2013年6月14日　統計数理研究所　オープンハウス

# Dependence Structure of Bivariate Order Statistics and its Applications

Xiaoling Dou　新領域融合研究センター　融合プロジェクト特任研究員

## Abstract

We study the dependence structure of bivariate order statistics, and prove that if the underlying bivariate distribution $H$ is positive quadrant dependent (PQD) then so is each pair of bivariate order statistics. As an application, we show that if $H$ is PQD, the bivariate distribution $K_+^{(n)}$, proposed by Bairamov and Bayramoglu (2012), is greater than or equal to Baker's (2008) distribution $H_+^{(n)}$. We also show that if $H$ is PQD, $K_+^{(n)}$ converges weakly to the Fréchet–Hoeffding upper bound as $n$ tends to infinity.

## Introduction

**Bivariate order statistics**

Suppose that $(X_1, Y_1), \dots, (X_n, Y_n) \sim^{i.i.d.} H(x,y) = \Pr(X \le x, Y \le y)$.

$$\text{Marginals}: F(x) := \Pr(X \le x); \quad G(y) := \Pr(Y \le y)$$
$$\text{Order statistics}: X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}; \quad Y_{1,n} \le Y_{2,n} \le \cdots \le Y_{n,n}$$

Distribution functions:

$$F_{r,n}(x) := \Pr(X_{r,n} \le x) = \sum_{i=r}^{n} \binom{n}{i} F^i(x)(1-F(x))^{n-i}$$
$$G_{s,n}(y) := \Pr(Y_{s,n} \le y) = \sum_{j=s}^{n} \binom{n}{j} G^j(y)(1-G(y))^{n-j}$$

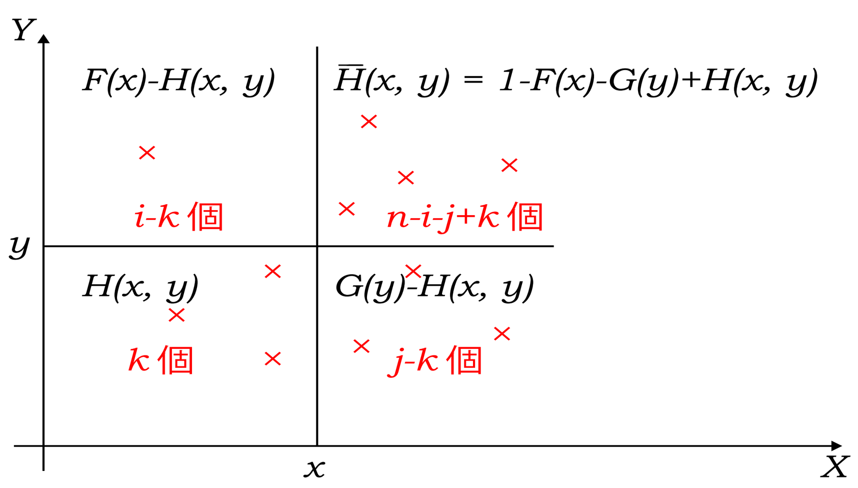
When $X$ and $Y$ are independent: $H(x,y) = F(x)G(y)$, the joint distribution of $(X_{r,n}, Y_{s,n})$:

$$\begin{aligned} K_{r,s}^{(n)}(x,y) &:= \Pr(X_{r,n} \le x, Y_{s,n} \le y) \\ &= \Pr(X_{r,n} \le x)\Pr(Y_{s,n} \le y) = F_{r,n}(x)G_{s,n}(y) \end{aligned}$$

