Distributions with Fixed Marginals and Related Topics
IMS Lecture Notes – Monograph Series Vol. 28, 1996

MIXTURES OF MULTIVARIATE QUASI-EXTREMAL DISTRIBUTIONS HAVING GIVEN MARGINALS

BY ENE-MARGIT TIIT*
University of Tartu

Let $F_i (i = 1, \ldots, k)$ be given univariate distributions and Π be the set of k-variate distributions having marginals F_i. In this paper the extremal and quasi-extremal multivariate distributions having the given marginals F_i are defined and their properties are examined. Since the set Π is convex, all mixtures of extremal distributions have the same marginals. Furthermore, the correlation matrices of extremal distributions are extremal in the set of correlation matrices, and there exists a one-to-one correspondence between the extremal distributions and the extremal correlation matrices. For a given correlation matrix R, its decomposition by extremal correlation matrices is proposed as an alternative model to factor analysis. The methods are compared and the conditions of their coincidence are indicated. All results obtained for the case of extremal distributions are generalised to quasi-extremal distributions.

1. Introduction. Let a set of univariate distribution functions F_i ($i = 1, \ldots, k$) be given. We are interested in the set $\Pi(F_1, \ldots, F_k) = \Pi$ of all k-variate distributions having the marginals F_i.

In the case $k = 2$ the set Π has a *minimal* and a *maximal* element – the so-called *Fréchet bounds* (see Hoeffding (1940) and Fréchet (1951)). If $k > 2$, then the maximal element (in the sense of stochastic ordering) of the set Π always exists (see Fèron (1956), Dall'Aglio (1972), Kemp (1973), Ruiz-Rivas (1979), Cuadras (1981), Tiit (1984)). In this paper some properties of the maximal distribution will be presented (see Section 2). In general, a smallest element of the set Π does not exist (see Fèron (1956), Dall'Aglio (1960), Ruiz-Rivas (1979), Tiit (1984), Kotz and Tiit (1992)). Some special cases in which a minimal distribution does exist, are indicated in the literature (see Dall'Aglio (1960) and (1991), Rüschendorf (1991)).

---

* This work was supported by Estonian Science Foundation Grant No. 551. The financial suport of the International Science Foundation is also gratefully acknowledged.

AMS 1991 Subject Classification: Primary 62E10; Secondary 62H25.

Key words and phrases: Correlation matrix, factor analysis, Fréchet bounds, extremal distribution, given marginals, convex decomposition, mixture of distributions.

For each set of marginal distributions $F_1, \ldots, F_k$ there exist 2^{k-1} distributions having some special properties that allow one to say that they are *extremal* (see Tiit (1984), (1986), (1992), Kotz and Tiit, (1992)). In this paper the definition and analytical rule for the construction of these extremal distributions for given marginals F_i $(i = 1, \ldots, k)$ is introduced (see Section 3). The rule involves *k-dimensional copulas* (see Sklar (1959), Schweizer (1991)), but in our paper another formalization will be used. Some properties of the extremal distributions will also be considered in the same Section 3.

The set of extremal distributions can be expanded by introducing the *quasi-extremal distributions* (see Tiit (1986), Tiit and Thetloff (1994)), i.e. distributions, having several independent uni- or multivariate marginals which are extremal. The definition and some properties of quasi-extremal distributions will be given in Section 4.

Let $\mathbf{E}(F_1, \ldots, F_k) = \mathbf{E}$ and $\mathbf{Q}(F_1, \ldots, F_k) = \mathbf{Q}$ be the sets of extremal and quasi-extremal distributions having the given marginals. Then the following obvious inclusions hold:

$$\mathbf{E}(F_1, \ldots, F_k) \subset \mathbf{Q}(F_1, \ldots, F_k) \subset \mathbf{\Pi}(F_1, \ldots, F_k).$$

Since the set $\mathbf{\Pi}$ is closed under convex combinations (see Kellerer (1964), Tiit (1984) etc.), it is of interest to investigate properties of *mixtures of extremal or quasi-extremal distributions* having the given marginals. These mixtures and their corresponding *convex combinations of extremal and quasi-extremal correlation matrices* will be considered in Section 5. In the same section, the one-to-one correspondence between the extremal (or quasi-extremal) distributions and extremal (or, correspondingly, quasi-extremal) correlation matrices will be deduced. The latter connection seems to have useful applications in multivariate statistical analysis.

Traditionally, the aim of building multivariate statistical models is to get an *approximation* of a *function of exploratory variables* for the given *dependent variable.* Often, this function is considered to be *linear.* In several cases, instead of the *model on the level of variables*, an alternative *model on the level of distributions* can be introduced. The latter model is obtained by approximating the given multivariate distribution by a *convex combination* of some so-called explanatory distributions. One possibility for these explanatory distributions is *the set of extremal distributions having the same marginals* (see Tiit, Thetloff (1994)). Since the multivariate extremal distributions are *singular*, i.e. equivalent to univariate distributions (which are either continuous or discrete, depending on the character of the given marginals), the exploitation of extremal distributions is especially effective in questions concerning *the reduction of dimensionality.*

In Section 6 the *convex-extremal decomposition of a correlation matrix* will be introduced as a methodological alternative to factor analysis (see Tiit and Thetloff (1994)). The similarities and differences between the factor decomposition (which forms the mathematical basis for factor and component analysis) and the convex-extremal decomposition are considered.

