ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XVI, 4 (1980)

TERESA KOWALCZYK, A. KOWALSKI, A. MATUSZEWSKI
and E. PLESZCZYŃSKA (Warszawa)

APPLICATIONS OF MONOTONIC DEPENDENCE FUNCTIONS: DESCRIPTIVE ASPECTS

1. INTRODUCTION

The importance of good descriptive statistics dealing with the strength of monotonic dependence of two random variables and with quantifying the form of a bivariate distribution is easily recognized by applied statisticians. Descriptive methods with impressive graphical representations are specially useful, since — as was remarked by Gnanadesikan and Wilk in [4] — "man is a geometrical animal and seems to need and want pictures for parsimony and to stimulate insight".

Monotonic dependence functions of bivariate distributions and their sample counterparts, introduced by Kowalczyk and Pleszczyńska in [5] and [6], are hoped to prove a good graphical device and an important tool of statistical data processing. Before recalling the definition, let us build up some intuitions by means of the following example.

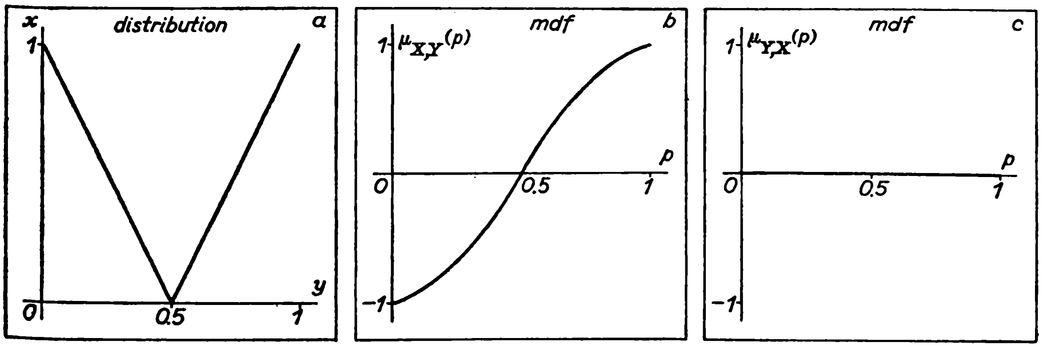

Fig. 1. V-shaped distribution and its mdf's

Suppose that the distribution of (X, Y) is concentrated so that it forms the shape of the letter V inscribed into the unit square (Fig. 1), both marginal distributions being uniform. It follows that, for any $p \in (0, 1)$, the p-quantiles of X and Y, say x_p and y_p, are both equal to p. With y's (or p's) increasing, the values of X decrease and increase in turn. Then one can look for a function, say $\mu_{X,Y}$, which maps $(0,1)$ into $[-1,1]$ and which takes negative (positive) values for $p < \frac{1}{2}$ $(p > \frac{1}{2})$, such that for any p the function $\mu_{X,Y}(p)$ measures in some way the tendency of large (small) values of X to associate with possibly large (small) values of Y. This function could be as that one represented by the graph in Fig. 1. A similar function for the ordered pair (Y, X) should identically be equal to zero, since Y is not monotonically related to X in any intuitive sense. The discrepancies between $\mu_{X,Y}$ and $\mu_{Y,X}$ reflect the asymmetry of the distribution.

Similar intuitive considerations can be repeated for any bivariate distribution, with p's corresponding to p-quantiles of the respective random variable. There is a lot of various possibilities of formalizing $\mu_{X,Y}$; one — chosen by Kowalczyk and Pleszczyńska [6] in the case of random variables with continuous marginal distribution functions — was to define $\mu_{X,Y}(p)$ as a suitably normalized expected value of X under the condition that Y exceeds its p-quantile. This conditional random variable is denoted here by $X\,|\,t_{p,Y}$, where $t_{p,Y}$ is the characteristic function of the set $\{y\colon y > y_p\}$. Thus $X\,|\,t_{p,Y}$ is a function of (X, Y) given in the form $(1-p)^{-1}Xt_{p,Y}(Y)$. Similarly we can introduce $X\,|\,t_{p,X}$, where $t_{p,X}$ is the characteristic function of $\{x\colon x > x_p\}$. Now, $\mu_{X,Y}$, called a *monotonic dependence function* (mdf) of X on Y, is defined by the formula

$$\mu_{X,Y}(p) = \begin{cases} \mu^{+}_{X,Y}(p) & \text{for } \mathrm{E}(X\,|\,t_{p,Y}) \geqslant \mathrm{E}X, \\ -\mu^{+}_{-X,Y}(p) & \text{otherwise,} \end{cases}$$

where

$$\mu^{+}_{X,Y}(p) = \frac{\mathrm{E}(X\,|\,t_{p,Y}) - \mathrm{E}X}{\mathrm{E}(X\,|\,t_{p,X}) - \mathrm{E}X} \tag{1}$$

and $p \in (0,1)$. Clearly, $\mu_{X,Y}(p)$ does not depend on the choice of x_p and y_p.