When $X$ and $Y$ are not independent, the joint distribution of $(X_{r,n}, Y_{s,n})$:

$$\begin{aligned} K_{r,s}^{(n)}(x,y) &:= \Pr(X_{r,n} \le x, Y_{s,n} \le y) \\ &= \Pr(\text{at least } r \text{ of the } X_\ell' s \text{ are } \le x, \text{ at least } s \text{ of the } Y_\ell' s \text{ are } \le y) \\ &= \sum_{i=r}^{n}\sum_{j=s}^{n}\sum_{k} f_{k,i,j}^{(n)}(x,y), \end{aligned}$$

$$f_{k,i,j}^{(n)}(x,y) = \frac{n!}{k!(i-k)!(j-k)!(n-i-j+k)!}(H(x,y))^k \times (F(x)-H(x,y))^{i-k}(G(y)-H(x,y))^{j-k}(\overline{H}(x,y))^{n-i-j+k},$$




$$K_{r,s}^{(n)}(x,y) = \sum_{i=r}^{n}\sum_{j=s}^{n}\sum_{k} f_{k,i,j}^{(n)}(x,y) = K_{r,s}^{(n)}(F,G,H)$$

**Positive quadrant dependence (PQD)**:

$$H(x,y) \ge F(x)G(y) \quad \text{for all } x,\ y.$$

**Negative quadrant dependence (NQD)**:

$$H(x,y) \le F(x)G(y) \quad \text{for all } x,\ y.$$

## Dependence Structure

**Theorem 1.**
For $1 \le r,\ s \le n$, the distribution $K_{r,s}^{(n)}$ is increasing in $H$.

**Proof:** $\frac{\partial}{\partial H} K_{r,s}^{(n)}(x,y) = n f_{r-1,s-1}^{(n-1)}(x,y) \ge 0.$ ■

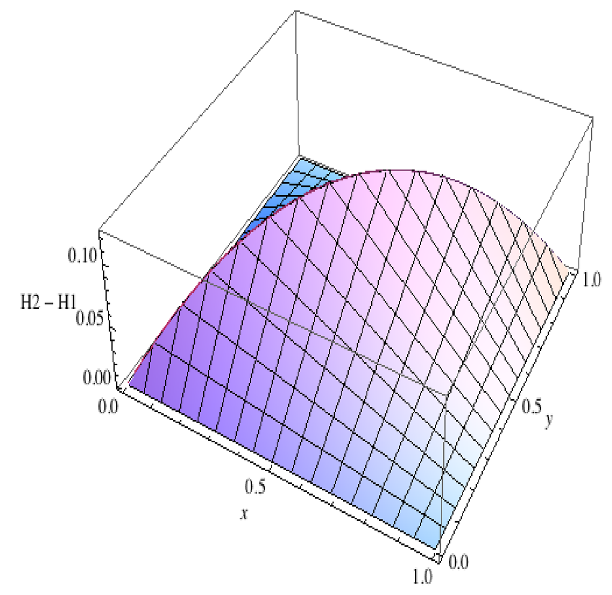



Figure 1: $H_2 - H_1$

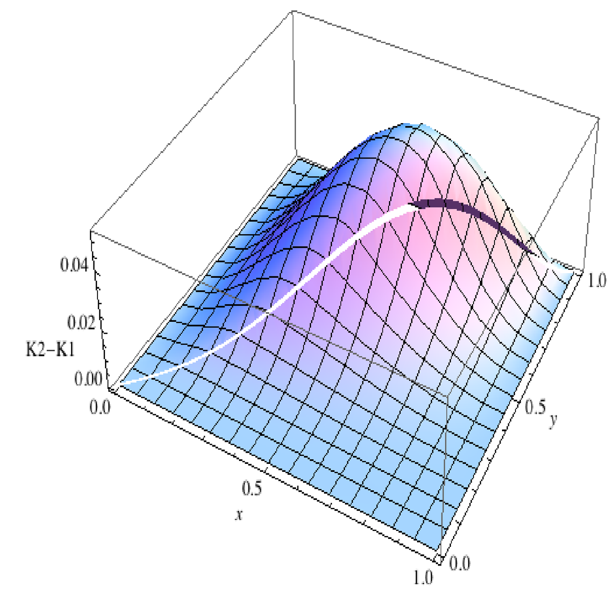



Figure 2: $K_{2,3}^{(3)}(f,G,H_2) - K_{2,3}^{(3)}(f,G,H_1)$

**Corollary 1.**
For $1 \le r,\ s \le n$, the joint distribution of $(X_{r,n}, Y_{s,n})$, $K_{r,s}^{(n)}$, is PQD if $H$ is PQD, and is NQD if $H$ is NQD.

