Ann. Inst. Statist. Math.
33 (1981), Part A, 347-359

CHARACTERIZATION OF DEPENDENCE CONCEPTS IN NORMAL DISTRIBUTIONS

LUDGER RÜSCHENDORF

(Received Oct. 31, 1980; revised Mar. 9, 1981)

Summary

In the present paper we deal with the characterization of some dependence concepts for the multivariate normal distribution. It turns out that normal distributions have some special properties w.r.t. these dependence concepts and, furthermore, that the characterizations are closely connected to some interesting problems on matrices. Some applications to simultaneous confidence bounds are discussed.

1. Normal distributions and positive orthant dependence

A basic concept of dependence was introduced by Lehmann [13]. Let $X=(X_1, \cdots, X_n)$ be a random vector on a probability space $(M, \mathfrak{A}, P)$. X is called positively orthant dependent (POD) if

$$(1) \qquad \mathrm{P}\left(\bigcap_{i=1}^{n} \{X_i \geqq \alpha_i\}\right) \geqq \prod_{i=1}^{n} \mathrm{P}(X_i \geqq \alpha_i), \qquad \text{for all } \alpha_1, \cdots, \alpha_n \in R^1 .$$

Similarly, X is called negatively orthant dependent (NOD) if

$$(2) \qquad \mathrm{P}\left(\bigcap_{i=1}^{n} \{X_i \leqq \alpha_i\}\right) \geqq \prod_{i=1}^{n} \mathrm{P}(X_i \leqq \alpha_i), \qquad \text{for all } \alpha_1, \cdots, \alpha_n \in R^1 .$$

It was shown by Rüschendorf [17], Theorem 2, that POD-distributions share with normal distributions the important property, that the independent elements in the class of all POD-distributions can be identified by some mixed moment conditions. In the present section we extend some properties of POD-distributions and discuss applications to normal distributions.

The characterization of POD in normal distributions is immediate from a theorem due to Slepian [21] in combination with a result of Lehmann [13]. Let X be $N(\mu, \Sigma)$-distributed and let $S_n=\{A \in R^{n\times n};\ A=(a_{ij})_{1\leq i,j\leq n}$ be positive semidefinite, $a_{ij}\geqq 0,\ \forall\, i, j\leqq n\}$. Then

$$(3) \qquad X \text{ is POD} \Longleftrightarrow X \text{ is NOD} \Longleftrightarrow \Sigma \in S_n\,.$$

The equivalence of POD and weak association was shown in Theorem 1 of Rüschendorf [17]. For applications of the POD-concept also the following closedness properties are useful.

PROPOSITION 1. Let $Y_1, \cdots, Y_k$ be POD, n-dimensional random variables, $Y_i=(Y_{i1}, \cdots, Y_{in})$, and let $\{Y_1, \cdots, Y_k\}$ be stochastically independent.

a) If $f_i: R^k \to R^1$, $1 \leqq i \leqq n$, are monotonically nondecreasing and measurable, then

$$(4) \qquad Y=(f_1(Y_{11}, \cdots, Y_{k1}), \cdots, f_n(Y_{1n}, \cdots, Y_{kn})) \text{ is POD.}$$

b) $\sum_{i=1}^{k} Y_i$ is POD.

c) If $Y_2 \geqq 0$, then $(Y_{11}Y_{21}, \cdots, Y_{1n}Y_{2n})$ is POD.

PROOF. In order to avoid technicalities we only give the proof of b). The proof of a) and c) is similar. Let $\alpha_1, \cdots, \alpha_n \in R^1$, and $X=Y_1$, $Y=Y_2$, then

$$\begin{aligned} P\Big(\bigcap_{i=1}^{n} \{X_i+Y_i \geqq \alpha_i\}\Big) &= \int P\Big(\bigcap_{i=1}^{n} \{X_i+y_i \geqq \alpha_i\}\Big) dP^Y(y_1, \cdots, y_n) \\ &\geqq \int \prod_{i=1}^{n} P(X_i+y_i \geqq \alpha_i) dP^Y(y_1, \cdots, y_n) \\ &\geqq \prod_{i=1}^{n} \int P(X_i+y_i \geqq \alpha_i) dP^{Y_i}(y_i) \\ &= \prod_{i=1}^{n} P(X_i+Y_i \geqq \alpha_i)\,. \end{aligned}$$

The second inequality follows from Theorem 1 of Rüschendorf [17] since $f_i(X_i)=I_{\{X_i+y_i \geqq \alpha_i\}}$ is monotonically nondecreasing in X_i. The case $k \geqq 2$ follows from induction.