In the example discussed above we get $\mu_{Y,X}(p) = 0$ and

$$\mu_{X,Y}(p) = \begin{cases} (2p-1)/(1-p) & \text{for } p \leqslant 1/2, \\ (2p-1)/p & \text{for } p > 1/2, \end{cases}$$

which is in accordance with the intuition (see Fig. 1). Obviously, in this case $\operatorname{corr}(X, Y)$ is equal to zero.

In [5] the definition of $\mu_{X,Y}$ is extended to a set C of random variables (X, Y) with finite expectations and marginal distributions not concentrated on a point. The only change is a suitable "randomization" of characteristic functions appearing in (1). Namely, for any $(X, Y) \in C$

$$t_{p,Y}(a) = \begin{cases} 0 & \text{for } a < y_p, \\ \gamma_{Y,p} & \text{for } a = y_p, \\ 1 & \text{for } a > y_p, \end{cases}$$

where

$$\gamma_{Y,p} = \begin{cases} 0 & \text{for } \mathrm{P}(Y = y_p) = 0, \\ \big(1-p-\mathrm{P}(Y > y_p)\big)/\mathrm{P}(Y = y_p) & \text{otherwise,} \end{cases}$$

and $t_{p,X}$ is defined analogously.

A pair of monotonic dependence functions $\mu_{X,Y}$ and $\mu_{Y,X}$ is treated as a characteristic describing a shape of the distribution of (X, Y) and as such it should be invariant on linear increasing transformations of both variables. It follows from Theorem 2.1 (iv) in [5] that this requirement is fulfilled while any of mdf's taken apart, say $\mu_{X,Y}$, is invariant on linear increasing transformations of X and on increasing transformations of Y, which seems desirable when dealing with a characteristic indicating the monotonic dependence of X on Y.

2. MDF'S FOR SOME ESSENTIAL CLASSES OF DISTRIBUTIONS

The example concerning the V-shaped distribution illustrates the fact that one has to expect a change of sign of at least one of the mdf's if intuitive requirements of monotonicity between the two variables are not fulfilled. To express this statement in a formal way, we recall (Theorem 2.1 (vi) in [5]) that if X and Y are positively (negatively) quadrant dependent, i.e.,

$$\mathrm{P}(X \leqslant x,\ Y \leqslant y) \geqslant (\leqslant) \mathrm{P}(X \leqslant x)\mathrm{P}(Y \leqslant y),$$

then both $\mu_{X,Y}$ and $\mu_{Y,X}$ are non-negative (non-positive). Hence a change of sign of any of the mdf's implies that X and Y are not quadrant dependent.

The class of quadrant dependent random variables contains many important families of random variables considered in the statistical literature. A review of such families of (X, Y) is given in [7]; it includes, e.g., contaminated independence models: $X = U + aZ$, $Y = V + bZ$ for any independent random variables U, V, Z and any a, b. Another example is that of linear regression dependence models: $Y = \alpha + \beta X + U$ for any independent X and U and any α, β. The class of quadrant dependent random variables is also closed under the operation of forming mixtures of any two independent pairs of quadrant dependent random variables under the condition that the dependence is either positive or negative for both pairs. In all these cases both mdf's are of a constant identical sign.

Returning to the class of distributions with constant mdf's, it is convenient to remind the concept of boundary distributions. Let F and G denote marginal distribution functions of X and Y, respectively; then there exist two distribution functions $H^+_{F,G}$ and $H^-_{F,G}$,

$$H^+_{F,G}(x, y) = \min\big(F(x), G(y)\big),$$
$$H^-_{F,G}(x, y) = \max\big(F(x)+G(y)-1, 0\big),$$

such that any other distribution function $H_{F,G}$ with marginal distribution functions F and G satisfies the inequalities

$$H^-_{F,G}(x, y) \leqslant H_{F,G}(x, y) \leqslant H^+_{F,G}(x, y) \quad \text{for every } x, y.$$

H^+ and H^- correspond to the so-called *boundary distributions* which are concentrated on a non-decreasing and non-increasing curve, respectively. For continuous F and G the curves are described by $F(x) = G(y)$ and $F(x) = 1-G(y)$.

Suppose that there exists an increasing function $h\colon R \to R$ such that the boundary distribution corresponding to $H^+_{F,G}$ or to $H^-_{F,G}$ is concentrated on $x = h(y)$ or on $-x = h(y)$, respectively, i.e., either X and $h(Y)$ or $-X$ and $h(Y)$ have the same distribution. Then, by Theorem 2.2 in [5], for any $\varrho \in [-1, 0) \vee (0, 1]$ and $(X, Y) \in C$ the equivalence

(2) $\mu_{X,Y}(p) \equiv \varrho \Leftrightarrow \mathrm{E}X \mid Y = \varrho h(Y)+(1-|\varrho|)\mathrm{E}X$ almost everywhere

holds, where ϱ is equal to $\operatorname{corr}\big(X, h(Y)\big)$ whenever the latter exists.

