Statistics
A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: https://www.tandfonline.com/loi/gsta20

Order statistics from overlapping samples: bivariate densities and regression properties

F. López-Blázquez, Nan-Cheng Su & Jacek Wesołowski

STATISTICS
https://doi.org/10.1080/02331888.2019.1599378

Order statistics from overlapping samples: bivariate densities and regression properties

F. López-Blázquez[a], Nan-Cheng Su [b] and Jacek Wesołowski[c]

[a]Departamento Estadistica e Investigatión Operativa, Universidad de Sevilla, Sevilla, Spain; [b]Department of Statistics, National Taipei University, New Taipei City, Taiwan; [c]Faculty of Mathematics and Information Science, Warsaw University of Technology, Warszawa, Poland

ABSTRACT

In this paper, we are interested in the joint distribution of two order statistics from overlapping samples. We give an explicit formula for the distribution of such a pair of random variables under the assumption that the parent distribution is absolutely continuous. We are also interested in the question to what extent conditional expectation of one of such order statistic given another determines the parent distribution. In particular, we provide a new characterization by linearity of regression of an order statistic from the extended sample given the one from the original sample, special case of which solves a problem explicitly stated in the literature. It appears that to describe the correct parent distribution it is convenient to use quantile density functions. In several other cases of regressions of order statistics we provide new results regarding uniqueness of the distribution in the sample.

ARTICLE HISTORY
Received 11 June 2018
Accepted 20 March 2019

KEYWORDS
Characterization; order statistics; overlapping samples; linearity of regression; uniqueness theorem

MSC 2010 SUBJECT CLASSIFICATIONS
Primary: 60E05; Secondary: 62E10

1. Introduction

Properties of order statistics (os's) $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ based on the sample $\{X_1, X_2, \ldots, X_n\}$ of independent and identically distributed (iid) random variables with absolutely continuous distribution are widely known, see, e.g. the monographs David and Nagaraja [1] or Arnold et al. [2] for excellent reviews. Much less is known for os's which arise from different samples which have common elements. There are two special cases which until now have been studied in the literature: (1) moving os's, when the subsequent samples are of the same size and have the same size of the overlap – see, e.g. Inagaki [3], David and Rogers [4], Ishida and Kvedaras [5] or Balakrishnan and Tan [6]; (let us mention that moving samples have a long history in quality control and time series analysis – in particular, the moving median is a simple robust estimator of location and the moving range is a current measure of dispersion complementing the moving average); (2) special cases of os's from the original and extended sample which except the original sample contains a number of additional observations – see, e.g. Siddiqui [7], Tryfos and Blackmore [8], Ahsanullah and Nevzorov [9] or López-Blázquez and Salamanca-Miño [10].

CONTACT Nan-Cheng Su sunanchen@gmail.com

In the latter paper the authors introduced a reference measure ν with respect to which the joint distribution of $(X_{n-k+1:n},\ X_{n-k+2:n+1})$ (which they were interested in) has a density. This measure, ν, defined by

$$\nu(B) = \mu_2(B) + \mu_1(\pi(B)), \quad B \in \mathcal{B}(\mathbb{R}^2), \tag{1}$$

where $\pi(B) = \{x \in \mathbb{R} :\ (x,x) \in B\}$ and μ_i is the Lebesgue measure in $\mathbb{R}^i$, $i=1,2$, will be of special interest for us here since ν will serve as the reference measure for bivariate densities of os's from overlapping samples.

In this paper we consider iid random variables $X_1, X_2, \ldots$ with the cumulative distribution function (cdf) denoted by F, its tail denoted by $\bar{F} := 1 - F$ and the density with respect to μ_1 denoted by f.

Let $\emptyset \neq A \subset \{1, 2, \ldots\}$ be such that $n_A := |A| < \infty$, where $|A|$ denotes the number of elements in A. By $X_{i:A}$ denote the ith os from the sample $\{X_k,\ k \in A\}$, $i = 1, \ldots, n_A$. In case $A = \{1, \ldots, n\}$ we have $X_{i:A} = X_{i:n}$, $i = 1, \ldots, n$. Consider additionally $\emptyset \neq B \subset \{1, 2, \ldots\}$ such that $n_B := |B| < \infty$. Our aim is to study the joint distribution of $(X_{i:A},\ X_{j:B})$, $i = 1, \ldots, n_A$, $j = 1, \ldots, n_B$. Of course, when $A \cap B = \emptyset$ the samples $\{X_k,\ k \in A\}$ and $\{X_k,\ k \in B\}$ are independent and the joint distribution of $(X_{i:A},\ X_{j:B})$ is just a product of marginal distributions of $X_{i:A}$ and $X_{j:B}$. We will only consider the case when $A \cap B \neq \emptyset$. Due to the permutation invariance of the distribution of $(X_1, X_2, \ldots)$ it suffices to take $A = \{1, \ldots, m\}$ and $B = \{r+1, r+2, \ldots, r+n\}$ with $r < m \leq n$. Then we denote $X_{j:n}^{(r)} := X_{j:B}$. In Section 2 we will study the joint density (with respect to the reference measure ν) of the pair $(X_{i:m},\ X_{j:n}^{(r)})$. The case $r=0$ is technically much simpler but the main idea of the approach is the same as in the general case. Therefore we first derive the joint distribution of $(X_{i:m},\ X_{j:n})$ in Section 2.1 while the general case of an arbitrary $r \geq 0$ is considered in Section 2.2 (with some technicalities moved to Appendix).

In Section 3 we are interested in regressions $\mathbb{E}(X_{i:m}|X_{j:n}^{(r)})$, $\mathbb{E}(X_{j:n}^{(r)}|X_{i:m})$ and related characterizations or identifiability questions. The main tools are representations of these regressions in terms of combinations of $\mathbb{E}(X_{k:n+r}|X_{\ell:n+r})$, $k, l \in \{1, \ldots, n+r\}$. Since for $r > 0$ such representations are rather complex, our considerations in this case will be restricted to the simplest cases of regressions of $X_{1:2}$, $X_{2:2}$ given $X_{1:2}^{(1)}$ or given $X_{2:2}^{(1)}$. They are studied in Section 3.1.

The case of $r=0$ is much more tractable, though since $m < n$ the analysis of each of two dual regressions $\mathbb{E}(X_{i:m}|X_{j:n})$ and $\mathbb{E}(X_{j:n}|X_{i:m})$ is quite different. In particular, Dołegowski and Wesołowski [11] (DW in the sequel) proved that

$$\mathbb{P}(X_{i:m} = X_{k:n}) = \frac{\binom{k-1}{i-1}\binom{n-k}{m-i}}{\binom{n}{m}} I_{\{i,\ldots,n-m+i\}}(k), \tag{2}$$

and, consequently, obtained the following representation

$$\mathbb{E}(X_{i:m}|X_{j:n}) = \sum_{k=i}^{n-m+i} \frac{\binom{k-1}{i-1}\binom{n-k}{m-i}}{\binom{n}{m}} \mathbb{E}(X_{k:n}|X_{j:n}). \tag{3}$$

Here and everywhere below equations involving conditional expectations are understood in the $\mathbb{P}$-almost sure sense.

It was proved in DW with the help of (3) that the condition

$$\mathbb{E}(X_{i:m}|X_{j:n}) = aX_{j:n} + b, \tag{4}$$

characterizes the parent distributions (exponential, Pareto and power) when $j \le i$ and $j \ge n - m + i$. The case of $i = j$ had been considered earlier in Ahsanullah and Nevzorov [9] even for an arbitrary shape of the regression function. The characterization through condition (4) given in DW is a direct generalization of characterizations by linearity of $\mathbb{E}(X_{i:n}|X_{j:n})$. Analysis of such problems has a long history – see, e.g. references in DW, in particular, Ferguson [12]. In this case the complete answer was given in Dembińska and Wesołowski [13] through an approach based on the integrated Cauchy functional equation (see also López-Blázquez and Moreno-Rebollo [14] who used instead differential equations). Actually, the question of determination of the parent distribution by the (nonlinear) form of regression $\mathbb{E}(X_{i:n}|X_{j:n})$ for non-adjacent i and j has not been completely resolved until now – see, e.g. Bieniek and Maciag [15].

In Section 3.2 we investigate characterizations by linearity of regression (4) in the remaining unsloved cases, i.e. when $i < j < n - m + i$. In particular, we solve the easiest non-trivial open problem explicitly formulated in DW. The dual case of regressions of an os from the extended sample given an os from the original sample, i.e. $\mathbb{E}(X_{j:n}|X_{i:m})$ is considered in Section 3.3. The main results in this subsection identify several new situations in which the shape of the regression function determines uniquely the parent distribution. Finally, some conclusions are discussed in Section 4.

2. Bivariate distribution of os's from overlapping samples

In this section we will derive joint distribution of the pair $(X_{i:m}, X^{(r)}_{j:n})$. This will be given through the density $f_{X_{i:m}, X^{(r)}_{j:n}}$ with respect to the measure ν introduced in Section 1. This density will be expressed as a linear combination of densities of pairs of os's $(X_{k:n+r}, X_{\ell:n+r})$, $1 \le k, \ell \le n + r$. The general formula is quite complicated technically as can be seen in Section 2.2, however the basic ideas are the same as in the simple case of $r = 0$ which, as a warm up, is considered first in Section 2.1.

2.1. Original sample and its extension – the case of $r = 0$

Let $\mathbb{R}^n_{\neq} = \{\underline{x} \in \mathbb{R}^n : x_i \neq x_j \text{ for } i \neq j\}$ and $\mathbb{R}^n_{\uparrow} = \{\underline{x} \in \mathbb{R}^n : x_1 < \cdots < x_n\}$. A vector with increasingly sorted components of $\underline{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_{\neq}$ will be denoted by $\mathrm{sort}_n(\underline{x}) := (x_{1:n}, \ldots, x_{n:n}) \in \mathbb{R}^n_{\uparrow}$ and $\sigma_n(\underline{x}) = \tau \in \mathcal{S}_n$ (set of permutations of $\{1, \ldots, n\}$) defined by $\tau(i) = j$ if $x_{i:n} = x_j$. The correspondence $\underline{x} \in \mathbb{R}^n_{\neq} \leftrightarrow (\mathrm{sort}_n(\underline{x}), \sigma_n(\underline{x})) \in \mathbb{R}^n_{\uparrow} \times \mathcal{S}_n$ is bijective.

For $\underline{x} \in \mathbb{R}^n_{\neq}$ denote $\underline{x}^{(m)} := (x_1, \ldots, x_m) \in \mathbb{R}^m_{\neq}$, $m = 1, \ldots, n$. Then $\mathrm{sort}_m(\underline{x}^{(m)}) = (x_{1:m}, \ldots, x_{m:m})$ and $\sigma_m(\underline{x}^{(m)})$, $m = 1, \ldots, n$, are sequences of increasing lengths that keep track of the sorting up to the sequential observation of the mth component of $\underline{x}$. For instance, if $\underline{x} = (2.3,\ 1.7,\ 3.4,\ 2.5,\ 1.2)$ then

$$\mathrm{sort}_1(\underline{x}^{(1)}) = (2.3), \quad \sigma_1(\underline{x}^{(1)}) = (1)$$
$$\mathrm{sort}_2(\underline{x}^{(2)}) = (1.7,\ 2.3), \quad \sigma_2(\underline{x}^{(2)}) = (21)$$
$$\mathrm{sort}_3(\underline{x}^{(3)}) = (1.7,\ 2.3,\ 3.4), \quad \sigma_3(\underline{x}^{(3)}) = (213)$$
$$\mathrm{sort}_4(\underline{x}^{(4)}) = (1.7,\ 2.3,\ 2.5,\ 3.4), \quad \sigma_4(\underline{x}^{(4)}) = (2143)$$
$$\mathrm{sort}_5(\underline{x}^{(5)}) = (1.2,\ 1.7,\ 2.3,\ 2.5,\ 3.4), \quad \sigma_5(\underline{x}^{(5)}) = (52143).$$

Observe that for $m < n$, $\sigma_m(\underline{x}^{(m)})$ is obtained from $\sigma_n(\underline{x})$ by deletion of $m+1, \ldots, n$.

Given a permutation $\tau \in \mathcal{S}_n$, let us denote by $\tau^{(m)} \in \mathcal{S}_m$ the permutation obtained from τ by deletion of the elements $m+1, \ldots, n$. For (fixed) values $i,\ k,\ m,\ n$ such that $1 \le i \le m$, $1 \le k \le n$ and $m \le n$, let us define

$$A_{i:m;\,k:n} = \{\tau \in \mathcal{S}_n : \tau(k) = \tau^{(m)}(i)\} \in \mathcal{S}_n.$$

Note that for any $\underline{x} \in \mathbb{R}^n_{\neq}$

$$\sigma_n(\underline{x}) \in A_{i:m;\,k:n} \quad \Leftrightarrow \quad x_{i:m} = x_{k:n}. \tag{5}$$

For instance, in the previous example, $\sigma_5(\underline{x}) = (52143) \in A_{2:4;\,3:5}$ because $x_{2:4} = x_{3:5} = 2.3$.

