Joint distribution of new sample rank of bivariate order statistics

Gulder Kemalbay\textsuperscript{a}\textsuperscript{*} and Ismihan Bayramoglu (Bairamov)\textsuperscript{b}

\textsuperscript{a}Department of Statistics, Yildiz Technical University, Istanbul, Turkey; \textsuperscript{b}Department of Mathematics, Izmir University of Economics, Izmir, Turkey

(Received 24 February 2014; accepted 24 February 2015)

Let \((X_k, Y_k), \ k = 1, \ldots, n,\) be independent copies of bivariate random vector \((X, Y)\) with joint cumulative distribution function \(F(x, y)\) and probability density function \(f(x, y)\). For \(1 \leq r, s \leq n,\) the vector of order statistics of \(X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}\) and \(Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n},\) respectively, is denoted by \((X_{r:n}, Y_{r:n}).\) Let \((X_{n+i}, Y_{n+i}), i = 1, 2, \ldots, m,\) be a new sample from \(F(x, y),\) which is independent from \((X_k, Y_k), \ k = 1, 2, \ldots, n.\) Let \(\xi_1\) be the rank of order statistics \(X_{r:n}\) in a new sample \(X_{n+1}, X_{n+2}, \ldots, X_{n+m}\) and \(\xi_2\) be the rank of order statistics \(Y_{s:n}\) in a new sample \(Y_{n+1}, Y_{n+2}, \ldots, Y_{n+m}.\) We derive the joint distribution of discrete random vector \((\xi_1, \xi_2)\) and a general scheme wherein the distributions of new and old samples are different is considered. Numerical examples for given well-known distribution are also provided.

**Keywords:** bivariate order statistics; probability density function; rank; bivariate binomial distribution; discrete random vector

\[ \sum_{i=r}^{n} \sum_{j=s}^{n} \sum_{k=a}^{b} \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!} \pi_{11}^{i-k} \pi_{12}^{j-k} \pi_{21}^{n-i-j+k} \]

where \(\pi_{11} = F(x, y), \pi_{12} = F_X(x) - F(x, y), \pi_{21} = F_Y(y) - F(x, y), \pi_{22} = 1 - F_X(x) - F_Y(y) + F(x, y)\) and \(a = \max(0, i + j - n), b = \min(i, j)\) (see [6]).

\textsuperscript{*}Corresponding author. Email: kemalbay@yildiz.edu.tr

\textcopyright 2015 Taylor & Francis
The distribution of bivariate order statistics can be easily obtained from the bivariate binomial distribution, which was first introduced by Aitken and Gonin [1]. Although bivariate binomial distribution is closely related to multinomial distribution, they differ greatly. Modifying the bivariate binomial model allows us to obtain conditional distributions of bivariate order statistics, which are important in several applications. Recently, Bairamov and Kemalbay [3] introduced new modifications of bivariate binomial distribution, which can be applied to derive the distribution of bivariate order statistics if a certain number of observations are within the given threshold set. These conditional distributions can be applied in studying the dependence among financial markets and in reliability analysis of complex systems. Bairamov and Bayramoglu [2] introduced new bivariate distributions based on Baker’s distributions, using the joint distribution of bivariate order statistics if a certain number of observations are within the given threshold set.

Considering a bivariate sample, David et al. [7] studied the distribution of the sample rank for a concomitant of an order statistic. Eryilmaz and Bairamov [9] studied joint distribution of ranks of an order statistic and its concomitant among a new sample. In this paper, we consider a random sample \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) with absolutely continuous cdf \(F(x, y)\), which can be considered a training sample, and another random sample \((X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \ldots, (X_{n+m}, Y_{n+m})\) with absolutely continuous cdf \(G(x, y)\), which can be considered the control. Let \(\xi_1\) be the rank of \(X_{rn}\) in the new sample \(X_{n+1}, X_{n+2}, \ldots, X_{n+m}\) and \(\xi_2\) be the rank of \(Y_{rn}\) in \(Y_{n+1}, Y_{n+2}, \ldots, Y_{n+m}\). The joint distribution of the discrete random vector \((\xi_1, \xi_2)\) is obtained using the bivariate binomial distribution. Numerical examples with well known bivariate underlying distribution are provided. The motivation for this study may be the following example: Denote by \((X_i, Y_i), i = 1, 2, \ldots, n\), two test results taken by \(n\) individuals, where \(X\) and \(Y\) are assumed to be dependent random variables with joint cdf \(F(x, y)\) and \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) are independent random vectors. Assume that an individual under consideration occupies the \(r\)th place in the first test and \(s\)th place in the second test, and that the individual’s test results are being compared with the test results of new individuals with different conditions. What is the probability that the order of the first test for the individual is \(p\) and that the order of the second test is \(q\), in the new sample results of a new test? The required probability is, \(P[\xi_1 = p, \xi_2 = q]\).

