POSITIVELY QUADRANT DEPENDENT BIVARIATE DISTRIBUTIONS WITH GIVEN MARGINALS

J. REIMANN

Department of Civil Engineering Mathematics,
Technical University, H-1521 Budapest

Received June 20, 1987

Abstract

Several measures for the dependence of two random variables are investigated in the case of given marginals and assuming positively quadrant dependence. Beyond known quantities (Spearman, Pearson correlation coefficient, etc.) three new measures are introduced and compared with the others. In detail are investigated the λ-dependent variables (Konijn) moreover a special type of bivariate distributions; a practical application in the hydrology of flood peaks is included.

1. Introduction

Let $\Pi^+ = \Pi^+(F, G)$ be the set of all continuous cdf's cumulative distribution functions (cdf's) H on R^2 having continuous, strictly increasing marginal cdf's F and G. It will be assumed that F and G have finite variances.

Let H be a positively quadrant dependent, (Lehmann [6]), in which case, it is well known that for $H(x, y)$ the following inequality holds:

$$F(x)G(y) \leq H(x, y) \leq \min\,[F(x), G(y)] \tag{1}$$

for all x, y.

For positively quadrant dependent cdf's we introduce the following connection — function:

$$\lambda(x, y) = \frac{H(x, y) - F(x)G(y)}{\min[F(x), G(y)] - F(x)G(y);} \tag{2}$$

due to (1) $0 \leq \lambda(x, y) \leq 1$. At the same time this is the deviation of H and F, G relative to its maximum value. Let $H^+(x, y) = \min\,[F(x), G(y)]$ which is the "largest" bivariate cdf with marginals $F(x)$ and $G(y)$ resp. namely

$$\begin{aligned} H^+(\infty, y) &= \min\,[1, G(y)] = G(y) \\ H^+(x, \infty) &= \min\,[F(x), 1] = F(x) \end{aligned} \tag{3}$$

In the case of a positive quadrant dependence, $H_0(x, y) = F(x)G(y)$ is the "smallest" bivariate cdf, with marginals $F(x)$ and $G(y)$ resp.

If the random variables X and Y have a joint cdf. $H(x, y) \geq F(x)G(y)$ then for a measure of their degree the following measure seems natural:

$$\lambda^* = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(x, y) f(x) g(y) dx dy =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H(x, y) - F(x)G(y)}{\min[F(x)G(y)] - F(x)G(y)} f(x)g(y)dxdy = E_{fg}[\lambda(X, Y)] \tag{4}$$

which is the expected "relative" deviation between H and FG.

Proposition 1. If $H \geq FG$ and further F and G are strictly increasing functions of x and y respectively, then the measure λ^* has the following properties:

(I) $\lambda^* = 0$ if X and Y are independent
$\lambda^* = 1$ if there is a monotonically increasing functional — relation between X and Y: $Y = G^{-1}[F(X)]$ or $X = F^{-1}[G(Y)]$

(II) λ^* is a monotonically increasing function of H, in the sense that if $H_2 \geq H_1$, then $\lambda_2^* \geq \lambda_1^*$

(III) λ^* is invariant under the concordant monotonic transformations of the r. variables X and Y.

Proof: *(I)* From (4) it follows, that $\lambda^* \equiv 0$ if $H = FG$ and $\lambda^* \equiv 1$ if $H = \min(F, G)$. Let $H = \min(F, G)$, i.e.

$$H = \begin{cases} F & \text{if } F \leq G \\ G & \text{if } F > G \end{cases}$$

Let now $\beta \geq \alpha$ and $F \geq G$, then $H = G$ i.e.

$H(\tilde{x}_\beta, \tilde{y}_\alpha) = G(\tilde{y}_\alpha) = \alpha = F(x_\alpha)$ where $\tilde{x}_\alpha$, $\tilde{y}_\alpha$ are the α-quantiles of F and G resp.)

Hence: $\tilde{y}_\alpha = G^{-1}[F(\tilde{x}_\alpha)]$ for all $\alpha \in [0, 1]$. Similarly if $F < G$, then $H = F$ and $H(\tilde{x}_\alpha, \tilde{y}_\beta) = F(\tilde{x}_\alpha) = \alpha = G(\tilde{y}_\alpha)$ i.e. $\tilde{x}_\alpha = F^{-1}[G(\tilde{y}_\alpha)]$, $\tilde{y}_\alpha = G^{-1}[F(\tilde{x}_\alpha)]$ $\alpha \in [0, 1]$.

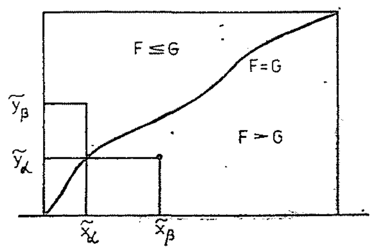

Fig. 1

Let now $Y = \varphi(X)$ be a strictly increasing continuous function, then

$$\alpha = G(\tilde{y}_\alpha) = P(Y < \tilde{y}_\alpha) = P(\varphi(X) < \tilde{y}_\alpha) =$$
$$= P[X < \varphi^{-1}(\tilde{y}_\alpha)] = P(X < \tilde{x}_\alpha) = F(\tilde{x}_\alpha)$$

i.e. $G(\tilde{y}_\alpha) = F(\tilde{x}_\alpha), \tilde{y}_\alpha = G^{-1}[F(\tilde{x}_\alpha)]\ \ \alpha \in [0, 1]$