An important special case is that where h is linear, which implies that $\mathrm{E}X \mid Y$ in (2) is almost everywhere linear in Y. It is convenient to introduce the following definition:

Definition 1. For any $(X, Y) \in C$ we say that X is *strongly linearly dependent* on Y if the regression function of X on Y is a linear non-constant function and there exists a linear $l\colon R \to R$ such that $l(Y)$ and X have the same distribution.

It follows from (2) that if X is strongly linearly dependent on Y, then there exists a unique $\varrho \neq 0$ such that $\mu_{X,Y}(p) \equiv \varrho$, with ϱ equal to $\operatorname{corr}(X, Y)$ whenever the latter exists. It is sometimes convenient to use the expression "X is ϱ-strongly linearly dependent on Y" in order to indicate the corresponding ϱ. The class of distributions with mutual ϱ-strong linearity (of X on Y and of Y on X), exemplified by binormal distributions, is that for which the correlation coefficient, if any, is a perfect measure of monotonic dependence, since both mdf's are identically equal to it.

For $(X, Y) \in C$, if X is ϱ-strongly linearly dependent on Y and if $f\colon R \to R$ is linear almost everywhere (with respect to the distribution of X), then $f(X)$ is ϱ'-strongly linearly dependent on Y with $\varrho' = \varrho$ when f is almost everywhere increasing and with $\varrho' = -\varrho$ when f is almost everywhere decreasing. This statement, which follows from Theorem 2.1 (iv) in [5], implies that the mdf of X on Y remains constant under increasing linear transformation of X. It would be interesting to know how some other increasing transformations of X affect the shape of mdf's when X is strongly linearly dependent on Y. This problem is approached here by means of some tentative examples.

Let (X, Y) be binormal $N(m_X, m_Y, \sigma_X, \sigma_Y, \varrho)$ and let $f(x) = \exp(x)$ so that $(f(X), Y)$ has a lognormal-normal distribution. It follows from the definition of the mdf that

$$\mu_{\exp(X),Y}(p) = \begin{cases} \{p - \Phi(\xi_p^* - \varrho\sigma_X)\}/\{p - \Phi(\xi_p^* - \sigma_X)\} & \text{for } \varrho \geqslant 0, \\ \{p - \Phi(\xi_p^* - \varrho\sigma_X)\}/\{\Phi(\xi_p^* + \sigma_X) - p\} & \text{for } \varrho < 0, \end{cases}$$

where Φ is the distribution function of a univariate standard normal variate and $\xi_p^* = \Phi^{-1}(p)$.

For $\varrho > 0$ the function $\mu_{\exp(X),Y}$ decreases from 1 to 0 as p increases from 0 to 1 (cf. Fig. 2a). This is in accordance with our intuition: the function $\exp(X)$ increases rapidly as $x \nearrow +\infty$ which spoils the positive relation between $\exp(X)$ and Y for large values of Y. On the other hand, $\exp(x)$ decreases slowly when $x \searrow -\infty$ and the positive relation between $\exp(X)$ and Y is stronger for small values of Y than that in the binormal case.