Example 1. Let $X_1, \cdots, X_k$ be $N(\mu, \Sigma)$-distributed with unknown μ, Σ and let $\{X_1, \cdots, X_k\}$ be stochastically independent. A confidence interval for μ proposed by Dunn [4] and Scott [19] is the following. Let $\bar{X}=\frac{1}{k}\sum_{i=1}^{k} X_i$ and $S=\frac{1}{k-1}\sum_{i=1}^{k}(X_i-\bar{X})^T(X_i-\bar{X})$ be the canonical estimators of μ, Σ; then consider

$$(5) \qquad R=\{x \in R^n;\ |x_i-\bar{X}_i| \leqq d_i\sqrt{S_{ii}},\ 1 \leqq i \leqq n\}\,,$$

where $\bar{X}_i$ is the ith component of $\bar{X}$, S_{ii} is the ith diagonal element of the random matrix S and $d_1, \cdots, d_n \in R^1_+$ are given.

If $\Sigma=I_n$, then

$$P_{\mu, I_n}(\mu \in R) = \prod_{i=1}^{n} P(|t_i| \leqq \sqrt{k}\, d_i) =: \gamma_0$$

where t_i are Student t-variables with $k-1$ degrees of freedom. The question is now for which normal distributions γ_0 can be used as conservative bound for the confidence interval R (the proof of the claim of Scott [19] that this should be true for all normal distributions was shown to be in error by Sidak [20]).

By means of Proposition 1 we obtain the following result:

(6) Let X_1 be $N(\mu, \Sigma)$-distributed and let $|X_1-\mu|=(|X_{11}-\mu_1|, \cdots, |X_{1n}-\mu_n|)$ be POD, then

$$P_{\mu, \Sigma}(\mu \in R) \geqq \gamma_0 .$$

PROOF. It is well known that one can choose $Y_i \sim N(0, \Sigma)$, $1 \leqq i \leqq k-1$, which are independent from each other and from $\{X_j, 1 \leqq j \leqq k\}$ such that S and $\tilde{S} = \frac{1}{k-1} \sum_{i=1}^{k-1} Y_i^T Y_i$ have the same distribution. Since by assumption $|Y_1|$ is POD we obtain from Proposition 1, a), that $(\tilde{S}_{11}, \cdots, \tilde{S}_{nn})$ is POD where $\tilde{S}_{ii} = \left(\frac{1}{k-1} \sum_{j=1}^{k-1} Y_{ji}^2\right)^{1/2}$, $1 \leqq i \leqq n$.

Using a result of Khatri [12] that $|X|$ is NOD for each $N(0, \Sigma)$-distributed X, we obtain by Proposition 1, c) that $\left(\frac{|\bar{X}_1-\mu_1|}{\tilde{S}_{11}}, \cdots, \frac{|\bar{X}_n-\mu_n|}{\tilde{S}_{nn}}\right)$ is NOD which implies (6).

Some conditions implying that $|X_1-\mu|$ is POD have been given by Sidak [20], Jogdeo [11] and Abdel-Hameed, Sampson [1]. A general lower bound for the probability of translated positive orthants of $|X_1-\mu|$ is given by the following proposition.

PROPOSITION 2. Let $X=(X_1, \cdots, X_n)$ be $N(0, \Sigma)$-distributed. Then for any $\alpha_1, \cdots, \alpha_n \in R_+^1$

$$P\left(\bigcap_{i=1}^{n} \{|X_i| \geqq \alpha_i\}\right) \geqq \prod_{i=1}^{n} P\left(|X_i| \geqq \frac{\sqrt{\sigma_{ii}}}{\gamma_i} \alpha_i\right), \tag{7}$$

where

$$\Sigma = (\sigma_{ij}), \quad \gamma_1 = \sqrt{\sigma_{11}}, \quad \gamma_m^2 = \frac{\det(\Sigma(m))}{\det(\Sigma(m-1))},$$

$$2 \leqq m \leqq n, \quad \text{and} \quad \Sigma(m) = (\sigma_{ij})_{1 \leqq i, j \leqq m} .$$