Since $(X_1, \ldots, X_n)$ has absolutely continuous distribution $\underline{X} \in \mathbb{R}^n_{\neq}$ $\mathbb{P}$-a.s. Therefore, $\mathrm{sort}_n(\underline{X})$ and $\sigma_n(\underline{X})$ are well defined $\mathbb{P}$-a.s. In particular, $\mathrm{sort}_n(\underline{X}) = (X_{1:n}, \ldots, X_{n:n})$ are the os's from the sample of size n.

Lemma 2.1: *Random elements* $\mathrm{sort}_n(\underline{X})$ *and* $\sigma_n(\underline{X})$ *are independent.*

The result follows immediately from the fact that the distribution of $(X_1, \ldots, X_n)$ is invariant under permutation and that ties appear with probability zero.

For $n \ge 1$, and $k \ne j$ with $1 \le j,\ k \le n$, it is well known that $(X_{k:n},\ X_{j:n})$ has a density with respect to μ_2. This density, denoted here by $f_{k,j:n}$, see, e.g. David and Nagaraja [1], p.12 for the explicit expression in terms of F, $\bar{F}$ and f, satisfies

$$\begin{aligned}\mathbb{P}(X_{k:n} \le x,\ X_{j:n} \le y) &= \iint_{(-\infty,x]\times(-\infty,y]} f_{k,j:n}(s,t)\,\mathrm{d}\mu_2(s,t) \\ &= \iint_{(-\infty,x]\times(-\infty,y]} f_{k,j:n}(s,t)\,\mathrm{d}\nu(\mathrm{s},\mathrm{t}),\end{aligned} \tag{6}$$

where for the last equality to hold we chose a version of the density $f_{k,j:n}$ satisfying $f_{k,j:n}(s,s) = 0$, $s \in \mathbb{R}$. We also denote the density of $X_{j:n}$ by $f_{j:n}$ for more simplification.

If $k = j$, the random vector $(X_{k:n},\ X_{j:n})$ assumes values on the diagonal of $\mathbb{R}^2$ so that it does not have a density with respect to μ_2, but it has a density with respect to ν of the form $f_{j,j:n}(s,t) = f_{j:n}(s)\delta_{s,t}$ (with $\delta_{s,t}$ the Kronecker's delta). Indeed, we have

$$\begin{aligned}\mathbb{P}(X_{j:n} \le x,\ X_{j:n} \le y) &= \mathbb{P}(X_{j:n} \le \min(x,y)) = \int_{-\infty}^{\min(x,y)} f_{j:n}(s)\,\mathrm{d}\mu_1(s) \\ &= \iint_{(-\infty,x]\times(-\infty,y]} f_{j:n}(s)\delta_{s,t}\,\mathrm{d}\nu(s,t) = \iint_{(-\infty,x]\times(-\infty,y]} f_{j,j:n}(s,t)\,\mathrm{d}\nu(s,t).\end{aligned} \tag{7}$$

Theorem 2.2: *For integers* $1 \le i \le m, 1 \le j \le n, m \le n$, *the random vector* $(X_{i:m}, X_{j:n})$ *has an absolutely continuous distribution with respect to* ν *and the density function is of the form*

$$f_{X_{i:m},X_{j:n}}(x,y) = \sum_{k=i}^{i+m-n} \frac{\binom{k-1}{i-1}\binom{m-k}{n-i}}{\binom{m}{n}} f_{k,j:n}(x,y). \tag{8}$$

Proof: A consequence of (5) and (2) is

$$\mathbb{P}(\sigma_n(\underline{X}) \in A_{i:m;k:n}) = \mathbb{P}(X_{i:m} = X_{k:n}) = \frac{\binom{k-1}{i-1}\binom{n-k}{m-i}}{\binom{n}{m}} I_{\{i,\ldots,n-m+i\}}(k). \tag{9}$$

Using Lemma 2.1, (9) and expressions (6) and (7), we get

$$\begin{aligned}
\mathbb{P}(X_{i:m} \le x, X_{j:n} \le y) &= \sum_{k=i}^{i+n-m} \mathbb{P}(X_{i:m} \le x, X_{j:n} \le y, X_{i:m} = X_{k:n}) \\
&= \sum_{k=i}^{i+n-m} \mathbb{P}(X_{k:n} \le x, X_{j:n} \le y, \sigma_n(\underline{X}) \in A_{i:m;\,k:n}) \\
&= \sum_{k=i}^{i+n-m} \mathbb{P}(X_{k:n} \le x, X_{j:n} \le y)\mathbb{P}(\sigma_n(\underline{X}) \in A_{i:m;\,k:n}) \\
&= \sum_{k=i}^{i+n-m} \frac{\binom{k-1}{i-1}\binom{m-k}{n-i}}{\binom{m}{n}} \mathbb{P}(X_{k:n} \le x, X_{j:n} \le y) \\
&= \iint_{(-\infty,x]\times(-\infty,y]} \sum_{k=i}^{i+m-n} \frac{\binom{k-1}{i-1}\binom{m-k}{n-i}}{\binom{m}{n}} f_{k,j:n}(s,t)\,\mathrm{d}\nu(s,\ t),
\end{aligned}$$

which proves the assertion. ■

Note that for $j \notin \{i, \ldots, i+n-m\}$ the distribution $(X_{i:m}, X_{j:n})$ is absolutely continuous with respect to the bivariate Lebesgue measure, μ_2. On the contrary, for $j \in \{i, \ldots, i+n-m\}$, it has a singular part, so that there is no density function with respect to μ_2. The advantage of the measure ν introduced in Section 1 is that the joint distribution of $(X_{i:m}, X_{j:n})$ is absolutely continuous with respect to ν in any case.

Formula (8) implies that conditional distribution $\mathbb{P}_{X_{i:m}|X_{j:n}=y}$ has the density, with respect to the measure ν_y defined by $\nu_y(B) = \mu_1(B) + \delta_B(y)$, $B \in \mathcal{B}(\mathbb{R})$, which reads

$$f_{X_{i:m}|X_{j:n}=y}(x) = \sum_{k=i}^{i+n-m} \frac{\binom{k-1}{i-1}\binom{n-k}{m-i}}{\binom{n}{m}} f_{X_{k:n}|X_{j:n}=y}(x), \tag{10}$$

where $f_{X_{j:n}|X_{j:n}=y}(x) = I_{\{y\}}(x)$. Consequently, the formula for the conditional expectation of $X_{i:m}$ given $X_{j:n}$ as given in (3) follows.

2.2. Overlapping samples – the general case of $r \geq 0$

In order to derive the formula for density of $(X_{i:m}, X^{(r)}_{j:n})$ in the general case, $r \geq 0$, we need first to do a little bit of combinatorics of permutations, which will allow us to find the probabilities

$$\mathbb{P}(X_{i,m} = X_{k:n+r}, X^{(r)}_{j:n} = X_{\ell:n+r}), \quad k, \ell \in \{1, \dots, n+r\}.$$

Consider three disjoint sets

$$\mathcal{A} = \{1, \dots, r\}, \quad \mathcal{B} = \{r+1, r+2, \dots, r+s\},$$
$$\mathcal{C} = \{r+s+1, r+s+2, \dots, r+s+t\}.$$

Denote $r+s+t=n$ and consider the set $\mathcal{S}_n$ of permutations of $\{1, \dots, n\}$. We will be interested in the subset D of permutations from $\mathcal{S}_n$ for which there are exactly i elements from the set $\mathcal{C}$ at the first k positions and there are exactly j elements from the set $\mathcal{A}$ at the first $\ell + k$ positions. That is,

$$D = \{\sigma \in \mathcal{S}_n : |\sigma(\{1, \dots, k\}) \cap \mathcal{C}| = i \text{ and } |\sigma(\{1, \dots, k+\ell\}) \cap \mathcal{A}| = j\}.$$

We assume $i \leq \min\{t, k\}$ and $j \leq \min\{r, k+\ell\}$, since otherwise $D = \emptyset$.

Lemma 2.3: *Let $\mathfrak{D}_{r,s,t,k,\ell,i,j} = |D|$, the number of elements in D. Then*

$$\mathfrak{D}_{r,s,t,k,\ell,i,j} = \frac{n!}{\binom{n}{k,\ell}} \binom{t}{i} \binom{r}{j} \sum_{m=\max\{0,j-\ell\}}^{\min\{j,k-i\}} \binom{j}{m} \binom{s}{k-i-m} \binom{s+t+m-k}{\ell+m-j}. \tag{11}$$

Proof: We denote $(a)_b = a(a-1)\dots(a-b+1)$, where b is positive integer, and $(a)_0 = 1$. Moreover, we follow the rule: $\binom{a}{b} = 0$ if $b < 0$ or $a < b$.

To obtain $\sigma \in D$ we perform the following four steps:

(1) Choose i positions out of $\{1, \dots, k\}$ in $\binom{k}{i}$ ways and fill these positions with elements from $\mathcal{C}$ in $(t)_i$ ways.
(2) For any $m = 0, \dots, j$ choose m out of remaining $k-i$ positions in $\{1, \dots, k\}$ in $\binom{k-i}{m}$ ways and fill them with elements of $\mathcal{A}$ in $(r)_m$ ways. Remaining $k-i-m$ positions out of $\{1, \dots, k\}$ fill with elements of $\mathcal{B}$ in $(s)_{k-i-m}$ ways.
(3) Choose $j-m$ positions for elements of $\mathcal{A}$ from $\{k+1, \dots, k+\ell\}$ in $\binom{\ell}{j-m}$ ways and fill them with elements of $\mathcal{A}$ in $(r-m)_{j-m}$ ways. Remaining $\ell - j + m$ positions out of $\{k+1, \dots, \ell\}$ fill with elements of $\mathcal{B} \cup \mathcal{C}$ in $(s-k+i+m+t-i)_{\ell-j+m} = (s+t+m-k)_{\ell-j+m}$ ways.
(4) The remaining $n-k-\ell$ positions fill with the rest of the elements of $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ in $(n-k-\ell)!$ ways.

Combining these four steps we get

$$|D| = \binom{k}{i}(t)_i \left(\sum_{m=0}^{j} \binom{k-i}{m} (r)_m \, (s)_{k-i+m} \binom{\ell}{j-m} (r-m)_{j-m} \, (s+t+m-k)_{\ell-j+m} \right) \times (n-k-\ell)!.$$

The formula (11) follows by simple transformations involving, e.g. $(r)_m \, (r-m)_{j-m} = (r)_j$. ■

Remark 2.1: Since the subset of permutations D as defined above can alternatively be written as

$$D = \{\sigma \in \mathcal{S}_n : \, |\sigma(\{k+\ell+1,\dots,n\}) \cap \mathcal{A}| = r-j \text{ and } |\sigma(\{k+1,\dots,n\}) \cap \mathcal{C}| = t-i\},$$

we have an equivalent formula for the number of elements in D:

$$|D| = \mathfrak{D}_{t,s,r,n-k-\ell,\ell,r-j,t-i}. \tag{12}$$

In the next result we give explicit forms for $\mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r})$ for all possible configurations of parameters i, m, k, n, r, j, ℓ.

Proposition 2.4: *Let* $A = \{1,\dots,r\}$, $B = \{r+1,\dots,m\}$ *and* $C = \{m+1,\dots,n+r\}$. *Probabilities*

$$p_{r,(i,m,k),(j,n,\ell)} := \mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r}).$$

are non-zero only if $i \le k \le i+n+r-m$ *and* $j \le \ell \le j+r$. *Then*

(i) *for* $k < \ell$

$$p_{r,(i,m,k),(j,n,\ell)} = \frac{(n-j+1)(|A|\mathfrak{D}_{|A|-1,|B|,|C|,k-1,\ell-k-1,k-i,\ell-j-1} + |B|\mathfrak{D}_{|A|,|B|-1,|C|,k-1,\ell-k-1,k-i,\ell-j})}{(r+n-\ell+1)(r+n)!};$$

(ii) *for* $k = \ell$

$$p_{r,(i,m,k),(j,n,k)} = \frac{|B|\mathfrak{D}_{|A|,|B|-1,|C|,k-1,0,k-i,k-j}}{(r+n)!};$$

(iii) *for* $k > \ell$

$$p_{r,(i,m,k),(j,n,\ell)} = \frac{(m-i+1)(|C|\mathfrak{D}_{|C|-1,|B|,|A|,\ell-1,k-\ell-1,\ell-j,k-i-1} + |B|\mathfrak{D}_{|C|,|B|-1,|A|,\ell-1,k-\ell-1,\ell-j,k-i})}{(r+n-k+1)(r+n)!}.$$

Proof of Proposition 2.4, due to its computational complexity, is given in Appendix.

Now we are ready to derive the formula for the density of $f_{X_{i:m},X_{j:n}^{(r)}}$.