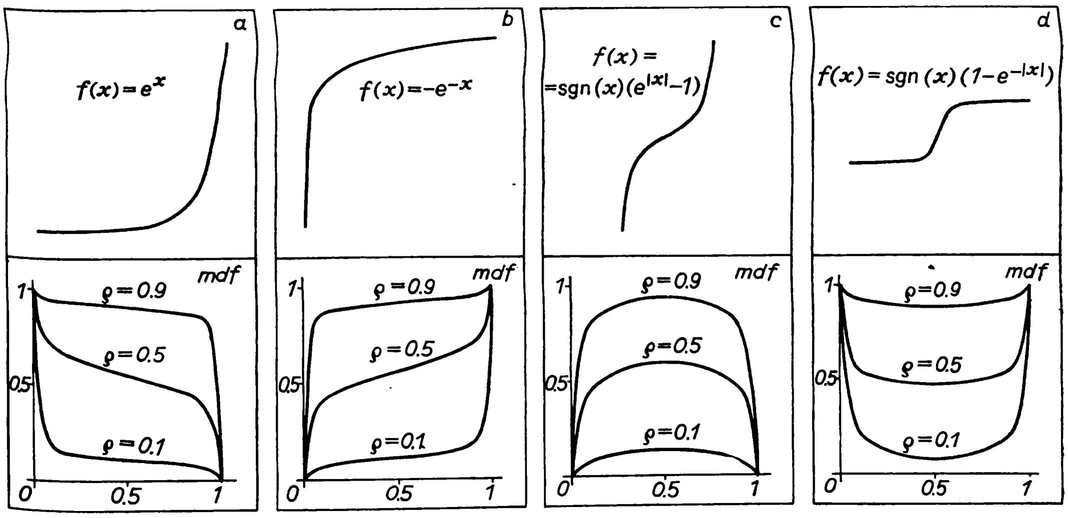

Fig. 2. Mdf's of $f(\xi)$ on η, where $(\xi, \eta) \sim N(0, 0, 1, 1, \varrho)$

For a function f which decreases rapidly as $x \searrow -\infty$ and increases slowly as $x \nearrow +\infty$ the situation is reversed. This case is illustrated in Fig. 2b where $f(x) = -\exp(-x)$. It is easy to check that for a binormal (X, Y) we have

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

By a similar intuitive argument one can expect that functions of the types considered in Fig. 2c and Fig. 2d lead to concave and convex, respectively, mdf's of $f(X)$ on Y.

Summing up, we assert, though we are not able yet to prove it, that if an increasing function f of one of the four types considered above is applied to X and if X is strongly linearly dependent on Y, then the mdf's of $f(X)$ on Y are non-increasing, non-decreasing, concave and convex, respectively. This is adequate to the shape of the distribution of $(f(X), Y)$, since under the influence of the corresponding f the distribution of (X, Y) is appropriately "stretched" or "narrowed". To build up some intuition, consider four distributions being uniform over the areas indicated in Figs. 3a-d. All these distributions represent a positive relation between X and Y. Moreover, its strength decreases gradually with y in the first distribution, increases gradually in the second one, increases and decreases in turn in the third one and, finally, decreases and increases in the last one. These intuitions, suggested by the forms of the distributions, are confirmed by the shapes of the corresponding mdf's presented in Fig. 3. It is easily seen that these mdf's correspond to those presented in Fig. 2.

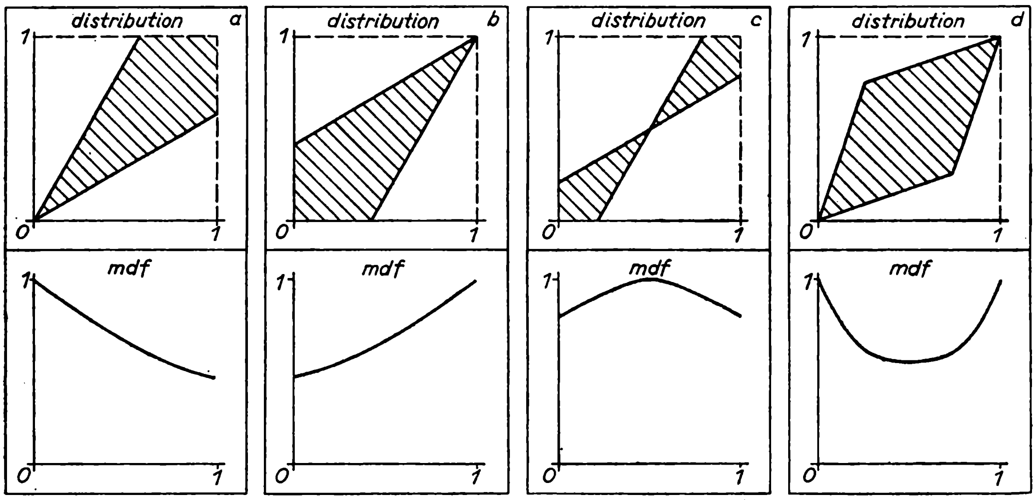

Fig. 3. Distributions uniform over shaded areas and the corresponding mdf's

The considerations above give some rough idea of the class of distributions with mdf's which are increasing, decreasing and unimodal.

Generally, different kinds of mdf's considered in this section could be useful in statistical modelling, since qualitative information of this type is often available in practice. The results of the section could also be used in data analysis and informal inference based on the graphs of empirical mdf's.

3. MDF'S IN THE NEIGHBOURHOOD OF BINORMAL DISTRIBUTIONS

3.1. General remarks. The methods of multivariate statistical analysis are developed almost exclusively for multivariate normal distributions and, therefore, the "neighbourhood" of this distribution is specially interesting in the investigations of robustness of particular methods and in statistical modelling. Some families of distributions "close" to binormal distributions are considered here from the point of view of the type of dependence. The first step in this direction was already made in Section 2, where distributions convertible to binormal by some monotonic marginal transformations were discussed (cf. Fig. 2).

Throughout this section, (ξ, η), (ξ', η') and (ξ'', η'') denote independent pairs of random variables with the same standard binormal distribution $N(0, 0, 1, 1, \varrho)$, and Φ is the distribution function of $N(0, 1)$.

