MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

ON CONCEPTS AND MEASURES OF BIVARIATE STOCHASTIC DEPENDENCE

T. KOWALCZYK, A. MATUSZEWSKI, A. NALBACH-LENIEWSKA,
E. PLESZCZYŃSKA

Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

1. Introduction

Functional monotonic dependence is considered as important special case of functional dependence between two variates, while linear functional dependence is a special case of monotonic dependence. Similar gradation seems to exist in the case of stochastic dependence between two random variables X and Y. It should be reflected in some way by suitably defined measures of global, monotonic and linear dependence such that, loosely speaking, global measures reduce to monotonic ones under monotonic models and to linear ones under linear models. This natural idea is confused by an invasion of various concepts of stochastic dependence and their measures appearing in the statistical literature. Global and monotonic dependence is occasionally referred to as connection and concordance (cf. Kruskal [5]). Monotonic dependence is also sometimes called positive-negative or signed dependence; it expresses the fact that large values of X tend to associate with large (or, alternatively, small) values of Y.

The paper aims at directing the attention of the reader towards these problems, mainly by means of some functional measures of general, monotonic and linear dependence which have a special interpretation in terms of "goodness of screening". These measures are considered in Section 2 and they provide an intuitive basis for the discussion given in Section 3. It should be stressed that the discussion is only intended to stimulate further investigations since the problem is far from being exhausted or even satisfactorily approached.

2. Measures of dependence referring to screening

Let us start with specifying certain models of monotonic and linear stochastic relationship between random variables X and Y. These models are all contained in the set C of all random pairs (X, Y) such that the distribution of X is non-degenerated and EX is finite.

Let MRF denote the set of random pairs consisting of all elements of C such that the regression function of X on Y is strictly monotone; the symbol "MRF" refers to "monotonic regression function". MRF is the sum of disjoint subsets MRF^+ and MRF^- corresponding to increasing and decreasing regression functions.

Monotonicity of regression functions of X on Y is a special formalization of the intuitive notion of monotonic dependence of X on Y, and MRF is an example of models which were referred to in the introduction as "monotonic models".

Let H denote the set of all increasing functions $\boldsymbol{R} \to \boldsymbol{R}$ and let Q denote the set of (X, Y)'s such that there exists an $h \in H$ for which the distribution of $h(Y)$ is equal to that of X or $-X$ (this means that one of the boundary distributions for (X, Y) is concentrated on $\{(x, y)\colon x = h(y)\}$ or on $\{(x, y)\colon -x = h(y)\}$, the notion of boundary distributions being given e.g. in Mardia [7]). For any $h \in H$, we denote by $\mathrm{LRF}(h)$ a subset of MRF such that the regression function of X on $h(Y)$ is linear. Obviously, any $\mathrm{LRF}(h)$ can also be partitioned onto $\mathrm{LRF}^+(h)$ and $\mathrm{LRF}^-(h)$.

Important special cases of $\mathrm{LRF}(h)$'s are those with h's linearly increasing. The sum of all such sets will be denoted here by SLD ("strong linear dependence"). $(X, Y) \in \mathrm{SLD}$ implies that the regression function of X on Y is linear but the converse is not always true: to be strongly linearly dependent on Y, X must fulfil the condition that h is linear, i.e. at least one of the boundary distributions is concentrated on a line. Strong linear dependence could be exemplified by binormal random vectors.

We have thus defined a sequence of models $\mathrm{SLD} \subset \mathrm{MRF} \subset C$ and a family of models $\mathrm{LRF}(h) \subset \mathrm{MRF}$ such that MRF is a formalization of stochastic monotonic dependence of X on Y while $\mathrm{LRF}(h)$ and SLD are formalizations of stochastic linear dependence of X on $h(Y)$ and of X on Y. Now we turn to some function measures of dependence referring to screening (cf. Kowalczyk, Kowalski, Matuszewski, Pleszczyńska [4]).

Suppose that items of some population are characterized by values (x, y) of an unobservable variable X and an observable variable Y. Screening is defined as rejecting a fraction p ($p \in (0, 1)$) of items with possibly small values of X on the basis of values of Y. For any $p \in (0, 1)$, we shall say that a function $s_{p,Y}\colon \boldsymbol{R} \to [0, 1]$ is a *p-screening* function if it is measurable and $Es_{p,Y}(Y)$ exists and is equal to $1-p$. The value $s_{p,Y}(y)$ for any $y \in \boldsymbol{R}$ is interpreted as the probability of acceptance of an item characterized by (x, y) for any $x \in \boldsymbol{R}$. Hence screening under $s_{p,Y}$ transforms the original random variable X onto $X|s_{p,Y}$ interpreted as X in the population of accepted items.