(II) it follows, that for $H_2 \geq H_1$

$$\lambda_2^* - \lambda_1^* = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H_2 - FG}{\min(F, G) - FG} fg\, dx\, dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H_1 - FG}{\min(F, G) - FG} fg\, dx\, dy =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H_2 - H_1}{\min(F, G) - FG} fg\, dx\, dy \geq 0$$

(III) Let $U = \varphi(X)$, $V = \psi(Y)$ where φ and ψ are both monotonically increasing or both monotonically-decreasing then $X = \varphi^{-1}(U)$, $Y = \psi^{-1}(V)$ and

$$F_1(u) = P(U < u) = P\big(\varphi(X) < u\big) = P\big(X < \varphi^{-1}(u)\big) = F[\varphi^{-1}(u)] = F(x)$$
$$G_1(v) = P(V <) = P\big(\psi(Y) < v\big) = P\big(Y < \psi^{-1}(v)\big) = G[\psi(v)] = G(y)$$
$$H_1(u, v) = P(U < u, V < v) = P\big(\varphi(X) < u, \psi(Y) < v\big) =$$
$$= P[X < \varphi^{-1}(u), Y < \psi^{-1}(v)] = H[\varphi^{-1}(u), \psi^{-1}(v)] = H(x, y)$$

$$F_1'(u) = f_1(u) = f[\varphi^{-1}(u)] \frac{d\varphi(u)}{du} = f(x) \frac{dx}{du}$$

$$G_1'(v) = g_1(v) = g[\psi^{-1}(v)] \frac{d\psi^{-1}(v)}{dv} = g(y) \frac{dy}{dv}$$

$$\lambda^*(U, V) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H_1(u, v) - F_1(u)G_1(v)}{\min[F_1(u), G_1(v)] - F_1(u)G(v)} f_1(u)g_1(v)du\, dv =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H[\varphi^{-1}(u), \psi^{-1}(v)] - F[\varphi^{-1}(u)]G[\psi^{-1}(v)]}{\min[F[\varphi^{-1}(u), G[\psi^{-1}(v)]] - F[\varphi^{-1}(u)]G[\psi^{-1}(v)]} f_1(u)g_1(v)du\, dv =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H(x, y) - F(x)G(y)}{\min[F(x), G(y)] - F(x)G(y)} f(x)g(y)dxdy = \lambda^*(X, Y)$$

q.e.d

2. Investigation of some nonparametric measurns of association in case of a positively quadrant dependence

There are very many possibilities to construct measures of association and a lot of them have been proposed. Among the most familiar measures we mention the following nonparametric ones:

$$r = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)dx\,dy}{\sigma_1\sigma_2} \quad \text{(correlation coefficient, Pearson)} \qquad (2.1.)$$

$$\varrho = 12 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)fg\,dx\,dy = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)fg\,dx\,dy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(F, G) - FG]fg\,dx\,dy} \qquad (2.2)$$

(Spearman)

$$\tau = 4 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Hh\,dx\,dy - 1 = \frac{1}{3}\,\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (Hh - FGfg)dx\,dy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(F, G) - FG]fg\,dydx} \qquad (2.3)$$

(Kendall)

$$\mu = 90 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)^2 fg\,dx\,dy = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)^2 fg\,dx\,dy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(F, G) - FG]^2 fg\,dx\,dy} \qquad (2.4)$$

(Hoeffding)

$$\gamma = \sqrt{\mu} \quad \text{(Blum—Kiefer—Rosenblatt)} \qquad (2.5)$$

$$q = 4\,H(\tilde{x}_{\frac{1}{2}}, \tilde{y}_{\frac{1}{2}}) - 1 = \frac{H(\tilde{x}_{\frac{1}{2}}, \tilde{y}_{\frac{1}{2}}) - F(\tilde{x}_{\frac{1}{2}})G(\tilde{y}_{\frac{1}{2}})}{\min F(\tilde{x}_{\frac{1}{2}}, \tilde{y}_{\frac{1}{2}}) - F(\tilde{x}_{\frac{1}{2}})G(\tilde{y}_{\frac{1}{2}})} \qquad (2.6)$$

(Blomqvist)

$$K = 4 \sup_{(x,y)} | H(x, y) - F(x)G(y) | \quad \text{(Schweizer-Wolff)} \qquad (2.7)$$

It is not difficult to construct other measures. For the case of a positively quadrant dependence beyond λ^* we propose the following further measures:

$$\nu = 90 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)[\min(F, G) - FG]fg\,dx\,dy =$$

$$= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)[\min(F, G) - FG]\,fg\,dx\,dy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(F, G) - FG]^2 fg\,dx\,dy} \qquad (2.8)$$

$$\omega = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H - FG}{\sqrt{F(1 - F)G(1 - G)}} fg\, dx\, dy \tag{2.9}$$

$$\lambda^{**} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)dx\, dy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(F, G) - FG]dx\, dy} = \frac{r}{r_+} \tag{2.10}$$

where r_+ is the correlation coefficient if the joint distribution of X and Y is $H^+(x, y) = \min [F(x), G(y)]$. For different H's the values of the mentioned measures depend on H in a fairly simple way. Some relations among them are contained in the following proposition.

Proposition 2.1.