3.2. Bivariate *t* Student's distributions. First, we start with a more general class of distributions containing a bivariate *t*.

Let (U, V) be a pair of random variables with zero means, the covariance matrix

$$K = \begin{Vmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{Vmatrix},$$

and a continuous distribution with the same marginals and linear regression functions. It follows that U and V are mutually ϱ-strongly linearly dependent. Let Z be a continuous random variable which is independent of (U, V) and has a positive density function on $(0, +\infty)$ and a finite expectation. We are interested in the form of the dependence of $X = UZ + m_1$ and $Y = VZ + m_2$, where $m = (m_1, m_2)$ is a pair of real numbers. It is easy to check by straightforward calculations that the regression function of X on Y is linear. Consequently, X and Y are mutually ϱ-strongly linearly dependent, which means that the form of the dependence of U and V is preserved and the role of Z is analogous to that of a positive constant influencing scale parameters only. It follows that ϱ is the correlation coefficient of U and V, and of X and Y, whenever they exist.

According to the terminology used in [2], a bivariate random variable has the *t* Student distribution with q degrees of freedom, location vector m and precision matrix $T = K^{-1}$ if it is defined as X and Y above with U and V replaced by $(\sigma_1\xi, \sigma_2\xi)$ and with $Z = S_q^{-1}$, where qS_q^2 is distributed as χ^2 with q degrees of freedom. Then for any $q > 1$ the bivariate t preserves the mutual strong linear dependence of the binormal (ξ, η) and, moreover, for $q > 2$ the coefficient ϱ is equal to $\operatorname{corr}(X, Y)$. For $q = 2$, ϱ can be interpreted as a generalized correlation coefficient. For $q = 1$, i.e., for a bivariate Cauchy distribution, EZ is not finite and the considerations above are not applicable. One can only suggest an extension of the definition of strong linear dependence of X on Y as well as the corresponding extension of the definition of the mdf to cover the case where X and Y are defined as at the beginning of this section while $\mathrm{E}Z$ is not finite. With such an extension one would have $\mu_{X,Y}(p) \equiv \varrho$ for any bivariate t distribution including the Cauchy one.

3.3. Contaminated binormal distributions. There is plenty of different possibilities of introducing the contamination to bivariate distributions and the terminology is rather unstable. We consider the following mixtures:

(i) $(1-\varepsilon)P_{\xi,\eta}+\varepsilon P_{\varkappa\xi',\varkappa\eta'}$,

(ii) $(1-2\varepsilon)P_{\xi,\eta}+\varepsilon P_{\varkappa\xi',\eta'}+\varepsilon P_{\xi'',\varkappa\eta''}$,

(iii) $(1-2\varepsilon)P_{\xi,\eta}+\varepsilon P_{\xi'+d,\eta'+d}+\varepsilon P_{\xi''-d,\eta''-d}$,

(iv) $(1-2\varepsilon)P_{\xi,\eta}+\varepsilon P_{\xi'+d,\eta'-d}+\varepsilon P_{\xi''-d,\eta''+d}$.

Here $\varkappa > 1$, $d > 0$, $0 < \varepsilon < \frac{1}{2}$, and $P_{\xi,\eta}$ denotes the probability distribution of (ξ, η). According to Devlin et al. [3], case (i) is called *contaminated normal* and cases (iii) and (iv), in which outliers appear along the principal axes of $P_{\xi,\eta}$, are called *major axis* and *minor axis outliers*, respectively. Case (ii), introduced by us, is called *independently contaminated normal.* In the sequel we discuss cases (i)-(iv) in turn denoting by (X, Y) the corresponding mixture.

In case (i), $\mu_{X,Y}(p) \equiv \operatorname{corr}(X, Y) = \varrho$, since it is easy to check that any mixture of two independent pairs of ϱ-strongly linearly dependent random variables is also ϱ-strongly linearly dependent.

In case (ii) we have

$$\mu_{X,Y}(p) = \varrho\frac{1-2\varepsilon+\varepsilon\left[\varkappa+\exp\left\{\left(y_p^2(\varkappa^2-1)\right)/2\varkappa^2\right\}\right]}{1-2\varepsilon+\varepsilon\left[1+\varkappa\exp\left\{\left(y_p^2(\varkappa^2-1)\right)/2\varkappa^2\right\}\right]}, \tag{3}$$

where y_p is the root of the equation

$$(1-\varepsilon)\Phi(y_p)+\varepsilon\Phi(y_p/\varkappa) = p.$$

The mdf's defined by (3) are concave as presented in Fig. 4a. Consequently, this type of contamination acts on (ξ, η) in a similar way as the function in Fig. 2c applied to ξ. This is in accordance with our intuition since the contamination introduces dispersion which weakens the monotonic dependence for p distant from 0.5, $\mu_{X,Y}(0.5)$ being still equal to ϱ. However, we see in Fig. 4a that the mdf's are not too strongly affected by the contamination even for considerably large $\varkappa$ and ε. On the other hand, the absolute values of $\operatorname{corr}(X, Y)$ given by $\varrho\left(1-2\varepsilon(\varkappa-1)\right)/(1-2\varepsilon\varkappa^2)$ are much smaller than the corresponding values of $|\varrho|$. For instance, for $\varrho = 0.5$, $\varepsilon = 0.05$ and $\varkappa = 3$ we have $\operatorname{corr}(X, Y) = 0.316$.

