. Then, the rth D-order function $F_{[r]}(x)$, $x \in (a, b)$, $1 \le r \le n$, where $a = a_i$ and $b = b_i$ (i.e., all CDFs have the same support) is defined as $F_{[r]}(x) = F_j(x)$, *if* $x \in D_r^j$ such that

$$D_r^j = \{ x : x \in [a, b], F_{i_1}(x) \ge F_{i_2}(x) \ge \cdots \ge F_{i_{r-1}}(x) \ge F_{i_j}(x) \ge F_{i_{r+1}}(x) \ge \ldots F_{i_n}(x) \} ,$$

for all

$$(i_1, i_2, \ldots, i_{r-1}, i_j, i_{r+1}, i_n) \in \Pi_{j,r}(1, 2, \ldots, n), i_k \ne j \quad for \quad k \ne r,$$

where $\Pi_{j,r}(1, 2, \ldots, n)$ is the class of all sequences of length n, in which the rth place is occupied by j, and the remaining $(n - 1)$ places are occupied by the elements of possible $(n - 1)!$ permutations of numbers $1, 2, \ldots, j - 1, j + 1, \ldots, n$. In addition, for the heterogeneous random domains, it is clear that without loss of generality, we can extend this condition by defining a and b as $a = min(a_i)$ and $b = min(b_i)$, respectively.

The rest of the paper is organized as follows. We use only $F_{[1]}$ and $F_{[n]}$ for constructing an improved approximations of $F_{1:n}$ and $F_{n:n}$ and provide lower and upper bounds for $F_{r:n}$ in Section 2. Unfortunately, these bounds are not CDFs, in the general case, but they have very important and useful features, which means that they are not uniform and sensitive to the given values and vary in terms of changing the domain points of the bound functions. Moreover, these features are described extensively and relatively, and the corresponding formulas for the most of the theoretical properties of these order statistics, such as moments and the mathematical expectations, are presented.

In Section 3, we compare our approximation of $F_{1:n}$ and $F_{n:n}$ with the approximation that is presented in Bayramoglu Bayramoglu (2018), and our approximation superiority will also be clearly represented throughout the figures. We also show, by another example, that even the bounds obtained in this study, work better than the D-order functions in many places.

2 Lower Bound, Upper Bound, and Approximation for $F_{i:n}$ in INID Case

In this section, we find the lower and upper bounds for the CDFs of order statistics in INID case. Some of their important properties are discussed. We also provide a proper approximation of $F_{i:n}$ utilizing these bounds.

Following two definitions of the first and the last of D-order functions ($F_{[1]}(x)$ and $F_{[n]}(x)$), it is clear that

$$F_{[n]}(x) \leq F_i(x) \leq F_{[1]}(x),$$

and

$$\bar{F}_{[1]}(x) \leq \bar{F}_i(x) \leq \bar{F}_{[n]}(x).$$

These bounds motivate us to construct new bounds for $F_{1:n}(x)$ and $F_{n:n}(x)$ as

$$1 - \prod_{i=1}^{n} \bar{F}_{[1]}(x) \leq F_{1:n}(x) = 1 - \prod_{i=1}^{n} \bar{F}_i(x) \leq 1 - \prod_{i=1}^{n} \bar{F}_{[n]}(x),$$

and

$$\prod_{i=1}^{n} F_{[n]}(x) \leq F_{n:n}(x) = \prod_{i=1}^{n} F_i(x) \leq \prod_{i=1}^{n} F_{[1]}(x),$$

or equivalently

$$G_{1:n;[1]}(x) \leq F_{1:n}(x) \leq G_{1:n;[n]}(x),$$

and

$$G_{n:n;[n]}(x) \leq F_{n:n}(x) \leq G_{n:n;[1]}(x),$$

where $G_{i:j;[k]}(x)$ ($1 \leq i \leq j \leq n$, $k = 1, n$) denote the CDFs of ith smallest order statistics arising from j, IID random variables with the same CDF $F_{[k]}(x)$. Note that, because each of the presented bound alone is a CDF, we approximate $F_{1:n}$ and $F_{n:n}$, as

$$F_{(1)}(x) = \frac{G_{1:n;[1]}(x) + G_{1:n;[n]}(x)}{2}, \tag{2.1}$$

$$F_{(n)}(x) = \frac{G_{n:n;[1]}(x) + G_{n:n;[n]}(x)}{2}. \tag{2.2}$$

The functions (2.1), and (2.2) are CDF, which we can use them for approximating $F_{1:n}$ and $F_{n:n}$,. In these approximations after utilizing $G_{i:j;[k]}(x)$, we can use the strong condition IID instead of the weaker condition INID.

Similar to the preceding Theorem, we construct bounds for $F_{i:n}, i = 1, \ldots, n$.