$EX|s_{p,Y}$ corresponds to the goodness of screening under $s_{p,Y}$. Here, the family of $EX|s_{p,Y}$ for all possible p's and p-screening functions is used to construct a measure of dependence of X on Y in the following way:

For any random variable U, any $p \in (0, 1)$ and any pth quantile u_p of U, let $t_{p,U}\colon \boldsymbol{R} \to [0, 1]$ be given by

$$t_{p,U}(a) = \begin{cases} 0 & \text{if} \quad a < u_p, \\ \gamma_{p,U} & \text{if} \quad a = u_p, \\ 1 & \text{if} \quad a > u_p, \end{cases}$$

where

$$\gamma_{p,U} = \begin{cases} \dfrac{1-p-P(U > u_p)}{P(U = u_p)} & \text{if} \quad P(U = u_p) > 0, \\ 0 & \text{otherwise}. \end{cases}$$

Obviously, $t_{p,U}$ is a p-screening function based on U. It will be called a *p-truncation* function.

For $(X, Y) \in C$, the expression $\big(E(X|s_{p,Y})-EX\big)/\big(E(X|t_{p,X})-EX\big)$ is well-defined and does not exceed 1. The closer it is to 1, the better is $s_{p,Y}$ as a p-screening function. Let $S_{p,Y}$ denote the set of all p-screening functions related to Y. Now, for any $(X, Y) \in C$ we define a screening dependence function $\nu_{X,Y}\colon (0, 1) \to [0, 1]$ such that for any $p \in (0, 1)$

$$\nu_{X,Y}(p) = \sup_{s_{p,Y} \in S_{p,Y}} \big(E(X|s_{p,Y})-EX\big)/\big(E(X|t_{p,X})-EX\big). \tag{2.1}$$

It was shown in Kowalczyk et al. [4] that

$$\nu_{X,Y}(p) \equiv \big(E(X|t_{p,EX|Y})-EX\big)/\big(E(X|t_{p,X})-EX\big), \tag{2.2}$$

which means that for any $p \in (0, 1)$ $t_{p,EX|Y}$ is the best p-screening function concerning X and based on Y. Consequently, $E(X|t_{p,EX|Y})$ is a mixture of $E(X|t_{p,X})$ (i.e. the upper bound available only when X is observable) and of EX, with coefficients $\nu_{X,Y}(p)$ and $1-\nu_{X,Y}(p)$.

Screening dependence function $\nu_{X,Y}$ is a function-valued measure of global dependence of X on Y, defined for $(X, Y) \in C$ and interpretable in terms of screening. Some properties of $\nu_{X,Y}$ are stated in Kowalczyk et al. [4]. We shall quote here only one of them, using the notation "a.e. Y" to replace "almost everywhere with respect to the distribution of Y":

$$\nu_{X,Y}(p) = \begin{cases} 1 & \text{iff} \quad \text{there exists an } f\colon \boldsymbol{R} \to \boldsymbol{R} \text{ such that} \\ & \qquad (X, f(Y)) \in C \text{ and } X = f(Y) \text{ a.e. } Y, \\ 0 & \text{iff} \quad EX|Y = EX \text{ a.e. } Y. \end{cases} \tag{2.3}$$

Now, let us turn to an analogous measure $\mu_{X,Y}$ of monotonic dependence, called a *monotonic dependence function*, which maps $(0, 1)$ onto $[-1, 1]$ for any $(X, Y) \in C$. It was first introduced in Kowalczyk and Pleszczyńska [3] for a subset of C and subsequently extended to C in Kowalczyk [2]. Using truncation functions, we define $\mu_{X,Y}(p)$ for any $p \in (0, 1)$ by

$$\mu_{X,Y}(p) \stackrel{\mathrm{df}}{=} \begin{cases} \mu^{+}_{X,Y}(p) & \text{if} \quad \mu^{+}_{X,Y}(p) \geqslant 0, \\ -\mu^{+}_{-X,Y}(p) & \text{otherwise}, \end{cases} \tag{2.4}$$

where

$$\mu^{\pm}_{X,Y}(p) \stackrel{\text{df}}{=} (E(X|t_{p,Y})-EX)/(E(X|t_{p,X})-EX). \tag{2.5}$$