Throughout this paper, we use the following notations: \(f(x, y) = \partial^2 F(x, y)/\partial x \partial y = F^{1, 1}(x, y), f_X(x) = \partial F_X(x)/\partial x\) and \(f_Y(y) = \partial F_Y(y)/\partial y\). Furthermore, \(F^{1, -}(x, y) = \partial F(x, y)/\partial x\) and \(F^{-1}(x, y) = \partial F(x, y)/\partial y\).

2. Auxiliary results

**Lemma 1** [5] The joint pdf of the bivariate order statistics \(X_{rn}\) and \(Y_{sn}\), \(1 \leq r, s \leq n\), is

\[
f_{r,s,n}(x, y) = \sum_{t_1=a_1}^{a_2} C_1[F(x, y)]^{t_1}[F_X(x) - F(x, y)]^{r-1-t_1}[F_Y(y) - F(x, y)]^{s-1-t_1}
\]

\[
\times \left[ \frac{F(x, y)}{t_1} \right]^{n-r-s+t_1+1} f(x, y)
\]

\[
+ \sum_{t_2=d_1}^{d_2} \sum_{t_3=c_1}^{c_2} \sum_{t_4=b_1}^{b_2} C_2[F(x, y)]^{t_1}[F_X(x) - F(x, y)]^{r-1-t_1-t_2}
\]

\[
\times [F_Y(y) - F(x, y)]^{s-1-t_1-t_2}[F(x, y)]^{p-r-s+t_1+t_2+t_4}
\]

\[
\times [F^{1, -}(x, y)]^{t_2} [f_Y(y) - F^{1, -}(x, y)]^{1-t_2}
\]

\[
\times [F^{-1}(x, y)]^{t_4} [f_X(x) - F^{-1}(x, y)]^{1-t_4},
\]

(1)
where the lower and upper bounds for the summations are

\[
\begin{align*}
a_1 &= \max(0, r + s - n - 1), \quad a_2 = \min(r - 1, s - 1), \\
b_1 &= \max(0, r + s - n - t_2 - t_4), \quad b_2 = \min(r - t_2 - 1, s - t_4 - 1), \\
c_1 &= \max(0, r - n + 1), \quad c_2 = \min(1, r - 1), \\
d_1 &= \max(0, s - n + 1), \quad d_2 = \min(1, s - 1)
\end{align*}
\]

and the constants $C_1, C_2$ are

\[
\begin{align*}
C_1 &= \frac{n!}{t_1! (r - 1 - t_1)!(s - 1 - t_1)!(n - r - s + t_1 + 1)!}, \\
C_2 &= \frac{n!}{t_1! (r - 1 - t_1 - t_2)!(s - 1 - t_1 - t_2)!(n - r - s + t_1 + t_2 + t_4)!}.
\end{align*}
\]

Formula (1) is from a paper by Barakat [5], who used this pdf for the derivation of product moments of bivariate order statistics from any arbitrary continuous distribution function. Barakat’s formula is correct, but he provided only the key component of the proof. Because this pdf is important in many calculations concerning bivariate order statistics, we provide the complete proof of this formula in the supplementary appendix. In the proof, we follow the same method followed by Barakat and use a multinomial-distribution based scheme.

3. Joint distribution of the new sample rank of $X_{r:n}$ and $Y_{s:n}$

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be independent copies of random vector $(X, Y)$ with absolutely continuous cdf $F(x, y)$ and joint pdf $f(x, y)$. The marginal cdfs and pdfs of $X$ and $Y$ are denoted by $F_X(x)$; $F_Y(y)$ and $f_X(x)$; $f_Y(y)$, respectively. Let $X_{r:n}$ and $Y_{s:n}$ be the $r$th and $s$th order statistics of $X$ and $Y$ samples, respectively. Further, assume that $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \ldots, (X_{n+m}, Y_{n+m})$ ($m \geq 1$) is another random sample with absolutely continuous cdf $G(x, y)$ and joint pdf $g(x, y)$. We assume that the two samples $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \ldots, (X_{n+m}, Y_{n+m})$ ($m \geq 1$) and $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are independent.