In case (iii) we have

$$\mu_{X,Y}(p) = \varrho + [d(1-\varrho)\varepsilon\{\Phi(y_p+d)-\Phi(y_p-d)\}]/M, \tag{4}$$

where

$$M = (1-2\varepsilon)\varphi(y_p)+\varepsilon[\varphi(y_p+d)+\varphi(y_p-d)+d\{\Phi(y_p+d)-\Phi(y_p-d)\}]$$

and y_p is the root of the equation

$$(1-2\varepsilon)\Phi(y_p)+\varepsilon\Phi(y_p-c)+\Phi(y_p+c) = p.$$

Obviously, for $\varrho > 0$, outliers appearing along the line $x = y$ should increase considerably the strength of the monotonic dependence for p

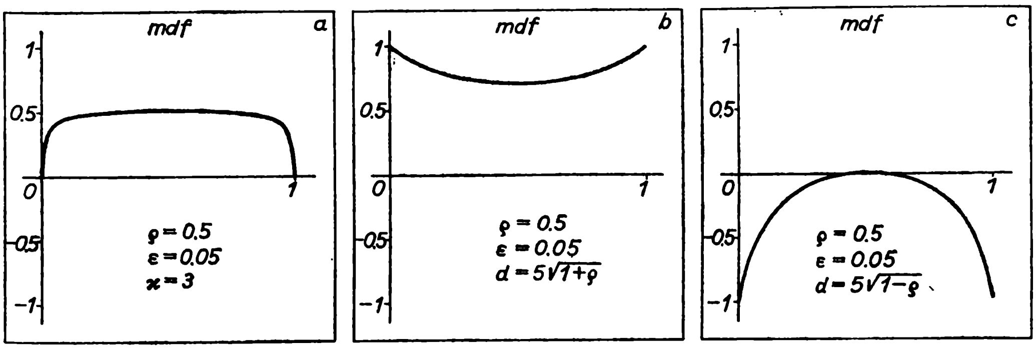

Fig. 4. Mdf's of contaminated binormal distributions
a — independently contaminated, *b* — major axis outliers, *c* — minor axis outliers

distant from 0.5. This means that major axis outliers act on (ξ, η) in a similar way as the function in Fig. 2d applied to ξ, which coincides with our intuition and is confirmed by the graphs given in Fig. 4b. Moreover,

$$\operatorname{corr}(X, Y) = (\varrho+2\varepsilon d^2)/(1+2\varepsilon d^2)$$

increases quickly with ε and d. For instance, for $\varrho = 0.5$, $\varepsilon = 0.05$ and $d = 5\sqrt{1.5}$ we have $\operatorname{corr}(X, Y) = 0.894$. Then for ε and d large enough the values of mdf's as well as of the correlation coefficient are very distant from ϱ.

In case (iv), the formula for the mdf corresponding to the minor axis outliers can easily be derived from the respective formula (4) for major axis outliers. Indeed, let $(X(\varrho), Y(\varrho))$ and $(X'(\varrho), Y'(\varrho))$ denote the pairs of random variables in the case of major and minor axis outliers, respectively. Then the distributions of $(X'(\varrho), Y'(\varrho))$ and of $(-X(-\varrho), Y(-\varrho))$ are identical and it follows from Theorem 2.1 (iv) in [5] that

$$\mu_{X',Y'}(p;\varrho) = -\mu_{X,Y}(p;-\varrho).$$

For $\varrho > 0$ and ε and d large enough, minor axis outliers spoil entirely the mutual monotonic dependence of ξ and η and make ϱ meaningless.

This explains the exceptional behaviour of the estimators of ϱ for this particular distribution which was pointed out in [3]. The corresponding graph is presented in Fig. 4c. The correlation coefficient given by the formula

$$\operatorname{corr}(X', Y') = (\varrho - 2\varepsilon d^2)/(1 + 2\varepsilon d^2)$$

is also strongly affected. For instance, for $\varrho = 0.5$, $\varepsilon = 0.05$ and $d = 5\sqrt{0.5}$ we get the value -0.333.

