The Annals of Statistics
1977, Vol. 5, No. 6, 1221-1227

MONOTONIC DEPENDENCE FUNCTIONS OF BIVARIATE DISTRIBUTIONS

BY T. KOWALCZYK AND E. PLESZCZYŃSKA

Polish Academy of Sciences

A new characterization of monotonic dependence is given here proceeding in a natural way from the consideration of a type of dependence weaker than quadrant dependence. More precisely, each bivariate distribution of (X, Y) is transformed onto a pair of functions $\mu_{X,Y}$ and $\mu_{Y,X}$ defined on the interval $0 < p < 1$ and taking values from $[-1, 1]$, with $\mu_{X,Y}(p)$ being a suitably normalized expected value of X under the condition that Y exceeds its pth quantile. The usefulness of these functions as a kind of measures of the strength of monotonic dependence as well as their close relation to regression functions is demonstrated. It is also suggested that these functions and their sample analogues could serve as useful tools in modelling and solving some statistical decision problems.

1. Definition and interpretation.

The intuitive notion of monotonic dependence between X and Y is that large values of X tend to associate with large values of Y (positive dependence) or with small values of Y (negative dependence). Lehmann (1966) gives three successively stronger definitions of monotonic dependence, starting from quadrant dependence characterized as follows: X is said to be positively quadrant dependent on Y if

$$\forall\, x, y \in R^2 \quad P(X < x \mid Y > y) \leqq P(X < x) \tag{1}$$

and is said to be negatively quadrant dependent on Y if (1) holds with "$\leqq$" replaced by "$\geqq$."

Positive quadrant dependence is equivalent to the property that for any $y \in R$ the random variable X_y with distribution defined by $P(X_y < x) = P(X < x \mid Y > y)$ is stochastically larger than X: $X_y \geqq_{st} X$. Similarly, negative quadrant dependence is equivalent to $X_y \leqq_{st} X$.

The set of all pairs of quadrant dependent real-valued random variables will be denoted here by QD, with obvious notations QD^+ and QD^- corresponding to positive and negative quadrant dependence.

We now introduce a weaker type of monotonic dependence, based on expectations of random variables X_y (provided that these expectations exist): instead of (1), we consider the inequality

$$\forall\, y \in R \quad EX_y \geqq EX\,, \tag{2}$$

with an obvious change of "$\geqq$" on "$\leqq$" in the negative dependence case. The set of all ordered pairs (X, Y) satisfying (2) will be denoted by EQD^+. Then

Received March 1976; revised November 1976.

AMS 1970 *subject classifications*. Primary 62G99; Secondary 62G05, 62H20.

Key words and phrases. Quadrant dependence, regression functions, measures of monotonic dependence, correlation coefficient, robustness, selection.

EQD^+ is a counterpart of QD^+ and it contains all pairs (X, Y) in QD^+ with finite expectations. EQD^- and EQD are defined in a similar way.

For the sake of simplicity, attention is restricted here to the set B of bivariate rv's with finite expectations and continuous marginal distribution functions. Therefore in further considerations the symbols QD, $QD^\pm$, EQD and $EQD^\pm$ will denote suitable subsets of B. To simplify the notation we shall use also these symbols to denote the corresponding sets of distributions.

Let x_p and y_p denote pth quantiles of X and Y, respectively, with $0 < p < 1$. The definition of EQD suggests that a characterization of monotonic dependence between X and Y for $(X, Y) \in B$ can be based on the suitably normalized differences $L_{X,Y}(p) = E(X \mid Y > y_p) - EX$, $0 < p < 1$ (for $(X, Y) \in B$ x_p and y_p are uniquely determined except for p's corresponding to the intervals of constancy of respective marginal distribution functions: thence any pth quantile y_p can be used in the definition of $L_{X,Y}$). Accordingly, we define

$$\begin{aligned}
\mu^+_{X,Y}(p) &= L_{X,Y}(p)/E(X \mid X > x_p) - EX\,, \\
\mu^-_{X,Y}(p) &= L_{X,Y}(p)/(EX - E(X \mid X < x_{1-p}))\,, \\
\mu_{X,Y}(p) &= \mu^+_{X,Y}(p) \qquad \text{if} \quad L_{X,Y}(p) \geqq 0\,, \\
&= \mu^-_{X,Y}(p) \qquad \text{if} \quad L_{X,Y}(p) \leqq 0
\end{aligned} \tag{3}$$

for any $0 < p < 1$ and $(X, Y) \in B$.