$\lambda^* \geq \frac{\varrho}{3}$	(2.1)	$\tau \geq \frac{\varrho}{3}$	(2.5)
$\lambda^{**} \geq r$	(2.2)	$\mu \geq 0.625\ \varrho^2$	2.6)
$\mu \leq \nu \leq \gamma = \sqrt{\mu}$	(2.3)	$\gamma \geq \frac{\sqrt{90}}{12} \varrho$	(2.7)
$\lambda^* \geq \omega$	(2.4)	$\lambda^* \geq 0.625\ \varrho^2$	(2.8)

Proof: (2.1) follows from the fact, that $\min (F, G) - FG \leq \frac{1}{4}$; namely

in case $F \leq G$, $\min(F, G) - FG = F(1 - G) \leq F(1 - F) \leq \frac{1}{4}$

in case $F > G$, $\min(F, G) - FG = G(1 - F \leq G(1 - G) \leq \frac{1}{4}$

$$\lambda^* = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H - FG}{\min(F, G) - FG} fg\, dx\, dy \geq 4 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG) fg\, dx\, dy = \frac{\varrho}{3}$$

(2.2) follows from the fact that $r_+ = \frac{1}{\sigma_1\sigma_2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\min F, G - FG)\, dx\, dy \leq 1$

(2.3) is a consequence of the inequality of Schwarz. Namely

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H - FG)[\min(F, G) - FG] fg\, dx\, dy \leq$$

$$\leq [\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(H - FG)^2 fg\, dx\, dy]^{\frac{1}{2}} [\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(F, G) - FG]^2 fg\, dx\, dy]^{\frac{1}{2}},$$

hence

$$\frac{\nu}{90} \leq \frac{\sqrt{\mu}}{\sqrt{90}} \cdot \frac{1}{\sqrt{90}}, \text{ i.e. } \nu \leq \sqrt{\mu} = \gamma,$$

further

$$\frac{\nu}{90} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H - FG)[\min(F, G) - FG] fg \, dx \, dy \geq$$

$$\geq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H - FG)^2 fg \, dx \, dy = \frac{\mu}{90}$$

(2.4) follows from the fact, that if $F \leq G$, then $1 - F \geq 1 - G$, i.e.

$$\sqrt{F(1-G)} \leq \sqrt{G(1-F)}$$

$$F(1-G) \leq \sqrt{F(1-F)G(1-G)}$$

and if $F \geq G$

$$G(1-F) \leq \sqrt{F(1-F)G(1-G)}$$ consequently

$$\lambda^* = \iint_{F \leq G} \frac{H - FG}{F(1-G)} fg \, dx \, dy + \iint_{F > G} \frac{H - FG}{G(1-F)} fg \, dx \, dy \geq$$

$$\geq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{H - FG}{\sqrt{F(1-F)G(1-G)}} fg \, dx \, dy = \omega$$

To see (2.5) we have to compare

$$\tau = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Hh \, dx \, dy - 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} FG \, fg \, dx \, dy$$

and

$$\frac{\varrho}{3} = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H \, fg \, dx \, dy - 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} FG \, fg \, dx \, dy$$

For $H \geq FG$, the relation

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H \, fg \, dx \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} FG \, h \, dx \, dy \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Hh \, dx \, dy \quad \text{(Konijn [4])}$$

is valid and it follows that

$$\tau \leq \frac{\varrho}{3}.$$

(2.6) is a consequence of Schwarz in inequality according to which

$$\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(H-FG)fg\,dx\,dy\right]^2 \le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(H-FG)^2 fg\,dx\,dy \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}1^2\cdot fg\,dx\,dy$$

and

$$\left(\frac{\varrho}{12}\right)^2 \le \frac{\mu}{90}$$

hence

$$\mu \ge \frac{90}{144}\varrho^2 = 0.625\,\varrho^2$$

and

$$\gamma = \sqrt{\mu} \ge \frac{\sqrt{90}}{12}\varrho$$

3 Investigation of a positive λ-dependence

Konijn [4] investigated the following type of cdf.

$$H_\lambda = \lambda \min(F, G) + (1-\lambda)FG \quad (0 \le \lambda \le 1) \tag{3.1}$$

It is obvious, that $H_\lambda \ge FG$ i.e. H_λ is positively quadrant dependent.

Proposition 3.1. If the r.v's X and Y have joint distribution function H_λ, then

$$\lambda^* = \varrho = \gamma = \lambda^{**} = \nu = K = q = \lambda \tag{3.2}$$

$$\tau \le \lambda; \tag{3.3}$$

$$r \le \lambda; \tag{3.4}$$

$$\mu = \lambda^2 \le \lambda \tag{3.5}$$

Proof: For the statement (3.1) we have

$$\lambda^* = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{H_\lambda - FG}{\min(F,G) - FG}fg\,dx\,dy = \lambda\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}fg\,dx\,dy = \lambda.$$

$$\varrho = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(H_\lambda - FG)fg\,dx\,dy = 12\,\lambda\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\min(F,G) - FG]fg\,dx\,dy = \lambda$$

being the second integral $\frac{1}{12}$ which follows from the sequence of equalities:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\min(F, G) - FG]fg\,dx\,dy = \iint_{F\le G} F(1-G)fg\,dx\,dy +$$

$$+ \iint_{F<G} G(1-F)fg\,dx\,dy = \int_{y=-\infty}^{\infty}(1-G)\Big[\int_{x=-\infty}^{F^{-1}(G)} F f\,dx\Big]g\,dy +$$

$$+ \int_{x=\wr}^{\infty}(1-F)\left(\int_{y=-\infty}^{G^{-1}(F)}(G g\,dy\right)f\,dx = \frac{1}{2}\int_{-\infty}^{\infty}(G^2 - G^3)g\,dy +$$

$$+ \frac{1}{2}\int_{-\infty}^{\infty}(F^2 - F^3)f\,dx = \frac{1}{12}$$

$\gamma = \sqrt{\mu}$ being according to 2.4 and 3.1 we have

$$\mu = 90\lambda^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\min(F, G) - FG]^2 fg\,dx\,dy = \lambda^2$$

as the relation

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\min(F, G) - FG]^2 fg\,dx\,dy = \frac{1}{90}$$

is well known.