2. Maximal and Minimal Distributions Having Given Marginals. In the bivariate case Hoeffding (see Hoeffding, (1940)) and Fréchet (see Fréchet (1951)) determined the maximal and minimal distributions F^+ and F^- having the given marginals F_1, F_2,

$$F^+(x_1, x_2) = \min(F_1(x_1), F_2(x_2)), \tag{1}$$

and

$$F^-(x_1, x_2) = \max(0, F_1(x_1) + F_2(x_2) - 1). \tag{2}$$

The distributions F^+ and F^- are the so-called *Fréchet bounds* (see Fréchet (1957), Dall'Aglio (1991), Kotz and Tiit (1992)).

In the case of bivariate distributions minimality and maximality follow from the stochastic ordering

$$F^- \leq F \leq F^+, \tag{3}$$

where F is an arbitrary distribution from the set $\mathbf{\Pi}(F_1, F_2)$. The same ordering (3) holds for several functionals depending on the bivariate distribution F. The most important of these is the linear correlation coefficient, which satisfies

$$r^+ \leq r \leq r^-, \tag{4}$$

as proved by Fréchet (see Fréchet (1951)).

The definition (1) of the maximal distribution F^+ can easily be generalized to any number k of marginals.

DEFINITION 1. *Let F_i be given univariate distributions ($i = 1, \ldots, k$). Then the k-variate distribution F^+, defined by*

$$F^+(x_1, \ldots, x_k) = \min_{1 \leq i \leq k} F_i(x_i) \tag{5}$$

is said to be the maximal k-variate distribution having the marginals $F_1, \ldots, F_k$.

The expression (5) is an immediate generalization of the expression (1). The maximal distribution belongs to the set $\mathbf{\Pi}$ and is its maximum in the sense of stochastic ordering, i.e.,

$$F^+ \geq F, \tag{6}$$

for any F in $\mathbf{\Pi}$. The maximal distribution F^+ has the following well-known properties (see, for instance, Tiit (1986), (1992) and Rüschendorf (1991)).

1^0. F^+ is singular: its support lies on a curve in R^k.

2^0. When all marginal distributions are identical and have support on $[0, 1]$, the support of F^+ lies on the main diagonal of the k-dimensional unit cube.

3^0. The curve supporting the maximal distribution is monotonically non-decreasing in each coordinate.

4^0. All q-variate marginal distributions $(1 < q < k)$ of F^+ are maximal.

5^0. F^+ is unique.

One important property of the maximal distribution is expressed in the following theorem:

THEOREM 1. *The multivariate maximal distribution F^+ is uniquely defined by its bivariate marginals $F_{ij}(i, j = 1, \ldots, k,\ i \neq j)$.*

PROOF. The proof for the case $k = 3$ is due to Dall-Aglio (1960). The extension to higher dimensions is straightforward and well-known.

Theorem 1 has the following consequences:

COROLLARY 1. *A k-variate distribution having the given univariate marginal distributions $F_1, \ldots, F_k$, is maximal, iff all its bivariate marginals are maximal (in the sense of Hoeffding-Fréchet).*

DEFINITION 2. Let $F_1, \ldots, F_k$ be given univariate distributions of the second order (all variances exist). The correlation matrix R^+, consisting of maximal (in the sense of inequality (4)) correlation coefficients $r_{ij}^+ = r^+(F_i, F_j)$ $(i = 1, \ldots, k-1,\ j = i, \ldots, k)$ is said to be *maximal.*

COROLLARY 2. *The k-variate distribution, having the given univariate marginal distributions $F_1, \ldots, F_k$, is maximal, iff its correlation matrix is maximal.*

The proof of Corollary 2 follows from Corollary 1 and the fact that the only bivariate distribution with maximal correlation r_{ij}^+ is the maximal bivariate distribution.

From Corollary 2 and the defining formula (1) the following simple conclusion follows:

COROLLARY 3. *If all given marginals $F_1, \ldots, F_k$ are of second order and identical, then all elements of the maximal correlation matrix are equal to one.*

3. Extremal Multivariate Distributions with given Marginals. The concept of extremal distributions forms a generalization of the concept of maximal distributions. The extremal distributions exist for all (finite) sets of given marginals, no matter if they are equal or non-equal, continuous, discrete or mixed. This concept depends essentially on the *partition of the index-set.*

DEFINITION 3. *Let* $I^o = \{1, \dots, k\}$ *be a given index-set, and* $I = \{i_1, \dots, i_q\}$ *an arbitrary subset of* I^o, *satisfying the condition*

$$1 \in I. \tag{7}$$

Let $I^c = I^o \setminus I = \{j_1, \dots, j_s\}$, *where* $s = k - q$. *Then we say that the pair* (I, I^c) *defines a partition of the index-set* I^o.

We now make the following

DEFINITION 4. *Let* $F_1, \dots, F_k$ *be given univariate distributions, and let* (I, I^c) *be a given partition of the index-set* $\{1, \dots, k\}$. *Then the distribution* F^{I,I^c}, *defined via*

$$F^{I,I^c}(x_1, \dots, x_k) = \max(0, (\min_{i \in I} F_i(x_i) + \min_{j \in I^c} F_j(x_j) - 1)), \tag{8}$$

is said to be extremal.

The correctness of Definition 4 is a consequence of

THEOREM 2. *The function* F^{I,I^c}, *defined by the formula (8), is a* k*-variate distribution function belonging to the set* $\mathbf{\Pi}$.

PROOF. A. We first prove that F^{I,I^c} is a k-variate distribution function.

1^0. F^{I,I^c} is a function of k real arguments $x_1, \dots, x_k$.

2^0. From Definition 1 it follows that $\min_{i \in I} F_i(x_i)$ is a distribution function, belonging to the set $\mathbf{\Pi}(F_i), i \in I$, and similarly for $\min_{j \in I^c} F_j(x_j)$. Both of them are singular (see property 1^0 of a maximal distribution) and hence equivalent to a univariate distribution function.