## Theoretical Applications

**Baker's (2008) distribution:**

$$H_R^{(n)}(x,y) = \sum_{r=1}^{n}\sum_{s=1}^{n} r_{rs}F_{r,n}(x)G_{s,n}(y), \quad \sum_{r=1}^{n} r_{sr} = \sum_{s=1}^{n} r_{sr} = \frac{1}{n}, \quad r_{sr} \ge 0.$$
$$H_+^{(n)}(x,y) = \frac{1}{n}\sum_{r=1}^{n} F_{r,n}(x)G_{r,n}(y); \quad H_-^{(n)}(x,y) = \frac{1}{n}\sum_{r=1}^{n} F_{r,n}(x)G_{n-r+1,n}(y).$$

**Bairamov and Bayramoglu's (2013) distribution:**

$$K_R^{(n)}(x,y) = \sum_{r=1}^{n}\sum_{s=1}^{n} r_{rs}\Pr(X_{r,n} \le x, Y_{s,n} \le y), \sum_{r=1}^{n} r_{sr} = \sum_{s=1}^{n} r_{sr} = \frac{1}{n}, \ r_{sr} \ge 0.$$
$$K_+^{(n)}(x,y) = \frac{1}{n}\sum_{r=1}^{n}\Pr(X_{r,n} \le x, Y_{r,n} \le y), K_-^{(n)}(x,y) = \frac{1}{n}\sum_{r=1}^{n}\Pr(X_{r,n} \le x, Y_{n-r+1,n} \le y),$$

**Theorem 2.**
(i) For $n \ge 1$, $K_+^{(n)} \ge H_+^{(n)}$ or $K_+^{(n)} \le H_+^{(n)}$ depending on $H$ is PQD or NQD.
(ii) For $n \ge 1$, $K_-^{(n)} \ge H_-^{(n)}$ or $K_-^{(n)} \le H_-^{(n)}$ depending on $H$ is PQD or NQD.

**Monotonicity of $K_+^{(n)}(x,y)$**

**Fact:** As $n \to \infty$, $H_+^{(n)}(x,y) \to \min\{F(x), G(y)\}$ (Dou et al. 2013)

**Problem:** As $n \to \infty$, $H_+^{(n)}(x,y) \to \min\{F(x), G(y)\}$ monotonically increases in $n$?

**Theorem 3.** (i) For $n \ge 2$, the distribution $K_+^{(n)}$ is of the form

$$K_+^{(n)} = H + \frac{1}{n}\sum\sum_{i+j\le n} \min\{i,j\}\binom{n}{i,j,n-i-j}(F-H)^i(G-H)^j(H+\overline{H})^{n-i-j}.$$

(ii) Let $W = (F-H)(G-H)$ and $V = H + \overline{H}$. Then

$$K_+^{(n)} = H + \sum_{m=2}^{n}\frac{1}{m-1}\sum_{i=1}^{[m/2]}\binom{m-1}{i,i-1,m-2i}W^iV^{m-2i}, \qquad n \ge 2,$$

where $[a]$ is the largest integer less than or equal to $a$. Equivalently,

$$K_+^{(n)} - K_+^{(n-1)} = \frac{1}{n-1}\sum_{i=1}^{[n/2]}\binom{n-1}{i,i-1,n-2i}W^iV^{n-2i} \ge 0, \qquad n \ge 2.$$

## References


- Bairamov, I. and Bayramoglu, K. (2013). From the Huang–Kotz FGM distribution to Baker's bivariate distribution, *Journal of Multivariate Analysis*, **113**, 106–115.
- Baker, R. (2008). An order-statistics-based method for constructing multivariate distributions with fixed marginals, *Journal of Multivariate Analysis*, **99**, 2312–2327.
- Dou, X., Kuriki, S., Lin, G. D. (2013). Dependence structures and asymptotic properties of Baker's distributions with fixed marginals, *Journal of Statistical Planning and Inference*, **143**, 1343-1354.