The independence property given in Lemma 2.1 allows to write the density of $(X_{i:m}, X_{j:n}^{(r)})$ as a linear combination of densities of bivariate os's from the sample $(X_1, \ldots, X_{n+r})$.

Theorem 2.5:

$$f_{X_{i:m},X_{j:n}^{(r)}}(x,y) = \sum_{k,\ell=1}^{n+r} p_{r,(i,m,k),(j,n,\ell)} f_{k,\ell:n+r}(x,y) \tag{13}$$

with coefficients $p_{r,(i,m,k),(j,n,\ell)}$ *as given in Proposition 2.4.*

Proof: Note that

$$\mathbb{P}(X_{i:m} \le x, , X_{j:n}^{(r)} \le y) = \sum_{k,\ell=1}^{n+r} \mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r}, X_{k:n+r} \le x, X_{\ell:n+r} \le y).$$

From the proof of Proposition 2.4 (see Appendix) it follows that the event $\{X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r}\}$ is a union of events of the form $\{X_{\sigma(1)} \le \ldots \le X_{\sigma(n+r)}\}$, where the union is with respect to permutations from special subsets of $\mathcal{S}_{n+r}$ (these subsets are different in each of three cases: $k < \ell$, $k = \ell$ and $k > \ell$). By Lemma 2.1 it follows that

$$\begin{aligned}
&\mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r}, X_{k:n+r} \le x, X_{\ell:n+r} \le y) \\
&= \mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r})\, \mathbb{P}(X_{k:n+r} \le x, X_{\ell:n+r} \le y).
\end{aligned}$$

Therefore the density $f_{X_{i:m},X_{j:n}^{(r)}}$ of $(X_{i:m}, X_{j:n}^{(r)})$ with respect to the measure ν (introduced in Section 1) assumes the form

$$f_{X_{i:m},X_{j:n}^{(r)}} = \sum_{k,\ell=1}^{n+r} \mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r}) f_{k,\ell:n+r}.$$

Now the result follows by inserting in the above expression correct forms of probabilities $\mathbb{P}(X_{i:m} = X_{k:n+r}, X_{j:n}^{(r)} = X_{\ell:n+r})$ which are given in Proposition 2.4 ■

Below we derive joint densities (with respect to ν) of $(X_{i:m}, X_{j:n}^{(r)})$ in several cases of special interest.

(i) **Order statistics from the original and extended samples.** Without any loss of generality we can assume that $m \le n$. Since

$$\mathbb{P}(X_{i:m} = X_{k:n}) = \mathbb{P}(X_{i:m} = X_{k:n}, X_{j:n}^{(0)} = X_{j:n}). \tag{14}$$

Proposition 2.4 with $r=0$, $j = \ell$, $|A| = 0$, $|B| = m$, $|C| = n - m$ applies and since the left-hand side of (14) does not depend on ℓ we can choose the case $k = \ell$. Therefore,

$$\begin{aligned}
\mathbb{P}(X_{i:m} = X_{k:n}) &= \frac{m\mathfrak{D}_{0,m-1,n-m,k-1,0,k-i,0}}{n!} = \frac{m(n-k)!(k-1)!}{n!}\binom{n-m}{k-i}\binom{m-1}{i-1} \\
&= \frac{\binom{k-1}{i-1}\binom{n-k}{m-i}}{\binom{n}{m}},
\end{aligned}$$

and thus the formula for the density of $(X_{i:m}, X_{j:n})$ agrees with (8).

(ii) **Moving maxima.** We consider $(X_{n:n}, X_{n:n}^{(r)})$. From Proposition 2.4 we get

$$\mathbb{P}(X_{n:n} = X_{k:n+r},\ X_{n:n}^{(r)} = X_{n+r:n+r}) = \frac{\binom{k-1}{n-1}}{\binom{n+r}{r}}, \quad k = n, n+1, \ldots, n+r-1,$$
$$\mathbb{P}(X_{n:n} = X_{n+r:n+r},\ X_{n:n}^{(r)} = X_{n+r:n+r}) = \frac{n-r}{n+r},$$
$$\mathbb{P}(X_{n:n} = X_{n+r:n+r},\ X_{n:n}^{(r)} = X_{\ell:n+r}) = \frac{\binom{\ell-1}{n-1}}{\binom{n+r}{r}}, \quad \ell = n, n+1, \ldots, n+r-1.$$

Consequently, Theorem 2.5 gives

$$f_{X_{n:n}, X_{n:n}^{(r)}}(x, y) = \begin{cases} \displaystyle\sum_{k=n}^{n+r-1} \frac{\binom{k-1}{n-1}}{\binom{n+r}{r}} f_{k,n+r:n+r}(x, y), & x < y, \\ \displaystyle\frac{n-r}{n+r} f_{n+r:n+r}(x), & x = y, \\ \displaystyle\sum_{k=n}^{n+r-1} \frac{\binom{k-1}{n-1}}{\binom{n+r}{r}} f_{k,n+r:n+r}(y, x), & x > y. \end{cases}$$

(iii) **Moving minima.**We consider $(X_{1:n}, X_{1:n}^{(r)})$. Then from Proposition 2.4 we get

$$\mathbb{P}(X_{1:n} = X_{1:n+r}, \quad X_{1:n}^{(r)} = X_{\ell:n+r}) = \frac{\binom{n+r-\ell}{n-1}}{\binom{n+r}{r}}, \quad \ell = 2, 3, \ldots, r+1,$$
$$\mathbb{P}(X_{1:n} = X_{1:n+r}, \quad X_{1:n}^{(r)} = X_{1:n+r}) = \frac{n-r}{n+r},$$
$$\mathbb{P}(X_{1:n} = X_{k:n+r}, \quad X_{1:n}^{(r)} = X_{1:n+r}) = \frac{\binom{n+r-k}{n-1}}{\binom{n+r}{r}}, \quad k = 2, 3, \ldots, r+1.$$

Consequently, Theorem 2.5 gives

$$f_{X_{1:n}, X_{1:n}^{(r)}}(x, y) = \begin{cases} \displaystyle\sum_{k=2}^{r+1} \frac{\binom{n+r-k}{n-1}}{\binom{n+r}{r}} f_{1,k:n+r}(x, y), & x < y, \\ \displaystyle\frac{n-r}{n+r} f_{1:n+r}(x), & x = y, \\ \displaystyle\sum_{k=2}^{r+1} \frac{\binom{n+r-k}{n-1}}{\binom{n+r}{r}} f_{1,k:n+r}(y, x), & x > y. \end{cases}$$

(iv) **Moving ith os's.** We consider $(X_{i:m}, X^{(1)}_{i:m})$. Then from Proposition 2.4 we get

$$\mathbb{P}(X_{i:m} = X_{i:m+1}, \quad X^{(1)}_{i:m} = X_{i:m+1}) = \frac{(m-i+1)(m-i)}{(m+1)m},$$

$$\mathbb{P}(X_{i:m} = X_{i:m+1}, \quad X^{(1)}_{i:m} = X_{i+1:m+1}) = \frac{i(m-i+1)}{(m+1)m},$$

$$\mathbb{P}(X_{i:m} = X_{i+1:m+1}, \quad X^{(1)}_{i:m} = X_{i:m+1}) = \frac{i(m-i+1)}{(m+1)m},$$

$$\mathbb{P}(X_{i:m} = X_{i+1:m+1}, \quad X^{(1)}_{i:m} = X_{i+1:m+1}) = \frac{i(i-1)}{(m+1)m}.$$

Consequently, Theorem 2.5 gives

$$f_{X_{i:m},X^{(1)}_{i:m}}(x,y) = \begin{cases} \dfrac{i(m-i+1)}{(m+1)m} f_{i,i+1:m+1}(x,y), & x<y, \\ \dfrac{(m-i+1)(m-i)}{(m+1)m} f_{i:m+1}(x) + \dfrac{i(i-1)}{(m+1)m} f_{i+1:m+1}(x), & x=y, \\ \dfrac{i(m-i+1)}{(m+1)m} f_{i,i+1:m+1}(y,x), & x>y. \end{cases}$$

3. Regression of overlapping os's

From Proposition 2.4 we know that $p_{r,(i,m,k),(j,n,l)}$ are non-zero only if $i \le k \le i+n+r-m$ and $j \le l \le j+r$. This together with (13) implies

$$f_{X_{i:m},X^{(r)}_{j:n}}(x,y) = \sum_{k=i}^{n+r-m+i} \sum_{\ell=j}^{j+r} p_{r,(i,m,k),(j,n,l)} f_{k,\ell:n+r}(x,y).$$

Consequently, the conditional distribution $\mathbb{P}_{X_{i:m}|X^{(r)}_{j:n}=y}$ has a density with respect to $\nu_y(dx) = \mu_1(dx) + \delta_y(dx)$ of the form

$$\begin{aligned} f_{X_{i:m}|X^{(r)}_{j:n}=y}(x) &= \sum_{k=i}^{n+r-m+i} \sum_{\ell=j}^{j+r} p_{r,(i,m,k),(j,n,l)} f_{X_{k:n+r}|X_{\ell:n+r}=y}(x) \frac{f_{\ell:n+r}(y)}{f_{j:n}(y)} \\ &= \sum_{k=i}^{n+r-m+i} \sum_{\ell=j}^{j+r} p_{r,(i,m,k),(j,n,l)} \frac{\ell\binom{n+r}{\ell}}{j\binom{n}{j}} F^{\ell-j}(y)\bar{F}^{j+r-\ell}(y) f_{X_{k:n+r}|X_{\ell:n+r}=y}(x). \end{aligned}$$

and the conditional distribution $\mathbb{P}_{X^{(r)}_{j:n}|X_{i:m}=x}$ has a density with respect to $\nu_x(dy) = \mu_1(dy) + \delta_x(dy)$ of the form

$$\begin{aligned} f_{X^{(r)}_{j:n}|X_{i:m}=x}(y) &= \sum_{k=i}^{n+r-m+i} \sum_{\ell=j}^{j+r} p_{r,(i,m,k),(j,n,l)} f_{X_{\ell:n+r}|X_{k:n+r}=x}(y) \frac{f_{k:n+r}(x)}{f_{i:m}(x)} \\ &= \sum_{k=i}^{n+r-m+i} \sum_{\ell=j}^{j+r} p_{r,(i,m,k),(j,n,l)} \frac{k\binom{n+r}{k}}{i\binom{m}{i}} F^{k-i}(x)\bar{F}^{n+r-m-k+i}(x) f_{X_{\ell:n+r}|X_{k:n+r}=x}(y). \end{aligned}$$

Therefore,

$$\mathbb{E}(X_{i:m}|X_{j:n}^{(r)}=y)=\sum_{k=i}^{n+r-m+i}\sum_{\ell=j}^{j+r}p_{r,(i,m,k),(j,n,l)}$$
$$\times\frac{\ell\binom{n+r}{\ell}}{j\binom{n}{j}}F^{\ell-j}(y)\bar{F}^{j+r-\ell}(y)\,\mathbb{E}(X_{k:n+r}|X_{\ell:n+r}=y) \quad (15)$$

and

$$\mathbb{E}(X_{j:n}^{(r)}|X_{i:m}=x)=\sum_{k=i}^{n+r-m+i}\sum_{\ell=j}^{j+r}p_{r,(i,m,k),(j,n,l)}$$
$$\times\frac{k\binom{n+r}{k}}{i\binom{m}{i}}F^{k-i}(x)\bar{F}^{n+r-m-k+i}(x)\,\mathbb{E}(X_{\ell:n+r}|X_{k:n+r}=x). \quad (16)$$

That is, both regressions we are interested in are represented through rather complicated expressions (15) and (16). Thus characterizations or identifiability questions for parent distributions through the form of $\mathbb{E}(X_{i:m}|X_{j:n}^{(r)})$ or $\mathbb{E}(X_{j:n}^{(r)}|X_{i:m})$ seems to be a difficult task in such a general framework. Therefore we will concentrate rather on the special cases of $r=0$ (when $m<n$) distinguishing two quite different subcases: in Section 3.2 we will consider characterizations by linearity of $\mathbb{E}(X_{i:m}|X_{j:n})$ while in Section 3.3 we will study identification through $\mathbb{E}(X_{j:n}|X_{i:m})$. For $r>0$ we will consider only the simplest case of $r=1$ and $m=n=2$ in Section 3.1.