3^0. From formula (2) it follows that if $\min_{i \in I} F_i$ and $\min_{j \in I^c} F_j$ are singular distribution functions, then F^{I,I^c}, defined by the formula (8), is a distribution function, as well.

B. From the construction it follows that the distribution F^{I,I^c} has the given marginals $F_i (i = 1, \dots, k)$, hence belongs to the set $\mathbf{\Pi}$.

The number of extremal distributions defined by the formula (8) is equal to the number of partitions of the index-set that satisfy the condition (7). Assuming that the trivial partition $(I^o, \emptyset)$, defining the maximal distribution, belongs to the set of all partitions, we see that the number of such partitions is 2^{k-1}, whence the cardinality of the set $\mathbf{E}(F_1, \ldots, F_k)$ is also 2^{k-1}.

Extremal distributions have several properties that are quite close to those of maximal distributions (see Tiit (1992), Tiit (1994)).

1^0. Any extremal distribution is singular. Its support lies on a curve in R^k.

2^0. In the special case, when all marginal distributions are symmetric, identical and have support in $[0,\ 1]$, the support of each extremal distribution lies on one of the diagonals of the k-dimensional unit cube.

3^0. The support of every extremal distribution F^{I,I^c} is a curve in R^k which is monotonic, non-decreasing in coordinates $x_1, \ldots, x_q$ and non-increasing in the rest.

4^0. Any multivariate marginal distribution F^J of the extremal distribution F^{I,I^c}, defined by an index-set $J = \{g_1, \ldots, g_t\}$ is maximal, if $J \subset I$ or $J \subset I^c$.

5^0. Bivariate marginals F_{ij} of the extremal distribution F^{I,I^c} are minimal if $i \in I$ and $j \in I^c$ or if $i \in I^c$ and $j \in I$.

6^0. If all given marginals are of the second order (that means, all variances exist), then the correlation matrix $R^{I,I^c} = R(F^{I,I^c})$ of the extremal distribution F^{I,I^c} is defined by its elements in the following way:

$$r_{ij}^{I,I^c} = \begin{cases} r_{ij}^+, & \text{if } i \in I \text{ and } j \in I \text{ or } i \in I^c \text{ and } j \in I^c, \\ r_{ij}^-, & \text{if } i \in I \text{ and } j \in I^c \text{ or } i \in I^c \text{ and } j \in I, \end{cases} \tag{9}$$

where r_{ij}^+ and r_{ij}^- are, respectively, the maximal and minimal correlation coefficients of the marginals F_i, F_j $(i, j = 1, \ldots, k)$, see the formula (5).

DEFINITION 5. *A correlation matrix consisting of maximal and minimal correlations only is said to be an extremal correlation matrix.*

From Definition 5 and property 6^0 the following is immediate:

COROLLARY 4. *The correlation matrix of an extremal distribution is extremal.*

We will denote the set of all extremal correlation matrices, defined by the marginals $F_1, \ldots, F_k$, by the symbol $\mathbf{R}^I(F_1, \ldots, F_k)$. If all marginal distributions F_i $(i = 1, \ldots, k)$ are equal and symmetric, then the extremal correlation matrices consist solely of ones and minus ones. In this case the extremal correlation matrices depend only on the partition (I, I^c), and we will denote this set by the symbol $\mathbf{R}^I$. As regards the connection between the extremal distributions and extremal correlation matrices, we have the following theorem (see Tiit (1986)):

THEOREM 3. *Let the univariate distributions $F_1, \ldots, F_k$ be given. Then the set of extremal distributions $\mathbf{E}(F_1, \ldots, F_k)$ and the set of extremal correlation matrices $\mathbf{R}^I(F_1, \ldots, F_k)$ are in one-to-one correspondence.*

PROOF. From Corollary 4 the correspondence $\mathbf{E} \Rightarrow \mathbf{R}^I$ follows immediately.

To establish the opposite correspondence, we show that a given extremal correlation matrix R^* uniquely defines a partition (I, I^c) of the index-set I^o. For, given the extremal correlation matrix $R^* = (r^*_{ij})$, it is easy to recover the partition (I, I^c) in the following way:

$$i \in I, \text{ if } r^*_{1i} > 0, \quad i \in I^c, \text{ if } r^*_{1i} < 0 \quad (i = 1, \ldots, k).$$

Since no element of the extremal correlation matrix can be equal to zero, the definition of the partition (I, I^c) is unique. Using the partition (I, I^c) the distribution F^{I,I^c} can be defined in the standard way with the aid of the formula (8).

The number of different extremal correlation matrices R^{I,I^c} is naturally equal to the number of extremal distributions, hence 2^{k-1}. From this it follows that the number of different correlation matrices of order k consisting of ones and minus ones, hence having minimal rank, is equal to 2^{k-1} as well.

4. Quasi-Extremal Multivariate Distributions. To enlarge the set $\mathbf{E}$ of extremal distributions having given univariate marginal distributions $F_1, \ldots, F_k$, we have to use more complicated partitions of the initial index-set I^o.