Obviously,

$$\mu_{X,Y} = \begin{cases} \mu^{\pm}_{X,Y} > 0 & \text{for} \quad (X, Y) \in \mathrm{MRF}^{+}, \\ -\mu^{\pm}_{-X,Y} < 0 & \text{for} \quad (X, Y) \in \mathrm{MRF}^{-}, \end{cases} \tag{2.6}$$

and, as shown in Kowalczyk [2],

$$\mu_{X,Y}(p) \equiv \begin{cases} 1\ (-1) & \text{iff} \quad \text{there exists a nondecreasing (nonincreasing)} \\ & \qquad \text{mapping } f\colon \boldsymbol{R} \to \boldsymbol{R} \text{ such that } X = f(Y) \text{ a.e. } Y, \\ 0 & \text{iff} \quad EX|Y = EX \text{ a.e. } Y. \end{cases} \tag{2.7}$$

By (2.2), (2.4) and (2.5), the screening dependence function of X on Y is equal to the monotonic dependence function of X on $EX|Y$:

$$\nu_{X,Y} = \mu_{X,EX|Y}. \tag{2.8}$$

We are interested in relations between $\mu_{X,Y}$ and $\nu_{X,Y}$. Obviously,

$$-\nu_{-X,Y} \leqslant \mu_{X,Y} \leqslant \nu_{X,Y}, \tag{2.9}$$

where $\forall p \in (0, 1)$ $\nu_{-X,Y}(p) = \nu_{X,Y}(1-p)$. Furthermore,

$$\mu_{X,Y} = \begin{cases} \nu_{X,Y} & \text{for} \quad (X, Y) \in \mathrm{MRF}^{+}, \\ -\nu_{-X,Y} & \text{for} \quad (X, Y) \in \mathrm{MRF}^{-}. \end{cases} \tag{2.10}$$

Referring to the latter formula, we shall say that (X, Y) *satisfies the condition* "$\mu = \nu$" if $\mu_{X,Y}$ is equal either to $\nu_{X,Y}$ or to $-\nu_{-X,Y}$. Hence any $(X, Y) \in \mathrm{MRF}$ satisfies "$\mu = \nu$" but the converse is not true. Obviously, the condition "$\mu = \nu$" expresses rather strong monotonicity existing between X and Y and can be fulfilled only when $\mu_{X,Y}$ is of constant sign. It is intuitively evident that in most cases $\mu_{X,Y}$ is not of constant sign when X is not monotonically dependent on Y, e.g. in the case of a U-shaped distribution.

Let $\preccurlyeq^{\nu}$ and $\preccurlyeq^{\mu}$ be defined on C by

$$\begin{aligned} (X, Y) \preccurlyeq^{\nu} (X', Y') \quad &\text{iff} \quad \forall p \in (0, 1) \nu_{X,Y}(p) \leqslant \nu_{X',Y'}(p), \\ (X, Y) \preccurlyeq^{\mu} (X', Y') \quad &\text{iff} \quad \forall p \in (0, 1) |\mu_{X,Y}(p)| \leqslant |\mu_{X',Y'}(p)| \end{aligned} \tag{2.11}$$

for any $(X, Y), (X', Y') \in C$. Then, for any such (X, Y) and (X', Y') satisfying "$\mu = \nu$",

$$(X, Y) \preccurlyeq^{\nu} (X', Y') \quad \text{iff} \quad (X, Y) \preccurlyeq^{\mu} (X', Y'). \tag{2.12}$$

Finally, we turn to a certain real-valued measure of linear stochastic dependence of X on Y, namely to the correlation coefficient $\varrho_{X,Y}$ defined on the subset of C containing distributions with finite non-zero second central moments. Here, it will be convenient to extend the definition of $\varrho_{X,Y}$ to any $(X, Y) \in \mathrm{SLD}$: whenever $(X, Y) \in \mathrm{SLD}$ and second moments are not finite (e.g., for a bivariate t-Student distribution with two degrees of freedom), ϱ in the formula

$$EX|Y = \varrho h(Y) + (1-|\varrho|)EX \tag{2.13}$$

will be interpreted as $\varrho_{X,Y}$. Hence in the sequel we shall use the symbols $\varrho_{X,h(Y)}$ in the extended sense for any $(X, Y) \in \bigcup_{h\in H} \mathrm{LRF}(h)$ (since $(X, Y) \in \mathrm{LRF}(h)$ implies $(X, h(Y)) \in \mathrm{SLD}$).