3.1. *Identifiability through regression functions when $r=1$ and $m=n=2$*

Here we only consider the simplest case of os's from overlapping samples (X_1,X_2) and (X_2,X_3), that is the case of $r=1$, $m=n=2$. Then (15) gives

(i) $\mathbb{E}(X_{2:2}|X_{2:2}^{(1)}=y)=\frac{y}{2}F(y)+\int_y^\infty xf(x)\,\mathrm{d}x+\frac{1}{F(y)}\int_{-\infty}^{y}xF(x)f(x)\,\mathrm{d}x,$

(ii) $\mathbb{E}(X_{1:2}|X_{1:2}^{(1)}=y)=\frac{y}{2}\bar{F}(y)+\int_{-\infty}^{y}xf(x)\,\mathrm{d}x+\frac{1}{\bar{F}(y)}\int_y^\infty x\bar{F}(x)f(x)\,\mathrm{d}x,$

(iii) $\mathbb{E}(X_{1:2}|X_{2:2}^{(1)}=y)=\frac{y}{2}\bar{F}(y)+\frac{1}{2F(y)}\int_{-\infty}^{y}xf(x)\,\mathrm{d}x+\int_{-\infty}^{y}xf(x)\,\mathrm{d}x-\frac{1}{F(y)}\int_{-\infty}^{y}xF(x)f(x)\,\mathrm{d}x,$

(iv) $\mathbb{E}(X_{2:2}|X_{1:2}^{(1)}=y)=\frac{y}{2}F(y)+\frac{1}{2\bar{F}(y)}\int_y^\infty xf(x)\,\mathrm{d}x+\int_y^\infty xf(x)\,\mathrm{d}x-\frac{1}{\bar{F}(y)}\int_y^\infty x\bar{F}(x)f(x)\,\mathrm{d}x.$

We will show that each of these four regressions determines uniquely the parent distribution. Note that for $Y_i=-X_i$ and $u=-y$ we have $\mathbb{E}(X_{1:2}|X_{1:2}^{(1)}=y)=-\mathbb{E}(Y_{2:2}|Y_{2:2}^{(1)}=u)$ and $\mathbb{E}(X_{1:2}|X_{2:2}^{(1)}=y)=-\mathbb{E}(Y_{2:2}|Y_{1:2}^{(1)}=u)$. Consequently, (i) and (ii) as well as (iii) and (iv) above are equivalent.

Theorem 3.1: *Let the parent distribution be absolutely continuous distribution with the interval support (a,b). Then regression function $\mathbb{E}(X_{2:2}|X_{2:2}^{(1)}=y)$ (alternatively,*

$\mathbb{E}(X_{1:2}|X_{1:2}^{(1)} = y))$, $y \in (a,b)$, determines uniquely the distribution of X if $(a,b) \subsetneq \mathbb{R}$. If $(a,b) = \mathbb{R}$ it determines the distribution up to a shift.

Proof: It suffices to consider only the case of $\mathbb{E}(X_{2:2}|X_{2:2}^{(1)})$.

Let us denote $K(y) = \mathbb{E}(X_{2:2}|X_{2:2}^{(1)} = y)$, $y \in (a,b)$. Performing integration by parts in (i) we get

$$K(y) = y + \int_y^b \bar{F}(x)\,\mathrm{d}x - \frac{1}{2F(y)} \int_a^y F^2(x)\,\mathrm{d}x.$$

Consequently, if $K(y)$ is the same for two distribution functions F and G, which are strictly increasing on (a,b) and thus have differentiable quantile functions Q_F and Q_G, then

$$2t\int_t^1 (1-w)\,\mathrm{d}Q_F(w) - \int_0^t w^2\,\mathrm{d}Q_F(w) = 2t\int_t^1 (1-w)\,\mathrm{d}Q_G(w) - \int_0^t w^2\,\mathrm{d}Q_G(w).$$

For $H := (Q_F - Q_G)'$ we obtain

$$L := 2t\int_t^1 (1-w)\,H(w)\,\mathrm{d}w = \int_0^t w^2\,H(w)\,\mathrm{d}w =: R, \quad t \in (0,1). \tag{17}$$

Differentiating (17) with respect to t twice we obtain

$$4(1-t)H(t) + t(2-t)H'(t) = 0, \quad t \in (0,1).$$

Consequently, there exists a constant c such that

$$H(t) = \frac{c}{(2-t)^2t^2}, \quad t \in (0,1).$$

Now, assuming that $c \neq 0$ we plug H back into (17). After cancelling c at the left-hand side we get

$$L = 2t\int_t^1 \frac{1-w}{w^2(2-w)^2}\,dw < 2t\int_t^1 w^{-2}\,\mathrm{d}w = 2(1-t),$$

while at the right-hand side we have

$$R = \int_0^t (2-w)^{-2}\,\mathrm{d}w > \frac{t}{4}.$$

Therefore, for any $t \in (0,1)$ we have $t < 8(1-t)$, which is impossible for t sufficiently close to 1. Consequently, $c = 0$, and thus $Q_F - Q_G = B$ for some constant B. It implies that either $B = 0$ or $a = -\infty$ and $b = \infty$. In the latter case for any $y \in \mathbb{R}$ there exists unique $x \in \mathbb{R}$ such that $y = Q_G(F(x))$. Therefore

$$G(y) = F(x) = F(Q_F(F(x))) = F(Q_G(F(x)) + B) = F(y+B).$$

This completes the proof. ■

Theorem 3.2: *Let X has absolutely continuous distribution with the interval support (a,b). Then the regression function $\mathbb{E}(X_{1:2}|X_{2:2}^{(1)} = y)$ (alternatively, $\mathbb{E}(X_{2:2}|X_{1:2}^{(1)} = y)$), $y \in (a,b)$,* *determines uniquely the distribution of X if $(a, b) \subsetneq \mathbb{R}$. If $(a, b) = \mathbb{R}$ it determines the distribution up to a shift.*

Proof: Again we consider only the case of $\mathbb{E}(X_{1:2}|X_{2:2}^{(1)})$.

Let us denote $K(y) = \mathbb{E}(X_{1:2}|X_{2:2}^{(1)} = y)$, $y \in (a, b)$. Performing integration by parts in (iii) we get

$$K(y) = y - \left(\frac{1}{2F(y)} + 1\right) \int_a^y F(x)\,\mathrm{d}x + \frac{1}{2F(y)} \int_a^y F^2(x)\,\mathrm{d}x.$$

Consequently, if $K(y)$ is the same for two distribution functions F and G, which are strictly increasing on (a, b) and thus have differentiable inverses Q_F and Q_G, then from the above formula we get

$$(1 + 2t) \int_0^t w\,\mathrm{d}Q_F(w) - \int_0^t w^2\,\mathrm{d}Q_F(w) = (1 + 2t) \int_0^t w\,\mathrm{d}Q_G(w) - \int_0^t w^2\,\mathrm{d}Q_G(w).$$

As in the proof above we denote $H = (Q_F - Q_G)'$. Then we have

$$L := (1 + 2t) \int_0^t w\,H(w) \quad \mathrm{d}w = \int_0^t w^2\,H(w) \quad \mathrm{d}w =: R, \quad t \in (0, 1). \tag{18}$$

Differentiating (18) twice with respect to t we obtain

$$(1 + 4t)H(t) + t(1 + t)H'(t) = 0, \quad t \in (0, 1).$$

Consequently, there exists a constant c such that

$$H(t) = \frac{c}{t(1 + t)^3}, \quad t \in (0, 1).$$

Now, assuming that $c \neq 0$ we plug H back into (18). After cancelling c at the left-hand side we have

$$L = (1 + 2t) \int_0^t (1 + w)^{-3}\,dw = \frac{1 + 2t}{2}\left(1 - \frac{1}{(1 + t)^2}\right) = \frac{(1 + 2t)((1 + t)^2 - 1)}{2(1 + t)^2}$$

while at the right-hand side we have

$$R = \int_0^t \frac{w}{(1 + w)^3}\,\mathrm{d}w = \frac{t^2}{2(1 + t)^2}.$$

Obviously, $L \neq R$. Consequently, $c = 0$, and thus $Q_F - Q_G = B$ for some constant B. To complete the proof we proceed as in the end of the proof of the previous theorem. ■

3.2. Linearity of regression of $X_{i:m}$ given $X_{j:n}$, $m < n$

In this subsection we consider linearity of regression as given in (4) when $i < j < n - m + i$ since, as mentioned before, the cases $j \leq i$ and $j \geq n - m + i$ have already been discussed in DW.

It seems that the only results available in the literature are for $m = i = 1$, i.e. with $X_{i:m} = X_1$. In particular, the case $n = 2k+1, j = k+1$ was considered in Wesołowski and Gupta [16] (referred to by WG in the sequel), see also Nagaraja and Nevzorov [17]. In WG it was shown that if $\mathbb{E}\,X_1 = 0$ the relation

$$\mathbb{E}(X_{1:1}|X_{k+1:2k+1}) = aX_{k+1:2k+1}, \tag{19}$$

implies $a \geq (k + 1/2k + 1)$ and up to a scaling factor uniquely determines the parent distribution. Examples of such distributions (up to scale factors) are:

- The uniform distribution on $[-1, 1]$ when $a = (k + 1/2k + 1)$.
- The Student t_2 distribution when $a = 1$. In fact, the same characterization of the Student t_2 distribution, sometimes with a different phrasing, e.g. writing $(1/n)\sum_{j=1}^n X_j$ instead of $X_{1:1} = X_1$, under the conditional expectation in (19) with $a = 1$ is given in Nevzorov [18] and Nevzorov et al. [19]. (These references apparently missed the result of WG.)
- The distribution with the cdf $F(x) = \frac{1}{2}\left(1 + (\sqrt{2}x)/(\sqrt{\sqrt{4+x^4}+x^2})\right)$, $x \in \mathbb{R}$, when $a = (4k + 3/3(2k + 1))$.

Related regression characterizations can be found also, e.g. in Balakrishnan and Akhundov [20], Akhundov et al. [21] (referred to by ABN in the sequel), Nevzorova et al. [22], Marudova and Nevzorov [23], Akhundov and Nevzorov [24], Yanev and Ahsanullah [25].

For describing our results in the sequel (as well as results from the literature) it is convenient to introduce the family of complementary beta distributions defined in Jones [26] for a restricted range of parameters α, β and then extended to any $\alpha, \beta \in \mathbb{R}$ in Jones [27]. Slightly changing the original Jones' formulation we say that an absolutely continuous distribution with distribution function F and density f belongs to the family of complementary beta distributions $CB(\alpha, \beta)$, $\alpha, \beta \in \mathbb{R}$, if

$$F^{\alpha}(x)\,\bar{F}^{\beta}(x) \propto f(x), \quad x \in \mathbb{R}.$$

Kamps [28] introduced $CB(-p, 1 + p - q)$ for integer p and q in the context of characterization of distributions by recurrence relations between moments of os's from the original and extended samples. Nevzorov et al. [19] observed that $CB(\alpha, \alpha)$ includes several interesting special cases: Student t_2 distribution for $\alpha = 3/2$, logistic distribution for $\alpha = 1$, squared sine distribution with $F(x) = \sin^2(x)I_{[0,\pi/2]}(x)$ for $\alpha = 1/2$. Extensive discussion of properties of $CB(\alpha, \beta)$ family is given in Jones [27]. In particular, it is observed there that if $\alpha = 0$ then $\beta = 1$ gives the exponential distribution, $\beta > 1$ gives the Pareto laws and $\beta < 1$ are power distributions on a bounded interval.

The family $CB(\alpha, \beta)$ is a subclass of the family $GS(\alpha, \beta, \gamma)$, the latter being defined through the equation

$$F^{\alpha}(x)(1 - F^{\gamma}(x))^{\beta} \propto f(x),$$

was introduced in Muiño et al. [29], i.e. $CB(\alpha, \beta) = GS(\alpha, \beta, 1)$. Another subclass of GS distributions was characterized by linearity of regression of sum of $X_{k-j:k-j}$ and $X_{k+r:k+r}$ given $X_{k:k}$ in Marudova and Nevzorov [23]. Basic properties of os's from a sample with the parent distribution belonging to the GS family were studied in Mohie El-Din et al. [30]. In a recent paper Hu and Lin [31] considered an extension of $GS(1,2,\gamma)$ class defined by the equation

$$F(x)(1-F^{\gamma}(x)) \propto x^a f(x),$$

for $\gamma > 0$ and $a \in [0,1]$ in the context of characterization by random exponential shifts of os's. All these equations can be treated as generalized versions of the logistic growth model or even as a more flexible growth model introduced in Richards [32].