3.4. Discretized binormal distributions. Let $s_1, \ldots, s_k$ be a sequence of real numbers such that

$$-\infty \stackrel{\mathrm{df}}{=} s_0 < s_1 < \ldots < s_k < s_{k+1} \stackrel{\mathrm{df}}{=} +\infty.$$

Let f be a mapping from R onto $\{0, 1, \ldots, k\}$ defined by

$$f(x; s_1, \ldots, s_k) = i \quad \text{if } s_i \leqslant x < s_{i+1}, \; i = 0, 1, \ldots, k.$$

We restrict ourselves to discretizations by means of the mapping f defined above. Such a transformation should preserve the type of monotonic dependence. For instance, it follows from Lemma 1 (iii) in [7] that the positive quadrant dependence is not affected by f.

The cases of only one and of both variables being discretized are considered in turn. For the pair $(\xi, f(\eta))$ we have

$$\mu_{\xi, f(\eta)}(p) = \varrho\{\varphi(s_{j-1})(\Phi(s_j) - p) + \varphi(s_j)(p - \Phi(s_{j+1}))\}/M \qquad (5)$$

$$(j = 1, \ldots, k+1),$$

where $0 \neq p \in [\Phi(s_{j-1}), \Phi(s_j)]$ and

$$M = \varphi(\Phi^{-1}(p))\{\Phi(s_j) - \Phi(s_{j-1})\}.$$

The formulas for $\mu_{f(\xi), \eta}$ and $\mu_{f(\xi), f(\eta)}$ were also easily derived from the definition of the mdf but they have a complicated form and, therefore, are not presented in this paper. The corresponding graphs are given in Figs. 5a, b, c. In each case three sets of the discretization points

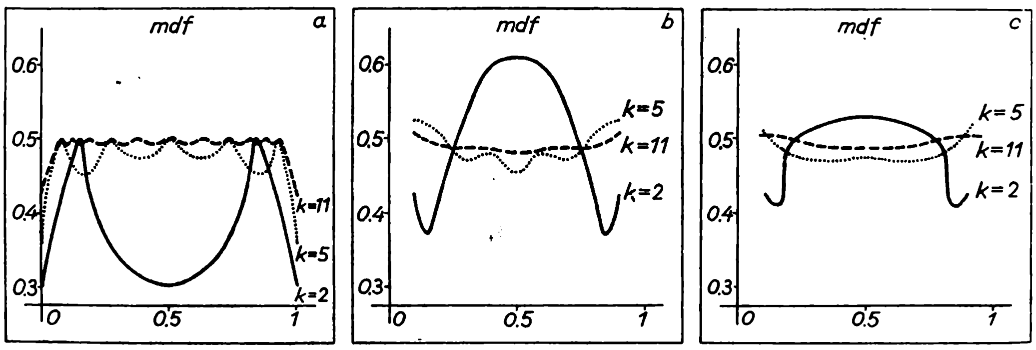

Fig. 5. Mdf's of discretized binormal distributions

a – the second component discretized, *b* – the first component discretized, *c* – both components discretized

$s_1, \ldots, s_k$ were considered:

(i) $k = 2$, $(s_1; s_2) = (-1; 1)$;

(ii) $k = 5$, $(s_1, \ldots, s_5) = (-1.5; -0.6; 0; 0.6; 1.5)$;

(iii) $k = 11$, $(s_1, \ldots, s_{11}) = (-1.5; -1.2; \ldots; 1.2; 1.5)$.

In the sequel, the three cases are referred to as $k = 2$, $k = 5$ and $k = 11$. For any of them, $\varrho = 0.5$.

Let us discuss first the pair $(\xi, f(\eta))$. Obviously, the strength of the monotonic dependence is not affected by discretization in the case where p's correspond to $s_1, \ldots, s_k$ and, consequently, $\mu_{\xi,f(\eta)}(\Phi(s_i)) = \varrho$ for $i = 1, \ldots, k$, as is seen from (5). Generally, however, the discretization of η results in the fact that the strong linearity of ξ on η is spoiled, since the values of ξ are more dispersed for $\eta \in (s_i, s_{i+1})$ than for η taking any particular value from this interval. This leads to the graphs in Fig. 5a. Since the discrepancies from ϱ increase with the lengths of the intervals (s_i, s_{i+1}), they are extreme for $k = 2$ while the mdf for $k = 11$ provides a good approximation for ϱ.

The case $k = 1$ is not presented in Fig. 5 in order to achieve greater clarity but it is evident that any dichotomization of η results in a concave $\mu_{\xi,f(\eta)}$ so that the effect is similar to that produced by the function in Fig. 2c applied to ξ.