PROOF. The proof of Proposition 2 is similar to the proof of

Theorem 3.2 of Das Gupta, Eaton, Olkin, Perlman, Savage and Sobel [3]. Let Σ have a factorization TT^T where T is an upper-triangular $n\times n$ matrix and let Y be $N(0, I_n)$-distributed. Then

$$\begin{aligned}\mathrm{P}\left(\bigcap_{i=1}^{n}\{|X_i|\geqq\alpha_i\}\right)&=\mathrm{P}\left(\bigcap_{i=1}^{n}\left\{\left|\sum_{j=i}^{n}t_{ij}Y_j\right|\geqq\alpha_i\right\}\right)\\&=\int_B\left(\int_H g\left(\sum_{i=1}^{n}y_i^2\right)dy_1\right)\prod_{i=2}^{n}dy_i\,,\end{aligned}$$

where $g\left(\sum_{i=1}^{n}y_i^2\right)$ is the density of Y,

$$B=\left\{(y_2,\cdots,y_n);\ \left|\sum_{j=i}^{n}t_{ij}y_j\right|\geqq\alpha_i,\ 2\leqq i\leqq n\right\}$$

and

$$H=H(y_2,\cdots,y_n)=\left\{y_1;\ \left|t_{11}y_1+\sum_{j=2}^{n}t_{1j}y_j\right|\geqq\alpha_1\right\}.$$

Let, furthermore, $H_0=\{y_1;\,|t_{11}y_1|\geqq\alpha_1\}$, then as a consequence of Winter's theorem (cf. Das Gupta, $\cdots$[3])

$$\int_H g\left(\sum_{j=1}^{n}y_j^2\right)dy_1\geqq\int_{H_0}g\left(\sum_{j=1}^{n}y_j^2\right)dy_1\,.$$

So we obtain by an inductive argument

$$\begin{aligned}\mathrm{P}\left(\bigcap_{i=1}^{n}\{|X_i|\geqq\alpha_i\}\right)&\geqq\mathrm{P}\left(\bigcap_{i=1}^{n}\{|t_{ii}Y_i|\geqq\alpha_i\}\right)\\&=\prod_{i=1}^{n}\mathrm{P}\left(|X_i|\geqq\frac{\sqrt{\sigma_{ii}}}{|t_{ii}|}\alpha_i\right).\end{aligned}$$

Since $\det\Sigma(m)=\det(T(m))\det(T(m))^T=\prod_{i=1}^{m}t_{ii}^2$, we obtain

$$t_{11}=\sqrt{\sigma_{11}}\,,\quad t_{ii}^2=\frac{\det\Sigma(i)}{\det\Sigma(i-1)}\,,\quad i\geqq 2\,.$$

Remark 1. Proposition 2 can be used to give a conservative bound for a larger class of distributions than those considered in Example 1.

2. Normal distribution and association

Association of random variables has been introduced by Esary, Proschan and Walkup [6]. Association has many useful statistical applications. Its definition is as follows: $X=(X_1,\cdots,X_n)$ is called associated, if

(8) $\operatorname{Cov}(f(X),g(X))\geqq 0$ for all monotonically nondecreasing functions f, g for which the integrals exist.

Clearly association of X implies POD. The following proposition shows that again the normal distribution has special properties concerning this dependence concept.

PROPOSITION 3. If there exists an associated random variable X with $\mathrm{Cov}(X)=\Sigma$, then

(9) $N(\mu, \Sigma)$-distributed variables are associated.

PROOF. In this proof we shall repeatedly make use of some results of Esary, Proschan and Walkup [6]. Let $X_1, \cdots, X_k$ be independent, n-dimensional with $P^{X_i}=P^X$, $1\leqq i\leqq k$, then the nk-dimensional vector $(X_1, \cdots, X_k)$ is associated and, therefore, $S_k=\frac{1}{\sqrt{k}}\sum_{i=1}^{k}(X_i-\mu)$, ($\mu=\mathrm{E}\,X$) is associated. By the central limit theorem $S_n\xrightarrow{\mathcal{D}}N(0, \Sigma)$. Since associated random variables are closed w.r.t. weak convergence we obtain, that $N(0, \Sigma)$-distributed random variables are also associated.

The following definition is due to Hall, Newman [9] and Markham [14].

DEFINITION 1. An $n\times n$ matrix $\Sigma\in R^{n\times n}$ is called completely positive if there exists a $k\in\mathcal{N}$ and an $A\in R^{n\times k}$,

(10) $$A=(a_{ij})_{\substack{1\leqq i\leqq n\\ 1\leqq j\leqq k}} \text{ with } A\geqq 0 \text{ (i.e. } a_{ij}\geqq 0 \text{ for all } i, j\text{) and}$$
$$\Sigma=AA^T.$$

Let C_n denote the set of all completely positive $n\times n$ matrices.

C_n defines a subset of S_n. We have the following result.

THEOREM 1. *If $\Sigma\in C_n$, then $N(\mu, \Sigma)$-distributed random variables are associated.*

PROOF. Let $\Sigma\in C_n$; then there exists an $n\times k$-matrix $A\geqq 0$ with $\Sigma=AA^T$. If Y is $N(0, I_k)$-distributed, then $\tilde{Y}=AY$ has the same distribution as X. But Y is associated and $A\geqq 0$ defines a monotonically nondecreasing function. So also $\tilde{Y}$ is associated.