To describe results in this section it will be also convenient to use the quantile function Q and quantile density function q. If f is a strictly positive density on some (possibly unbounded) interval (a,b) then the respective distribution function F is invertible on (a,b) and thus its inverse, quantile function Q, is well defined on $(0,1)$. Moreover it is absolutely continuous with respect to the Lebesgue measure on $(0,1)$, that is $Q(y) = \int_{y_0}^{y} q(u)\,du$ for some $y_0 \in [0,1]$. The function q is called the quantile density function. Note that

$$f = T(F) \quad \Leftrightarrow \quad q = 1/T, \tag{20}$$

and thus q together with y_0 uniquely determine F. In particular, the quantile density q for a distribution in $CB(\alpha,\beta)$ has the form

$$q(u) \propto u^{-\alpha}(1-u)^{-\beta}, \quad u \in (0,1).$$

Remark 3.1: Note that the regression condition

$$\mathbb{E}(X_{1:1}|X_{j:n}) = aX_{j:n}, \tag{21}$$

has been reduced in WG to the condition $M_\lambda(x) = Ax$ where $A = (na - 1/n - 1)$ and

$$M_\lambda(x) := \lambda\mathbb{E}(X|X < x) + (1-\lambda)\mathbb{E}(X|X > x)$$

with $\lambda = (j - 1/n - 1)$. In particular, formula (4) in WG says that for a positive A (necessarily, $A \geq 1/2$) condition (21) holds if and only if the quantile function Q satisfies

$$Q(y) = cy^{\frac{\lambda}{A}-1}(1-y)^{\frac{1-\lambda}{A}-1}(\lambda - y), \quad y \in (0,1). \tag{22}$$

Differentiating (22) we get the following expression for the quantile density

$$q(y) \propto y^{\frac{\lambda}{A}-2}(x)(1-y)^{\frac{1-\lambda}{A}-2}((\lambda - A)\lambda - 2\lambda(1-A)y + (1-A)y^2).$$

Therefore, for $A = 1$

$$q(y) \propto y^{-1-\frac{n-j}{n-1}}(1-y)^{-1-\frac{j-1}{n-1}}, \quad y \in (0,1).$$

Since $A = 1$ implies $a = 1$ we conclude that $\mathbb{E}(X_{1:1}|X_{j:n}) = X_{j:n}$ characterizes the $CB(1 + (n - j/n - 1), 1 + (j - 1/n - 1))$ family. This result was independently proved in Balakrishnan and Akhundov [20], see also Corollary 2.1 in ABN and Nevzorov [33].

ABN characterized also the family $CB(1 + (1-\lambda)i,\ 1 + \lambda i)$ for positive integer i and $\lambda \in (0,1)$ through the condition

$$\mathbb{E}(\lambda X_{i:2i+1} + (1-\lambda)X_{i+2:2i+1}|X_{i+1:2i+1}) = X_{i+1:2i+1}.$$

This result, stated as Theorem 3.1 in ABN, includes the result of Nevzorov [18] who characterized the family $CB(1 + (i/2),\ 1 + (i/2))$ by the above condition with $\lambda = 1/2$.

Remark 3.2: Before we proceed further with a new related characterization let us add here a small complement to Theorem 3.1 of ABN (also valid for Theorem. 2 of Nevzorov [18]). The proof as given in ABN (similarly as that of the main result from Balakrishnan and Akhundov [20]), exploits the ideas originating from the proof of Theorem. 1 in Nevzorov [18]. In particular, an important part of the proof of Theorem 3.1 in ABN is integration by parts in which the following two identities

$$\lim_{u\to-\infty} uF^i(u) = 0 \quad \text{and} \quad \lim_{u\to\infty} u\bar{F}^i(u) = 0 \tag{23}$$

are necessary. Therefore either these conditions or suitable moments conditions, see Lemma 3.3, have to strengthen the assumptions of Theorem 3.1 of ABN. It is of some interest to note that distributions from $CB(1+(1-\lambda)i,\ 1+\lambda i)$ appearing in the statement of Theorem 3.4 given below do not possess finite expectations.

The result below shows that natural integrability assumptions are responsible for speed of convergence to zero of suitable powers of the left and right tails of the cdf F. In particular, it follows that integrability of both $X_{i:2i+1}$ and $X_{i+2:2i+1}$ imply (23).

Lemma 3.3: *If* $\mathbb{E}|X_{k:n}| < \infty$ *then*

$$\lim_{x\to-\infty} xF^k(x) = 0 \quad \textit{and} \quad \lim_{x\to\infty} x\bar{F}^{n-k+1}(x) = 0. \tag{24}$$

Proof: Note that, for $x < 0$,

$$\int_{-\infty}^{x} |t|\, F^{k-1}(t)\bar{F}^{n-k}(t)f(t)\mathrm{d}t \geq |x|\bar{F}^{n-k}(x)\int_{-\infty}^{x} F^{k-1}(t)f(t)\,\mathrm{d}t = \frac{1}{k}|z|\bar{F}^{n-k}(x)F^k(x).$$

Integrability of $X_{k:n}$ implies that the left-hand side above converges to 0 as $x\to-\infty$. The first limit in (24) follows since $\lim_{x\to-\infty}\bar{F}^{n-k}(x) = 1$.

Similarly, for $x > 0$,

$$\begin{aligned}&\int_{x}^{\infty} |t|\, F^{k-1}(t)\bar{F}^{n-k}(t)f(t)\mathrm{d}t\\ &\quad\geq |x|F(x)^{k-1}\int_{x}^{\infty}\bar{F}^{n-k}(t)f(t)\,\mathrm{d}t = \frac{1}{n-k+1}|x|F^{k-1}(x)\bar{F}^{n-k+1}(x).\end{aligned}$$

Since $\mathbb{E}|X_{k:n}| < \infty$ the left-hand side above converges to 0 as $x\to\infty$. Since $\lim_{x\to\infty} F^{k-1}(x) = 1$ the second limit in (24) follows. ■

Basically, we have described the state of art of the characterizations by linearity of regression of an os from the original sample given an os from the extended sample. In the next result we present a new contribution whose proof borrows some ideas from Nevzorov [18].

Theorem 3.4: *Let $2 \le j \le n-1$. Assume that $\mathbb{E}|X_{j-1:n-2}| < \infty$ and that $f > 0$ on some interval (a, b) (possibly unbounded). If*

$$\mathbb{E}(X_{j-1:n-2}|X_{j:n}) = X_{j:n}, \tag{25}$$

then the parent distribution has the quantile density q of the form

$$q(u) \propto \frac{j-1+(n-2j+1)u}{u^{1+(j-1)\lambda}\,(1-u)^{1+(n-j)(1-\lambda)}}, \quad u \in (0,1), \tag{26}$$

where $\lambda = j(j-1)/(n-j+1)(n-j) + j(j-1)$.

Proof: Due to (3) we can write

$$\mathbb{E}(X_{j-1:n-2}|X_{j:n}) = \frac{(n-j+1)(n-j)}{n(n-1)}\mathbb{E}(X_{j-1:n}|X_{j:n}) + \frac{2(n-j)(j-1)}{n(n-1)}X_{j:n} + \frac{j(j-1)}{n(n-1)}\mathbb{E}(X_{j+1:n}|X_{j:n}).$$

Combining the above equation with (25) we obtain

$$(1-\lambda)\mathbb{E}(X_{j-1:n}|X_{j:n}) + \lambda\mathbb{E}(X_{j+1:n}|X_{j:n}) = X_{j:n}. \tag{27}$$

Note that

$$\mathbb{E}(X_{j-1:n}|X_{j:n} = x) = \int_{-\infty}^{x} t\frac{dF^{j-1}(t)}{F^{j-1}(x)}.$$

Now, Lemma 3.3 implies $\lim_{x\to-\infty} xF^{j-1}(x) = 0$ and thus integration by parts gives

$$\mathbb{E}(X_{j-1:n}|X_{j:n} = x) = x - \frac{\int_{-\infty}^{x} F^{j-1}(t)\,\mathrm{d}t}{F^{j-1}(x)}.$$

On the other hand,

$$\mathbb{E}(X_{j+1:n}|X_{j:n} = x) = -\int_{x}^{\infty} t\,\frac{d\bar{F}^{n-j}(t)}{\bar{F}^{n-j}(x)}.$$

Lemma 3.3 implies $\lim_{x\to\infty} x\bar{F}^{n-j}(x) = 0$ and thus integration by parts gives

$$\mathbb{E}(X_{j+1:n}|X_{j:n} = x) = x + \frac{\int_{x}^{\infty} \bar{F}^{n-j}(t)\,\mathrm{d}t}{\bar{F}^{n-j}(x)}.$$

Consequently, (27) assumes the form

$$\frac{\int_{-\infty}^{x} F^{j-1}(t)\,\mathrm{d}t}{F^{j-1}(x)} = \frac{\lambda}{1-\lambda}\frac{\int_{x}^{\infty} \bar{F}^{n-j}(t)\,\mathrm{d}t}{\bar{F}^{n-j}(x)}.$$

Let us introduce two functions G and H defined as follows: $G(x) = \int_{-\infty}^{x} F^{j-1}(t)\,\mathrm{d}t$ and $H(x) = \int_{x}^{\infty} \bar{F}^{n-j}(t)\,\mathrm{d}t$. Consequently, the above equation can be written as

$$\frac{G'}{G} = -\frac{1-\lambda}{\lambda}\frac{H'}{H}. \tag{28}$$

Upon integration we get $GH^{\frac{1-\lambda}{\lambda}} = K$ for some constant K. Then after multiplication of (28) by $GH^{1/\lambda}$ we get

$$H^{1/\lambda} = -\frac{1-\lambda}{\lambda}\left(GH^{\frac{1-\lambda}{\lambda}}\right)\frac{H'}{G'} = K\frac{1-\lambda}{\lambda}\frac{\bar{F}^{n-j}}{F^{j-1}}$$

and thus

$$H \propto F^{-(j-1)\lambda}\bar{F}^{(n-j)\lambda}.$$

Now we differentiate the above equation and obtain

$$f \propto \frac{F^{1+(j-1)\lambda}\,\bar{F}^{1+(n-j)(1-\lambda)}}{j-1+(n-2j+1)F}.$$

Thus the final result follows from (20). ■

Now we will present two corollaries of the above result which are closely connected to regression characterizations considered in literature.

Corollary 3.5: *Assume that* $\mathbb{E}|X_{i:2i-1}| < \infty$. *If*

$$\mathbb{E}(X_{i:2i-1}|X_{i+1:2i+1}) = X_{i+1:2i+1}, \tag{29}$$

then the parent distribution is $CB(1+(i/2),\, 1+(i/2))$.

Proof: It follows directly from Theorem 3.4 by taking there $j=i+1$ and $n=2i+1$. Note that then $\lambda = \frac{1}{2}$. ■

The above corollary is also a consequence of a characterization of $CB(1+(1-\lambda)i,\ 1+\lambda i)$ distribution for positive integer i and $\lambda \in (0,1)$ through the condition

$$\mathbb{E}(\lambda X_{i:2i+1} + (1-\lambda)X_{i+2:2i+1}|X_{i+1:2i+1}) = X_{i+1:2i+1}, \tag{30}$$

which is stated as Theorem 3.1 in ABN (it also includes the result of Nevzorov [18] who characterized the family $CB(1+(i/2),\, 1+(i/2))$ by the above condition with $\lambda = 1/2$). Note that (29) combined with (3) gives (30) with $\lambda = 1/2$.

The second corollary is related to an open problem stated in DW. In the concluding remarks of that paper, the authors suggested that possibly the easiest open questions in characterizations by linearity of regression of an os from a restricted sample with respect to an os from an extended sample are the following two cases:

$$\mathbb{E}(X_{1:2}|X_{2:4}) = aX_{2:4} + b, \quad \text{and} \quad \mathbb{E}(X_{2:2}|X_{3:4}) = aX_{3:4} + b.$$

Actually each of these two conditions was written in DW in the expanded integral form. Unfortunately there are misprints in those formulas: "y" is missing under all integrals and in the second equation the coefficients of two integrals should be: 1/3 instead of 1/6 for the first integral and 1/2 instead of 1/3 for the second.

We are able to solve these problems only when $a=1$ and $b=0$.

Corollary 3.6:

(1) *If* $\mathbb{E}|X_{1:2}| < \infty$ *and* $\mathbb{E}(X_{1:2}|X_{2:4}) = X_{2:4}$, *then* $q(u) \propto \frac{1+u}{u^{5/4}(1-u)^{5/2}}$, $u \in (0,1)$.
(2) *If* $\mathbb{E}|X_{2:2}| < \infty$ *and* $\mathbb{E}(X_{2:2}|X_{3:4}) = X_{3:4}$, *then* $q(u) \propto \frac{2-u}{u^{5/2}(1-u)^{5/4}}$, $u \in (0,1)$.

Proof: These results follow directly from Theorem 3.4 by taking: in the first case $j=2$ and $n=4$ and thus $\lambda = 1/4$; in the second case $j=3$ and $n=4$ and thus $\lambda = 3/4$. ■

From the proof of Theorem 3.4 it follows that if $\mathbb{E}|X_{j-1:n}| < \infty$ and $\mathbb{E}|X_{j+1:n}| < \infty$ and (27) holds for an arbitrary (but fixed) $\lambda \in (0,1)$ then the quantile density of X_1 has the form given in (26). This is a direct extension of Theorem 3.1 of ABN (and Theorem 2 of Nevzorov [18]) which follows by taking $n=2i+1$ and $j=i+1$. Note that this is the only case among possible forms of q in (26) when the distribution of X_1 is of the complementary beta form.

3.3. OS from the extended sample given os from the original sample

In this subsection we still keep the assumption that $m<n$ but the conditioning now will be with respect to $X_{i:m}$.