For the statement concerning ν we have

$$\nu = 90\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(H - FG)[\min(F, G) - FG]fg\,dx\,dy =$$

$$= 90\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\min(F, G) - FG]^2 fg\,dx\,dy = \lambda$$

$$\varkappa = 4\sup_{(xy)}(H_\lambda - FG) = 4\lambda\sup_{(xy)}[\min(F, G) - FG] = 4\lambda\cdot\frac{1}{4} = \lambda$$

$$q = 4\left[H_\lambda(\tilde{x}_{\frac{1}{2}}, \tilde{y}_{\frac{1}{2}}) - \frac{1}{4}\right] = 4\left[\lambda\cdot\frac{1}{2} + (1-\lambda)\frac{1}{4} - \frac{1}{4}\right] = \lambda$$

To see that 3.3 holds let us denote min (F, G) by H^+. Konijn [4] has shown that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H^+h^+\,dx\,dy = \frac{1}{2},\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}FGh^+\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H^+fg\,dx\,dy = \frac{1}{3}$$

hence

$$\tau = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_\lambda h_\lambda \, dx \, dy - 1 = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H_\lambda - FG) h_\lambda \, dx \, dy +$$

$$+ 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} FG(h_\lambda - fg) dx \, dy$$

where $h_\lambda = \lambda h^+ + (1 - \lambda) fg$.

This way by simple computation we get:

$$\tau = \frac{\lambda^2}{3} + \frac{2}{3} \lambda \leq \lambda \quad \text{(equality holds in the case of } \lambda = 1 \text{ only).}$$

(3.4) Can be seen by direct computation:

$$r = \frac{1}{\sigma_1 \sigma_2} (H_\lambda - FG) dx \, dy = \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{[\min(F, G) - FG]}{\sigma_1 \sigma_2} dx \, dy = \lambda r_+ \leq \lambda$$

$$(r_+ = 1 \text{ iff } G^{-1}[F(x)] = ax + b, \text{ where } a \geq 0)$$

Relation (3.5) is obvious:

$$\mu = \lambda^2 \leq \lambda.$$

Remark: Let $H_{\lambda_i} = \lambda_i \min(F, G) + (1 - \lambda_i) FG \quad i = 1, 2$, then

$$r_\lambda = \frac{1}{\sigma_1 \sigma_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H\lambda_i - FG) dx \, dy =$$

$$= \frac{\lambda_i}{\sigma_1 \sigma_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\min(F, G) - FG] dx \, dy = \lambda_i r_+ \quad i = 1, 2$$

Here

$$r_+ = \frac{1}{\sigma_1 \sigma_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\min(F, G) - FG] dx \, dy =$$

$$= \frac{1}{\sigma_1 \sigma_2} \left[\int_{-\infty}^{\infty} x G^{-1}[F(x)] f(x) dx - \int_{-\infty}^{\infty} x f(x) dx \int_{-\infty}^{\infty} y g(y) dy \right]$$

i.e.

$$\frac{r_{\lambda_1}}{r_{\lambda_2}} = \frac{\lambda_1}{\lambda_2}$$

Theorem 1. shows that if the joint distribution of the random variables X and Y is $H_\lambda = \lambda \min(F, G) + (1 - \lambda) FG$, then the coefficient λ expresses itself the degree of positive association between the r.v.s.

Corollary 3.1.: If for the cdf $H \geq FG$ holds and we calculate the measures

$$\alpha_1 = \lambda^*,\ \alpha_2 = \varrho,\ \alpha_3 = \gamma,\ \alpha_4 = \lambda^{**},\ \alpha_5 = \nu,\ \alpha_6 = \varkappa,\ \alpha_7 = q\ \alpha_8 = \tau,$$
$$\alpha_9 = r,\ \alpha_{10} = \mu$$

then we can find the linear combination of $\min(F, G)$ and FG,

$$H_\lambda^+ = \lambda_i \min(F, G) + (1 - \lambda_i)FG \qquad (i = 1, 2, \ldots, 10)$$

for which $\alpha_i^+ = \alpha_i \quad (i = 1, 2, \ldots, 10)$

where

$$\lambda_1 = \lambda^+,\ \lambda_2 = \varrho,\ \lambda_3 = \gamma,\ \lambda_4 = \lambda^{**},\ \lambda_5 = \nu,\ \lambda_6 = \varkappa,\ \lambda_7 = q,$$
$$\lambda_8 = \sqrt{1 + 3\tau} - 1,\ \lambda_9 = \frac{r}{r_+},\ \lambda_{10} = \sqrt{\mu}$$

Theorem 3.3: If r.v.s. X and Y have the joint cdf $H = \lambda \min(F, G) + (1 - \lambda)FG$ then the regression curve of Y with respect to X has the form

$$E(Y|X) = x) = \bar{y}(x) = \lambda G^{-1}[F(x)] + (1 - \lambda)E(Y) \qquad (3.6)$$

Proof: The conditional cdf of Y under the condition $X = x$.