For $1 \leq r, s \leq n; m \geq 1$, we define the random variables $\eta_1$ and $\eta_2$ as follows:

\[
\begin{align*}
\eta_1 &= \sum_{i=1}^{m} I_{(X_{n:i} - X_{n:i})}, \\
\eta_2 &= \sum_{i=1}^{m} I_{(Y_{n:i} - Y_{n:i})},
\end{align*}
\]

where $I_{(x)} = 1$ if $x \geq 0$ and $I_{(x)} = 0$ if $x < 0$ is an indicator function. Random variables $\eta_1$ and $\eta_2$ are referred to as exceedance statistics. Clearly, $\eta_1$ shows the total number of new $X$ observations $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$, which does not exceed a random threshold based on the $r$th order statistic $X_{r:n}$, and is constructed from the previous $X$ sample $X_1, X_2, \ldots, X_n$. Similarly, $\eta_2$ is the number of new observations $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+m}$, which does not exceed $Y_{s:n}$. The random variable $\xi_1 = \eta_1 + 1$ indicates the rank of $X_{r:n}$ in the new sample $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$, and the random variable $\xi_2 = \eta_2 + 1$ indicates the rank of $Y_{s:n}$ in the new sample $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+m}$.
The joint probability mass function (pmf) of the random variables $\xi_1$ and $\xi_2$ is given in the following theorem.

**Theorem 1**  
The joint pmf of $\xi_1$ and $\xi_2$ is

$$
P[\xi_1 = p, \xi_2 = q] = P[\eta_1 = p - 1, \eta_2 = q - 1]$$

$$= \sum_{l=\max(0,p+q-m-2)}^{\min(p-1,q-1)} C_{m,l,p,q} \left\{ \sum_{i=1}^{d_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x,y)]^l \times [G_X(x) - G(x,y)]^{q-l-1} [G_Y(y) - G(x,y)]^{q-l-1} \right.$$  

$$\times [G_Y(y) - G(x,y)]^{q-l-1} \sum_{i=1}^{d_2} \int_{-\infty}^{\infty} [G(x,y)]^l \times [F(x,y)]^n [F_X(x) - F(x,y)]^{n-r-s+l+1}$$

$$\times [F(x,y)]^n [F_Y(y) - F(x,y)]^{s-1-l} dx\ dy$$

$$+ \sum_{i=1}^{d_2} \sum_{i=1}^{d_2} C_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x,y)]^l \times [G_X(x) - G(x,y)]^{q-l-1} [G_Y(y) - G(x,y)]^{q-l-1} \left[ G_X(x) - G(x,y) \right]^{q-l-1}$$

$$\times [F(x,y)]^n [F_X(x) - F(x,y)]^{n-r-s+l+1}$$

$$\times [F_Y(y) - F(x,y)]^{s-1-l} dx\ dy \right\};$$

where

$$C_{m,l,p,q} = \frac{m!}{l!(p-l-1)!(q-l-1)!(m-p-q+l+2)!},$$

the upper and lower bounds $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$ and the constants $C_1, C_2$ are as given in Equations (2) and (3), respectively.

**Proof**  
Assume that in a fourfold sampling scheme, the outcome of the random experiment is one of the events $A_iB_i, A_iB^c_i, A^c_iB_i$ and $A^c_iB^c_i$, $i = 1, 2, \ldots, m$, where $A_i = \{X_{n+i} \leq X_{i,n}\}$, $B_i = \{Y_{n+i} \leq Y_{i,n}\}$, and the events $A^c_i$ and $B^c_i$ are the complements of the events $A_i$ and $B_i$, respectively. Assuming that the event $\{\eta_1 = p - 1, \eta_2 = q - 1\}$ occurs. Furthermore, assuming that $l$ of the events $A_iB_i, i = 1, 2, \ldots, m$, occur, then $p - l - 1, q - l - 1$ and $m - p - q + l + 2$ of the events $A_iB^c_i, A^c_iB_i$ and $A^c_iB^c_i$, $1, 2, \ldots, m$, respectively, will occur. On the other hand, by conditioning on $X_{i,n} = x$ and $Y_{i,n} = y$, we get $P(AB) = P(A_iB_i) = P(X_{n+i} \leq x, Y_{n+i} \leq y)$, $P(AB^c) = P(A_iB^c_i) = P(X_{n+i} \leq x, Y_{n+i} > y)$, $P(A^cB) = P(A^c_iB_i) = P(X_{n+i} > x, Y_{n+i} \leq y)$ and $P(A^cB^c) = P(A^c_iB^c_i) = P(X_{n+i} > x, Y_{n+i} > y)$. Therefore, we get