The function $\mu_{X,Y}$ will be called a monotonic dependence function (mdf) of X on Y; thus, to every distribution in B there corresponds a pair of mdf's.

The interpretation of $\mu_{X,Y}(p)$ for any chosen $p \in (0, 1)$ is clear. The value $E(X \mid Y > y_p)$ compared with EX characterizes a tendency of values of Y greater than y_p to associate with possibly large values of X. Moreover, for any two (X, Y) and (X', Y') in B which have marginal distributions respectively equal, the tendency can be considered to be stronger for (X, Y) than for (X', Y') if $E(X \mid Y > y_p) > E(X' \mid Y' > y_p)$ which is evidently equivalent to $\mu_{X,Y}(p) > \mu_{X',Y'}(p)$.

Similarly, $\mu_{X,Y}$ as a whole can be considered a measure of strength of monotonic dependence between X and Y in the following sense. For any two elements of B with marginal distributions respectively equal, a positive monotonic dependence between X and Y is said to be stronger than that between X' and Y' (symbolically, $(X, Y) \succ^+ (X', Y')$) if $\mu_{X,Y}(p) \geqq \mu_{X',Y'}(p)$ for all $p \in (0, 1)$, with strict inequalities for some p's. It follows that positive monotonic dependence between X and Y is strongest (related to that of all elements in B which are comparable to (X, Y) with respect to $\succ^+$) iff $\mu_{X,Y}(p) \equiv 1$, and it is weakest iff $\mu_{X,Y}(p) \equiv -1$; thence, by Property 4 proved in Section 2, it is extreme iff X is equivalent to a strictly monotonic function of Y (more precisely, to a Y-a.e. increasing or decreasing function of Y, the exact definitions being given in Section 2). If instead of B we consider EQD^+ then the dependence is weakest iff $\mu_{X,Y}(p) \equiv 0$ or equivalently iff a regression function of X on Y is constant (cf. Property 4).

Some equivalent representations of $\mu^+_{X,Y}$ and $\mu^-_{X,Y}$ can be found by easy transformations of (3), namely

$$\begin{aligned}\mu^+_{X,Y}(p) &= (E(X\mid Y > y_p) - E(X\mid Y < y_p))/(E(X\mid X > x_p) - E(X\mid X < x_p))\\ &= (EX - E(X\mid Y < y_p))/(EX - E(X\mid X < x_p))\end{aligned}$$

and similarly for $\mu^-_{X,Y}(p)$.

2. Properties of $(\mu_{X,Y}, \mu_{Y,X})$. It is assumed throughout this section that $(X, Y) \in B$ and $p \in (0, 1)$.

PROPERTY 1. $-1 \leqq \mu_{X,Y}(p) \leqq 1$.

PROOF. It suffices to show that $\mu^+_{X,Y}(p) \leqq 1$ and $\mu^-_{X,Y}(p) \geqq -1$. The first of these inequalities follows from $E(X\mid Y > y_p) \leqq E(X\mid X > x_p)$ and the second from $E(X\mid Y > y_p) \geqq E(X\mid X \leqq x_{1-p})$.

Let ϕ be the joint distribution function of (X, Y) and let $E(X;\, Y > y_p) = \int_{-\infty}^{\infty}\int_{y_p}^{\infty} x\, d\phi(x, y)$; we shall also use similar notations as $E(X;\, X > x_p)$ and so on.

PROPERTY 2.

$$\begin{aligned}\mu_{X,Y}(p) &= 1 \qquad \text{iff} \quad P(X < x_p,\, Y > y_p) = P(X > x_p,\, Y < y_p) = 0\,,\\ &= -1 \qquad \text{iff} \quad P(X < x_{1-p},\, Y < y_p) = P(X > x_{1-p},\, Y > y_p) = 0\,.\end{aligned}$$

PROOF.

$$\begin{aligned}\mu_{X,Y}(p) = 1 &\Leftrightarrow E(X;\, Y > y_p,\, X < x_p) = E(X;\, Y < y_p,\, X > x_p)\\ &\Leftrightarrow P(X < x_p,\, Y > y_p) = P(X > x_p,\, Y < y_p) = 0\,;\end{aligned}$$

the proof of the second part is similar.

In the following ϕ_Z will denote the distribution function of rv Z. Let $F_Z{}^+$ denote a set of real-valued functions on R called Z-a.e. increasing: $f \in F_Z{}^+$ iff for any $a, b \in R$ $\phi_Z(a) < \phi_Z(b) \Rightarrow f(a) < f(b)$. Similarly a set $F_Z{}^-$ consisting of Z-a.e. decreasing functions is defined.