$$G(y|x) = \frac{1}{f(x)} \frac{\partial H(x, y)}{\partial x}$$

As

$$H_\lambda = \begin{cases} \lambda F + (1 - \lambda)FG & \text{if} \quad F \leq G \\ \lambda G + (1 - \lambda)FG & \text{if} \quad F > G \end{cases}$$

it follows that in this case

$$G(y|x) = \lambda + (1 - \lambda)G \qquad \text{if} \quad F \leq G$$
$$(1 - \lambda)G \qquad \text{if} \quad F > G$$

and we obtain:

$$\bar{y}(x) = \int_{-\infty}^{\infty} y dG(y, x) = \lambda y + (1 - \lambda) \int_{-\infty}^{\infty} y\, G(y) dy =$$
$$= G^{-1}[F(x)] + (1 - \lambda)E(Y)$$

which is true as in case

$$F = G,\ y = G^{-1}[F(x)]\,.$$

Remark 1: Let the joint cdf of r.v.s. X and Y be a two-dimensional normal cdf, with marginals:

$$N(m_1, \sigma_1) \quad \text{and} \quad N(m_1, \sigma_2)$$

and with correlation coefficient r.

Then $F(x) = \Phi\left(\frac{x - m_1}{\sigma_1}\right)$, $G(y) = \Phi\left(\frac{y - m_2}{\sigma_2}\right)$ where $\Phi(x)$ is the standard normal cdf.

The equation of the quantile curve is:

$$\tilde{y}(x) = G^{-1}[F(x)] = \frac{\sigma_2}{\sigma_1}(x - m_1) + m_2$$

The equation of the regression line of Y with respect to X is

$$\bar{y}(x) = r\frac{\sigma_2}{\sigma_1}(x - m_1) + m_2$$

Hence

$$\bar{y}(x) = rG^{-1}[F(x)] + (1 - r)E(Y) \tag{3.7}$$

i.e. the relation (3.6) holds for bivariate normal distributions as well substituting $\lambda = r$.

From this fact we get the following theorem:

Theorem 3.4: Let H be a two dimensional normal cdf with correlation coefficient r and with marginals F and G; let further

$$H_r = r\min(F, G) + (1 - r)FG\,. \tag{3.8}$$

Then for H and H_r the correlation coefficients, as also the regression lines coincide.

Remark 2: This way (3.7) is a necessary condition for a two dimensional cdf. $H(x, y)$ with normal marginals F and G resp. to be two a dimensional normal cdf.

The fact, that (3.7) is not a sufficient condition for two a dimensional normality shows the following example: (Rényi [8]. pp. 317—318).

Let $H(x, y)$ be a two variate cdf having density function:

$$h(x, y) = \frac{1}{2\pi}\left(\sqrt{2}\,e^{-\frac{x^2}{2}} - e^{-x^2}\right)e^{-y^2} + \left(\sqrt{2}\,e^{-\frac{y^2}{2}} - e^{-y^2}\right)e^{-x^2}$$

The marginal densities are:

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \text{ and } g(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} \quad \text{resp.}$$

A simple calculation shows that $r(X, Y) = 0$, but X and Y are not independent, as $h(x, y) \neq f(x)g(y)$.

The conditional density function of Y under the condition $X = x$ is

$$g(y/x) = \frac{h(x,y)}{f(x)} = \frac{1}{\sqrt{2\pi}}(\sqrt{2} - e^{-\frac{x^2}{2}})e^{-y^2} + \sqrt{2\pi}(\sqrt{2}\, e^{-\frac{y^2}{2}} - e^{-y^2})\, e^{-\frac{x^2}{2}}$$

The regression function of Y with respect to X is:

$$\bar{y}(x) = \int_{-\infty}^{\infty} y\, g(y|x)dy = \frac{1}{\sqrt{2\pi}}(\sqrt{2} - e^{-\frac{x^2}{2}}) \int_{-\infty}^{\infty} y e^{-y^2}\, dy +$$
$$+ \sqrt{2\pi}\, e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} (\sqrt{2}\, e^{-\frac{y^2}{2}} - e^{-y^2})\, dy = 0 = E(Y)$$

The equation of the quantile curve is: $y = G^{-1}[F(x)] = x$. Since $r = 0$

$$y(x) = 0 \cdot x + (1 - 0)E(Y)$$

i.e. (3.7) holds.

The relation (3.8) is a somewhat more attractive example for the fact that (3.7) is not a sufficient condition for the bivariate normality.

4. Realition for q, ρ, τ, and λ* in some special type of distributions

(4.1) Let us consider the bivariate distribution H_0 which has the general appearance of Fig. 2.

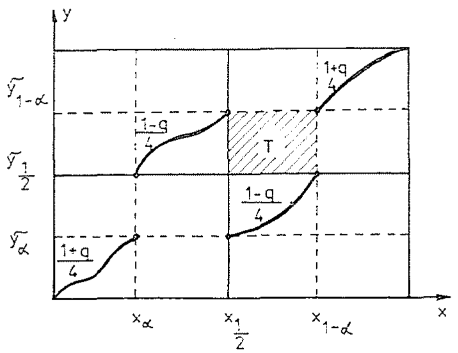

Fig. 2

For this bivariate distribution

$$q_0 = 4\,\alpha - 1, \quad \text{i.e.} \quad \alpha = \frac{1+q}{4} \tag{4.1}$$

Kruskal [4] has shown, that in this case

$$\varrho_0 = 1 - \frac{3}{16}(1-q)^3; \quad \tau_0 = 1 - \frac{(1-q)^2}{4} = \\ = 1 - \frac{4}{16}(1-q)^2; \quad \text{i.e.} \quad \varrho \geq \tau \tag{4.2}$$