$$P[\xi_1 = p, \xi_2 = q]$$

$$= P[\eta_1 = p - 1, \eta_2 = q - 1]$$

$$= P\left\{ \sum_{i=1}^{m} I(X_{i,n} - X_{n+i}) = p - 1, \sum_{i=1}^{m} I(Y_{i,n} - Y_{n+i}) = q - 1 \right\}$$


Using Lemma 1 in Equation (7), we obtain Equation (4).

By the definition of a bivariate binomial distribution with fourfold sampling scheme described above, we obtain

\[
P\left\{ \sum_{i=1}^{m} I(x_{ni}-x_{ni}) = p - 1, \sum_{i=1}^{m} I(y_{ni}-y_{ni}) = q - 1 \right\} = \min(p-1,q-1) \sum_{l=\max(0,p+q-m-2)}^{\min(p-1,q-1)} C_{m,l,p,q} [P(AB)]^{l} [P(ABc)]^{p-l-1} \times [P(A^c B)]^{q-l-1} [P(A^c Bc)]^{m-p-q+l+2}. \tag{6}
\]

Substituting Equation (6) in Equation (5), we then have

\[
P\{\xi_1 = p, \xi_2 = q\} \equiv P\{\eta_1 = p - 1, \eta_2 = q - 1\} \]

\[
= \sum_{l=\max(0,p+q-m-2)}^{\min(p,q-1)} C_{m,l,p,q} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G(x,y)]^{l} [G_X(x) - G(x,y)]^{p-l-1} \times [G_Y(y) - G(x,y)]^{q-l-1} [\tilde{G}(x,y)]^{m-p-q+l+2} f_{r,s,n}(x,y) \, dx \, dy.
\]

\[
p, q = 1, \ldots, m + 1. \tag{7}
\]

Using Lemma 1 in Equation (7), we obtain Equation (4).

**Corollary 1** Under the hypothesis \(H_0 : F(x, y) = G(x, y)\), the joint pmf of \(\xi_1\) and \(\xi_2\) is:

\[
P\{\xi_1 = p, \xi_2 = q\} = \sum_{l=\max(0,p+q-m-2)}^{\min(p,q-1)} C_{m,l,p,q} \left\{ \sum_{t_1=0}^{a_2} C_{t_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x,y)]^{l+t_1} \times [F_X(x) - F(x,y)]^{p+r-l-t_1-2} [F_Y(y) - F(x,y)]^{q+s-l-t_1-2} \times [\tilde{F}(x,y)]^{m+n-p-s-q-r+l+t_1+3} f(x,y) \, dx \, dy \right. \\
+ \sum_{t_2=0}^{d_2} \sum_{t_3=0}^{c_2} \sum_{t_4=0}^{b_2} C_{t_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x,y)]^{l+t_3} \times [F_X(x) - F(x,y)]^{p+r-l-t_2-2} [F_Y(y) - F(x,y)]^{q+s-l-t_1-t_2-2} \times [\tilde{F}(x,y)]^{m+n-p-s-q-r+l+t_1+t_3+t_4+2} f(x,y) \, dx \, dy \\
\left. + \sum_{t_5=0}^{d_1} \sum_{t_6=0}^{c_1} \sum_{t_7=0}^{b_1} C_{t_5} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x,y)]^{l+t_7} \times [F_X(x) - F(x,y)]^{p+r-l-t_5-2} [F_Y(y) - F(x,y)]^{q+s-l-t_1-t_5-2} \times [\tilde{F}(x,y)]^{m+n-p-s-q-r+l+t_1+t_5+t_7+2} f(x,y) \, dx \, dy \\
+ \left. \sum_{t_8=0}^{d_1} \sum_{t_9=0}^{c_1} \sum_{t_{10}=0}^{b_1} C_{t_8} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x,y)]^{l+t_{10}} \times [F_X(x) - F(x,y)]^{p+r-l-t_8-2} [F_Y(y) - F(x,y)]^{q+s-l-t_1-t_8-2} \times [\tilde{F}(x,y)]^{m+n-p-s-q-r+l+t_1+t_8+t_{10}+2} f(x,y) \, dx \, dy \right\} ;
\]

\[
p, q = 1, \ldots, m + 1. \tag{8}
\]