Theorem 2.1. *If $X_1, \ldots, X_n$ are INID random variables with CDFs $F_1, \ldots, F_n$, respectively, then, for all $1 \le i \le n$,*

$$LF_{i:n}(x) \le F_{i:n}(x) \le UF_{i:n}(x), \tag{2.3}$$

where

$$LF_{i:n}(x) = \max\left\{\sum_{k=i}^{n}\binom{n}{k}F_{[n]}^k(x)\bar{F}_{[1]}^{n-k}(x), 1 - \sum_{k=0}^{i-1}\binom{n}{k}F_{[1]}^k(x)\bar{F}_{[n]}^{n-k}(x)\right\}, \tag{2.4}$$

and

$$UF_{i:n}(x) = \min\left\{\sum_{k=i}^{n}\binom{n}{k}F_{[1]}^k(x)\bar{F}_{[n]}^{n-k}(x), 1 - \sum_{k=0}^{i-1}\binom{n}{k}F_{[n]}^k(x)\bar{F}_{[1]}^{n-k}(x)\right\}. \tag{2.5}$$

Proof. Consider, for $i = 0, 1, \ldots, n$,

$$P_i(x) = P(\text{exactly } i \text{ of } X_1, \ldots, X_n \text{ are } \le x).$$

Thus

$$\binom{n}{i}F_{[n]}^i(x)\bar{F}_{[1]}^{n-i}(x) \le P_i(x) \le \binom{n}{i}F_{[1]}^i(x)\bar{F}_{[n]}^{n-i}(x).$$

Now, apply the following formula (see Bairamov and Tavangar (2015))

$$F_{i:n}(x) = \sum_{k=i}^{n} p_k(x) = 1 - \sum_{k=0}^{i-1} p_k(x).$$

Then, we get

$$\sum_{k=i}^{n}\binom{n}{k}F_{[n]}^k(x)\bar{F}_{[1]}^{n-k}(x) \le F_{i:n}(x) \le \sum_{k=i}^{n}\binom{n}{k}F_{[1]}^k(x)\bar{F}_{[n]}^{n-k}(x),$$

and

$$1 - \sum_{k=0}^{i-1}\binom{n}{k}F_{[1]}^k(x)\bar{F}_{[n]}^{n-k}(x) \le F_{i:n}(x) \le 1 - \sum_{k=0}^{i-1}\binom{n}{k}F_{[n]}^k(x)\bar{F}_{[1]}^{n-k}(x).$$

Finally, take minimum from upper bounds and maximum from lower bounds, to complete the proof. □

The $LF_{i:n}(x)$ and $UF_{i:n}(x)$ have some interesting properties, which are listed below.

<div align="center">Bounds Properties :</div>

- One of the important advantages of these bounds is the point to point dependence to the fixed given value x, which means that $LF_{i:n}(x_0)$, and $UF_{i:n}(x_0)$ have the different form of $LF_{i:n}(x_1)$ and $UF_{i:n}(x_1)$, respectively, where $x_0 \neq x_1$. This property makes the bounds functions flexible and causes the difference between the main function and the related bounds as small as possible. Moreover, these bounds enable us to find a rigorous and tight confidence interval for many of the theoretical properties of $X_{i:n}$, which are directly related to the $F_{i:n}$, such as bounds for moment generating functions, characteristic functions, and so on.

- It is obvious that the approximation of the $F_{i:n}$ in INID case, must also be satisfied in the IID situations. Therefore, it is logical to expect that the alternative of $F_{i:n}$ in the INID case satisfies also the IID samples. However, it is easy to check that $F_{[i]} \neq F_{i:n}$ in the INID case as well as in the IID form. It is shown that our bound and consequently our approximations of $F_{i:n}$ becomes equality for the IID case.

Proposition 2.1. *Let $X_1, \ldots, X_n$ be IID random variables with the same CDFs F; then for all $1 \leq i \leq n$,*

$$
\begin{aligned}
LF_{i:n}(x) &= UF_{i:n}(x) \\
&= \sum_{k=i}^{n} \binom{n}{k} F^k(x) \bar{F}^{n-k}(x).
\end{aligned}
$$

Proof. In the IID case, it is straightforward to check that $F_{[i]} = F$, and due to the relation $F + \bar{F} = 1$, we have

$$
\sum_{k=0}^{i-1} \binom{n}{k} F_{[n]}^k(x) \bar{F}_{[1]}^{n-k}(x) + \sum_{k=i}^{n} \binom{n}{k} F_{[1]}^k(x) \bar{F}_{[n]}^{n-k}(x)
$$

$$
= \sum_{k=0}^{i-1} \binom{n}{k} F_{[1]}^k(x) \bar{F}_{[n]}^{n-k}(x) + \sum_{k=i}^{n} \binom{n}{k} F_{[n]}^k(x) \bar{F}_{[1]}^{n-k}(x)
$$

$$= \sum_{k=0}^{i-1} \binom{n}{k} F^k(x) \bar{F}^{n-k}(x) + \sum_{k=i}^{n} \binom{n}{k} F^k(x) \bar{F}^{n-k}(x)$$

$$= \sum_{k=0}^{n} \binom{n}{k} F^k(x) \bar{F}^{n-k}(x) = 1.$$

Finally, with some slight calculations, the proof is completed. □

- Since we are looking for an alternative representation of a CDF, it can be rational to expect that our formula must have the CDF properties and consequently also our bounds too. According to the flexibility of these bounds, it is reasonable that we expect their first and last values are 0 and 1, respectively. In the following, we prove the accuracy of this statement. The continuousness of these bounds is clear, and only it is remained to check the increasing behavior of these bounds to show the CDF feature of $LF_{i:n}(x)$ and $UF_{i:n}(x)$. It is easy to achieve these properties by plotting all bounds, so if these bounds are CDF, we can use a convex linear combinations of these functions or alternatively as a special case, its average as a supersede of $F_{i:n}$. Otherwise, we can use one of these bounds (preferred one, which is CDF) as an approximation.