PROPERTY 3. *For any real a and b, $a \neq 0$*

$$\begin{aligned}\mu_{aX+b,f(Y)}(p) &= (\operatorname{sgn} a)\mu_{X,Y}(p) &&\textit{if}\ \ f \in F_Y{}^+\,,\\ &= (-\operatorname{sgn} a)\mu_{X,Y}(1-p) &&\textit{if}\ \ f \in F_Y{}^-\,.\end{aligned}$$

The proof of Property 3 is obvious.

It is also easily seen that $\mu_{X,Y}$ is continuous.

A notation $\doteq$ will be used in this paper to denote the equivalence of random variables, defined as: $U \doteq V$ if the probability of the event $U = V$ is equal to 1.

PROPERTY 4.

$$\begin{aligned}\mu_{X,Y}(p) &\equiv 1(-1) &&\text{iff} \quad \exists f \in F_Y{}^+(F_Y{}^-) \quad \textit{such that} \quad X \doteq f(Y)\,;\\ &\equiv 0 &&\text{iff} \quad E(X\mid Y) \doteq EX\,;\\ &\equiv \mu^\pm_{X,Y}(p) &&\text{iff} \quad (X, Y) \in EQD^\pm\,.\end{aligned}$$

PROOF.

(i) By Property 2,

$$\mu_{X,Y}(p) \equiv 1 \Leftrightarrow \forall\, p \in (0, 1) \quad P(X < x_p \mid Y < y_p) = 1$$
$$\Leftrightarrow \exists f \in F_Y{}^+ \quad \text{such that} \quad X \doteq f(Y)\,.$$

A similar result for $\mu_{X,Y}(p) \equiv -1$ follows from Property 3.

(ii) $$\mu_{X,Y}(p) \equiv 0 \Leftrightarrow \forall\, p \in (0, 1) \quad \textstyle\int_{y_p}^{\infty} E(X \mid Y = t)\, d\phi_Y(t) = \int_{y_p}^{\infty} EX\, d\phi_Y(t)$$
$$\Leftrightarrow E(X \mid Y) \doteq EX\,.$$

The third equality appearing in the thesis follows from the definitions of EQD^+ and EQD^-.

PROPERTY 5. *If* $(X, Y) \in QD^+$ *and* $EX < \infty$, $EY < \infty$ *then* $\mu_{Y,X} \geqq 0$.

PROOF. Immediate.

According to Property 5, both mdf's are nonnegative in case of a distribution from QD^+. It is clear that they are both nonpositive in case of a distribution from QD^-.

PROPERTY 6. *If* $(X, Y) \in QD$ *then* $\mu_{X,Y}(p) \equiv 0$ iff X *and* Y *are independent.*

PROOF. It is known (Lehmann (1966)) that for $(X, Y) \in QD$ X and Y are independent iff Cov $(X, Y) = 0$. Thence Property 6 follows from Property 4.

THEOREM 1. *Let* (X, Y), $(X', Y') \in B$ *and* $\phi_X = \phi_{X'}$, $\phi_Y = \phi_{Y'}$. *Then* $\mu_{X,Y} = \mu_{X',Y'}$ iff $E(X \mid Y)$ *and* $E(X' \mid Y')$ *have the same distribution.*

PROOF.

$$\mu_{X,Y} = \mu_{X',Y'} \Leftrightarrow \forall\, p \in (0, 1) \quad \textstyle\int_{y_p}^{\infty} E(X \mid Y = t)\, d\phi_Y(t) = \int_{y_p}^{\infty} E(X' \mid Y' = t)\, d\phi_{Y'}(t)\,,$$

which was to be proved.

For any $(U, V) \in B$, let $\gamma_{U,V}$ be a function in $F_V{}^+$ defined by $\gamma_{U,V}(a) = \max\{b \in R : \phi_U(b) = \phi_V(a)\}$. Given (X, Y), consider a random variable $Y^* = \gamma_{X,Y}(Y)$ which obviously has the same distribution as X. Then by Theorem 1 $\mu_{X,Y^*} = \mu_{Y^*,X}$ iff $E(X \mid Y^*)$ and $E(Y^* \mid X)$ have the same distribution.

Theorem 1 establishes some connections between mdf's and regression functions in the general case. Theorem 2 below specifies these connections in some important special case.