DEFINITION 6. *Let I^o be the index-set $\{1, \ldots, k\}$. We say that $(L_1, \ldots, L_t) = L$ is a t-partition $(1 \le t \le k)$ of the index-set I^o if the subsets $L_f = \{i_1^f, \ldots, i_{q_f}^f\} (f = 1, \ldots, t)$ of I^o are non-overlapping and fulfill the following conditions:*

$$i_1^{f_1} < i_1^{f_2}, \text{ if } f_1 < f_2 \ (f_1 = 1, , \ldots, f_2 - 1, \ f_2 = 2, \ldots, t) \tag{10}$$

and

$$\bigcup_{f=1}^{t} L_f = I^o.$$

Naturally, the cardinalities q_f of the subsets L_f satisfy the condition

$$\sum_{f=1}^{t} q_f = k.$$

The *quasi-extremal multivariate distribution* corresponding to a t-partition L is defined in the following way:

1^0. All q_f-variate marginals, corresponding to subsets L_f of indices ($f = 1, \dots, t$), are extremal.

2^0. The q_f-variate marginals, corresponding to *different subsets* $L_f (f = 1, \dots t)$ are independent.

For the precise formulation of this concept we need the definition of a *t-double partition* of the initial index-set I^o.

DEFINITION 7. *Let L be a t-partition of the index-set I^o. We say that (L, J) is a t-double partition of the index-set I^o if, for every subset L_f, there is defined a two-partition (J_f, J_f^c) so that the subsets $J_f = (j_1^f, \dots, j_{q_f}^f)$ and $J_f^c = (l_1^f, \dots, l_s^f)$ of L_f are non-overlapping and satisfy the following conditions:*

$$j_1^f = i_1^f, \tag{11}$$

$$J_f \cup J_f^c = L_f \ (f = 1, \dots, t).$$

It is easy to see that condition (11) generalizes condition (7) to the case of a t-partition of the initial subset I_0.

DEFINITION 8. *Let univariate distributions $F_1, \dots, F_k$ and a t-double partition (L, J) of the index-set I^o be given. Then, the distribution $F^{(L,J)}$ defined via*

$$F^{(L,J)}(x_1, \dots, x_k) = \prod_{f=1}^{t} (\max(0, (\min_{i \in J_f} F_i(x_i) + \min_{j \in J_f^c} F_j(x_j) - 1))), \tag{12}$$

is said to be quasi-extremal.

Definition (12) is an immediate generalization of definition (8).

We shall use the symbol $\mathbf{Q}(F_1, \dots, F_k) = \mathbf{Q}$ to denote the set of all quasi-extremal distributions having the marginals $F_1, \dots, F_k$. The cardinality of the set $\mathbf{Q}$, which depends on k, can be calculated (see Tiit (1986), Tiit, Tammet (1994)) and increases rapidly with k. The correlation matrix of a quasi-extremal distribution is said to be a *quasi-extremal correlation matrix.* The quasi-extremal correlation matrix can be defined immediately by the t-double partition (L, J) and the given marginals in the following way.

DEFINITION 9. *Let (L, J) be a t-double partition of the given index-set I^o. The correlation matrix $R^{(L,J)} = (r_{ij}^{(L,J)})$ is said to be quasi-extremal if it* *is defined element-wise in the following way:*

$$r_{ij}^{(L,J)} = \begin{cases} 0, & \text{if } i \in L_{f_1} \text{ and } j \in L_{f_2}, f_1 \neq f_2, \\ r_{ij}^{+}, & \text{if } i \in J_f \text{ and } j \in J_f \text{ or } i \in J_f^c \text{ and } j \in J_f^c \ (f = 1, \ldots, t), \\ r_{ij}^{-}, & \text{if } i \in J_f \text{ and } j \in J_f^c \text{ or } i \in J_f^c \text{ and } j \in J_f \ (f = 1, \ldots, t). \end{cases} \quad (13)$$

where r_{ij}^{+} and r_{ij}^{-} are the maximal and minimal correlation coefficients for the given marginals F_i, F_j, respectively.

Definition (13) is an immediate generalization of definition (9) for the case in which the correlation matrix consists of t independent blocks.

When all the given marginals in the formula (12) are equal and symmetric, all minimal correlation coefficients r_{ij}^{-} are equal to -1 and all maximal correlation coefficients are equal to 1.

We shall use the symbol $\mathbf{R}^{(\mathbf{L},\mathbf{J})}(F_1, \ldots, F_k)$ to denote the set of all quasi-extremal correlation matrices for the given marginals $F_1, \ldots, F_k$. If the marginals are equal and symmetric, we shall simply write $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$.

There is a correspondence between quasi-extremal distributions and quasi-extremal correlation matrices which is similar to the one between extremal distributions and extremal correlation matrices (see Theorem 3). Specifically, we have

COROLLARY 5. *Let univariate distributions of the second order $F_1, \ldots, F_k$ be given. Then the set $\mathbf{Q}$ of quasi-extremal distributions and the set $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$ of quasi-extremal correlation matrices (with the same given marginals) are in one-to-one correspondence.*

The important subclasses of *quasi-maximal distributions* and *quasi-maximal correlation matrices* can be extracted from the classes of quasi-extremal distributions and quasi-extremal correlation matrices.

DEFINITION 10. *Let L be a t-partition of the given index-set I^o. The distribution F^L is said to be quasi-maximal if it is defined via*

$$F^L(x_1, \ldots, x_k) = \prod_{f=1}^{t} (\min_{i \in L_f} F_i(x_i)).$$

The correlation matrix R^L of a quasi-maximal distribution is said to be *quasi-maximal.* The quasi-maximal correlation matrix can be defined element-wise by the t-partition L and the given marginal distributions $F_1, \ldots, F_k$ in the following way:

$$r_{ij}^{L} = \begin{cases} 0, \text{ if } i \in L_{f_1} \text{ and } j \in L_{f_2}, f_1 \neq f_2, \\ r_{ij}^{+}, \text{ if } i \in L_f \text{ and } j \in L_f, (f = 1, \ldots, t). \end{cases} \tag{14}$$

Formula (14) is a special case of formula (13), when all 'negative' subsets J_f^c of the parts L_f are empty ($f = 1, \ldots, t$). In the case of equal marginal distributions all maximal correlation coefficients equal to one.