From (8), we derive the conditional density of $X_{j:n}|X_{i:m}=x$ with respect to ν_x as

$$f_{X_{j:n}|X_{i:m}=x}(y) = \sum_{k=i}^{i+n-m} \frac{\binom{k-1}{i-1}\binom{n-k}{m-i}}{\binom{n}{m}} \frac{f_{k:n}(x)}{f_{i:m}(x)} f_{X_{j:n}|X_{k:n}=x}(y)$$
$$= \sum_{k=i}^{i+n-m} \binom{n-m}{k-i} F^{k-i}(x)\bar{F}^{n-m-(k-i)}(x) f_{X_{j:n}|X_{k:n}=x}(y),$$

and consequently we have the representation

$$\mathbb{E}(X_{j:n}|X_{i:m}=x) = \sum_{\ell=0}^{n-m} \binom{n-m}{\ell} F^{\ell}(x)\bar{F}^{n-m-\ell}(x)\, \mathbb{E}(X_{j:n}|X_{i+\ell:n}=x). \qquad (31)$$

It is known that if $j > i+\ell$, then the conditional distribution of $\mathbb{P}_{X_{j:n}|X_{i+\ell:n}=x}$ is the same as the distribution of the $(j-i-\ell)$th os obtained from an i.i.d. sample of size $(n-i-\ell)$ from a parent distribution function $(F(t)-F(x))/((1-F(x)), x<t<\infty$; if $j<i+\ell$, it is the same as the distribution of the jth os in a sample of size $i+\ell-1$ from the parent distribution function $(F(t))/((F(x)), -\infty<t<x$. Using these facts, after some algebraic manipulation, $\mathbb{E}(X_{j:n}|X_{i:m}=x)$ can be expressed in a more explicit form. In general, characterizations (or identifiability question) through the form of $\mathbb{E}(X_{j:n}|X_{i:m})$ seem to be difficult and in some cases linearity of such conditional expectation is plainly non-admissible. Consequently, while discussing characterizations we will restrict our considerations only to several tractable cases.

In the following lemma we derive relatively simple representations of $\mathbb{E}(X_{j:n}|X_{i:m})$ in special cases which will be used in characterizations later on in this subsection. For these special cases we provide straightforward proofs which is an alternative to derivations based on the general formula (31).

Lemma 3.7: *For $m < n$, $1 \le i \le m$ and $1 \le j \le n$, we have*

(i) $\mathbb{E}(X_{1:n}|X_{1:m} = x) = x\bar{F}^{n-m}(x) + (n-m)\int_{-\infty}^{x} t\bar{F}^{n-m-1}(t)f(t)\mathrm{d}t$,
(ii) $\mathbb{E}(X_{n:n}|X_{m:m} = x) = xF^{n-m}(x) + (n-m)\int_{x}^{\infty} tF^{n-m-1}(t)f(t)\mathrm{d}t$,
(iii) $\mathbb{E}(X_{i:m+1}|X_{i:m} = x) = x\bar{F}(x) + \frac{i}{F^{i-1}(x)}\int_{-\infty}^{x} tF^{i-1}(t)f(t)\mathrm{d}t$,
(iv) $\mathbb{E}(X_{j:n}|X_{1:1} = x) = \int_{-\infty}^{x} tf_{j:n-1}(t)dt + x\binom{n-1}{j-1}F^{j-1}(x)\bar{F}^{n-j}(x) + \int_{x}^{\infty} tf_{j-1:n-1}(t)\mathrm{d}t$.

Proof: First we have

$$\begin{aligned}\mathbb{E}(X_{1:n}|X_{1:m} = x) &= x\mathbb{P}(\min\{X_{m+1},\dots,X_n\} > x|X_{1:m} = x)\\ &\quad + \mathbb{E}(\min\{X_{m+1},\dots,X_n\}I_{\{\min\{X_{m+1},\dots,X_n\}<x\}}|X_{1:m} = x)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(X_{n:n}|X_{m:m} = x) &= x\mathbb{P}(\max\{X_{m+1},\dots,X_n\} < x|X_{m:m} = x)\\ &\quad + \mathbb{E}(\max\{X_{m+1},\dots,X_n\}I_{\{\max\{X_{m+1}\dots,X_n\}>x\}}|X_{m:m} = x)\end{aligned}$$

and the assertions (i) and (ii) follow immediately.

Next, note that

$$X_{i:m+1} = X_{i:m}I_{\{X_{m+1}>X_{i:m}\}} + X_{m+1}I_{\{X_{i-1:m}<X_{m+1}<X_{i:m}\}} + X_{i-1:m}I_{\{X_{m+1}<X_{i-1:m}\}}.$$

Consequently,

$$\begin{aligned}\mathbb{E}(X_{i:m+1}|X_{i:m} = x) &= x\bar{F}(x) + \int_{-\infty}^{x} t\mathbb{P}(X_{i-1:m} < t|X_{i:m} = x)f(t)\mathrm{d}t\\ &\quad + \mathbb{E}(X_{i-1:m}F(X_{i-1:m})|X_{i:m} = x)\\ &= x\bar{F}(x) + \int_{-\infty}^{x} t\left(\frac{F(t)}{F(x)}\right)^{i-1} f(t)\mathrm{d}t\\ &\quad + \int_{-\infty}^{x} yF(y)\,(i-1)\frac{F^{i-2}(y)}{F^{i-1}(x)}f(y)\,\mathrm{d}y\end{aligned}$$

and this yields the assertion (iii).

Finally, the assertion (iv) follows from

$$\mathbb{E}(X_{j:n}|X_{1:1} = x) = \mathbb{E}(X_{j:n}|X_n = x)$$

and

$$X_{j:n} = X_{j:n-1}I_{\{X_n>X_{j:n-1}\}} + X_nI_{\{X_{j-1:n-1}<X_n<X_{j:n-1}\}} + X_{j-1:n-1}I_{\{X_n<X_{j-1:n-1}\}}.$$

The proof is completed. ■

Our main objective in this subsection is to show that the shape of the regresion curves studied in Lemma 3.7 determine the parent distribution F.

In the remaining part of this subsection we assume that the density $f = F'$ is strictly positive on $(a, b) := \{x \in \mathbb{R} : 0 < F(x) < 1\}$ with $-\infty \le a < b \le \infty$.

First we will consider the cases (i) and (ii) of Lemma 3.7. It appears that under mild conditions $\mathbb{E}(X_{1:n}|X_{1:m})$ and $\mathbb{E}(X_{n:n}|X_{m:m})$ determine the parent cdf F.

Theorem 3.8: *Suppose that g is a differentiable function on (a, b).*

(A) *If $\mathbb{E}|X_{1:n}| < \infty$ and*

$$\mathbb{E}(X_{1:n}|X_{1:m}) = g(X_{1:m}), \tag{32}$$

then

(A.1) $\lim_{x\to a+} g(x) = a$;

(A.2) *g' is a decreasing function with* $\lim_{x\to a^+} g'(x) = 1$ *and* $\lim_{x\to b^-} g'(x) = 0$;

(A.3) $F(x) = 1 - \sqrt[n-m]{g'(x)}$, *for $x \in (a, b)$.*

(B) *If $\mathbb{E}|X_{n:n}| < \infty$ and*

$$\mathbb{E}(X_{n:n}|X_{m:m}) = g(X_{m:m}), \tag{33}$$

then

(B.1) $\lim_{x\to b^-} g(x) = b$;

(B.2) *g' is an increasing function with* $\lim_{x\to a^+} g'(x) = 0$ *and* $\lim_{x\to b^-} g'(x) = 1$;

(B.3) $F(x) = \sqrt[n-m]{g'(x)}$, *for $x \in (a, b)$.*

Proof: Since there is an obvious duality between the two cases (it suffices to consider negative of the original observations to move between (A) and (B)) we provide only the proof of (A).

From (i) of Lemma 3.7 and (32) we obtain the equation

$$x\bar{F}^{n-m}(x) + (n-m)\int_a^x t\bar{F}^{n-m-1}(t)f(t)\mathrm{d}t = g(x), \quad x \in (a, b), \tag{34}$$

then (A.1) follows easily by taking limits $x \to a^+$ in both sides of (34). Differentiating (34), after elementary algebra, we get

$$\bar{F}^{n-m}(x) = g'(x), \quad x \in (a, b),$$

then (A.2) follows from the well-known properties of a cdf and (A.3) is immediate. ■

Note that linearity of regression in (32) (or (33)) is impossible since it would lead to F being constant.

As an illustration of Theorem 3.8 we provide two examples:

- if either

$$\mathbb{E}(X_{1:n}|X_{1:m} = x) = \frac{1-(1-x)^{n-m+1}}{n-m+1}, \quad 0 \le x \le 1,$$

or

$$\mathbb{E}(X_{n:n}|X_{m:m} = x) = \frac{x^{n-m+1} + n - m}{n-m+1}, \quad 0 \le x \le 1.$$

then the parent distribution is uniform on $(0, 1)$;

- if

$$\mathbb{E}(X_{1:n} \mid X_{1:m} = x) = \frac{1 - \exp(-(n-m)x)}{n-m}, \quad x \geq 0,$$

  then the parent distribution is standard (i.e. with mean 1) exponential.

Now we will consider case (iii) of Lemma 3.7.

Theorem 3.9: *Suppose* $\mathbb{E}|X_{i:m+1}| < \infty$ *for certain integers* $2 \leq i \leq m$ *and let* h *be a differentiable function on* (a, b). *If*

$$\mathbb{E}(X_{i:m+1}|X_{i:m}) = X_{i:m} - h(X_{i:m}).$$

Then

$$F(x) = \frac{h^{-\frac{1}{i-1}}(x)}{h^{-\frac{1}{i-1}}(b^-) + \frac{1}{i-1}\int_x^b h^{-\frac{i}{i-1}}(t)\,\mathrm{d}t}, \quad x \in (a,\ b). \tag{35}$$

Proof: Using Lemma 3.7(iii) we have

$$x\overline{F}(x) + \frac{i}{F^{i-1}(x)}\int_a^x tF^{i-1}(t)f(t)\,\mathrm{d}t = x - h(x),\ x \in (a, b)$$

and, after simple algebra,

$$F^{i-1}(x)h(x) = xF^i(x) - i\int_a^x tF^{i-1}(t)f(t)\,\mathrm{d}t, \quad x \in (a, b).$$

Integrating by parts (observe that if $a = -\infty$ then $\lim_{x \to a^+} xF^i(x) = 0$ due to the hypothesis $\mathbb{E}|X_{i:m+1}| < \infty$ and Lem. 3.3),

$$F^{i-1}(x)h(x) = \int_a^x F^i(t)\,\mathrm{d}t, \quad x \in (a, b), \tag{36}$$

from which we conclude that $h(x) > 0$ for $x \in (a, b)$.

Let $G(x) = \int_a^x F^i(t)\,\mathrm{d}t$, $x \in (a, b]$, so

$$F(x) = G'(x)^{\frac{1}{i}}, \quad x \in (a, b), \tag{37}$$

and, after some algebra, (36) yields

$$G^{-\frac{i}{i-1}}(x)\,G'(x) = h^{-\frac{i}{i-1}}(x), \quad x \in (a, b).$$

Therefore,

$$\int_x^b G^{-\frac{i}{i-1}}(t)\,G'(t)\,\mathrm{d}t = \int_x^b h^{-\frac{i}{i-1}}(t)\,\mathrm{d}t, \quad x \in (a, b)$$

or equivalently

$$G^{-\frac{1}{i-1}}(x) - G^{-\frac{1}{i-1}}(b) = \frac{1}{i-1}\int_x^b h^{-\frac{i}{i-1}}(t)\,\mathrm{d}t, \quad x \in (a, b).$$

From (36) it follows that $h(b^-) = G(b)$, and thus

$$G(x) = \left(h^{-\frac{1}{i-1}}(b^-) + \frac{1}{i-1}\int_x^b h^{-\frac{i}{i-1}}(t)\,\mathrm{d}t \right)^{-(i-1)}, \quad x \in (a,b) \tag{38}$$

(if $h(b^-) = \infty$, which is possible, we take $h(b^-)^{-\frac{1}{i-1}} = 0$). The result follows now easily by (37). ■

As an illustration of Theorem 3.9 we provide some examples:

- if for $\alpha > 0$

$$\mathbb{E}\left(X_{i:m+1} \mid X_{i:m} = x\right) = x - \frac{x^{\alpha+1}}{i\alpha + 1}, \quad 0 \le x \le 1,$$

  then the parent distribution is power with the cdf $F(x) = x^{\alpha}$, $x \in [0,1]$.
- if for $A > 0$ and $r > 0$ such that $ri > 1$

$$\mathbb{E}\left(X_{i:m+1} \mid X_{i:m} = x\right) = x - \frac{(1 + A(b-x))^{-r+1}}{A(ri-1)}, \quad x \le b,$$

  then the parent distribution is (negative) Type IV Pareto distribution with the cdf $F(x) = (1 + A(b-x))^{-r}$, $x \in (-\infty, b]$;
- if we specialize the above example by fixiing $r = 1$ we obtain characterization of cdf $F(x) = 1/(1 + A(b-x))$, $x \le b$, by the linearity of regression

$$\mathbb{E}(X_{i:m+1} | X_{i:m} = x) = x - \frac{1}{A(i-1)}, \quad x \le b;$$

- if for $\lambda > 0$

$$\mathbb{E}\left(X_{i:m+1} \mid X_{i:m} = x\right) = x - \frac{\exp(\lambda x)}{i\lambda}, \quad x \le 0.$$

  then the parent distribution is negative exponential with the cdf $F(x) = \exp(\lambda x)$, $x \in (-\infty, 0]$.

