Joint distribution of new sample rank of bivariate order statistics

Gülder Kemalbay* and Ismihan Bayramoglu (Bairamov)†
*Yildiz Technical University, Department of Statistics, Turkey
†Izmir University of Economics, Department of Mathematics, Turkey

Appendix

Proof of Lemma 1. We consider the limit

$$\begin{aligned} f_{r,s:n}(x,y) &= \lim_{\delta_x,\delta_y\to 0}\frac{P\{x<X_{r:n}\le x+\delta_x, y<Y_{s:n}\le y+\delta_y\}}{\delta_x\delta_y}, \\ -\infty &< x,y<\infty \end{aligned} \quad (1)$$

and use the description of compound event $\{x<X_{r:n}\le x+\delta_x, y<Y_{s:n}\le y+\delta_y\}$ with the multinomial distribution to obtain:

$$\begin{aligned} f_{r,s:n}(x,y)\delta_x\delta_y &\approx P\{x<X_{r:n}\le x+\delta_x, y<Y_{s:n}\le y+\delta_y\} \\ &= P\{r-1\ of\ X's\in(-\infty,x]\,,\ 1\ of\ X's\in(x,x+\delta_x]\,, \\ &\quad n-r\ of\ X's\in(x+\delta_x,\infty),\ s-1\ of\ Y's\in(-\infty,y]\,, \\ &\quad 1\ of\ Y's\in(y,y+\delta_y]\,,\ n-s\ of\ Y's\in(y+\delta_y,\infty)\}. \end{aligned}$$

The compound event $\{x<X_{r:n}\le x+\delta_x, y<Y_{s:n}\le y+\delta_y\}$ can be realized by the following configuration:

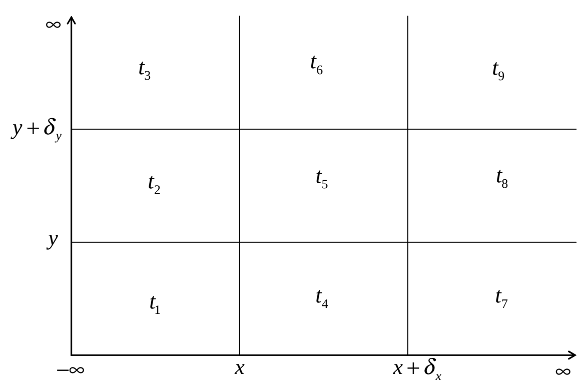

Figure 1. Realization of the compound event

*Corresponding author. E-mail: kemalbay@yildiz.edu.tr (Kemalbay, G.)
†E-mail:ismihan.bayramoglu@ieu.edu.tr (Bayramoglu, I.)

This means that the total number of observations (X,Y) falls in nine regions denoted by I_i, where $I_i \cap I_j = \varnothing$ and $I_i \cup I_j = \overline{\mathbb{R}}^2$; $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ for all $i \neq j$, $i,j = 1,2,...,9$. All possible regions $I_1, I_2, ..., I_9$ are defined as follows:

$$\begin{array}{ll}
I_1 = (-\infty, x] \cap (-\infty, y], & I_6 = (x, x+\delta_x] \cap (y+\delta_y, \infty), \\
I_2 = (-\infty, x] \cap (y, y+\delta_y], & I_7 = (x+\delta_x, \infty) \cap (-\infty, y], \\
I_3 = (-\infty, x] \cap (y+\delta_y, \infty), & I_8 = (x+\delta_x, \infty) \cap (y, y+\delta_y], \\
I_4 = (x, x+\delta_x] \cap (-\infty, y], & I_9 = (x+\delta_x, \infty) \cap (y+\delta_y, \infty). \\
I_5 = (x, x+\delta_x] \cap (y, y+\delta_y], &
\end{array}$$

Let t_i observations fall in the region I_i, $i = 1,2,...,9$. For example, t_1 of (X,Y) are observed in $I_1 = (-\infty, x] \cap (-\infty, y]$ and t_2 of (X,Y) are observed in $I_2 = (-\infty, x] \cap (y, y+\delta_y]$.