We shall use the symbol $\mathbf{R}^{\mathbf{L}}(F_1, \ldots, F_k)$ to denote the set of all quasi-maximal correlation matrices having the given marginals. If the marginals are equal, we shall simply write $\mathbf{R}^{\mathbf{L}}$.

It is clear that the sets of correlation matrices satisfy

$$\mathbf{R}^{\mathbf{L}} \subset \mathbf{R}^{(\mathbf{L},\mathbf{J})}, \ \mathbf{R}^{I} \subset \mathbf{R}^{(\mathbf{L},\mathbf{J})}.$$

5. Mixtures of Extremal and Quasi-Extremal Distributions. It is well-known that the set $\mathbf{\Pi}$ is closed under the formation of mixtures. That is why the use of mixtures of extremal distributions seems to be quite promising in solving different multivariate problems.

DEFINITION 11. *Let the k-variate distributions $G_1, \ldots, G_m$ and the non-negative numbers (weights) $w_1, \ldots, w_m$ such that*

$$\sum_{g=1}^{m} w_g = 1 \tag{15}$$

be given. Then the distribution G^*, defined via

$$G^* = \sum_{g=1}^{m} w_g G_g, \tag{16}$$

is said to be a mixture of the distributions G_g.

The mixture of k-variate distributions is, naturally, k-variate and has the following useful properties (see Tiit (1986) and Tiit (1992)):

1^0. If all components of the mixture have the same marginals, then the mixture has these same marginals as well.

2^0. All initial (marginal and mixed) moments ν_h of the mixture are expressed as convex combinations of the same moments of the components and with the same weights as in the mixture.

3^0. If all components of the mixture have the same marginal distributions (of the second order), then the correlation matrix of the mixture is the convex combination of the correlation matrices of the components, where the weights are again the same as in the mixture.

Using the property 3^0 for the mixture of extremal (quasi-extremal) distributions, defined by the formulae (15, 16), we obtain the following:

COROLLARY 6. *Let G^* be a mixture* (16) *of extremal (quasi-extremal) distributions, having the fixed marginal distributions (of the second order) $F_1, \dots, F_k$, i.e., let*

$$G_g := F^{I_g, I_g^c} \text{ or } G_g := F^{(L_g, J_g)} \ (g = 1, \dots, m).$$

Then the correlation matrix of the mixture G^* can be expressed as a convex combination of the extremal (quasi-extremal) correlation matrices:

$$R(G^*) = \sum_{g=1}^{m} w_g R(G_g). \tag{17}$$

The correspondence between the extremal (quasi-extremal) distributions and the corresponding correlation matrices, established in Theorem 3 and Corollary 5, remains valid in the case of their mixtures and convex combinations having the same weights.

COROLLARY 7. *Let the univariate distributions $F_1, \dots, F_k$ be given and let R^* be a convex combination (17) of some extremal (quasi-extremal) correlation matrices R_g corresponding to the marginals F_i $(i = 1, \dots, k)$. Then the mixture G^*, defined by formula (16) using the extremal (quasi-extremal) distributions corresponding to R_g and the same weights w_g, belongs to the set $\mathbf{\Pi}(F_1, \dots, F_k)$ and has the correlation matrix R^*.*

6. Convex-Extremal Decomposition and Factor Decomposition of a Given Correlation Matrix. Formula (17) can also be considered as a *decomposition of the correlation matrix $R(G^*)$* in the following sense:

DEFINITION 12. *Let R be a correlation matrix of order $k \geq 2$, and let $\mathbf{R}$ be a class of correlation matrices of the same order. If there exist non-negative coefficients w_g, fulfilling condition (15), so that the equation*

$$R = \sum_{g=1}^{m} w_g R_g, \tag{18}$$

holds, with

$$R_g \in \mathbf{R} \ (g = 1, \dots, m),$$

then we say that the correlation matrix R has a convex decomposition in the class $\mathbf{R}$, *or, briefly, that the matrix R is decomposable in the class* $\mathbf{R}$.

Property 3^0 of the mixture of distributions gives us the key for the interpretation of the convex decomposition: we can regard the initial distribution as a *mixture* of distributions, characterized by correlation matrices R_g, in which the marginal distributions of all components of the mixture are the same. The weights of the mixture are equal to the coefficients w_g of the convex decomposition (18).

To us it seems to be very appropriate to use as the class $\mathbf{R}$ either the class of extremal correlation matrices $\mathbf{R}^I$, the class of quasi-extremal correlation matrices $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$, or the class of quasi-maximal matrices $\mathbf{R}^{\mathbf{L}}$. The reason is the fact that distributions having these correlation matrices, i.e. extremal, quasi-extremal and quasi-maximal distributions, are in some sense the simplest possible ones. This methodology seems particularly fruitful when all marginals are equal (up to linear transformations) and symmetric. In this case all elements of the sets $\mathbf{R}^I$, $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$ and $\mathbf{R}^{\mathbf{L}}$ depend on the dimension k only, but do not depend on the marginal distributions. Hence, in this case, the matrices are in some sense *standard*.

We shall introduce the term *convex-extremal decomposition* to indicate the decomposition (18) when $\mathbf{R} = \mathbf{R}^I$ or $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$.

The following are some properties of the convex-extremal decomposition of a correlation matrix (see Tiit (1986)):

1^0. If the convex-extremal decomposition of a correlation matrix R exists, then there exists a decomposition having not more than $m = k(k-1)/2+1$ terms.