To consider the question how large C_n is, we need the following definition.

DEFINITION 2.

a) An element $\Sigma\in S_n$ is called diagonally dominant, if $\sigma_{ii}\geqq\sum_{j\neq i}\sigma_{ij}$, $\forall i\leqq n$.

b) For a convex cone $A\subset R^m$ let $\mathfrak{E}(A)$ denote the set of extreme directions of A.

c) $P_n=\{\Sigma \in C_n;$ there exists an $A\in R^{n\times n}$, $A\geqq 0$ with $\Sigma=AA^T\}$

d) For an $n\times n$ matrix A and $\alpha=(i_1,\cdots,i_k)$, $\beta=(j_1,\cdots,j_k)$ where $1\leqq i_1<i_2<\cdots<i_k\leqq n$, $1\leqq j_1<j_2<\cdots<j_k\leqq n$ let $A(\alpha|\beta)$ denote the minor of A with rows $i_1,\cdots,i_k$ and columns $j_1,\cdots,j_k$.

The results of the following proposition are partially known. We include them for the reason of completeness.

PROPOSITION 4.

a) If Σ is diagonally dominant, then $\Sigma\in C_n$.

b) If $n\leqq 4$, then $P_n=C_n=S_n$.

c) If $n\geqq 5$, then $P_n\subsetneq C_n\subsetneq S_n$ (where $\subsetneq$ means strict inclusion).

d) C_n is a convex cone with $\mathfrak{E}(C_n)=\{cc^T;\ c\in R^n,\ c\geqq 0\}$.

e) $C_n=\mathrm{con}\,(P_n)$ (convex hull) and P_n, C_n are pathwise connected and closed subsets of $R^{n\times n}$.

f) The parameter k from Definition 1 can be chosen $\leqq n^2+1$.

g) If $\Sigma\in S_n$, then Σ has a factorization LL^T, where $L\geqq 0$, L is a lower triangular $n\times n$ matrix if and only if $\Sigma(1,\cdots,k,i\,|\,1,\cdots,k,j)\geqq 0$ for all $k\leqq i$, $j\leqq n$.

PROOF.

a) Let $\Sigma\in R^{n\times n}$ be diagonally dominant. Then define the $n\times \dfrac{n(n+1)}{2}$ matrix A by $A=(a_{k,\{i,j\}})_{1\leqq i,j,k\leqq n}$ with

$$a_{k,\{i,j\}}=\begin{cases}0\,, & \text{if } i,j\neq k\\ \sqrt{\sigma_{ik}}\ , & \text{if } j=k,\ i\neq k\\ \sqrt{\sigma_{jk}}\ , & \text{if } i=k,\ j\neq k\\ \sqrt{\sigma_{kk}-\sum\limits_{l\neq k}\sigma_{kl}}\ , & \text{if } i=j=k\,.\end{cases}$$

It is easy to check, that $\Sigma=AA^T$, so $\Sigma\in C_n$.

b) was proved by Gray, Wilson [8] and independently by Plesken, Rüschendorf, Krafft [15] using geometric arguments.

c) The inclusion $C_n\subsetneq S_n$ is due to Hall, Newman [9]. To prove $P_n\subsetneq C_n$ take

$$\Sigma_0=\begin{bmatrix} I_{n-2} & B\\ B^T & I_2\end{bmatrix}\in R^{n\times n}$$

with $B\in R^{(n-2)\times 2}$, $B>0$ (componentwise) and I_k denoting the unit matrix of dimension k. Using the arguments of Gray, Wilson [8] resp. Plesken, Rüschendorf, Krafft [15] we obtain that $\Sigma_0\notin P_n$. But if we choose Σ_0 diagonally dominant we obtain by part a)

$\Sigma_0 \in C_n$.

d) follows from Theorems 2.1, 3.1 of Hall, Newman [9].

e) Since $P_n \supset \mathfrak{E}(C_n)$ we clearly have $C_n = \operatorname{con}(P_n)$. Defining the closed and pathwise connected set $R_n = \{A \in R^{n\times n};\ A \geqq 0\}$ and the continuous map $\varphi: R_n \to R^{n\times n}$ by $\varphi(A) = AA^T$, we clearly have $\varphi(R_n) = P_n$. This implies that P_n is closed (since bounded subsets of P_n have bounded origins) and, furthermore, that P_n is pathwise connected (as continuous image of the pathwise connected set R_n). Since $C_n = \operatorname{con}(P_n)$ the same is true for C_n.