Proposition 4.1: For the cdf. H_0

$$\lambda_0^* = 1 - \frac{(1-q)^3}{16}$$

Proof: For any points $(x, y) \in R^2$ but the points of the rectangle T, $H_0(x, y) = \min(F, G)$ holds, from which follows that:

$$\lambda_0^* = 1 - \iint_T fg\, dx\, dy + \iint_T \frac{H - FG}{\min(F, G) - FG} fg\, dx\, dy \tag{4.3}$$

$$\iint_T fg\, dx\, dy = \int_{\tilde{x}_{\frac{1}{2}}}^{\tilde{x}_{1-\alpha}} f \left(\int_{\tilde{y}_{\frac{1}{2}}}^{\tilde{y}_{1-\alpha}} g\, dy \right) dx = \left(1 - \alpha - \frac{1}{2}\right)^2 = \frac{(1-q)^2}{16}$$

$$\iint_T \frac{H - FG}{\min(F, G) - FG} fg\, dx\, dy \geq \min \frac{H - FG}{\min(F, G) - FG} fg\, dx\, dy$$

A simple calculation shows that $\min\limits_{(x, y) \in T} \dfrac{H - FG}{\min(F, G) - FG} = q$

From (4.3) follows that

$$\lambda_0^* \geq 1 - \frac{(1-q)^2}{16} + q \frac{(1-q)^2}{16} = 1 - \frac{(1-q)^3}{16} \geq \varrho \geq \tau. \tag{4.4}$$

(4.2) Let us consider now the bivariate distribution H defined inside the unit square for which the probability mass is uniformly spread within the two squares T_1:

$$(0, 0), \left(\frac{1}{2}, 0\right) \left(\frac{1}{2}, \frac{1}{2}\right), \left(0, \frac{1}{2}\right) \quad \text{and} \quad T_3: \left(\frac{1}{2}, \frac{1}{2}\right), \left(1, \frac{1}{2}\right) (1, 1), \left(\frac{1}{2}, 1\right)$$

The support of this distribution can be seen in Fig. 3.

Kruskal [5] has given for this distribution the following values

$$q = 1, \quad \tau = \frac{1}{2}, \quad \varrho = \frac{3}{4} = \frac{3}{2}\tau$$

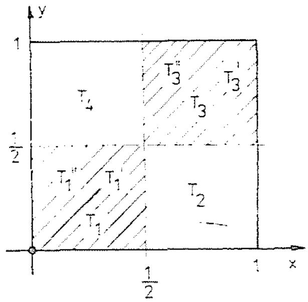

Fig. 3

We can show that for this distribution

$$\lambda^* = 4 \ln 2 - 2 \approx \frac{4}{5} \tag{4.5}$$

To see this we proceed as follows:
Within the square T_1:

$$H(x, y) = \frac{xy}{\frac{1}{2}} = 2\,xy, \quad F(x) = x, \quad G(y) = y,$$

$$f(x) = g(y) = 1$$

Hence: $H - FG = xy$ and

$$\iint\limits_{T_1} \frac{H - FG}{\min(F, G) - FG} fg\,dx\,dy = \iint\limits_{T_1'} \frac{y}{1 - y}\,dx\,dy + \iint\limits_{T_1''} \frac{x}{1 - x}\,dx\,dy =$$

$$= \int\limits_{y=0}^{\frac{1}{2}} \frac{y}{1 - y} \left(\int\limits_{x=0}^{y} dx \right) dy + \int\limits_{x=0}^{\frac{1}{2}} \frac{x}{1 - x} \left(\int\limits_{y=0}^{x} dy \right) dx =$$

$$= 2 \ln 2 - \frac{10}{8} = \iint\limits_{T_3} \frac{H - FG}{\min(F, G) - FG}\,dx\,dy$$

Within the square T_2 and T_4:
$H(x, y) = \min(F, G)$ therefore

$$\iint\limits_{T_2 + T_4} \frac{H - FG}{\min(F, G) - FG}\,dx\,dy = \frac{1}{2}$$

Hence:

$$\int\limits_0^1 \int\limits_0^1 \frac{H - FG}{\min(F, G) - FG} fg\,dx\,dy = 4 \ln 2 - 2 \approx \frac{4}{5}$$

(4.3) D. Morgenstern [7] investigated the following type of bivariate distribution:

$$H_1 = FG + \alpha F(1 - F)G(1 - G) \tag{4.6}$$

$$\text{where } -1 \leq \alpha \leq 1$$

In case of a positively quadrant dependence $0 \leq \alpha \leq 1$ must hold. For this distribution:

$$\lambda_1^* = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{H_1 - FG}{\min(F, G) - FG} fg\,dx\,dy = \alpha \iint_{F \leq G} G(1 - F)fg\,dx\,dy +$$

$$+ \alpha \iint_{F > G} F(1 - G)fg\,dx\,dy = \alpha \int_{y=-\infty}^{\infty} G\left(\int_{x=-\infty}^{F^{-1}(G)} (1 - F)f\,dx\right) g\,dy +$$

$$+ \alpha \int_{x=-\infty}^{\infty} F\left(\int_{y=-\infty}^{G^{-1}(F)} (1 - G)g\,dy\right) f\,dx = \alpha \int_{-\infty}^{\infty} \left(G^2 - \frac{G^3}{2}\right) g\,dy +$$

$$+ \alpha \int_{-\infty}^{\infty} \left(F^2 - \frac{F^3}{3}\right) f\,dx = \frac{5}{12}\alpha$$

$$\varrho_1 = 12 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H_1 - FG)fg\,dx\,dy = 12\alpha \int_{-\infty}^{\infty} (F - F^2)f\,dx \int_{-\infty}^{\infty} (G - G^2)g\,dy = \frac{\alpha}{3}$$

$$q_1 = 4\,H_1(\tilde{x}_{\frac{1}{2}}, \tilde{y}_{\frac{1}{2}}) - 1 = 4\left(\frac{1}{4} + \frac{\alpha}{16}\right) - 1 = \frac{\alpha}{4}$$