The marginal distributions of \(\xi_1\) and \(\xi_2\) can be obtained from the distribution of exceedance statistics \(\eta_1\) and \(\eta_2\), which have been well studied in the literature under different conditions.
These statistics are also useful for constructing a test criterion for testing the hypothesis $H_0 : F_X(x) = G_X(x)$ against some classes of alternatives. For more details see [4,8,10,12–16]. Using Corollary 1, one can construct a test criterion for testing the hypothesis $H_0 : F(x, y) = G(x, y)$ against some classes of alternatives.

The marginal distribution of $\xi_1$ is

$$P\{\xi_1 = p\} = P\{\eta_1 = p - 1\}$$

$$= \left(\begin{array}{c} m \\ p - 1 \end{array} \right) \int_{-\infty}^{\infty} [G_X(x)]^p - 1 [1 - G_X(x)]^m - p + 1 f_{r,n}(x) \, dx$$

$$= \left(\begin{array}{c} m \\ p - 1 \end{array} \right) \frac{1}{B(r, n - r + 1)} \int_{-\infty}^{\infty} [G_X(x)]^p - 1 [1 - G_X(x)]^m - p + 1$$

$$\times [F_X(x)]^{p - 1}[1 - F_X(x)]^{m - p + 1} f_X(x) \, dx;$$

$$p = 1, \ldots, m + 1,$$  \hspace{1cm} (9)

where $B(a, b)$ is Beta function and $f_{r,n}(x)$ is the pdf of $X_{r,n}$.

If $F_X(x)$ and $G_X(x)$ are continuous, then under the hypothesis $H_0 : F_X(x) = G_X(x)$, the finite distribution of the random variable $\xi_1$ is a negative hypergeometric distribution of the first kind

$$P\{\xi_1 = p\} = P\{\eta_1 = p - 1\}$$

$$= \left(\begin{array}{c} m \\ p - 1 \end{array} \right) \frac{1}{B(r, n - r + 1)} B(r + p - 1, n + m - p - r + 2)$$

$$= \left(\begin{array}{c} r + p - 2 \\ r - 1 \end{array} \right) \frac{(n + m - p - 1)}{(n - r)} \frac{1}{\binom{n + m}{n}} ; \quad p = 1, \ldots, m + 1,$$  \hspace{1cm} (10)

which can be found in Wesolowski and Ahsanullah [16]. By similar consideration, we can easily find the marginal distribution of $\xi_2$.

3.1 Correlation coefficient between $\xi_1$ and $\xi_2$

**Proposition 1** The expected value and variance of the random variable $\xi_1$ can be easily obtained from their expression in terms of the indicator function, namely,

$$E(\xi_1) = E(\eta_1 + 1) = E\left[ \sum_{i=1}^{m} I_{(X_{r,n} - X_{r,n})} \right] + 1$$

$$= \sum_{i=1}^{m} P(X_{r,n} \leq X_{r,n}) + 1$$

$$= \sum_{i=1}^{m} \int_{-\infty}^{\infty} P(X_{r,n} \leq x | X_{r,n} = x) P(X_{r,n} = x) \, dx + 1$$

$$= m \int_{-\infty}^{\infty} G_X(x) f_{r,n}(x) \, dx + 1.$$  \hspace{1cm} (11)
\[ \text{Var}(\xi_1) = \text{Var}(\eta_1) = \text{Var} \left[ \sum_{i=1}^{m} I_{(X_{r,n} - X_{n+i})} \right] + 2 \text{Cov} \left[ \sum_{1<i<j\leq m} I_{(X_{r,n} - X_{n+i})}, I_{(X_{r,n} - X_{n+j})} \right] \]

\[ = \sum_{i=1}^{m} P(X_{n+i} \leq X_{r,n}) - \left[ \sum_{i=1}^{m} P(X_{n+i} \leq X_{r,n}) \right]^2 + 2 \sum_{1<i<j\leq m} P(X_{n+i} \leq X_{r,n}, X_{n+j} \leq X_{r,n}) \]

\[ = m \int_{-\infty}^{\infty} G_X(x) f_{r,n}(x) \, dx - \left[ m \int_{-\infty}^{\infty} G_X(x) f_{r,n}(x) \, dx \right]^2 + 2 \frac{m(m-1)}{2} \int_{-\infty}^{\infty} G_X(x)^2 f_{r,n}(x) \, dx; \quad (12) \]

where

\[ f_{r,n}(x) = \frac{1}{B(r,n-r+1)} [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r} f_X(x) \]

is the pdf of \( X_{r,n} \).