f) If $\Sigma \in C_n$, then there exist by definition of C_n (or by d)) $c_i \in R^n$, $c_i \geqq 0$, $1 \leqq i \leqq k$, such that

$$\Sigma = \sum_{i=1}^{k} c_i c_i^T = \sum_{i=1}^{k} \frac{1}{k} d_i d_i^T$$

with $d_i := \sqrt{k}\, c_i$, $1 \leqq i \leqq n$. This implies that $\Sigma \in \operatorname{con}\{d_i d_i^T;\ 1 \leqq i \leqq k\}$ which is a compact, convex subset of $R^{n\times n}$. By a well known theorem of Caratheodory each point of a compact convex subset A of R^m has a representation as convex combination of $m+1$ extreme points of A. Therefore, $\Sigma = \sum_{j=1}^{m+1} \alpha_j d_{i_j} d_{i_j}^T$, where $1 \leqq i_j \leqq k$, $0 \leqq \alpha_j$, $\sum_{j=1}^{m+1} \alpha_j = 1$ and $m = \min\{k-1, n^2\}$.

g) has been proved by Markham [14].

Remark 2.

1) Proposition 4, b) implies that for $n \leqq 4$ association is equivalent with POD and positive correlation. It is not known to the author whether this result is true also for $n \geqq 5$. This question leads to the difficult and unsolved problem of determination of $\mathfrak{E}(S_n)$.

2) Proposition 4, f) improves on a bound for the index k given by Hall, Newman [9] who proved that k can be chosen smaller than 2^n.

3) The characterization in g) due to Markham [14] has a simple geometric interpretation. If $\Sigma = BB^T$ with $B \in R^{n\times n}$, where B has row vectors $b_1, \cdots, b_n$ and if $q_1, \cdots, q_n$ are the orthogonal vectors obtained from $b_1, \cdots, b_n$ by the Gram-Schmidt orthogonalization process, then the condition on the minors is equivalent to the condition that $b_1, \cdots, b_n$ lie in the convex cone

$$C(q_1, \cdots, q_n) = \left\{ \sum_{i=1}^{n} \alpha_i q_i;\ \alpha_i \geqq 0,\ 1 \leqq i \leqq n,\ \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

3. Positive likelihood ratio dependence and positive stochastic dependence

Within this section we discuss some concepts which are stronger than association. We need the following definition which is essentially due to Barlow, Proschan [2] and Dykstra, Hewett, Thompson [5]. Let for nonnegative function f and measure μ, $f\mu$ denote the measure with density f w.r.t. μ.

DEFINITION 3. Let X_1, X_2 be k, l-dimensional random variables and define:

1) X_1 is stochastically increasing in X_2 ($X_1 \uparrow_{\text{st.}} X_2$) if for all $x, y \in R^l$, $x \leqq y$ implies that

$$P^{X_1|X_2=x} \leqq_{\text{st.}} P^{X_1|X_2=y}$$

where $\leqq_{\text{st.}}$ means stochastic order (of the k-dim. conditional distributions).

2) If $P^{(X_1,X_2)} = f\lambda^n$ (i.e. the distribution of (X_1, X_2) has density f w.r.t. Lebesgue-measure λ^n), then (X_1, X_2) have positive likelihood ratio dependence (plrd) if for all $x_i \leqq y_i$, $i=1, 2$

$$f(x_1, x_2)f(y_1, y_2) \geqq f(x_1, y_2)f(y_1, x_2)\,.$$

Remark 3.

1) If $k=l=1$, plrd is equivalent to the notion TP_2 (totally positive of order 2) (cf. Barlow, Proschan [2], p. 143).
2) (X_1, X_2) plrd is equivalent to the condition that $P^{X_1|X_2=x}$ has a (multivariate) monotone likelihood ratio when x is considered as a parameter. (X_1, X_2) plrd is equivalent to (X_2, X_1) plrd.
3) One can avoid an inconsistency of Definition 3, 2) arising from different choices of f by including an a.s. condition.

The class of matrices which turn out to be central for these dependence concepts are the M-matrices which where introduced by Ostrowski.

DEFINITION 4. Let $A \in R^{n\times n}$, $A=(a_{ij})_{i,j\leqq n}$. Then A is called an M-matrix, if

(11) $\qquad a_{ij} \leqq 0,\ \forall\, i \neq j$, and if all principal minors are positive.

An important result on M-matrices is, that each M-matrix is of monotone kind i.e.

(12) $$A^{-1} \text{ exists and } A^{-1} \geqq 0\,.$$

For several properties of M-matrices see Poole, Boullion [16].