Let ρ be any number in $[-1, 1]$. Given (X, Y), define

$$\begin{aligned} Y_\rho &= Y^* \quad (\text{i.e., } \gamma_{X,Y}(Y)) && \text{for} \quad \rho > 0\,, \\ &= Y && \text{for} \quad \rho = 0\,, \\ &= \gamma_{-X,Y}(Y) && \text{for} \quad \rho < 0\,, \end{aligned}$$

so that the distribution of Y_ρ is the same as that of $(\operatorname{sgn} \rho)X$ for $|\rho| > 0$.

THEOREM 2. *For any* $\rho \in [-1, 1]$,

$$\mu_{X,Y}(p) \equiv \rho \qquad \text{iff} \quad E(X \mid Y_\rho) \doteq \rho Y_\rho + (1 - |\rho|)EX\,.$$

PROOF. Obviously, it suffices to prove Theorem 2 under the assumption that $EX = 0$.

Let $y_{\rho,p}$ denote pth quantile of Y_ρ. Assume first that $\rho > 0$. Then $E(X \mid Y_\rho) \doteq \rho Y_\rho$ implies that $\forall p \in (0, 1)$ $\operatorname{sgn} L_{X,Y}(p) = 1$. Hence $\mu_{X,Y}(p) = \mu_{X,Y_\rho}(p) = E(X \mid Y_\rho > y_{\rho,p})/E(X \mid X > x_p) = E(\rho Y_\rho \mid Y_\rho > y_{\rho,p})/E(X \mid X > x_p) = \rho$. On the other hand,

$$\begin{aligned}\mu_{X,Y}(p) \equiv \rho &\Rightarrow \mu^+_{X,Y}(p) \equiv \mu^+_{X,Y_\rho}(p) \equiv \rho \\ &\Rightarrow \forall p \in (0, 1) \quad E(X \mid Y_\rho > y_{\rho,p}) = E(\rho X \mid X > x_p) \\ &\Rightarrow \forall p \in (0, 1) \int_{y_{\rho,p}}^\infty E(X \mid Y_\rho = t)\, d\phi_{Y_\rho}(t) = \int_{x_p}^\infty \rho t\, d\phi_X(t) \\ &\Rightarrow E(X \mid Y_\rho) \doteq \rho Y_\rho .\end{aligned}$$

For $\rho < 0$ the proof is analogous while for $\rho = 0$ Theorem 2 follows from Property 4.

A necessary and sufficient condition for $\mu_{X,Y}(p) \equiv \mu_{Y,X}(p) \equiv \rho$ for any $\rho \in [-1, 1]$ can be easily deduced from Theorem 2. In particular

$$\begin{aligned}\mu_{X,Y}(p) \equiv \mu_{Y,X}(p) \equiv 1(-1) \quad & \text{iff} \quad \exists f \in F_Y{}^+(F_Y{}^-) \text{ such that } X \doteq f(Y) ; \\ \equiv 0 \quad & \text{iff} \quad E(X \mid Y) \doteq EX \wedge E(Y \mid X) \doteq EY ; \\ \equiv \rho \quad & \text{if the distribution of } (X, Y) \text{ is bivariate} \\ & \text{normal with the correlation coefficient equal} \\ & \text{to } \rho .\end{aligned}$$

COROLLARY 1. *Let (X, Y) be any element of B with a correlation coefficient $r_{X,Y}$ existing and equal to $\rho \neq 0$. Then $\mu_{X,Y}(p) \equiv \rho$* iff *$E(X \mid Y)$ is a linear function of Y and the distributions of standardized variables* (sgn ρ)X *and Y are identical.*

PROOF. It is easily seen that it is sufficient to prove Corollary 1 for standardized X and Y. By Theorem 2, $\mu_{X,Y}(p) \equiv r_{X,Y} = \rho$ implies that $E(X \mid Y_\rho)$ is a linear function of Y_ρ and that $r_{X,Y} = r_{X,Y_\rho}$. Hence $r_{X,Y} = E(YE(X \mid Y)) = \rho E(YY_\rho) \Rightarrow E(YY_\rho) = 1 \Rightarrow Y_\rho \doteq Y$ in view of the properties of the correlation coefficient.

The proof of sufficiency is immediate.