Similarly, we can easily obtain the expected value and variance of \( \xi_2 \).

**Corollary 2**  
Under the hypothesis \( H_0 : F_X(x) = G_X(x) \), the expected value and variance of the random variable \( \xi_1 \) is

\[ E(\xi_1) = \frac{B(r, 1, n-r+1)}{B(r, n-r+1)} + 1 = \frac{r}{n+1} + 1, \quad (13) \]

\[ \text{Var}(\xi_1) = \frac{B(r, 1, n-r+1)}{B(r, n-r+1)} - \left[ \frac{B(r, 1, n-r+1)}{B(r, n-r+1)} \right]^2 + m(m-1) \frac{B(r+2, n-r+1)}{B(r, n-r+1)} \]

\[ = m \frac{r}{n+1} \left[ 1 - \frac{r}{n+1} \right] + m(m-1) \frac{(r+1)\,r}{(n+2)(n+1)}, \quad (14) \]

**Proposition 2**  
The covariance between \( \xi_1 \) and \( \xi_2 \) is

\[ \text{Cov}(\xi_1, \xi_2) = \text{Cov}(\eta_1, \eta_2) = E \left[ \sum_{i=1}^{m} I_{(X_{r,n} - X_{n+i})} \sum_{i=1}^{m} I_{(Y_{s,n} - Y_{n+i})} \right] - E \left[ \sum_{i=1}^{m} I_{(X_{r,n} - X_{n+i})} \right] E \left[ \sum_{i=1}^{m} I_{(Y_{s,n} - Y_{n+i})} \right] \]

\[ = \sum_{i=1}^{m} \{ P(X_{n+i} \leq X_{r,n}, Y_{n+i} \leq Y_{s,n}) - P(X_{n+i} \leq X_{r,n})P(Y_{n+i} \leq Y_{s,n}) \} \]

\[ = m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_X(x) f_{r,s,n}(x,y) \, dx \, dy - m \int_{-\infty}^{\infty} G_X(x) f_{r,n}(x) \, dx \int_{-\infty}^{\infty} G_Y(y) f_{s,n}(y) \, dy; \quad (15) \]

where \( f_{r,n}(x) \) is the pdf of \( X_{r,n} \), \( f_{s,n}(y) \) is the pdf of \( Y_{s,n} \), and \( f_{r,s,n}(x,y) \) is the joint pdf of the bivariate order statistics \( X_{r,n} \) and \( Y_{s,n} \).
Corollary 3  Under the hypothesis $H_0 : F(x, y) = G(x, y)$, the covariance between $\xi_1$ and $\xi_2$ is

$$\text{Cov}(\xi_1, \xi_2) = m \left[ \sum_{l_1=a_1}^{d_1} \sum_{l_2=c_1}^{b_2} C_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x, y)]^{l_1+1} [F_X(x) - F(x, y)]^{r-1-l_1} \times [F_Y(y) - F(x, y)]^{s-1-l_2} \text{d}x \text{d}y + \sum_{l_1=d_1}^{d_2} \sum_{l_2=c_1}^{b_2} C_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x, y)]^{l_1+1} \times [F_X(x) - F(x, y)]^{r-1-l_1-l_2} [F_Y(y) - F(x, y)]^{s-1-l_2} \times [F^{1-}(x, y)]^{l_2} [f_X(x) - F^{1-}(x, y)]^{1-l_2} \text{d}x \text{d}y \right].$$

The upper and lower bounds $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$ and the constants $C_1, C_2$ are as given in Equations (2) and (3), respectively.