For the application to normal variables we need the following factorization properties.

LEMMA 1. *Let A be positive definite. Then A is an M-matrix*

$$(13)\qquad \begin{aligned} &\Longleftrightarrow \textit{There exists a lower triangular M-matrix } L \textit{ with } A=LL^T. \\ &\Longleftrightarrow \textit{There exists an upper triangular M-matrix } U \textit{ with } A=UU^T. \end{aligned}$$

PROOF. Fiedler and Ptak [7] have shown the existence of a lower triangular M-matrix L and an upper triangular M-matrix U such that

$$(14)\qquad A=LU.$$

Since $A=A^T$ we have $LU=U^TL^T$. Defining $D=L^{-1}U^T=U(L^T)^{-1}$, D is a diagonal matrix and

$$(15)\qquad LD=U^T.$$

(15) implies that $D\geqq 0$ since L, U are M-matrices. So we can define $\tilde{L}=LD^{1/2}$ and obtain $A=\tilde{L}\tilde{L}^T$ with an M-matrix $\tilde{L}$. This proves the first equivalence. For the second equivalence observe that with $A=(a_{ij})$ also $B=(a_{n-i+1,n-j+1})$ is an M-matrix. So there exists a lower triangular M-matrix L with $B=LL^T$. Let $L=(l_{ij})$ and define $U=(l_{n+1-i,n+1-j})$; then U is an upper triangular M-matrix and it is easy to see that $A=UU^T$.

Remark 4. The second equivalence of Lemma 1 was proved in a different way by Jacobsen [10].

Lemma 1, (12) and Proposition 4, g) imply the following corollary.

COROLLARY 1. *If Σ is positive definite and if Σ^{-1} is an M-matrix, then $\Sigma \in P_n$ and, especially, $N(\mu, \Sigma)$-distributed random variables are associated.*

From the following characterization of positive stochastic dependence in normal distributions we isolate the following lemma.

LEMMA 2. *Let $X=(X_1, X_2)$ be $N(0, \Sigma)$-distributed and Σ be positive definite. If Σ^{-1} is an M-matrix, then there exists a random variable Y_2 with independent components and independent of X and, furthermore, a monotonically nondecreasing function h such that X has the same distribution as*

$$(16)\qquad (X_1, h(X_1, Y_2)).$$

PROOF. By Lemma 1 there exists a lower triangular M-matrix L such that $\Sigma^{-1}=L^T L$ or, equivalently, $\Sigma=L^{-1}(L^{-1})^T$. If $Y=(Y_1, Y_2)$ is $N(0, I_n)$-distributed we, therefore, may assume that $X=L^{-1}Y$. Let $L=\begin{pmatrix} L_{11} & 0 \\ L_{12} & L_{22} \end{pmatrix}$ be the partition of L corresponding to X_1, X_2, then L_{11}, L_{22} are M-matrices and $L_{12} \leqq 0$ (componentwise). Therefore, using that

$$L^{-1}=\begin{pmatrix} L_{11}^{-1} & 0 \\ -L_{22}^{-1}L_{12}L_{11}^{-1} & L_{22}^{-1} \end{pmatrix}$$

we obtain

$$X_1=L_{11}^{-1}Y_1$$

and

$$X_2=-L_{22}^{-1}L_{12}L_{11}^{-1}Y_1+L_{22}^{-1}Y_2=-L_{22}^{-1}L_{12}X_1+L_{22}^{-1}Y_2 .$$

Defining $h(x_1, y_2):=-L_{22}^{-1}L_{12}x_1+L_{22}^{-1}y_2$ and using $L_{22}^{-1} \geqq 0$, $L_{12} \leqq 0$ we obtain that h is monotonically nondecreasing and $X=(X_1, h(X_1, Y_2))$.

For $x=(x_1, \cdots, x_n)$ and $R \subset \{1, \cdots, n\}$, $R=(r_1, \cdots, r_k)$ denote by $x_R := (x_{r_1}, \cdots, x_{r_k})$ and $x_{(i)}=(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$.