Returning to the pair $(f(\xi), \eta)$, let us notice first that for $k = 1$ the function f can be considered as the limit case of the functions in Fig. 2d, and hence the corresponding mdf is convex. For $k > 1$, since there is no dispersion of the values of $f(\xi)$ for $\eta \in (s_i, s_{i+1})$, the strength of the monotonic dependence of $f(\xi)$ on η increases for $p \in (\Phi(s_i), \Phi(s_{i+1}))$ as compared with that of ξ on η, and hence $\mu_{f(\xi),\eta}$ restricted to this interval is concave. This is best illustrated in Fig. 5b for $k = 2$ and p belonging to the interval $(\Phi(-1), \Phi(1)) = (0.159, 0.841)$.

The mdf's of ξ on $f(\eta)$ and of $f(\xi)$ on η are interesting examples of mdf's for a "mixed" continuous-discrete bivariate distribution. Both mdf's should be compared with the correlation coefficient of $f(\xi)$ and η, evaluated by the formula (cf. [1])

$$\operatorname{corr}(f(\xi), \eta) = \varrho \sum_{i=1}^{k} [\varphi(s_i)/\sigma_{f(\xi)}],$$

where

$$\sigma^2_{f(\xi)} = \sum_{i=1}^{k} [i^2\{\Phi(s_{i+1}) - \Phi(s_i)\}] - \{\mathrm{E}f(\xi)\}^2$$

and

$$\mathrm{E}f(\xi) = \sum_{i=1}^{k} [i\{\Phi(s_{i+1}) - \Phi(s_i)\}].$$

For $\varrho = 0.5$ and $k = 2, 5$ and 11 (cf. Fig. 5) $\operatorname{corr}\big(f(\xi), \eta\big)$ is equal to 0.39, 0.45 and 0.48, respectively.

Figs. 5a and 5b reflect the opposite tendencies which characterize mdf's when only the first or only the second component is discretized. It was interesting to find out "the effect of the battle" between these tendencies when both variables were discretized. The result of the compromise is as given in Fig. 5c. Evidently, the graphs are of similar shape to the corresponding ones in Fig. 5b but they are "smoother". This means that the discretization of the first component was "more powerful" than that of the second one, the latter being noticeable only in the smoothing.

Fig. 5c throws some light on the problem of presentation of monotonicity tendencies demonstrated by contingency tables in an arbitrary case.

4. CONCLUDING REMARKS

In this paper the mdf's are characterized as a measure of monotonicity of a bivariate distribution. However, practical usefulness of mdf's depends mostly on the properties of their sample counterparts. By Theorem 3.1 in [5], for any $p \in (0, 1)$ the empirical mdf's converge to the theoretical ones with probability 1. Asymptotic or finite sample distributions are not known yet. We obtained some results by Monte Carlo methods and we used them in procedures dealing with correlation estimating and binormality testing. This will be presented in a forthcoming paper. Moreover, empirical mdf's will be used in model choice problems when one wants to choose a distribution convertible to binormal by a monotonic marginal transformation belonging to a specified class of functions.

References

[1] N. R. Cox, *Estimation of the correlation between a continuous and a discrete variable*, Biometrics 30 (1974), p. 171-179.

[2] M. M. De Groot, *Optimal statistical decisions*, McGraw-Hill, New York 1970.

[3] S. J. Devlin, R. Gnanadesikan and J. R. Kettenring, *Robust estimation and outlier detection with correlation coefficients*, Biometrika 62 (1975), p. 531-545.

[4] R. Gnanadesikan and M. B. Wilk, *Data analytic methods in multivariate statistical analysis*, p. 593-637 in: P. R. Krishnaiah (editor), *Multivariate analysis*, vol. 2, New York 1969.

[5] T. Kowalczyk, *General definition and sample counterparts of monotonic dependence functions in the bivariate case*, Math. Operationsforschung und Statistik 8 (1977), p. 351-369.

[6] — and E. Pleszczyńska, *Monotonic dependence functions of bivariate distributions*, Ann. Statist. 5 (1977), p. 1221-1227.
[7] E. L. Lehmann, *Some concepts of dependence*, Ann. Math. Statist. 37 (1966), p. 1137-1153.

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
00-901 WARSZAWA

Received on 10. 10. 1977

TERESA KOWALCZYK, A. KOWALSKI, A. MATUSZEWSKI
i E. PLESZCZYŃSKA (Warszawa)

ZASTOSOWANIA FUNKCJI ZALEŻNOŚCI MONOTONICZNEJ W ANALIZIE DANYCH

STRESZCZENIE

Scharakteryzowano funkcję zależności monotonicznej jako statystykę opisową, służącą do badania monotoniczności i kształtu rozkładu dwóch zmiennych. Szczegółowej analizie poddano otoczenie rodziny dwuwymiarowych rozkładów normalnych, do którego zaliczono rozkłady t Studenta, rozkłady normalne z zakłóceniami i rozkłady normalne poddane dyskretyzacji.