Proposition 2.2. *If* X_1, X_2, ..., X_n *are random variables, respectively, supported in* $[a_1, b_1]$, ..., $[a_n, b_n]$, $a_i, b_i \in \mathbb{R}, i \in \mathbb{N}$, *and also relations (2.4) and (2.5) hold, then*

$$LF_{i:n}(a) \ = \ UF_{i:n}(a) = 0,$$
$$LF_{i:n}(b) \ = \ UF_{i:n}(b) = 1,$$

where $a = min(a_1, a_2, \ldots, a_n)$ *and* $b = max(b_1, b_2, \ldots, b_n)$.

Proof. The CDF property of D-order functions results in

$$F_{[1]}(a) = F_{[n]}(a) = 1 - F_{[1]}(b) = 1 - F_{[n]}(b) = 0.$$

On the other hand, for $y = \{a, b\}$,

$$F_{[1]}^i(y) \bar{F}_{[n]}^{n-i}(y) = F_{[n]}^i(y) \bar{F}_{[1]}^{n-i}(y) = I_{i=0}(i) I_{y=a}(y) + I_{i=n}(i) I_{y=b}(y) \,,$$

where

$$I_{i=0}(i) = \begin{cases} 1 & if \ i = 0, \\ 0 & if \ i \neq 0. \end{cases}$$

Now, it is obvious to see that, for all $i = 0, 1, \ldots, n$,

$$
\begin{aligned}
\sum_{k=i}^{n} \binom{n}{k} F_{[1]}^{k}(y) \bar{F}_{[n]}^{n-k}(y) &= \sum_{k=i}^{n} \binom{n}{k} F_{[n]}^{k}(y) \bar{F}_{[1]}^{n-k}(y) \\
&= I_{i=0}(i) I_{y=a}(y) + I_{i=n}(i) I_{y=b}(y) .
\end{aligned}
$$

Finally, with a little accuracy in relations (2.4) and (2.5) and some routine mathematical calculations, the proof is completed. □

- It is difficult to calculate all of the proposed CDFs by Bayramoglu, Bayramoglu (2018), and frequently time-consuming, especially when the sample size is large. But in this strategy, we only need the first and the last of D-order functions. These two functions can be easily calculated in R statistical software Team (2018), by two famous commands *pmin* and *pmax*, respectively, and the difficulty of derivation Bayramoglu's CDFs will be fixed. These features can increase the accuracy of the calculation and considerably reduce the time it takes to perform the calculation. Moreover, when the sample size becomes larger and larger, the CDFs of extreme values tend more and more to the horizontal axis ($Y = 0$ and $Y = 1$), and consequently, the precision of corresponding D-order alternatives for these random variables becomes smaller and smaller. Because in D-order approximations, the effect of sample size is nearly ignored, low accuracy results when one wants to approximate the CDF of order statistics for large sample sizes.

- As the last lack of D-order functions, it is possible to mention their inattention in approximation CDF of order statistics in the case that there exist some similar CDFs amog the original CDFs. In an analytical form, assume n independent random variables $X_1, X_2, \ldots, X_n$ such that $X_{k_{j-1}+1}, X_{k_j}, \ldots, X_{k_j} \sim F_j$, where $j = 1, 2, \ldots, m$ and $k_0 = 0, \sum_{j=1}^{m} k_j = n$. In this situation, there exist just m kinds of D-order function aiming to approximate n different CDFs of order statistics. The worst situation has occurred in the IID case since there is just one type of D-order functions for approximation n kinds of CDFs of order statistics. Finally, it is worth noting that our approximations fixed this problem as well as previous problems.

3   Examples

[1] In this section, we provide two examples, which are an excellent showcases for the previous theoretical backgrounds. The first example compares our result with Bayramoglu's upshots. The second example represents the biasedness of D-order functions for $n = 4$ in comparison with our alternative representation.

Examples 3.1 (Bayramoglu 2018, Bayramoglu (2018)). Let X_1 and X_2 be INID random variables supported in [0,1], respectively, with corresponding CDFs

$$F_1(x) = (2x - x^2)^2, \quad 0 \le x \le 1,$$
$$F_2(x) = x, \quad 0 \le x \le 1.$$

Then

$$F_{[1]}(x) = \begin{cases} x & \text{if } 0 \le x \le c, \\ (2x - x^2)^2 & \text{if } c \le x \le 1, \end{cases} \qquad F_{[2]}(x) = \begin{cases} (2x - x^2)^2 & \text{if } 0 \le x \le c, \\ x & \text{if } c \le x \le 1, \end{cases}$$

where $c = \frac{3 - \sqrt{5}}{2}$.