Characterization by $\mathbb{E}(X_{1:1}|X_{j:n})$ (of course, $X_{1:1} = X_1$) was studied in WG and Balakrishnan and Akhundov [20] in the linear case. In case (iv) of Lemma 3.7 we will consider the dual conditional expectation $\mathbb{E}(X_{j:n}|X_{1:1})$. We are not able to express the parent distribution in terms of this regression function in this case. Instead we solve a more modest question of identifiability of the distribution of X_1.

Theorem 3.10: *Suppose $\mathbb{E}|X_{j:n}| < \infty$ for certain integers $1 \le j \le n$. Then the conditional expectation $\mathbb{E}(X_{j:n}|X_{1:1})$ uniquely determines the parent cdf F.*

Proof: Denote $h(x) = \mathbb{E}(X_{j:n}|X_{1:1} = x)$, $x \in (a, b)$. From (iv) of Lemma 3.7,

$$h(x) = \int_{-\infty}^{x} tf_{j:n-1}(t)\mathrm{d}t + x\binom{n-1}{j-1}F^{j-1}(x)\bar{F}^{n-j}(x) + \int_{x}^{\infty} tf_{j-1:n-1}(t)\mathrm{d}t, \quad x \in (a,b).$$

Then differentiating the above equation with respect to x we get

$$h'(x) = \binom{n-1}{j-1}F^{j-1}(x)\bar{F}^{n-j}(x), \quad x \in (a,b). \tag{39}$$

Let us assume that the H is not unique, that is there exist two different distribution functions F and G with the same support such that H is the same for F and G. Hence

$$F^{\frac{j-1}{n-j}}(x)\bar{F}(x) = G^{\frac{j-1}{n-j}}(x)\bar{G}(x), \quad x \in (a,b),$$

which can be rewritten as

$$\int_{G(x)}^{F(x)} \left((j-1)t^{\frac{j-1}{n-j}-1} - (n-1)t^{\frac{j-1}{n-j}}\right)\mathrm{d}t = 0, \quad x \in (a,b). \tag{40}$$

Note that, for $0 < t < (j-1/n-1)$, the integrand in (40) is strictly positive. Therefore $F(x) = G(x)$ in a right neighbourhood of the left end of the support. Consequently, we have $x_0 = \sup\{x \geq a : F(x) = G(x)\} > a$ and by continuity, $F(x_0) = G(x_0)$.

Let us prove that $F(x_0) = G(x_0) \geq (j-1/n-1)$. Assume the opposite, $F(x_0) = G(x_0) < (j-1/n-1)$. Then, by continuity of F and G, there exists $\epsilon > 0$ such that $F(x_0 + \epsilon) < (j-1/n-1)$ and $G(x_0+\epsilon) < (j-1/n-1)$. Hence again the integrand in (40) is strictly positive and we get $F(x_0+\epsilon) = G(x_0+\epsilon)$ which contradicts the definition of x_0. Therefore, $F(x_0) = G(x_0) \geq (j-1/n-1)$. Consider now an arbitrary $x > x_0$. Since F and G are strictly increasing on (a,b) we see that $F(x) > (j-1/n-1)$ and $G(x) > (j-1/n-1)$. But for $t > (j-1/n-1))$ the integrand in (40) is strictly negative. Consequently $F(x) = G(x)$. ■

Note that due to (39) the derivative of the regression function can be useful for determining the parent cdf. In particular, it follows from (39) that (a) if $\mathbb{E}(X_{1:n}|X_{1:1} = x) = g(x)$ is differentiable then $F(x) = 1 - \sqrt[n-1]{g'(x)}$; (b) if $\mathbb{E}(X_{n:n}|X_{1:1} = x) = g(x)$ is differentiable then $F(x) = \sqrt[n-1]{g'(x)}$. Actually, these results are also covered by Theorem 3.8 for $m = 1$.

Finally, we use (39) to derive two new characterizations of the logistic distribution.

Corollary 3.11: *Assume that either*

$$\mathbb{E}(X_{2:3}|X_{1:1} = x) = \frac{2e^x}{1+e^x}, \quad x \in \mathbb{R},$$

or with unknown parent cdf F

$$\mathbb{E}(X_{2:3}|X_{1:1} = x) \propto F(x), \quad x \in \mathbb{R},$$

or

$$\mathbb{E}(X_{2:3}|X_{1:1} = x) \propto \bar{F}(x), \quad x \in \mathbb{R}.$$

Then X_1, X_2 and X_3 have the logistic distribution.

Proof: In the first case (39) implies

$$F(x)\bar{F}(x) = \frac{e^x}{(1+e^x)^2}.$$

There are two solutions of the above quadratic equation in the unknown $F(x)$. Only one of them, $F(x) = (e^x/1 + e^x)$, $x \in \mathbb{R}$, gives the valid (logistic) distribution function.

In the remaining two cases (39) yields $f \propto F\bar{F}$, i.e. we obtain a distribution from $CB(1,1)$ family, which is the family of logistic laws – see, e.g. Galambos [34]. ■

The last two cases in the above corollary can be easily generalized:

- if $\mathbb{E}(X_{j:n}|X_{1:1} = x) \propto F^s(x)$, $s \in \mathbb{R}$, then the parent distribution belongs to $CB(j - s, n - j)$;
- if $\mathbb{E}(X_{j:n}|X_{1:1} = x) \propto \bar{F}^s(x)$, $s \in \mathbb{R}$, then the parent distribution belongs to $CB(j - 1, n - j - s + 1)$.

4. Conclusion

The aim of this paper is two-fold: (1) derivation of bivariate distribution of os's from overlapping samples in the general overlapping scheme; (2) investigations of regression properties of os's from overlapping samples, in particular, extension of characterizations by linearity of regression of os's or identifiability results to the overlapping situation. Throughout the paper we assumed that the original observations are iid and their common distribution is absolutely continuous with respect to the Lebesque measure. The first task was fully resolved. Though the general formula is quite complicated, in several important special cases it gives quite transparent formulas and can be useful, e.g. in studying moving order statistics or analysing conditional structure of os's from overlapping samples. Regarding the second task we identified new settings in which linearity of regression or the general form of the regression function characterizes the parent distributions, in several other cases uniqueness results were obtained instead. However, the issue of characterizing of the parent cdf F by using a general relation

$$\mathbb{E}\left(X_{i:m}|X_{j:n}^{(r)}\right) = h(X_{j:n}^{(r)}),$$

where $h : \mathbb{R} \to \mathbb{R}$ is a Borel function, remains open and seems to be rather difficult to settle.

Finally let us mention that in the special case of $\mathbb{E}(X_{i:m}|X_{j:n})$, due to (3), the problem we studied embeds naturally in the question of characterization of the parent distribution by regression of L-statistics of the form $\sum_{i=1}^{n} a_i X_{i:n}$ on a single os $X_{j:n}$.

Acknowledgments

The authors are greatly indebted to the referees for valuable suggestions which helped significantly in improving the final version of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Support for this research was provided in part by the Ministry of Science and Technology, Taiwan [grant numbers 102-2118-M-305-003 and 104-2118-M-305-003], by Ministerio de Economia, Industria y Competitividad, Spain [grant number MTM2016-74983] and by the National Science Center, Poland [grant number 2016/21/B/ST1/00005].

ORCID

Nan-Cheng Su http://orcid.org/0000-0002-3629-3568

References

[1] David HA, Nagaraja HN. Order statistics, Wiley series in probability and statistics. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2003.
[2] Arnold BC, Balakrishnan N, Nagaraja HN. A first course in order statistics (classics in applied mathematics, vol. 54). Philadelphia: SIAM; 2008.
[3] Inagaki N, The distribution of moving order statistics. In: Matsusita K, editor. Recent developments in statistical inference and data analysis. Amsterdam: North Holland; 1980. p. 137–142.
[4] David HA, Rogers MP. Order statistics in overlapping samples, moving order statistics and U-statistics. Biometrika. 1983;70(1):245–249.
[5] Ishida I, Kvedaras V. Modeling autoregressive processes with moving-quantiles-implied nonlinearity. Econometrics. 2015;3:2–54.
[6] Balakrishnan N, Tan T. A parametric test for trend based on moving order statistics. J Statist Comput Simul. 2016;86(4):641–655.
[7] Siddiqui MM. Order statistics of a sample and of an extended sample. In: Puri ML, editor. Nonparametric techniques in statistical inference. London: Cambridge Univ. Press; 1970. p. 417–423.
[8] Tryfos P, Blackmore R. Forecasting records. JASA. 1985;80:46–50.
[9] Ahsanullah M, Nevzorov VB. Generalized spacings of order statistics from extended sample. J Statist Plann Inference. 2000;85:75–83.
[10] López-Blázquez F, Salamanca-Miño B. Maximal correlation in a non-diagonal case. J Multivariate Anal. 2014;131:265–278.
[11] Dołegowski A, Wesołowski J. Linearity of regression for overlapping order statistics. Metrika. 2015;78:205–218.
[12] Ferguson T. On a Rao–Shanbhag characterization of the exponential/geometric distribution. Sankhya A. 2002;64:246–255.
[13] Dembińska A, Wesołowski J. Linearity of regression for non-adjacent order statistics. Metrika. 1998;48:215–222.
[14] López Bláquez F, Moreno-Rebollo JL. A characterization of distributions based on linear regression of order statistics and record values. Sankhyā Ser A. 1997;59(3):311–323.
[15] Bieniek M, Maciag K. Uniqueness of characterization of absolutely continuous distributions by regressions of generalized order statistics. Adv Statist Anal. 2018;102(3):359–380.
[16] Wesołowski J, Gupta AK. Linearity of convex mean residual life time. J Statist Plann Inference. 2001;99:183–191.
[17] Nagaraja HN, Nevzorov VB. On characterizations based on records values and order statistics. J Statist Plann Inference. 1997;63:271–284.
[18] Nevzorov VB, Ona property of Student's distribution with two degrees of freedom. Zap. Nauch. Sem. POMI 294 (2002), 148–157. (in Russian; English translation: J Math Sci. 127(1) (2005), 757–1762).
[19] Nevzorov VB, Balakrishnan N, Ahsanullah M. Simple characterizations of Student t_2-distribution. JRSS D. 2003;52(3):395–400.
[20] Balakrishnan N, Akhundov IS. A characterization by linearity of the regression function based on order statistics. Statist Probab Lett. 2003;63:435–440.

[21] Akhundov IS, Balakrishnan N, Nevzorov VB. New characterizations by properties of midrange and related statistics. Comm Statist Theory Meth. 2004;33(12):3133–3134.
[22] Nevzorova L, Nevzorov VB, Akhundov I. A simple characterization of Student's t_2 distribution. Metron. 2007;55:53–57.
[23] Marudova NM, Nevzorov VB. A class of distributions that includes Student's t_2-distribution. J Math Sci. 2009;159(3):312–316.
[24] Akhundov I, Nevzorov V. From Student's t_2 distribution to Student's t_3 distribution through characterizations. Comm Statist Simulation Comput. 2012;41(6):710–716.
[25] Yanev G, Ahsanullah M. Characterizations of Student's t-distribution via regressions of order statistics. Statist. 2012;46(4):429–435.
[26] Jones MC. The complementary beta distribution. J Statist Plann Inference. 2002;104(2): 329–337.
[27] Jones MC. On a class of distributions defined by the relationship between their density and distribution functions. Comm Statist Theory Methods. 2007;36(10):1835–1843.
[28] Kamps U. A general recurrence relation for moments of order statistics in a class of probability distributions and characterizations. Metrika. 1991;38:215–225.
[29] Muiño JM, Voit EO, Sorribas A. GS-distributions: a new family of distributions for continuous unimodal variables. Comput Statist Data Anal. 2006;50:2769–2798.
[30] Mohie El-Din MM, Amein MM, Hamedani GG. On order statistics for GS-distributions. J Statist Theory Appl. 2012;11(3):237–264.
[31] Hu CY, Lin GD. Characterizations of the logistic and related distributions. J Math Anal Appl. 2018;463:79–92.
[32] Richards FJ. A flexible growth function for empirical use. J Exp Bot. 1959;10:290–300.
[33] Nevzorov VB. On some regression relations of connecting sample averages and order statistics. Vest Sankt Petersburg Gos Univ. 2015;60(2):364–368. (in Russian).
[34] Galambos J. Characterizations. In: Balakrishnan N, editor. Handbook of the logistic distribution. New York: Marcel Dekker; 1991. p. 169–188.