2^0. There exist correlation matrices that are not decomposable in the classes $\mathbf{R}^I$ and $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$.

3^0. If a correlation matrix R is decomposable in the class $\mathbf{R}^I$ or $\mathbf{R}^{\mathbf{L}}$, then it is, evidently, also decomposable in the class $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$.

4^0. If all given marginals are equal and symmetric, then every correlation matrix decomposable in the class $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$, is also decomposable in the class $\mathbf{R}^I$.

5^0. In general, the decomposition (18) is not unique in the class $\mathbf{R}^{(\mathbf{L},\mathbf{J})}$.

In the following discussion we assume for the sake of simplicity that the given univariate distributions F_i are of the second order, identical and symmetric. In this case all extremal correlation matrices consist solely of ones and minus ones, and quasi-extremal matrices consist of ones, minus ones and zeros.

Our aim is to compare the convex-extremal decomposition with the most commonly used decomposition of the correlation matrix – the *factor decomposition.*

As is well-known, factor analysis uses the decomposition of the given correlation matrix R of order k by its eigenvectors and eigenvalues:

$$R = H \Lambda H', \tag{19}$$

where H is the matrix of eigenvectors $h_g = (h_1^g, \ldots, h_k^g)$, $g = 1, \ldots, k$, Λ is the matrix of non-increasingly ordered eigenvalues, $\Lambda = diag(\lambda_1, \ldots, \lambda_k)$ and k is the order of the matrix R. We assume for the sake of simplicity that the eigenvalues are different, $\lambda_i \neq \lambda_j$, if $i \neq j$. In factor analysis, instead of the matrix H, the block of H, consisting of q first columns $(q < k)$, is used.

The convex-extremal decomposition of the correlation matrix has the form (18), where $R_g := R^{I_g}$. Here the number m of components satisfies the condition $1 \leq m \leq k(k-1)/2+1$ (see property 1^0 of the convex decomposition).

The factor decomposition of a correlation matrix can be interpreted as a *linear transformation of variables* – the initial variables are expressed as linear combinations of a small number of hypothetical factors having some good properties, e.g., standardization or orthogonality.

The convex-extremal decomposition of a correlation matrix can be interpreted as a *mixture of extremal matrices* and from this it follows that *the initial distribution can be expressed as a mixture of extremal distributions,* or – on the level of applications – *the population can be expressed as a sum of subpopulations having the same marginal distributions but a substantially different dependence structure.*

In certain cases the above mentioned decompositions give identical results. To determine these cases we proceed as follows.

Let $A = (a_{ij})$ and $B = (b_{ij})$ both be $p \times q$-matrices. We say that they are *orthogonal*, if the equation

$$\sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij} \cdot b_{ij} = 0$$

holds.

Let R be a correlation matrix and rewrite the factor decomposition (19) as follows:

$$R = \sum_{g=1}^{k} \lambda_g h_g h_g' = \sum_{g=1}^{k} v_g Q_g, \tag{20}$$

where $v_g = \lambda_g/k$ and the matrix $Q_g = kh_g h'_g$ is defined by its elements:

$$q^g_{ij} = kh^g_i h^g_j. \tag{21}$$

Concerning this decomposition, the following lemmas hold:

LEMMA 1. *The coefficients v_g of the sum* (20) *satisfy the condition* (15).

PROOF. This follows immediately from the well-known properties of the eigenvalues of the correlation matrix:

$$\lambda_j \geq 0, \quad \sum_{j=1}^{k} \frac{\lambda_j}{k} = 1.$$

LEMMA 2. *The matrix Q_g $(g = 1, \ldots, k)$, defined by the formula* (20), *has the following properties:*

1^0. *Q_g is positive definite;*

2^0. $\mathrm{Tr} Q_g = k$;

3^0. *If $g \neq f$, then Q_g and Q_f are orthogonal.*

PROOF. These facts follow from the definition of the matrices Q_g and from the properties of eigenvectors of a correlation matrix, exhibited in the following three equations:

$$xQx' = k \sum_{i=1}^{k} \sum_{j=1}^{k} x_i h^g_i h^g_j x_j = \big(\sum_{i=1}^{k} x_i h^g_i\big)^2 \geq 0;$$

$$\mathrm{Tr}\, Q_g = \sum_{i=1}^{k} q^g_{ii} = k \sum_{i=1}^{k} \big(h^g_i\big)^2 = k;$$

$$\sum_{i=1}^{k} \sum_{j=1}^{k} q^g_{ij} q^f_{ij} \; = \; k^2 \big(\sum_{i=1}^{k} h^g_i h^f_i\big)^2 \; = \; 0.$$

LEMMA 3. *The matrix Q_g, defined by the formula* (20), *is a correlation matrix if and only if all components of the eigenvector h_g have equal absolute values.*

PROOF. A. Let us suppose that $\mid h^g_j \mid = \max_{1 \leq i \leq k} \mid h^g_i \mid$ and that there exists an index l so that the inequality $\mid h^g_l \mid < \mid h^g_j \mid$ is valid. Let us use the following notation: $\bar{h}(2) = 1/k \sum_{i=1}^{k} (h^g_i)^2$. Then $\bar{h}(2)$ is the average of the squares of the components of the eigenvector h_g; and from the properties of an average we have the following inequality:

$$| h_j^g |^2 > 1/k \sum_{i=1}^{k} (h_i^g)^2 = 1/k.$$

Hence $| q_{jj}^g | > 1$ and Q_g cannot be a correlation matrix.