Now, we approximate $F_{1:2}(x)$. According to relation (2.1), we have

$$
\begin{aligned}
F_{(1)}(x) &= \frac{G_{1:2;[2]}(x) + G_{1:2;[1]}(x)}{2} \\
&= \frac{1 - (1 - F_{[2]}(x))^2 + 1 - (1 - F_{[1]}(x))^2}{2} \\
&= (2x - x^2)^2 + x - \frac{(2x - x^2)^4 + x^2}{2},
\end{aligned}
$$

and similarly by relation (2.2), we obtain

$$
\begin{aligned}
F_{(2)}(x) &= \frac{G_{2:2;[2]}(x) + G_{2:2;[1]}(x)}{2} \\
&= \frac{F_{[1]}^2(x) + F_{[2]}^2(x)}{2}
\end{aligned}
$$

---

[1]In this part, all of the numerical calculation with 7 digits of decimals and drawing all graphs are done with R statistical software; Team (2018).

$$= \frac{(2x - x^2)^4 + x^2}{2}.$$

Our first step is to compare these approximations with the one provided by Bayramoglu, Bayramoglu (2018). This comparison is shown in Figure 1. The accuracy of our approximation is clear, and also the adaptation of our results and the original function is surprising. This feature is because of the sensitivity of the bounds and its relative symmetry with respect to the original function. In fact, by using just two CDFs defined according to the D-order notion, we are able to improve the Bayramoglu's presentations for $F_{1:2}(x)$ and $F_{2:2}(x)$ with high accuracy. This note shows the importance of D-order definitions.
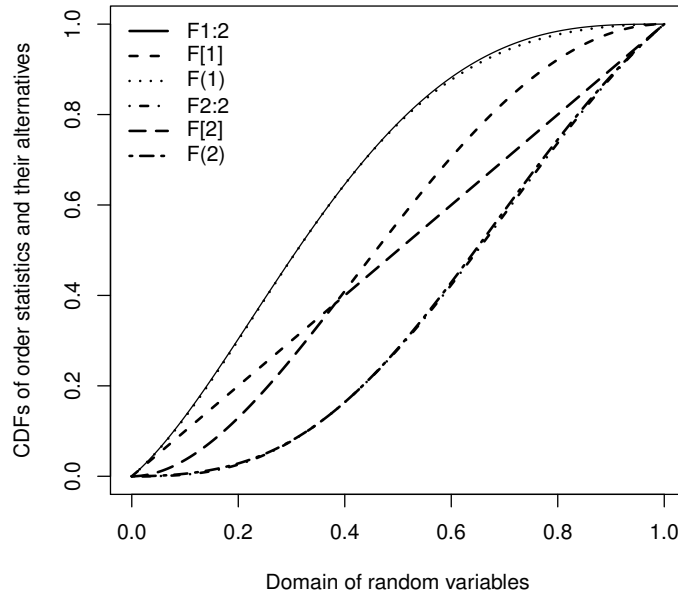

Figure 1: $F_{1:2}(x)$ and $F_{2:2}(x)$ and their alternatives

The next step is to calculate and display the bounds of $F_{1:2}(x)$ and $F_{2:2}(x)$. With respect to Theorem 2.1, bounds may be CDFs or not. When, their mean is not a CDF, we can not use the mean of two bounds as an approximation of $F_{i:n}$. It has two benefits. Firstly, it provides additional information about our bounds, which those of them are

CDF or not. Secondly, it can well evaluate the range of occurrence of many probabilities. Besides, we can consider lower and upper bounds of $\int h(x)dF_{i:n}(x)$ for any monotone function $h(.)$. This feature also can enable us to construct bounds for some theoretical properties of any ordinary order statistics arising from INID random variables such as moment generating function and moments for nonnegative and negative random variables.

By Theorem 2.1, we can find bounds for $F_{1:2}(x)$,

$$\max\left\{1 - \bar{F}_{[2]}^2(x), 2F_{[2]}(x)\bar{F}_{[1]}(x) + F_{[2]}^2(x)\right\} \leq F_{1:2} \leq \min\left\{1 - \bar{F}_{[1]}^2(x), 2F_{[1]}(x)\bar{F}_{[2]}(x) + F_{[1]}^2(x)\right\},$$

and similarly, for $F_{2:2}$,

$$\max\left\{F_{[2]}^2(x), 1 - \bar{F}_{[2]}^2(x) - 2F_{[1]}(x)\bar{F}_{[2]}(x)\right\} \leq F_{2:2} \leq \min\left\{F_{[1]}^2(x), 1 - \bar{F}_{[1]}^2(x) - 2F_{[2]}(x)\bar{F}_{[1]}(x)\right\}.$$