Appendix

Proof of Proposition 2.4: **(i)** $\boldsymbol{k < l}$. Note that

$$\{X_{i:m} = X_{k:n+r},\, X^{(r)}_{j:n} = X_{l:n+r}\} = \bigcup_{\substack{\alpha\in A\cup B\\ \beta\in B\cup C\\ \alpha\neq\beta}} \{X_{i:m} = X_{k:n+r} = X_\alpha,\, X^{(r)}_{j:n} = X_{l:n+r} = X_\beta\} \tag{A1}$$

and the sets under $\bigcup$ at the right-hand side are pair-wise disjoint. Moreover, for any distinct $\alpha \in A\cup B$ and $\beta \in B\cup C$

$$\{X_{i:m} = X_{k:n+r} = X_\alpha,\, X^{(r)}_{j:n} = X_{l:n+r} = X_\beta\} = \bigcup_{\sigma\in S(\alpha,\beta)} \{X_{\sigma(1)} \le \ldots \le X_{\sigma(r+n)}\}, \tag{A2}$$

where

$$\begin{aligned} S(\alpha,\beta) = \Big\{&\sigma \in \mathcal{S}_{n+r} : \sigma(k) = \alpha,\, \sigma(l) = \beta,\, |\sigma(\{1,\ldots,k-1\}) \cap (A_\alpha \cup B_{\alpha,\beta})| = i-1,\\ &\big|\sigma(\{1,\ldots,l-1\}\setminus\{k\}) \cap (B_{\alpha,\beta}\cup C_\beta)\big| = j-1-I_B(\alpha)\Big\}. \end{aligned}$$

Here and in the sequel we denote $U_{x_1,\ldots,x_K} := U\setminus\{x_1,\ldots,x_K\}$ for any set U. Since the sets A_α, $B_{\alpha,\beta}$ and C_β are disjoint and

$$A_\alpha \cup B_{\alpha,\beta} \cup C_\beta = (A\cup B\cup C)_{\alpha,\beta}$$

it follows that

$$\begin{aligned} S(\alpha,\beta) = \Big\{&\sigma \in \mathcal{S}_{n+r} : \sigma(k) = \alpha,\, \sigma(l) = \beta, |\sigma(\{1,\ldots,k-1\}) \cap C_\beta| = k-i,\\ &|\sigma(\{1,\ldots,l-1\}\setminus\{k\}) \cap A_\alpha| = l-j-1+I_B(\alpha)\Big\}. \end{aligned}$$

Therefore, by Lemma 2.3

$$|S(\alpha,\beta)| = \mathfrak{D}_{|A|-I_A(\alpha),\,|B|-I_B(\alpha)-I_B(\beta),\,|C|-I_C(\beta),\,k-1,\,l-k-1,\,k-i,\,l-j-I_A(\alpha)}.$$

Note that the sets under the $\bigcup$ sign in (A2) are pair-wise disjoint $\mathbb{P}$-a.s. and each of them has probability $1/(n+r)!$ Therefore

$$P(\alpha,\beta) := \mathbb{P}\left(\{X_{i:m} = X_{k:n+r} = X_\alpha,\, X^{(r)}_{j:n} = X_{l:n+r} = X_\beta\}\right) = \frac{|S(\alpha,\beta)|}{(n+r)!}.$$

There are four possible cases for the triplet $(A_\alpha, B_{\alpha,\beta}, C_\beta)$:

$$(A_\alpha, B_{\alpha,\beta}, C_\beta) = \begin{cases} (A_\alpha, B_\beta, C), & \text{if } \alpha\in A,\ \beta\in B,\\ (A_\alpha, B, C_\beta), & \text{if } \alpha\in A,\ \beta\in C,\\ (A, B_{\alpha,\beta}, C), & \text{if } \alpha,\beta\in B,\\ (A, B_\alpha, C_\beta), & \text{if } \alpha\in B,\ \beta\in C. \end{cases}$$

That is, following (A1) we obtain

$$\begin{aligned}
&\mathbb{P}\left(\{X_{i:m} = X_{k:n+r},\, X^{(r)}_{j:n} = X_{l:n+r}\right)\\
&= \sum_{\substack{\alpha\in A,\\ \beta\in B}} P(\alpha,\beta) + \sum_{\substack{\alpha\in A,\\ \beta\in C}} P(\alpha,\beta) + \sum_{\alpha,\,\beta\in B} P(\alpha,\beta) + \sum_{\substack{\alpha\in B,\\ \beta\in C}} P(\alpha,\beta)\\
&= \frac{|A||B||S(r,m)| + |A||C||S(r,n+r)| + |B|(|B|-1)|S(m-1,m)| + |B||C||S(m,n+r)|}{(r+n)!}\\
&= \frac{|A||B|\mathfrak{D}_{|A|-1,\,|B|-1,\,|C|,\,k-1,\,l-k-1,\,k-i,\,l-j-1}}{(r+n)!} + \frac{|A||C|\mathfrak{D}_{|A|-1,\,|B|,\,|C|-1,\,k-1,\,l-k-1,\,k-i,\,l-j-1}}{(r+n)!}\\
&\quad + \frac{|B|(|B|-1)\mathfrak{D}_{|A|,\,|B|-2,\,|C|,\,k-1,\,l-k-1,\,k-i,\,l-j}}{(r+n)!} + \frac{|B||C|\mathfrak{D}_{|A|,\,|B|-1,\,|C|-1,\,k-1,\,l-k-1,\,k-i,\,l-j}}{(r+n)!}.
\end{aligned}$$

Denote numerators in subsequent four fractions above by I_1, I_2, I_3 and I_4, respectively. Note that

$$\begin{aligned}
I_1 + I_2 &= |A|[|B|\mathfrak{D}_{|A|-1,\,|B|-1,\,|C|,\,k-1,\,l-k-1,\,k-i,\,l-j-1} + |C|\mathfrak{D}_{|A|-1,\,|B|,\,|C|-1,\,k-1,\,l-k-1,\,k-i,\,l-j-1}]\\
&= |A|\frac{(|A|+|B|+|C|-2)!}{\binom{|A|+|B|+|C|-2}{k-1,\,l-k-1}}\binom{|A|-1}{l-j-1}\Bigg[\binom{|C|}{k-i}\sum_{m=1}^{l-j-1}\binom{l-j-1}{m}|B|\\
&\quad\times\binom{|B|-1}{i-m-1}\binom{|B|+|C|+m-k}{j+m-k}\\
&\quad+|C|\binom{|C|-1}{k-i}\sum_{m=1}^{l-j-1}\binom{l-j-1}{m}\binom{|B|}{i-m-1}\binom{|B|+|C|+m-k}{j+m-k}\Bigg].
\end{aligned}$$

We will use several times the following elementary identity

$$s\binom{s-1}{r} = (s-r)\binom{s}{r}. \tag{A3}$$

Applying (A3) at the right-hand side above we get

$$\begin{aligned}
I_1 + I_2 &= |A|\frac{(|A|+|B|+|C|-2)!}{\binom{|A|+|B|+|C|-2}{k-1,\,l-k-1}}\binom{|A|-1}{l-j-1}\binom{|C|}{k-i}\\
&\quad\times\sum_{m=0}^{l-j-1}\binom{l-j-1}{m}\binom{|B|}{i-m-1}\binom{|B|+|C|+m-k}{j+m-k}(|B|+1+m+|C|-k)
\end{aligned}$$

$$
\begin{aligned}
&= |A|(|B|+|C|-j+1)\,\frac{(|A|+|B|+|C|-2)!}{\binom{|A|+|B|+|C|-2}{k-1,l-k-1}}\binom{|A|-1}{l-j-1}\binom{|C|}{k-i} \\
&\quad\times\sum_{m=0}^{l-j-1}\binom{l-j-1}{m}\binom{|B|}{i-m-1}\binom{|B|+|C|+m-k+1}{j+m-k} \\
&= \frac{|A|(|B|+|C|-j+1)}{|A|+|B|+|C|-l+1}\,\mathfrak{D}_{|A|-1,|B|,|C|,k-1,l-k-1,k-i,l-j-1}.
\end{aligned}
$$

Similarly to I_1+I_2, we can also obtain a explicit form of I_3+I_4. Then combining the expressions for I_1+I_2 and I_3+I_4 we get the final formula in this case.

(ii) $k=l$. Then

$$
\{X_{i:m}=X_{k:n+r}=X^{(r)}_{j:n}\}=\bigcup_{\alpha\in B}\{X_{i:m}=X_{k:n+r}=X^{(r)}_{j:n}=X_\alpha\}=\bigcup_{\alpha\in B}\bigcup_{\sigma\in S(\alpha)}\{X_{\sigma(1)}\le\ldots\le X_{\sigma(n+r)}\},
\tag{A4}
$$

where

$$
\begin{aligned}
S(\alpha)&=\{\sigma\in\mathcal{S}(n+r):\ \sigma(k)=\alpha,\ |\sigma(\{1,\ldots,k-1\})\cap(A\cup B_\alpha)|=i-1, \\
&\qquad |\sigma(\{1,\ldots,k-1\})\cap(B_\alpha\cup C)|=j-1\} \\
&=\{\sigma\in\mathcal{S}(n+r):\ \sigma(k)=\alpha,\ |\sigma(\{1,\ldots,k-1\})\cap C|=k-i, \\
&\qquad |\sigma(\{1,\ldots,k-1\})\cap A|=k-j\}.
\end{aligned}
$$

By Lemma 2.3 it follows that

$$
|S(\alpha)|=\mathfrak{D}_{|A|,|B|-1,|C|,k-1,0,k-i,k-j}.
$$

Since the right-hand side of (A4) is the union of pair-wise disjoint sets having the same probability $1/(n+r)!$ we get immediately the final formula in this case.

(iii) $\boldsymbol{k>l}$. Note that in this case (A1) and (A2) remain formally valid however this time the set $S(\alpha,\beta)$ is different:

$$
\begin{aligned}
S(\alpha,\beta)=\{\sigma\in\mathcal{S}_{n+r}:\ &\sigma(k)=\alpha,\ \sigma(l)=\beta,\ |\sigma(\{1,\ldots,l-1\})\cap(B_{\alpha,\beta}\cup C_\beta)|=j-1, \\
&\left|\sigma(\{1,\ldots,k-1\}\setminus\{l\})\cap(A_\alpha\cup B_{\alpha,\beta})\right|=i-1-I_B(\beta)\}.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
S(\alpha,\beta)=\{\sigma\in\mathcal{S}_{n+r}:\ &\sigma(k)=\alpha,\ \sigma(l)=\beta,\ |\sigma(\{1,\ldots,l-1\})\cap A_\alpha|=l-j, \\
&\left|\sigma(\{1,\ldots,k-1\}\setminus\{l\})\cap C_\beta\right|=k-i-1+I_B(\beta)\}.
\end{aligned}
$$

Therefore, according to Lem. 2.3

$$
|S(\alpha,\beta)|=\mathfrak{D}_{|C_\beta|,\,|B_{\alpha,\beta}|,|A_\alpha|,l-1,k-l-1,l-j,k-i-I_C(\beta)}.
$$

Thus, analogously as in Case (i) we obtain

$$
\begin{aligned}
&\mathbb{P}\left(X_{i:m}=X_{k:n+r},\,X^{(r)}_{j:n}=X_{l:n+r}\right) \\
&=|B||C|\frac{|S(m,n+r)|}{(r+n)!}+|A||C|\frac{|S(r,n+r)|}{(n+r)!}+|B|(|B|-1)\frac{|S(m-1,m)|}{(n+r)!}+|A||B|\frac{|S(r,m)|}{(n+r)!}
\end{aligned}
$$

$$= \frac{|B||C|\mathfrak{D}_{|C|-1,\,|B|-1,\,|A|,\,l-1,\,k-l-1,\,l-j,\,k-i-1}}{(r+n)!} + \frac{|A||C|\mathfrak{D}_{|C|-1,\,|B|,\,|A|-1,\,j-1,\,k-l-1,\,l-j,\,k-i-1}}{(r+n)!}$$
$$+ \frac{|B|(|B|-1)\mathfrak{D}_{|C|,\,|B|-2,\,|A|,\,l-1,\,k-l-1,\,l-j,\,k-i}}{(r+n)!} + \frac{|A||B|\mathfrak{D}_{|C|,\,|B|-1,\,|A|-1,\,l-1,\,k-l-1,\,l-j,\,k-i}}{(r+n)!}.$$

This formula is the analogue of the respective one from Case (i) with the roles of $|A|$ vs. $|C|$, k vs. l and i vs. j being exchanged. The final result follows again by combining first two and second two numerators above with the use of (A3), similarly as it was done in Case (i). ∎