It was shown in Kowalczyk [2] that for any $(X, Y) \in C \cap Q$

$$\mu_{X,Y}(p) \equiv \begin{cases} \varrho_{X,h(Y)} & \text{iff} \quad (X, Y) \in \mathrm{LRF}(h), \\ \varrho_{X,Y} & \text{iff} \quad (X, Y) \in \mathrm{SLD}, \end{cases} \tag{2.14}$$

and for $(X, Y) \notin \bigcup_{h\in H} \mathrm{LRF}(h)$ $\mu_{X,Y}$ is not constant. Moreover, by Kowalczyk *et al.* [4], for any $(X, Y) \in C \cap Q$

$$\nu_{X,Y}(p) \equiv |\varrho_{X,h(Y)}| \quad \text{iff} \quad (X, Y) \in \mathrm{LRF}(h). \tag{2.15}$$

Evidently, the latter statements provide some new interpretation of $\varrho_{X,h(Y)}$ for $(X, Y) \in \bigcup_{h\in H} \mathrm{LRF}(h)$, namely that referring to screening.

Let $\preccurlyeq^{\varrho}$ be defined by $(X, Y) \preccurlyeq^{\varrho} (X', Y')$ iff $|\varrho_{X,Y}| \leqslant |\varrho_{X',Y'}|$. Then, for any $(X, Y), (X', Y') \in \mathrm{SLD}$,

$$(X, Y) \preccurlyeq^{\nu} (X', Y') \Leftrightarrow (X, Y) \preccurlyeq^{\mu} (X', Y') \Leftrightarrow (X, Y) \preccurlyeq^{\varrho} (X', Y'). \tag{2.16}$$

It is known that $\varrho_{X,h(Y)}$ is equal to 1 (-1) iff the distribution of (X, Y) is concentrated on $\{(x, y)\colon x = h(y)\}$ or $\{(x, y)\colon -x = h(y)\}$, i.e. is equal to one of the boundary distributions. Specifically, $|\varrho_{X,Y}|$ is equal to 1 iff the distribution of (X, Y) is concentrated on a line.

3. Discussion

Let us start with a brief recapitulation of Section 2. Three different aspects of stochastic dependence of X on Y, global, monotonic and linear, were considered, measured respectively by ν, μ and ϱ. The sets of arguments and values of the three measures were different. Partial orderings $\preccurlyeq^{\nu}$, $\preccurlyeq^{\mu}$ and $\preccurlyeq^{\varrho}$ were introduced reflecting the corresponding partial orderings of the sets of values. The maximum elements of the sets of arguments according to these ordering relations were shown to be those with distributions concentrated on the graph of a certain arbitrary, monotonic and linear function, respectively; these elements belong, respectively, to C, MRF and SLD and achieve extreme values of ν, μ and ϱ. The minimum elements were shown to be those with constant regression functions of X on Y in the case of $\preccurlyeq^{\nu}$ and $\preccurlyeq^{\mu}$ and with uncorrelated X and Y in the case of $\preccurlyeq^{\varrho}$. Moreover, the three measures were "consistent" on appropriately chosen subsets of arguments according to the following general definition:

Let $(\mathscr{X}, \Sigma)$ be any measurable space and let K be the set of all probability measures on Σ. Let B_1 and B_2 be subsets of K, $B_1 \cap B_2 \neq \emptyset$. For $i = 1, 2$, let $\varphi_i\colon B_i \to \Gamma_i$ be a measure of some aspect of B_i, and let $\preccurlyeq^{B_i}$ and $\preccurlyeq^{\Gamma_i}$ be partial orderings of B_i and Γ_i, respectively, such that for any $P, P' \in B_i$

$$P \preccurlyeq^{B_i} P' \Rightarrow \varphi_i(P) \preccurlyeq^{\Gamma_i} \varphi_i(P').$$

Now, suppose that the two aspects measured by φ_1 and φ_2 are considered to be equivalent on some subset $A \subset B_1 \cap B_2$. Then we require φ_1 and φ_2 to be consistent on A, namely

(i) for any $P, P' \in A$

$$\varphi_1(P) \leqslant^{\Gamma_1} \varphi_1(P') \Leftrightarrow \varphi_2(P) \leqslant^{\Gamma_2} \varphi_2(P'),$$

(ii) there exists $\alpha\colon \varphi_1(A) \overset{\text{onto}}{\leftrightarrow} \varphi_2(A)$ such that for any $P \in A$ $\alpha \circ \varphi_1(P) = \varphi_2(P)$.