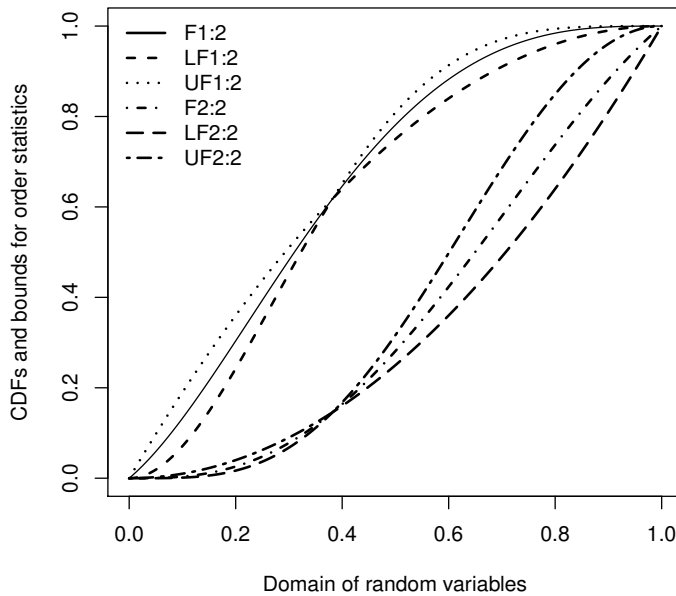In Figure 2, the bounds of $F_{1:2}(x)$ and $F_{2:2}(x)$ are shown.

Figure 2: $F_{1:2}(x)$ and $F_{2:2}(x)$ and their bounds

The goodness of these bounds can be immediately understood, which means that depending on any given value, these bounds also have different values and are not uniformly bounded.

Based on Figure 2, it is evident that the bounds are itself CDFs and in each case, we can utilize the mean of two bounds as an approximation of order statistics CDFs as we previously have done. For more understanding of the performance of these bounds, we compare them with D-order functions. This comparison is depicted in Figure 3.

It is clear that the upper bound of $F_{1:2}$ and the lower bound of $F_{2:2}$ are much better alternatives for $F_{1:2}(x)$ and $F_{2:2}(x)$ in comprised with the corresponding approximations of Bayramoglu's $F_{(1)}(x)$ and $F_{(2)}(x)$. Even the performance of the lower bound of $F_{1:2}$ and the upper bound of $F_{2:2}$ in the dominated part of the random variable domains are better than the corresponding approximations of D-order functions. (By approximating the CDFs of order statistics, it is clear, that our approximation all over the interval $[0, 1]$, are significantly closed to the original functions as it can be seen in Figure 3. It is noteworthy that, the performances of $LF_{1:2}(x)$ and $UF_{2:2}(x)$ is better than Bayramoglu's presentation, in the intervals $[0.152, 1]$ and $[0, 0.718]$ respectively, which means these bounds are better than Bayramoglu's approximation in 85 and 71 percent of situations.) This proposition itself can show the importance of these bounds, and there is even a reason to use these functions as an approximation of $F_{1:2}(x)$ and $F_{2:2}(x)$.

Finally, $F_{(1)}$ and $F_{(2)}$ are CDFs, hence there exist corresponding random variables $X_{(1)}$ and $X_{(2)}$, respectively. It is interesting to compare the expected values, variance, the first quartile (Q_1), median (Q_2), and the third quartile (Q_3) of these random variables with what calculated through D-order functions.

These measures can be found in Table 1. Based on this table, one can see the closeness of all tendency measures of $X_{i:2}$ and $X_{(i)}$.

Table 1: Measures of central and skewed tendency for presented random variables

| –	Expectation	Variance	Mode	Q_1	Q_2	Q_3
$X_{[1]}$	0.4483615	0.05908507	0.0000000	0.2500000	0.4588044	0.6340048
$X_{1:2}$	0.3333333	0.04126991	0.2254046	0.1698130	0.3103567	0.4746262
$X_{(1)}$	0.3365079	0.04311163	0.2309758	0.1713331	0.3107164	0.4756098
$X_{[2]}$	0.5183052	0.07124662	0.5000000	0.2928770	0.5000000	0.7500000
$X_{2:2}$	0.6333333	0.04650793	0.7101017	0.4756052	0.6496785	0.8083108
$X_{(2)}$	0.6301587	0.04655081	0.6855557	0.4746203	0.6451675	0.8035395
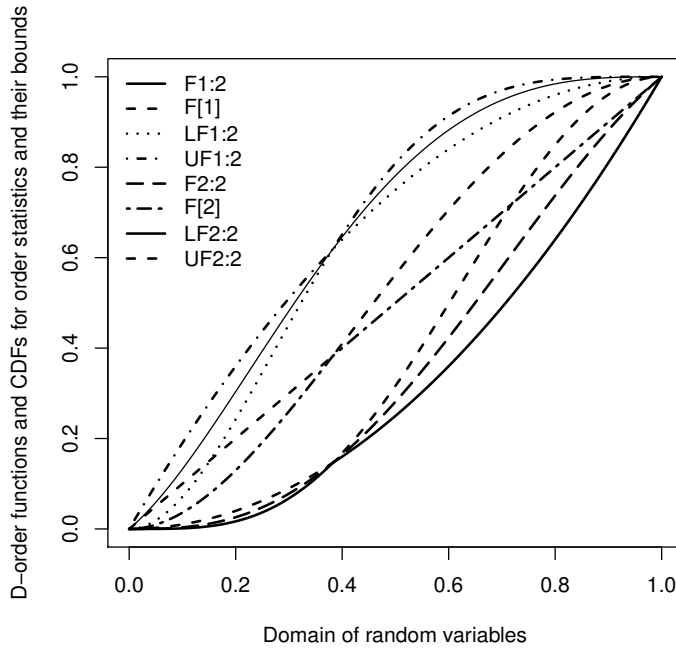
Figure 3: $F_{1:2}(x)$ and $F_{2:2}(x)$, their bounds and D-order functions

This example can be considered as a special case of the following Lemma.