It is easy to show, that

$$\mu = \frac{\alpha^2}{10}; \quad \gamma = \sqrt{\mu} = \frac{\alpha}{\sqrt{10}} \quad \text{and} \quad \nu = \frac{17}{56}\alpha = 0.3\alpha \quad \text{i.e.}$$

$$\frac{1}{2} > \lambda_1^* > \varrho_1 > \gamma_1 > \nu_1 > q_1 > \mu_1 \tag{4.7}$$

For the case of exponential marginals in (4.6) i.e. $F = 1 - e^{-x}$ and $G = 1 - e^{-y}$ Gumbel [2] has shown that the correlation coefficient has the value

$$r_1 = \frac{\alpha}{4} = q_1$$

In this case

$$\lambda_1^* > \varrho_1 < q_1 = r_1$$

(4.4) Let us now consider the following one parameter family of bivariate distributions:

$$H_2 = \min(F, G)[1 - \alpha(1 - F)(1 - G)] \quad \text{where} \quad 0 \leq \alpha \leq 1 \tag{4.8}$$

5. Approximate values of a two-dimensional *cdf* H in case of positively quadrant-dependence

Let H the joint *cdf* of the pair of random variables X and Y, and let the marginal *cdf*-s F and G respectively. We suppose, that

$$H \geq FG.$$

We shall compare the probability of any quadrant $X < x$, $Y < y$ under the distribution H with the corresponding probability under the distribution $H = \lambda \min(F, G) + (1 - \lambda) FG$ for suitable chosen value of λ.

First of all, we shall determine the value of λ, for which the relation:

$$\varphi(\lambda) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H_\lambda - H)^2 fg \, dx \, dy = \min \tag{5.1}$$

holds.

As $H_\lambda - H = (H_\lambda - FG) - (H - FG)$ the minimum-problem can be written in the following form:

$$\varphi(\lambda) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [(H_\lambda - FG) - (H - FG)]^2 fg \, dx \, dy = \tag{5.2}$$

$$= \lambda^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\min(F, G) - FG]^2 fg \, dx \, dy - 2\lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\min(F, G) - FG] \cdot$$

$$\cdot [H - FG] fg \, dx \, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H - FG)^2 fg \, dx \, dy = \min.$$

Due to (2.4) and (2.8) the equation (5.2) has the following form:

$$\varphi(\lambda) = \frac{\lambda^2}{90} - \frac{2\lambda\nu}{90} + \frac{\mu}{90}. \tag{5.3}$$

The function $\varphi(\lambda)$ takes its minimum if

$$\varphi'(\lambda) = \frac{2\lambda - 2\nu}{90} = 0 \quad \text{i.e. if} \quad \lambda = \nu \tag{5.4}$$

Then

$$\varphi(\nu) = \frac{\nu^2 - 2\nu^2 + \mu}{90} = \frac{\mu - \nu^2}{90}. \tag{5.5}$$

By (2.3)

$$\nu^2 \leq \mu \leq \nu \text{ therefore}$$

$$\varphi(\nu) \leq \frac{\nu - \nu^2}{90} \leq \frac{1}{360} \approx 0.0027. \tag{5.6}$$

From (5.5) it follows that the smaller the difference between μ and ν^2, the better the approximation of H by H_λ. If $H = H_\lambda$, then $\mu = \lambda^2$, $\nu = \lambda$ i.e. $\varphi(\nu) = 0$.

Remark 1.

As $H_\lambda - FG = \lambda\,[\min(F, G) - FG]$ we can say that H_λ keeps the proportion between $\min(F, G)$ and FG.

Let us now introduce the following functions of the random variables X and Y:

$$U(X, Y) = \min\,[F(X), G(Y)] - H(X, Y); \tag{5.7}$$

$$V(X, Y) = H(X, Y) - F(X)\,G(Y);$$

$$Z(X, Y) = \min\,[F(X), G(Y)] - F(X)\,G(Y)$$

If $H = H_\lambda$ $(0 \leq \lambda \leq 1)$ then

$$U_\lambda = (1 - \lambda)Z,\ V_\lambda = \lambda Z \text{ and } U_\lambda = \frac{1 - \lambda}{\lambda} V_\lambda \tag{5.8}$$

i.e. between the random variables U_λ, V_λ and Z there is a linear functional relationship. It follows, that the correlation coefficients between the pairs (U_λ, Z), (V_λ, Z), (U_λ, V_λ) all are equal to 1.

$$r(U_\lambda, Z) = r(V_\lambda, Z) = r(U_\lambda, V_\lambda) = 1 \tag{5.9}$$

Remark 2.