The Pearson correlation coefficient between $\xi_1$ and $\xi_2$ can be calculated using Equations (12) and (14).

| $n$ | $r$ | $s$ | $\alpha$ | $m$ | $p$ | $q$ | $P(\xi_1 = p, \xi_2 = q)$ |
|-----|-----|-----|---------|-----|-----|-----|---------------------------|
| 5   | 2   | 3   | 0.1     | 1   | 1   | 1   | 0.3382                    |
|     |     |     |         |     |     |     | 0.3284                    |
|     |     |     |         |     |     |     | 0.1618                    |
|     |     |     |         |     |     |     | 0.1716                    |
| 10  | 5   | 7   | 0.5     | 2   | 1   | 1   | 0.0606                    |
|     |     |     |         |     |     |     | 0.1403                    |
|     |     |     |         |     |     |     | 0.1164                    |
|     |     |     |         |     |     |     | 0.0653                    |
|     |     |     |         |     |     |     | 0.1945                    |
|     |     |     |         |     |     |     | 0.1943                    |
|     |     |     |         |     |     |     | 0.0249                    |
|     |     |     |         |     |     |     | 0.0888                    |
|     |     |     |         |     |     |     | 0.1135                    |

Table 2. $P(\xi_1 = p), P(\xi_2 = q)$ as a function of $n, r, s, m$.

| $n$ | $r$ | $m$ | $p$ | $P(\xi_1 = p)$ | $n$ | $s$ | $m$ | $q$ | $P(\xi_2 = q)$ |
|-----|-----|-----|-----|----------------|-----|-----|-----|-----|----------------|
| 5   | 2   | 1   | 1   | 0.6666         | 5   | 3   | 1   | 1   | 0.5            |
|     |     |     | 2   | 0.3333         |     |     | 2   | 1   | 0.1515         |
| 10  | 5   | 2   | 1   | 0.3182         | 10  | 7   | 2   | 1   | 0.4545         |
|     |     |     | 2   | 0.4545         |     |     | 2   | 1   | 0.4242         |
|     |     |     | 3   | 0.2273         |     |     | 3   | 2   | 0.4242         |
|     |     |     | 3   | 0.3049         |     |     | 3   | 3   | 0.3727         |
|     |     |     | 4   | 0.0932         |     |     | 4   | 3   | 0.1615         |
Table 3. Correlation coefficient values of $\xi_1$ and $\xi_2$.

| $n$ | $r$ | $s$ | $\alpha$ | $m$ | $\text{Cov}(\xi_1, \xi_2)$ | $\rho(\xi_1, \xi_2)$ |
|-----|-----|-----|----------|-----|--------------------------|---------------------|
| 5   | 2   | 3   | 0.1      | 1   | 0.0049                   | 0.0208              |
|     |     |     | -0.1     |      | -0.0049                  | -0.0208             |
|     | 0.5 |     |          |      | 0.0224                   | 0.0949              |
|     | 0.9 |     |          |      | 0.0441                   | 0.1871              |
| 10  | 5   | 7   | 0.1      | 2   | 0.4626                   | 0.8141              |
|     | 0.5 |     |          |      | 0.4946                   | 0.8958              |
|     | -0.5|     |          |      | -0.4946                  | -0.8958             |
|     | 0.9 |     |          |      | 0.5063                   | 0.9756              |
| 10  | 4   | 6   | 0.1      | 3   | 0.0159                   | 0.0189              |
|     | 0.5 |     |          |      | 0.0795                   | 0.0948              |
|     | 0.9 |     |          |      | 0.1299                   | 0.1549              |
|     | -0.9|     |          |      | -0.1299                  | -0.1549             |

3.2 Numerical example

For illustration of the joint distribution of $\xi_1$ and $\xi_2$, consider the bivariate Farlie–Gumbel–Morgenstern (FGM) distribution with uniform marginals. The distribution function and densities for the FGM distribution are as follows:

$$F(x, y) = xy[1 + \alpha(1 - x)(1 - y)],$$

$$f(x, y) = 1 + \alpha(1 - 2x)(1 - 2y),$$

$$F^{-1}_1(x, y) = \frac{\partial F(x, y)}{\partial x} = y[1 + \alpha(1 - y)(1 - 2x)],$$

$$F^{-1}_1(x, y) = \frac{\partial F(x, y)}{\partial y} = x[1 + \alpha(1 - x)(1 - 2y)],$$

$$0 \leq x, y \leq 1; \quad -1 \leq \alpha \leq 1.$$

Under the hypothesis $H_0 : F(x, y) = G(x, y)$, we illustrate some numerical results for joint and marginal distribution of $\xi_1$ and $\xi_2$ in Tables 1 and 2, respectively.