Lemma 3.1. *Let F_1 and F_2 be two CDFs with the same supported interval such that the equation $F_1(x) - F_2(x) = 0$ has at most three different real solutions. Then*

$$F_{(1)}(x) = F_1(x) + F_2(x) - \frac{F_1^2(x) + F_2^2(x)}{2},$$

$$F_{(2)}(x) = \frac{F_1^2(x) + F_2^2(x)}{2},$$

$$F_{(1)}^r(x) + F_{(2)}^r(x) = F_1^r(x) + F_2^r(x), \quad r \in \mathbb{R},$$

$$F_{(1)}^r(x) \times F_{(2)}^r(x) = F_1^r(x) \times F_2^r(x), \quad r \in \mathbb{R},$$

$$f_{(1)}(x) + f_{(2)}(x) = f_1(x) + f_2(x),$$

$$\mu^r_{(1)}(x) + \mu^r_{(2)}(x) \;\; = \;\; \mu^r_1(x) + \mu^r_2(x), \quad r \in \mathbb{R},$$

provided that the expectations exist.

Proof. Two obvious solutions are the first and the last point of the given interval, and in this case it is obvious to see that $F_{[1]}(x) = \max\{F_1(x), F_2(x)\}$ and $F_{[2]}(x) = \min\{F_1(x), F_2(x)\}$. Consequently, for any function $g(.)$, we have

$$g(F_{[2]}(x)) + g(F_{[1]}(x)) = g(F_1(x)) + g(F_2(x)), \tag{3.1}$$

and

$$g(F_{[2](x)})g(F_{[1](x)}) = g(F_1(x))g(F_2(x)). \tag{3.2}$$

Finally let $g(x) = 1 - (1-x)^2$ and $g(x) = x^2$, in relations (2.1) and (2.2), to complete the proof in this case.

In another case, we have an extra real solution in the interval, say c, and

$$F_{[1]}(x) = \begin{cases} F_1(x) & if \ x \le c, \\ F_2(x) & if \ c \le x, \end{cases} \qquad F_{[2]}(x) = \begin{cases} F_2(x) & if \ x \le c, \\ F_1(x) & if \ c \le x, \end{cases}$$

or

$$F_{[1]}(x) = \begin{cases} F_2(x) & if \ x \le c, \\ F_1(x) & if \ c \le x, \end{cases} \qquad F_{[2]}(x) = \begin{cases} F_1(x) & if \ x \le c, \\ F_2(x) & if \ c \le x; \end{cases}$$

Again it is straightforward to see that relations (3.1) and (3.2) hold for any function $g(.)$, as considered previously. Now, the proof is similar to the proof of the previous case. □

An interesting fact about this lemma is that for $n = 2$, we do not need to calculate the Bayramolu's CDFs and our approximations of $F_{i:n}$ can be easily derived from the original CDFs. In addition, some theoretical properties of corresponding random variables such as moments and probability density functions can be calculated in the same manner.

Here, we present an example for $n = 4$ to show the method of deriving alternatives for CDF of any order statistics. Likewise, it can be observed that the tendency behavior of CDFs of extreme values approaches to the horizontal axis, and it leads to the bad performance of corresponding D-order functions.

Examples 3.2. Let X_1, X_2, X_3, and X_4 be INID random variables supported in $[0,1]$, respectively, which have CDFs

$$
\begin{aligned}
F_1(x) &= (2x - x^2)^2, \quad 0 \le x \le 1, \\
F_2(x) &= x, \quad 0 \le x \le 1, \\
F_3(x) &= \frac{1 - e^{-x}}{1 - e^{-1}}, \quad 0 \le x \le 1, \\
F_4(x) &= \frac{\Phi(x) - 0.5}{\Phi(1) - 0.5}, \quad 0 \le x \le 1,
\end{aligned}
$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

Then the D-order of these function are obtained as

$$F_{[1]}(x) = F_1(x)I_{[0,0.6175849]}(x) + F_3(x)I_{[0.6175849,1]}(x),$$

$$F_{[2]}(x) = F_4(x)I_{[0,0.4966055]}(x) + F_3(x)I_{[0.4966055,0.6175849]}(x) + F_1(x)I_{[0.6175849,1]}(x),$$

$$F_{[3]}(x) = F_2(x)I_{[0,0.3819655]}(x) + F_3(x)I_{[0.3819655,0.4966055]}(x) + F_4(x)I_{[0.4966055,1]}(x),$$

and

$$F_{[4]}(x) = F_3(x)I_{[0,0.3819655]}(x) + F_2(x)I_{[0.3819655,1]}(x),$$

where

$$
I_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } o.w. \end{cases}
$$

As previously proved, if $LF_{i:n}$ and $UF_{i:n}$ are increasing functions, then they are also CDF's. Consequently, when $n > 2$, in the first step, all of the bound functions should be drawn in order to determine which of them have a nondecreasing feature. Then, for any fixed i, if both of corresponding bounds are nondecreasing, then we can approximate $F_{i:n}$, by taking mean of these bounds, or any arbitrary convex linear combinations. However, when one of the corresponding bounds do not have nondecreasing property, it is preferred to choose one bound function, which has a nondecreasing property as an alternative representation for corresponding $F_{i:n}$. Finally, if none of two bounds are CDFs, we can only bound $F_{i:n}$ and another related theoretical property.