B. Let $| h_i^g |= h$ for $i = 1, \ldots, k$, then $h = 1/k$ and $| q_{ij}^g |= 1$ for $i = 1, \ldots, k$; and from Lemma 2 it follows that in this case Q_g is a correlation matrix.

COROLLARY 8. *If all components of an eigenvector h_g of a correlation matrix R have identical absolute values, then the matrix Q_g defined by the formula (19) is an extremal correlation matrix. The partition (I, I^c) corresponding to Q_g is defined in the following way:*

$$i \in \begin{cases} I_g, & \text{if } h_i^g > 0, \\ I_g^c, & \text{if } h_i^g < 0. \end{cases}$$

The following result follows from Lemmas 1 – 3 and Corollary 8:

THEOREM 4. A. *The factor decomposition of a correlation matrix R is at the same time the convex-extremal decomposition of the given correlation matrix R whenever all components of every eigenvector used in the decomposition have, correspondingly, equal absolute values:*

$$| h_i^g |= h_g \ (i = 1, \ldots, k, \ g = 1, \ldots, q).$$

B. *The convex-extremal decomposition of a correlation matrix R is at the same time its factor decomposition, whenever all extremal correlation matrices used in the decomposition are orthogonal.*

From this it follows that, in this case, the number of components q of the decomposition cannot be greater than k.

7. Example. Analysing Changes in the Anthropological Structure of Estonian School Children. In this section a real data-set and real problem will be discussed.

With the aim of analysing the development of Estonian children's body structure with age, measurements were made on schoolchildren aged 6 to 18, both boys and girls. About 150 to 200 children in each age and sex group were measured. The choice of schools was representative for Estonia. For every schoolchild the following 12 most informative measurements were used:

(the number of each variable is given in parentheses): *weight* (1), *height* (2), *cervical height* (3), *foot length* (4), *upper limb length* (5), *lower limb length* (6), *chest circumference* (7), *pelvis circumference* (8), *biacromial breadth* (9), *chest breadth* (10), *chest depth* (11), *pelvis breadth* (12). On the basis of these data 26 correlation matrices of order 12 were calculated, one for each age-sex group. Using component or factor analysis in all subpopulations several principal components/factors can be found, but there is no effective procedure for establishing and modelling *the changes* in the factor structures of childen of different age groups.

Next, the convex decomposition of correlation matrices by quasi-extremal ones was used. The assumptions used in the study were the following: All anthropological measurements have distributions quite close to the normal distribution and the deviations from normality are almost the same – small positive skewness and very small kurtosis (see Kaarma, 1981). Hence the assumption about the equality of the distributions (up to a linear transformation) is valid. As all the correlations between the measured variables were positive, it was reasonable to use the decomposition by *quasi-maximal* distributions. In this case the assumption about symmetry of all marginals is not necessary. The main advantage of the extremal decomposition over the principal component decomposition is the existence of *the same standard base elements* for all correlation matrices to be considered – these are the quasi-maximal correlation matrices, consisting of ones and zeros only.

In this case the construction of the model of change can be carried out by analysing *the change of the coefficients* in the decompositions.

In order to obtain comparable decompositions, several quasi-maximal correlation matrices, which were common in decompositions of different age-groups, were chosen.

Let us describe the first step of the approximation used (further approximations are given in Tiit, Thetloff (1994)). At first the decomposition by three matrices R_1, consisting of all 1's, $R_0 = I$ (unit matrix) and R_2, was made. The last matrix corresponds to the following partition:

$$L = \{\{1,7,8\},\{2,3,5,6\},\{9\},\{10\},\{11\},\{12\}\}.$$

$$R_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In finding the approximations, the standard regression procedures were used, i.e., the elements of the given correlation matrix were taken as the values (measurements) of the dependent variable and the elements of extremal or quasi-extremal correlation matrices as explanatory variables. The model without constant term was used, and the estimated regression coefficients b_g were taken as weights w_g. In order to satisfy the condition (15), the following additional transformation was made: $Y := Y - X_h$; $X_g := X_g - X_h$, $(g = 1, \ldots, h-1)$. Then $w_h = 1 - \sum_{g=1}^{h-1} w_g$. The solution was acceptable only if all regression coefficients w_g were non-negative.

For example, in the case of the six-year old boys we obtained the following decomposition:

$$R = 0.158 R_0 + 0.613 R_1 + 0.229 R_2. \tag{22}$$

One possible interpretation of the model (22) is the following. On the basis of their body structure, the population of 6-year-old boys can be divided into three groups. The first group corresponds to R_1 – all body measurements of the boys in this group are linearly dependent (proportional). Hence the body structure of all boys in this group is quite similar, but their sizes can be different. From the decomposition formula (22) it follows, that about 61% of all boys belong to this group.

The second group, corresponding to R_2, can be characterized by two sets of proportional measurements – the first consisting of the measurements, characterizing the tallness of the boys – height, cervical height, upper and lower limb lengths; the second consisting of measurements characterizing the thickness of the boys – weight, chest and pelvis circumference. Typically, the variables belonging to different sets are uncorrelated (independent). These measurements are uncorrelated with all other measurements (foot length, biacromial breadth, etc.), as well. Hence the boys who can be either thick or thin and are characterized as pycnic and leptosomous anthropometrical types belong to this group. The frequency of 6-year old boys belonging to these types is about 23%.

The third group corresponds to the unit correlation matrix, where all body measurements can be considered as uncorrelated. In fact, we can assume that the correlations between body measurements are rather weak and that the higher correlations might be situated in random cells of the correlation matrix. It follows from formula (22) that the percentage of such 6-year old boys is about 16%.