We can easily verify the results in Tables 1 and 2. Let us consider the case $n = 10, r = 5, m = 2, p = 3$ in Table 2. Then $P[\xi_1 = 3] = P[\xi_1 = 3, \xi_2 = 1] + P[\xi_1 = 3, \xi_2 = 2] + P[\xi_1 = 3, \xi_2 = 2] = 0.0249 + 0.0888 + 0.1135 = 0.2272$ which coincides with the result in Table 1.

In addition to these results, in Table 3, we provide results for the correlation coefficient between $\xi_1$ and $\xi_2$.

4. Conclusion

We consider a bivariate random sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ with absolutely continuous cdf $F(x, y)$ and another random sample $(X_{n+1}, Y_{n+1}), (X_{n+2}, Y_{n+2}), \ldots, (X_{n+m}, Y_{n+m})$ with absolutely continuous cdf $G(x, y)$. The former is considered a training sample and the latter the control. We are interested in the joint distribution of the rank of $r^{th}$ order statistic of the $X$ sample $X_{r:n}$ in a new sample $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$ and the rank of $s^{th}$ order statistic $Y_{s:n}$ among a new sample $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+m}$. The joint pmf of these ranks is obtained. The moments of marginal distribution and correlation between ranks are calculated, and numerical examples are provided. The obtained joint distribution has several applications such as in analysing two test results of $n$ individuals and in investigating and predicting hydrological events, such as storms and floods, all of which can be described by two dependent random variables.
Acknowledgements

The authors are grateful to the Editor and anonymous reviewers for their valuable comments and suggestions which resulted in improvement of the presentation of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Supplemental data and research materials

Supplemental data for this article can be accessed at 10.1080/02664763.2015.1023705.

References

[1] A.C. Aitken and H.T. Gonin, On fourfold sampling with and without replacement, Proc. Roy. Soc. Edinburgh Sect. A 55 (1935), pp. 114–125.
[2] I. Bairamov and K. Bayramoglu, From the Huang–Kotz FGM distribution to Baker’s bivariate distribution, J. Multivariate Anal. 113 (2013), pp. 106–115.
[3] I. Bairamov and G. Kemalbay, Some novel discrete distributions under fourfold sampling schemes and conditional bivariate order statistics, J. Comput. Appl. Math. 248 (2013), pp. 1–14.
[4] I.G. Bairamov and S. Kotz, On distributions of exceedances associated with order statistics and record values for arbitrary distributions, Statist. Papers 42(2) (2001), pp. 171–185.
[5] H.M. Barakat, On moments of bivariate order statistics, Ann. Inst. Statist. Math. 51(2) (1999), pp. 351–358.
[6] H.A. David, Order Statistics, Wiley, New York, 1981.
[7] H.A. David, M.J. O’Connell, and S.S. Yang, Distribution and expected value of the rank of a concomitant of an order statistic, Ann. Statist. 5 (1977), pp. 216–223.
[8] B. Epstein, Tables for the distribution of the number of exceedances, Ann. Math. Statist. 25 (1954), pp. 762–768.
[9] S. Eryilmaz and I.G. Bairamov, On a new sample rank of an order statistics and its concomitant, Statist. Probab. Lett. 63 (2003), pp. 123–131.
[10] E.J. Gumbel and H. von Schelling, The distribution of the number of exceedances, Ann. Math. Statist. 21 (1950), pp. 247–262.
[11] J.S. Huang, X. Dou, S. Kuriki, and G.D. Lin, Dependence structure of bivariate order statistics with applications to Bayramoglu’s distributions, J. Multivariate Anal. 114 (2013), pp. 201–208.
[12] N.L. Johnson and S. Kotz, Some generalizations of Bernoulli and Polya-Eggenberger contagion models, Statist. Papers 32 (1991), pp. 1–17.
[13] W. Katzenbeisser, The distribution of two sample location exceedance test statistics under Lehmann alternatives, Statist. Hefte (now Papers) 26 (1985), pp. 131–138.
[14] W. Katzenbeisser, The exact power of two-sample location tests based on exceedances against shift alternatives, Math. Op. Forsch. (Statistics) 20 (1986), pp. 47–54.
[15] S.A. Matveychuk and Yu.I. Petunin, A generalization of the Bernoulli model arising in order statistics, I/II Ukr. Math. J. 42(518–528/43) (1991), pp. 779–785.
[16] J. Wesolowski and M. Ahsanullah, Distributional properties of exceedance statistics, Ann. Inst. Statist. Math. 50 (1998), pp. 543–565.