Now we have to examine if the bounds are CDFs or not. By drawing these bounds, it is clearly understood that all of them are CDFs. Nondecreasing property of all bounds can be obviously seen in Figure 4, and consequently, we can approximate each of $F_{i:n}$. These substitutes are the mean of two corresponded bounds.
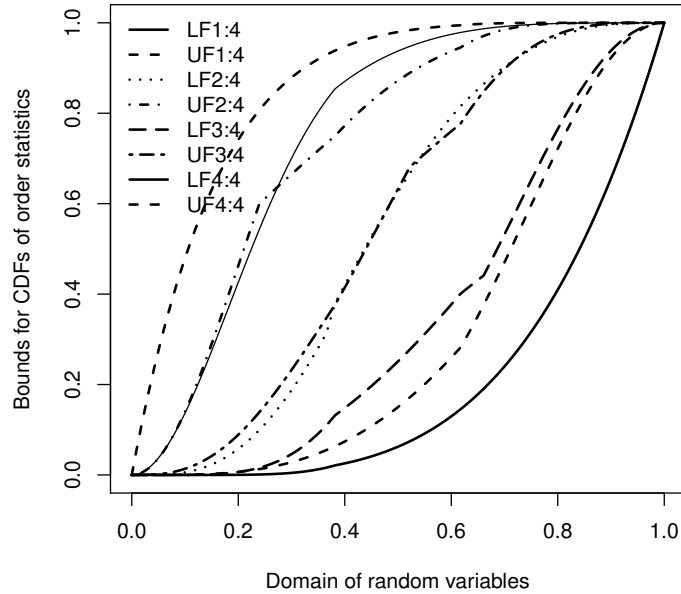
Figure 4: Lower and upper bounds for all CDFs of ordinary order statistics

From another point of view, due to the fact that our approximations are also CDFs, there exist corresponding random variables, for which we are interested in comparing some of their measures of central and skewed tendency. Their moments are collected in Table 2.

It is observed that for almost all of the provided measures, the accuracy of our results is much better than Bayramoglu's alternatives. Moreover, our approximations, D-order corresponding functions, and upper and lower bounds for CDFs of ordinary order statistics arising from these heterogeneous random variables, are shown in Figures 5a and 5b. The comparison results for extreme values of this example (first and last ordinary order statistics) can be seen in Figure 5a. As mentioned in the first item of bounds properties, the performance of prime and latter D-order functions get worse and worse by increasing the sample size. This note can be clearly understood in this figure.

D-order functions performance is far better in the approximation of middle ordinary order statistics as seen in Figure 5b, but still it is unfavorable in comparison with our representations.

4   Conclusion and Recommendation

We considered INID random variables $X_1, \dots, X_n$ with respectively CDFs $F_1, \dots, F_n$ and utilized the first and the last D-order functions to construct lower and upper bounds for any CDF of ordinary order statistics ($F_{i:n}$). We proved all of the necessary and sufficient conditions of these bounds to be a CDF except their nondecreasing behavior features. This property remains as an open problem for the researches to prove or contradict. This gap can be fixed by drawing all of the bounds with R statistical software (Team (2018)). The related command are also given in the text.

The new definition of D-order functions seems to be very appealing and therefore useful. Their main advantages are independence and the fact that they are CDFs. Unfortunately, when the sample size increases, the calculation of D-order functions becomes long and time-consuming. Furthermore, the performance of the first and the last of these functions get worse. Finally, the manner of fixing these deficiencies are widely expressed.

Table 2: Measures of central and skewed tendency of the random variables having CDFs described in Example 3.2.