THEOREM 2. *Let Σ be positive definite and X be $N(0, \Sigma)$-distributed. Then the following conditions are equivalent.*

a) *Σ^{-1} is an M-matrix*

(17) b) *for all $i \leqq n$: $X_i \uparrow_{\text{st.}} X_{(i)}$*

c) *for all $R, S \subset \{1, \cdots, n\}$, $R \cap S=\phi$: $X_R \uparrow_{\text{st.}} X_S$.*

PROOF. a) $\Longrightarrow$ c)

By a simple conditioning argument it holds for $S_1 \subset S_2$ that

(18) $$X_R \uparrow_{\text{st.}} X_{S_2} \text{ implies } X_R \uparrow_{\text{st.}} X_{S_1} .$$

((18) is independent from the normality assumption). Therefore, we may assume that $R+S=\{1, \cdots, n\}$. Furthermore, it is clear from the definition that the condition that Σ^{-1} is an M-matrix, implies that $(Q\Sigma Q^T)^{-1} = (Q^{-1})^T \Sigma^{-1} Q^{-1}$ is an M-matrix for all permutation matrices Q. Now using Lemma 2 with Q corresponding to the partition $R+S=\{1, \cdots, n\}$ (i.e. $Q(X_1, X_2)=(X_R, X_S)$) we obtain a representation $(X_R, X_S)=(X_R, h(X_R, Y_S))$ as in Lemma 2. But this is enough to imply $X_R \uparrow_{\text{st.}} X_S$ (cf. Barlow, Proschan [2], p. 147, Lemma 4.8).

c) $\Longrightarrow$ b)

Take specially $R=\{i\}$, $S=\{1, \cdots, i-1, i+1, \cdots, n\}$.

b) $\Longrightarrow$ a)

Let $\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}$ be a partition of Σ corresponding to $X=(X_1, X_2)$. Then the conditional distribution of X_1 given $X_2=x_2$ is given by

$$\mathrm{P}^{X_1|X_2=x_2}=N(\Sigma_{12}\Sigma_{22}^{-1}x_2, \Sigma_{1.2}) \quad \text{with} \quad \Sigma_{1.2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T \tag{19}$$

(cf. Theorem 2.2.7, p. 47 of Srivastava, Khatri [22]).

Therefore, $X_1 \uparrow_{\mathrm{st.}} X_2$ if and only if

$$\Sigma_{12}\Sigma_{22}^{-1}\geqq 0 . \tag{20}$$

Let now $\Sigma^{-1}=\begin{pmatrix}A_{11} & A_{12}\\ A_{12}^T & A_{22}\end{pmatrix}$, then we obtain from Corollary 1.4.2 of Srivastava, Khatri [22] $\Sigma_{22}^{-1}=A_{22}-A_{12}^TA_{11}^{-1}A_{12}$ and $\Sigma_{12}=-A_{11}^{-1}A_{12}\Sigma_{22}$, which implies that

$$\Sigma_{12}\Sigma_{22}^{-1}=-A_{11}^{-1}A_{12}\geqq 0 . \tag{21}$$

So from $X_1 \uparrow X_{(1)}$ we obtain, that $a_{11}>0$ and $a_{1j}\leqq 0$, $\forall j\neq 1$, where $A=(a_{ij})=\Sigma^{-1}$. Using the above given permutation argument we obtain similarly, that $a_{ii}>0$ and $a_{ij}\leqq 0$, $\forall i\neq j$, $1\leqq i\leqq n$, i.e. Σ^{-1} is an M-matrix.

Remark 5. (20) and (21) may be used to give a characterization of the notion that X is stochastically increasing in sequence, i.e. $X_i \uparrow_{\mathrm{st.}} (X_1, \cdots, X_{i-1})$, $2\leqq i\leqq n$. This notion is equivalent to the condition

$$X_R \uparrow_{\mathrm{st.}} X_S \text{ for all } R\geqq S,\ R\cap S=\phi \tag{22}$$

($R\geqq S$ means that each component of R is larger than each component of S).

For the proof of (22) apply Theorem 4.13 of Barlow, Proschan [2].

Concerning plrd we have the following result.

THEOREM 3. *Let X be $N(0, \Sigma)$-distributed, where Σ is positive definite. Then the following conditions are equivalent:*

(23)

a) *Σ^{-1} is an M-matrix*

b) *$(X_i, X_{(i)})$ are plrd, $1\leqq i\leqq n$*

c) *(X_R, X_S) are plrd for all $R+S=\{1, \cdots, n\}$*

d) *(X_i, X_j) are plrd for all $i\neq j$.*

PROOF. The equilvalence of a) and d) is due to Sarkar [18] and Barlow, Proschan [2]. Let $A=\begin{pmatrix}A_{11} & A_{12}\\ A_{12}^T & A_{22}\end{pmatrix}=\Sigma^{-1}$ be a partition corresponding to $X=(X_1, X_2)$ and let $f(x_1, x_2)=((2\pi)^n \det \Sigma)^{1/2} e^{-(x_1,x_2)A(x_1,x_2)^T/2}$ (we consider x_1, x_2 as row vectors). Then (X_1, X_2) is plrd