In view of (2.6), (2.10), (2.12) and (2.16), ν is consistent with μ on MRF$^+$ and on MRF$^-$ and is consistent with ϱ on SLD$^+$ and on SLD$^-$ while μ is consistent with ϱ on SLD. This illustrates possible formalizations of "reducibility of global measures to monotonic ones under monotonic models and to linear ones under linear models", which was vaguely postulated in the introduction.

However, the triple of measures ν, μ and ϱ, and the triple of models C, MRF and SLD are only illustrations of models and measures which could be introduced, and relations $\preccurlyeq^{\mu}$ and $\preccurlyeq^{\nu}$ are only illustrations of appropriate partial orderings.

The general problem of a synthetic description of some aspect of a set of distributions is very complicated. Bickel and Lehmann [1] suggested that the postulates concerning the aspects in question should deal in the first place with a partial ordering of the set of distributions, indicating cases in which one distribution possesses the attribute under consideration more strongly than another one. A real-valued measure of this attribute should then preserve this ordering. Further conditions considered by Bickel and Lehmann were that of invariance and of the effectivity with which the value of the measure can be estimated from a sample. Finally, the authors postulated that in the case of the existence of a "natural" real-valued parameter characterizing the given attribute for a subset of arguments, a measure should reduce to the parameter in question on this subset.

The latter requirement is related to that of the consistency of two measures characterizing two different aspects equivalent on the embedded model.

Bickel and Lehmann considered only real-valued measures. It seems to us, however, that when dealing with more complicated aspects of bivariate distributions, such as shape, monotonic dependence and so on, one has to use descriptive statistics with "richer" sets of values. Moreover, partial orderings seem insufficient. It seems that the whole problem should be approached by appropriately developed methods of measurement theory, such as proceeding from relation systems concerning sets of distributions to "simpler" relation systems, not necessarily numerical ones. Among the relations, that of "closeness" should be involved.

We close the discussion with a short reference to a very interesting approach to bivariate dependence problems proposed by Lissowski [6]. The main idea is to construct real-valued measures of dependence interpretable in terms of goodness of prediction of the value of unobservable X exploring the value of observable Y. Using some oversimplification, Lissowski proposes to consider non-negative loss functions describing the consequences of wrong prediction and to attach to any bivariate distribution and any loss function the normalized difference of expected losses under the "optimal" decision rule based on Y and under the "optimal" decision rule based on X. The choice of the loss function corresponds to the choice of the "type of dependence" of X on Y. This approach has made it possible to classify and provide a new interpretation for nearly all existing real-valued measures of dependence and has inspired some practical applications.

There are strong analogies between the above approach and our research on measures of dependence interpretable in terms of screening. The decision problem of predicting the value of unobservable X basing on observable Y is parallel to that of screening a fraction of values of unobservable X on the basis of Y.

A synthetic description of various types of stochastic dependence is certainly one of the most important problems for statistical theory and applications. It is to be stressed that it is the applicability of the various descriptions that should be the decisive criterion, and therefore many measures are needed to deal with particular situations. What causes anxiety at the present stage is the disproportion between the progress of inventing new measures and the progress of theoretical considerations on the subject. The latter are urgently needed.

References

[1] P. J. Bickel, E. L. Lehmann, *Descriptive statistics for nonparametric models. I. Introduction*, Ann. Statist. 3 (1975), pp. 1038–1041.

[2] T. Kowalczyk, *General definitions and sample counterparts of monotonic dependence functions of bivariate distributions*, Math. Operationsforschung Statist. 8 (1977), pp. 351–365.

[3] T. Kowalczyk, E. Pleszczyńska, *Monotonic dependence functions of bivariate distributions*, Ann. Statist. 5 (1977), pp. 1221–1227.

[4] T. Kowalczyk, A. Kowalski, A. Matuszewski, E. Pleszczyńska, *Screening and monotonic dependence functions in the multivariate case*. Ann. Statist. 7 (1979), pp. 607–614.

[5] W. Kruskal, *Ordinal measures of association*, J. Amer. Statist. Assoc. 53 (1958), pp. 814–861.

[6] G. Lissowski, *Statistical association and prediction*, in: *Problems of formalization of the social sciences*, Ossolineum, Wrocław 1977.

[7] K. V. Mardia, *Families of bivariate distributions*, Hafner, Darien Conn.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976