In practical problems the two-dimensional *cdf.* H is usually unknown, but in many cases we may suppose that its marginal *cdf*-s F and G are known. If we have a sample $(x_1 y_1)$, $(x_2, y_2), \ldots (x_n y_n)$ we have the empirical two-dimensional *cdf.* $H_n(x, y)$ and by means of F and G, we have a sample for U, V and Z:

$$U^{(i)} = \min\,[F(X_i), G(Y_i)] - H_n(X_i Y_i),$$

$$V^{(i)} = H_n(X_i, Y_i) - F(X_i)G(Y_i) \text{ and } Z^{(i)} = \min F(x_i)G(y_i) -$$

$$- F(X_i)G(Y_i), \qquad (i = 1, 2, \ldots, n)$$

From this sample we can estimate the correlation coefficients in (5.9) and if their values are close to 1 then we may expect, that the approximation of H by H_λ "good" or even we may accept that the null hipotesis H_0: $H = H_\lambda$ holds.

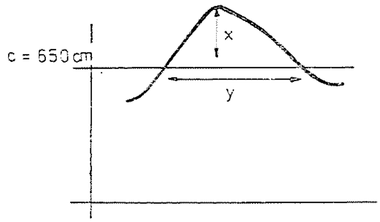

Fig. 4

Let us consider the following example taken from the flood-hydrology.

Example. For the River Tisza in the period 1900—1970 in the second quater of every year (1 Apr.—30 June) above the level $c = 650$ cm the following flood-Peaks were observed.

Table I

Year	X (cm)	Y (day)	Year	X (cm)	Y (day)
1901	29	5	1941	204	68
1902	14	3	1942	38	7
1907	108	42		51	11
1912	72	19		60	14
	34	10			
1914	128	22	1944	4	3
1915	110	35	1952	2	5
1916	73	13	1956	39	10
				37	7
1919	266	49	1958	66	25
1920	16	2	1962	170	33
1922	124	36	1964	114	19
1924	220	51	1965	98	15
1932	273	42	1967	134	41
1937	53	11	1970	309	91
1940	197	38			
	40	8			
	28	5			

Testing the goodness of fit show that the exendance X have the *cdf*: $F(x) = 1 - e^{-0.01x}$ and the duration of floods Y have the

$$cdf\colon G(y) = 1 - e^{-0.05y}$$

For the joint bivariate distribution of the pair (X, Y) the sample was obtained from Table 1.

The value of the correlation coefficient between $V = H_n - FG$ and $Z = \min(F, G) - FG$ is $r(V, Z) \approx 0.9$ so we may accept the validity of hypothesis H_0:

(5.10)
$$H = H_\nu = \nu \min\left[1 - e^{-0,01x}, 1 - e^{-0,05y}\right] + \\ + (1 - \nu)(1 - e^{-0,01x})(1 - e^{-0,05y})$$

Now the estimated value of ν is needed. For the *cdf.* H_ν the value of ν agrees with the value of $q = 4\,H_\nu - 1$. cf. (3.2). The estimation of the value of q is very easy from the sample

$$\hat{q} = 4 \cdot \frac{14}{31} - 1 = 0.8\,.$$

For comparison of the value of H_ν and the empirical *cdf.* H_n let us consider these values in the quartile-points $(\tilde{x}_{1/4}, \tilde{y}_{1/4})$, $(\tilde{x}_{1/2}, y_{1/4})$, ... $(\tilde{x}_{3/4}, y_{3/4})$:

	H	H_n	$(H-H_n)^2$
$(\tilde{x}_{1/4}, \tilde{y}_{1/4})$	0.2125	0.1935	0.000484
$(\tilde{x}_{1/2}, \tilde{y}_{1/4})$	0.225	0.1935	0.000992
$(\tilde{x}_{3/4}, \tilde{y}_{1/4})$	0.225	0.1935	0.000992
$(\tilde{x}_{1/4}, \tilde{y}_{1/2})$	0.2376	0.1935	0.001945
$(\tilde{x}_{1/2}, \tilde{y}_{1/2})$	0.2376	0.1935	0.001945
$(\tilde{x}_{3/4}, \tilde{y}_{1/2})$	0.45	0.4516	0.000000
$(\tilde{x}_{1/4}, \tilde{y}_{3/4})$	0.475	0.4838	0.00007
$(\tilde{x}_{1/2}, \tilde{y}_{3/4})$	0.475	0.4838	0.00007
$(\tilde{x}_{3/4}, \tilde{y}_{3/4})$	0.712	0.68	0.00102

Hence the mean-quadratical derivation between H_n and H_ν is:

$$\frac{\sum_{1}^{9}(H_\nu - H_n)^2}{9} \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (H_\nu - H)^2 fg\,dx\,dy = 0{,}00074\,.$$

In our example above the sample size ($n = 31$) is not large enough for carrying out a test exactly, but the high value of r along with the tabulation heuristically suggests the validity of our inference.

References

1. Blomqvist, N. (1950): On a measure of dependence between two random variables. Ann. Math. Statist. *35*, 138—149.
2. Gumbel, E. J. (1960): Bivariate exponential distributions. J. Amer. Statist. Assoc. *55*, 698—707.
3. Höeffding, W. (1948): A nonparametric test of independence Ann. Math. Statist. *19* 546—557.
4. Konijn, H. S. (1959): Positive and negative dependence of two random variables. Sankhya, *21*, 269—280.
5. Kruskal, W. H. (1958): Ordinal measures of association. American Statist. Association Journal Dec. 1958.
6. Lehmann, E. L. (1966): Some concepts of dependence. Ann. Math. Statist. *37*, 1137—1154.
7. Morgenstern, D. (1956): Einfache Beispiele zweidimensionaler Verteilungen. Mitteilungsblatt für Math. Stat. *8*, 234—5.
8. Rényi, A. (1954): Valószínűségszámítás. Tankönyvkiadó, Budapest. (in Hungarian)

Prof. Dr. József Reimann H-1521 Budapest