To illustrate the discrepancies between $\mu_{X,Y}(p)$ and $r_{X,Y}$ for various p's when $\mu_{X,Y}(p) \not\equiv r_{X,Y}$, consider a bivariate exponential distribution belonging to QD^-, with the distribution function ϕ defined as $\phi(x, y) = 1 - e^{-x} - e^{-y} + e^{-(x+y+xy)}$; $x \geqq 0, y \geqq 0$. It is easy to check that $r_{X,Y} = -.40365$ and

$$\mu_{X,Y}(p) = ((p - 1)\ln(1 - p))/(p \ln p(1 - \ln(1 - p))) .$$

Hence $\mu_{X,Y}$ is a decreasing function with negative values, convergent to 0 for $p \to 0^+$ and to -1 for $p \to 1^-$.

3. Discussion. An intuitive notion of monotonic dependence (which could possibly be called signed dependence: positive or negative) has been formalized in many ways in case of two or more dimensions (cf. for instance [1], [4], [5], [6], [9], [11], [12]), and a couple of parameters have been defined to serve as a measure of monotonic dependence (inter alia, it is worth mentioning here a survey of ordinal measures given by W. Kruskal (1958) and a survey of robust measures of correlation given by Devlin, Gnanadesikan and Kettenring (1975)).

Comparing $\mu_{X,Y}$ with other measures it should be emphasized that $\mu_{X,Y}$ is a function and in general takes on different values for different values of p. Corollary 1 states the conditions under which $\mu_{X,Y}$ can be replaced by the correlation coefficient $r_{X,Y}$ without any loss of information.

Another feature which differentiates $\mu_{X,Y}$ from other measures of monotonic dependence is its "hybrid" nature expressed by the fact that $\mu_{X,Y}$ is invariant under increasing transformations of Y and linearly increasing transformations of X, while ordinal measures of association are invariant under increasing transformations performed on both variables, and "interval" measures of association (e.g., correlation coefficient) are invariant under linearly increasing transformations on both variables. The importance of the hybrid nature follows from the fact that there exist practical situations in which metric is relevant just in the case of one variable (cf. J. B. Kruskal (1964)).

As stated before in Section 2, the assumption of continuity of marginal distributions has been adopted in this paper mostly for the sake of simplicity and convenience. An extension of the definition of $\mu_{X,Y}$ for all bivariate distributions with finite expectations can be easily constructed if one follows the idea given in Pleszczyńska (1970) where an archetype of $\mu_{X,Y}(.5)$ has been considered. Moreover, a modified more robust definition of $\mu_{X,Y}$ could be suggested, with expectations replaced by robust measures of location suitable in nonparametric and nonsymmetric case (see Bickel and Lehmann (1975)).

The authors believe that $\mu_{X,Y}$ and its multidimensional analogues will prove useful in various selection, discrimination and regression problems. Respective decision schemes are now being investigated.

Acknowledgment. We express our sincere thanks to the referee for his helpful comments.

REFERENCES

[1] ALAM, K. and WALLENIUS, K. T. (1976). Positive dependence and monotonicity in conditional distributions. *Comm. Statist. Theor. Math.* **A5(6)** 525–534.

[2] BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models. II. Location. *Ann. Statist.* **3** 1045–1069.

[3] DEVLIN, S. J., GNANADESIKAN, R. and KETTENRING, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika* **62** 531–545.

[4] ESARY, J. D. and PROSCHAN, F. (1972). Relationships among some concepts of bivariate dependence. *Ann. Math. Statist.* **43** 651–655.

[5] ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables with applications. *Ann. Math. Statist.* **38** 1466–1474.

[6] JOGDEO, KUMAR (1968). Characterizations of independence in certain families of bivariate and multivariate distributions. *Ann. Math. Statist.* **39** 433–441.

[7] KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29** 1–27.

[8] KRUSKAL, W. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **49** 732–64.

[9] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.

[10] PLESZCZYŃSKA, E. (1970). A sample coefficient of monotonic dependence, II. *Zastos. Mat.* **3** 281–297 (in Polish). (English translation in *Selected Transl. Math. Statist. Prob.* **13** 271–289.)

[11] SHAKED, M. (1975). On concepts of dependence for multivariate distributions. Ph.D. dissertation, Univ. of Rochester.

[12] YANAGIMOTO, T. (1972). Families of positively dependent random variables. *Ann. Inst. Statist. Math.* **24** 559–573.

MATHEMATICAL INSTITUTE
POLISH ACADEMY OF SCIENCES
SNIADECKICH 8, 00-950
WARSZAWA, POLAND

COMPUTATION CENTRE
POLISH ACADEMY OF SCIENCES
BOX 22, 00-901 WARSZAWA PKIN
POLAND