–	*Expectation*	*Variance*	*Mode*	Q_1	Q_2	Q_3
$X_{[1]}$	0.4051309	0.06896395	0.0000000	0.17198220	0.3798886	0.6339869
$X_{1:4}$	0.1843764	0.02073397	0.0000000	0.06961658	0.1517968	0.2674345
$X_{(1)}$	0.1946320	0.02103398	0.2309758	0.08069021	0.1674673	0.2772036
$X_{[2]}$	0.4468496	0.07314843	0.4966055	0.21556320	0.4417774	0.6426276
$X_{2:4}$	0.3628090	0.03116173	0.3045699	0.22742370	0.3479505	0.4835257
$X_{(2)}$	0.3605321	0.0388380	0.6855557	0.19532660	0.3504830	0.5006237
$X_{[3]}$	0.4742661	0.07446167	0.3819655	0.25000000	0.45880530	0.6935207
$X_{3:4}$	0.5461947	0.03506705	0.5502603	0.40914160	0.5478789	0.6856511
$X_{(3)}$	0.5463243	0.04452736	0.2309758	0.37910920	0.5405354	0.7203276
$X_{[4]}$	0.5183052	0.07124662	0.3819655	0.29287700	0.5000000	0.7500000
$X_{4:4}$	0.7511721	0.03001898	0.9362720	0.63736750	0.7797503	0.8929016
$X_{(4)}$	0.7435602	0.03071863	0.6855557	0.64086940	0.7702442	0.8822124
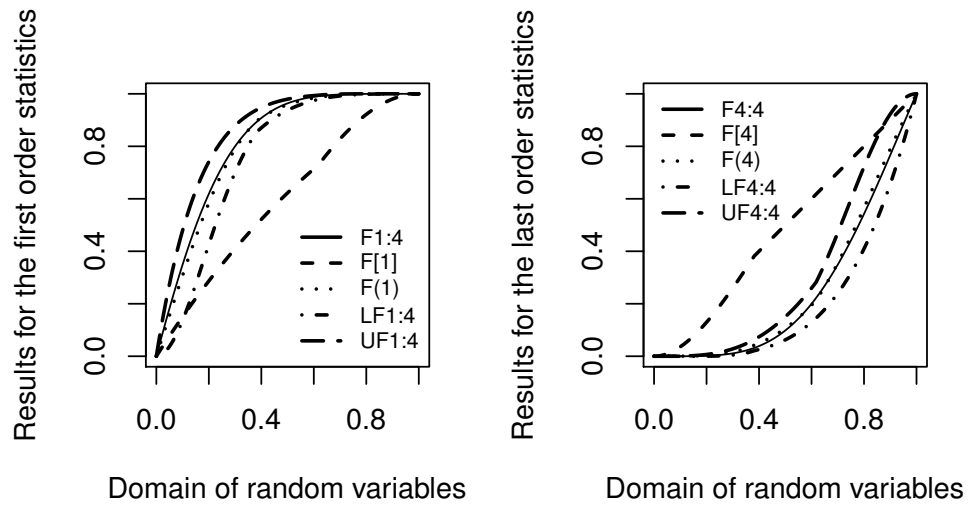
(a)
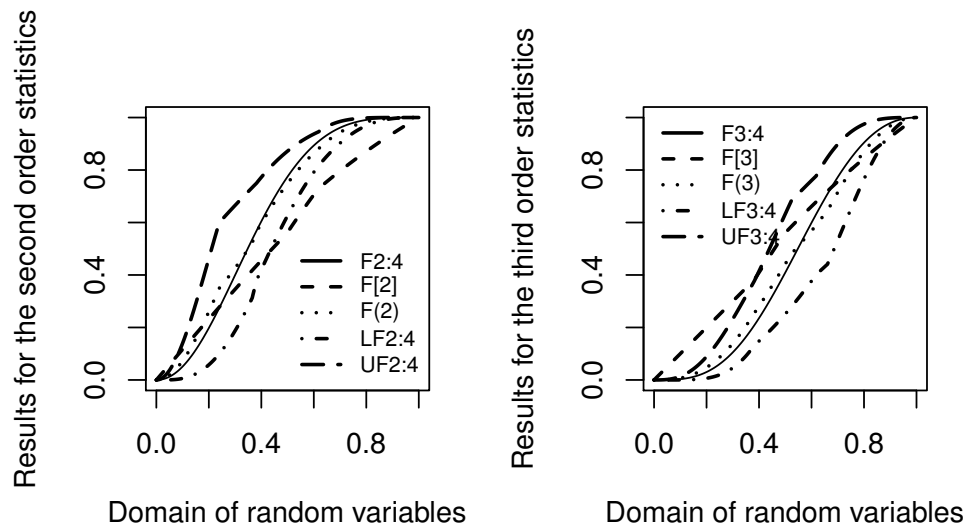
(b)

Figure 5: Comparison of different upshot for the middle order statistics

Acknowledgment

References

Ahsanullah, M., Nevzorov, V. B., and Shakil, M. (2013), An introduction to order statistics.

Arnold, B. C. and Balakrishnan, N. (2012), *Relations, bounds and approximations for order statistics*, Springer Science & Business Media, volume 53.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992), *A first course in order statistics*, volume 54, Siam.

Bairamov, I. and Tavangar, M. (2015), Residual lifetimes of k-out-of-n systems with exchangeable components. *Journal of The Iranian Statistical Society*, **14**(1), 63–87.

Balakrishnan, N., Bendre, S., and Malik, H. (1992), General relations and identities for order statistics from non-independent non-identical variables. *Annals of the Institute of Statistical Mathematics*, **44**(1), 177–183.

Balakrishnan, N. and Sultan, K. (1998), 7 recurrence relations and identities for moments of order statistics. *Handbook of Statistics*, **16**, 149–228.

Bayramoglu, I. (2018), A note on the ordering of distribution functions of inid random variables. *Journal of Computational and Applied Mathematics*, **343**, 49–54.

David, H. A. and Nagaraja, H. N. (2004), Order statistics. *Encyclopedia of Statistical Sciences*.

Reiss, R.-D. (2012), Approximate distributions of order statistics: with applications to nonparametric statistics. *Springer Science & Business Media*.

Team, R. C. (2018), R: A language and environment for statistical computing.