$$\begin{aligned}
&\Longleftrightarrow \text{ for all } x=(x_1, x_2)\leqq(y_1, y_2)=y \\
&\qquad\qquad f(x_1, x_2)f(y_1, y_2)\geqq f(x_1, y_2)f(y_1, x_2) \\
&\Longleftrightarrow x_1A_{12}x_2^T+y_1A_{12}y_2^T\leqq x_1A_{12}y_2^T+y_1A_{12}x_2^T \qquad \text{for } x\leqq y \\
&\Longleftrightarrow (x_1-y_1)A_{12}(y_2-x_2)^T\geqq 0 \qquad \text{for } x_i\leqq y_i,\ i=1, 2 \\
&\Longleftrightarrow A_{12}\leqq 0\,. \qquad (24)
\end{aligned}$$

Using a permutation argument (24) implies the equivalence of a), b), c).

Remark 6. Theorem 3 shows a difference between the notions of plrd and positive stochastic dependence. While (X_i, X_j) plrd for all $i \neq j$ is equivalent to the condition that Σ^{-1} is an M-matrix, it follows from (20), that the condition $X_i \uparrow_{\text{st.}} X_j$ for all $i\neq j$ is equivalent to the much weaker assumption that $\Sigma \in S_n$ in other words to the positive correlation assumption. In spite of that 'globally' both concepts are equal for normal distributions.

RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN

REFERENCES

[1] Abdel-Hameed, M. and Sampson, A. R. (1978). Positive dependence of the bivariate and trivariate absolute normal t, χ^2 and F distributions, *Ann. Statist.*, **6**, 1360-1368.

[2] Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*, Holt, New York.

[3] Das Gupta, S., Eaton, M. L., Olkin, I., Perlman, M. D., Savage, L. J. and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions, *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, **2**, Univ. of California Press, Berkeley California, 241-265.

[4] Dunn, O. J. (1958). Estimation of the means of dependent variables, *Ann. Math. Statist.*, **29**, 1095-1110.

[5] Dykstra, R. L., Hewett, J. E. and Thompson, W. A., Jr. (1973). Events which are almost independent, *Ann. Statist.*, **1**, 674-681.

[6] Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications, *Ann. Math. Statist.*, **38**, 1466-1474.

[7] Fiedler, M. and Ptak, V. (1962). On matrices with non-positive off-diagonal elements and positive principal minors, *Czech. Math. J.*, **12**, 382-400.

[8] Gray, L. J. and Wilson, D. G. (1980). Nonnegative factorization of positive semidefinite nonnegative matrices, *Linear Algebra and Appl.*, **31**, 119-127.

[9] Hall, M., Jr. and Newman, M. (1963). Copositive and completely positive quadratic forms, *Proc. Camb. Phil. Soc.*, **59**, 329-339.

[10] Jacobsen, D. H. (1974). Factorization of symmetric M-matrices, *Linear Algebra and Appl.*, **9**, 275-278.

[11] Jogdeo, K. (1977). Association and probability inequalities, *Ann. Statist.*, **5**, 495-504.

[12] Khatri, C. G. (1967). On certain inequalities for normal distributions and their ap-

plications to simultaneous confidence bouuds, *Ann. Math. Statist.*, **38**, 1853-1867.

[13] Lehmann, E. L. (1966). Some concepts of dependence, *Ann. Math. Statist.*, **37**, 1137-1153.

[14] Markham, T. L. (1971). Factorization of completely positive matrices, *Proc. Camb. Phil. Soc.*, **69**, 53-58.

[15] Plesken, W., Rüschendorf, L. and Krafft, O. (1980). Nonnegative factorization of nonnegative definite, nonnegative matrices, unpublished manuscript.

[16] Poole, G. and Boullion, T. (1974). A survey on *M*-matrices, *SIAM Rev.*, **16**, 419-427.

[17] Rüschendorf, L. (1980). Weak association of random variables, to appear in *J. Multivariate Anal.*

[18] Sarkar, T. K. (1969). Some lower bounds of reliability, *Technical Report* No. 124, Department of Operations Research and Statistics, Stanford University, California.

[19] Scott, A. (1967). A note on conservative regions for the mean of a multivariate normal, *Ann. Math. Statist.*, **38**, 278-280.

[20] Sidak, Z. (1975). A note on C. G. Khatri and A. Scott's papers on multivariate normal distributions, *Ann. Inst. Statist. Math.*, **27**, 181-184.

[21] Slepian, D. (1962). The one-sided barrier problem for Gaussian noise, *Bell Syst. Tech. J.*, **41**, 463-501.

[22] Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*, North Holland, New York.