Let us denote $C_1 = \{X \leq x, Y \leq y\}$, $C_2 = \{X \leq x, y < Y \leq y+\delta_y\}$, $C_3 = \{X \leq x, Y > y+\delta_y\}$, $C_4 = \{x < X \leq x+\delta_x, Y \leq y\}$, $C_5 = \{x < X \leq x+\delta_x, y < Y \leq y+\delta_y\}$, $C_6 = \{x < X \leq x+\delta_x, Y > y+\delta_y\}$, $C_7 = \{X > x+\delta_x, Y \leq y\}$, $C_8 = \{X > x+\delta_x, y < Y \leq y+\delta_y\}$, $C_9 = \{x < X \leq x+\delta_x, y < Y \leq y+\delta_y\}$.

In the experiment, assume that the outcomes are pairs C_i with probabilities $P(C_i) = p_i$, where $\sum_i p_i = 1$ and $i = 1,2,...,9$. Then we have

$$\begin{aligned}
&P(C_1) = F(x,y), \\
&P(C_2) = F(x, y+\delta_y) - F(x,y), \\
&P(C_3) = F_X(x) - F(x, y+\delta_y), \\
&P(C_4) = F(x+\delta_x, y) - F(x,y), \\
&P(C_5) = F(x+\delta_x, y+\delta_y) - F(x+\delta_x, y) - F(x, y+\delta_y) + F(x,y), \\
&P(C_6) = F_X(x+\delta_x) - F_X(x) - F(x+\delta_x, y+\delta_y) + F(x, y+\delta_y), \\
&P(C_7) = F_Y(y) - F(x+\delta_x, y), \\
&P(C_8) = F_Y(y+\delta_y) - F_Y(y) - F(x+\delta_x, y+\delta_y) + F(x+\delta_x, y), \\
&P(C_9) = 1 - F_X(x+\delta_x) - F_Y(y+\delta_y) + F(x+\delta_x, y+\delta_y).
\end{aligned}$$

Let ζ_i be the number of occurrences in which C_i appears out of n repetitions, $i = 1,2,...,9$. Clearly, the random vector $(\zeta_1, \zeta_2, ..., \zeta_9)$ is multinomial with pmf

$$\begin{aligned}
P\{\zeta_1 &= t_1, \zeta_2 = t_2, ..., \zeta_9 = t_9\} \\
&= \frac{n!}{t_1! t_2! \cdots t_9!} [P(C_1)]^{t_1} [P(C_2)]^{t_2} \cdots [P(C_9)]^{t_9}. \qquad (2)
\end{aligned}$$

Therefore,

$$\begin{aligned}
P\{x &< X_{r:n} \leq x+\delta_x, y < Y_{s:n} \leq y+\delta_y\} \\
&\equiv P\{\zeta_1 + \zeta_2 + \zeta_3 = r-1, \zeta_4 + \zeta_5 + \zeta_6 = 1, \\
&\qquad \zeta_1 + \zeta_4 + \zeta_7 = s-1, \zeta_2 + \zeta_5 + \zeta_8 = 1\} \\
&= \sum_{\substack{t_1+t_2+t_3=r-1 \\ t_4+t_5+t_6=1 \\ t_1+t_4+t_7=s-1 \\ t_2+t_5+t_8=1}} P\{\zeta_1 = t_1, \zeta_2 = t_2, ..., \zeta_9 = t_9\}. \qquad (3)
\end{aligned}$$

Let us consider the summation $t_1+t_2+t_3=r-1$, $t_4+t_5+t_6=1$, $t_1+t_4+t_7=s-1$ and $t_2+t_5+t_8=1$. Because $t_4+t_5+t_6=1$, this equality is valid if and only if t_4, t_5 and t_6 are 0 or 1. Similarly, t_2, t_5 and t_8 are 0 or 1, because $t_2+t_5+t_8=1$. Thus, all possible values of (t_4,t_5,t_6) and (t_2,t_5,t_8) are described in the following table:

t_4	t_5	t_6		t_2	t_5	t_8
1	0	0		1	0	0
0	1	0		0	1	0
0	0	1		0	0	1

Because the number of occurrences of C_5 is $t_5=1$, from the table, we have $t_4=t_6=0$, $t_2=t_8=0$. If $t_5=0$, $t_4+t_6=1$ and $t_2+t_8=1$, where $(t_4=1,t_6=0)$ or $(t_4=0,t_6=1)$ and $(t_2=1,t_8=0)$ or $(t_2=0,t_8=1)$. Therefore, this summation can be reduced to a simpler form according to cases of $t_5=1$ or $t_5=0$:

$$\begin{aligned}
&P\{x<X_{r:n}\le x+\delta_x, y<Y_{s:n}\le y+\delta_y\}\\
=&\sum_{\substack{t_1+t_2+t_3=r-1\\ t_4+t_5+t_6=1\\ t_1+t_4+t_7=s-1\\ t_2+t_5+t_8=1}} P\{\zeta_1=t_1,\zeta_2=t_2,...,\zeta_9=t_9\}\\
=&\sum_{t_1=a_1}^{a_2} P(t_1;r,s,n)+\sum_{t_4=d_1}^{d_2}\sum_{t_2=c_1}^{c_2}\sum_{t_1=b_1}^{b_2} P(t_1,t_2,t_4;r,s,n), \qquad (4)
\end{aligned}$$

where

$$\begin{aligned}
&P(t_1;r,s,n)\\
\equiv\ &P\{\zeta_1=t_1,\zeta_2=0,\zeta_3=r-1-t_1,\zeta_4=0,\zeta_5=1,\\
&\quad \zeta_6=0,\zeta_7=s-1-t_1,\zeta_8=0,\zeta_9=n-r-s+t_1+1\};
\end{aligned}$$

and

$$\begin{aligned}
&P(t_1,t_2,t_4;r,s,n)\\
\equiv\ &P\{\zeta_1=t_1,\zeta_2=t_2,\zeta_3=r-1-t_1-t_2,\zeta_4=t_4,\zeta_5=0,\\
&\quad \zeta_6=1-t_4,\zeta_7=s-1-t_1-t_4,\zeta_8=1-t_2,\\
&\quad \zeta_9=n-r-s+t_1+t_2+t_4\}.
\end{aligned}$$

Therefore, by using (2) we obtain

$$\begin{aligned}
&P(t_1;r,s,n)\\
=&\frac{n!}{t_1!(r-1-t_1)!(s-1-t_1)!(n-r-s+t_1+1)!}\\
&\times[F(x,y)]^{t_1}[F_X(x)-F(x,y+\delta_y)]^{r-1-t_1}\\
&\times[F(x+\delta_x,y+\delta_y)-F(x+\delta_x,y)-F(x,y+\delta_y)+F(x,y)]\\
&\times[F_Y(y)-F(x+\delta_x,y)]^{s-1-t_1}\\
&\times[1-F_X(x+\delta_x)-F_Y(y+\delta_y)+F(x+\delta_x,y+\delta_y)]^{n-r-s+t_1+1} \qquad (5)
\end{aligned}$$

and

$$
\begin{aligned}
& P(t_1, t_2, t_4; r, s, n) \\
= \quad & \frac{n!}{t_1!(r-1-t_1-t_2)!(s-1-t_1-t_4)!(n-r-s+t_1+t_2+t_4)!} \\
& \times [F(x,y)]^{t_1} [F(x, y+\delta_y) - F(x,y)]^{t_2} \\
& \times [F_X(x) - F(x, y+\delta_y)]^{r-1-t_1-t_2} [F(x+\delta_x, y) - F(x,y)]^{t_4} \\
& \times [F_X(x+\delta_x) - F_X(x) - F(x+\delta_x, y+\delta_y) + F(x, y+\delta_y)]^{1-t_4} \\
& \times [F_Y(y) - F(x+\delta_x, y)]^{s-1-t_1-t_4} \\
& \times [F_Y(y+\delta_y) - F_Y(y) - F(x+\delta_x, y+\delta_y) + F(x+\delta_x, y)]^{1-t_2} \\
& \times [1 - F_X(x+\delta_x) - F_Y(y+\delta_y) + F(x+\delta_x, y+\delta_y)]^{n-r-s+t_1+t_2+t_4}.
\end{aligned} \tag{6}
$$

Finally, by substituting the (5) and (6) in the equation (4) and taking the limit as $\delta_x, \delta_y \to 0$, we obtain the joint density function of bivariate order statistics $X_{r:n}$ and $Y_{s:n}$.