In a similar fashion, the correlation matrices characterizing the body structure of boys and girls of all age groups were obtained. As a result, two 3-dimensional time series – one $T(B)$ for boys and the second $T(G)$ for girls,

$$T(B) = (t_i^0(B), t_i^1(B), t_i^2(B)),\ T(G) = (t_i^0(G), t_i^1(G), t_i^2(G))$$
$$i = 6, \ldots, 18,$$

were constructed. Here the index i characterizes the age, and the following general condition,

$$\sum_{j=0}^{2} t_i^j(A) = 1\ (i = 6, \ldots, 18,\ A = B, G)$$

is satisfied.

Analysing the time series $t_i^j(\cdot)$ we see the following trends in the coefficients:

1^0. The rate of the *full dependence set*, characterized by the correlation matrix R_1, i.e., the *entirely proportional body types*, decreases, for both boys and girls, about 1.2 to 1.5 per cent per year, the change is statistically significant, $p < 0.05$.

2^0. The other characteristic subpopulation, described by the correlation matrix R_2, i.e., the children *belonging to different anthropometrical types*, increases, on average about one per cent per year, whence the change is strongly significant ($p < 0.01$).

Remarks.

A. The advantages of the convex-extremal decomposition are the following:

1) The set of elements of the basis (extremal/quasi-extremal correlation matrices and hence distributions) is standard and does not depend on the concrete sample.
2) The number h of parameters (weights) that one has to estimate is much less than in the case of principal factor decomposition, where $h \cdot k$ factor loadings must be estimated, h being the number of factors.
3) The mixture of distributions can be interpreted as a mixture of populations and hence is quite easy to understand.
4) No rotation procedure is needed to improve the solution.

B. The shortcomings of the convex-extremal decomposition are:

1) The level of the best possible approximation is not always 100%.
2) The solution is not always unique.
3) There is no standard software for this procedure.

As regards (1) and (2), we note, that they are more or less typical in most multivariate models, including factor analysis.

Acknowledgements. The author is thankful to Professors H. Sherwood and M. Taylor for their kind support and to two anonymous referees for valuable remarks.

References

CUADRAS, C. (1981) Metodos de analisis multivariante. Barcelona, 52 (In Spanish).

DALL'AGLIO, G. (1960) Les fonctions extremes de la classe de Fréchet a trois dimensions, *Publ. Inst. Statist. Univ. Paris 9*, 175–188.

DALL'AGLIO, G. (1972) Fréchet classes and compatibility of distribution functions. *Sympos. Math.* **9**, 131–150.

DALL'AGLIO, G. (1992) Fréchet classes: the beginnings. *Advances Probability Distributions with Given Marginals.* Dordrecht, 1–12.

FÉRON, R. (1956) Sur les tableaux de correlation dont les marges sont données, cas de l'espace à trois dimensions. *Publ. Inst. Statist. Univ. Paris 5*, 3–12.

FRÉCHET, M. (1951) Sur les tableaux de correlation dont les marges sont donnees. *Ann. Univ. Lyon Sec. A, fasc.* **14**, 53–77.

HOEFFDING, W. (1940) Masstabinvariante Korrelationstheorie. *Schr. Math. Inst. u Inst. Angew. Math. Univ. Berlin* **5**, 179–233.

KAARMA, H. (1981) Multivariate statistical investigation of the system of anthropometrical variables of women. *Tallinn, 'Valgus'* **1**, 1–200 (in Russian).

KELLERER, H. G. (1961) Verteilungsfunktionen auf Produkträumen mit gegebenen Marginal Funktionen, *Math. Annalen* **153**, 323–344.

KEMP, J. F. (1973) Advanced Problem #5894, *Amer. Math. Monthly* **80**, 83.

KOTZ, S., and TIIT, E.-M. (1992) Bounds in multivariate dependence. *Acta et Commentationes Universitatis Tartuensis* **942**, 35–45.

RUIZ-RIVAS, C. (1979) Contribuciones a la construccion de la distribuciones multivariantes con marginales fijas. Tesis Doctoral, Fac Mat, Univ. Barcelona (manuscript).

RÜSCHENDORF, L. (1991) Fréchet bounds and its applications. *Advances in Probability Distributions with Given Marginals*, Dordrecht, 151–188.

SCHWEIZER, B. (1991) Thirty years of copulas. (1991) *Advances in Probability Distributions with Given Marginals*, Dordrecht, 13–50.

SKLAR, A. (1959) Fonctions de repartition a n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

TIIT, E.-M. (1984) Definition of random vectors with given marginal distributions and given correlation matrix. *Acta et Commentationes Universitatis Tartuensis* **685**, 21–36.

TIIT, E.-M. (1986) Random vectors with given arbitrary marginals and given correlation matrix. *Acta et Commentationes Universitatis Tartuensis* **733**, 14–38.

TIIT, E.-M. (1992) Extremal multivariate distributions having given discrete marginals. *Acta et Commentationes Universitatis Tartuensis* **942**, 94–113.

TIIT, E.-M. (1994) Mixture of extremal distributions – a model, alternative to the log-linear model. In: *Probability Theory and Mathematical Statistics. Proceedings of the 6th Vilnius Conference.* Vilnius, 1994, 703–714.

TIIT, E.-M., and TAMMET, S. (1994) Using index-vectors and partitions in multivariate analysis, *Acta et Commentationes Universitatis Tartuensis* **968**, 55–75.

TIIT, E.-M., and THETLOFF, M. (1994) The convex-extremal decomposition of correlation matrix. An application in anthropometrical research. *Acta et Commentationes Universitatis Tartuensis* **968**, 77–97.

INSTITUTE OF MATHEMATICAL STATISTICS
FACULTY OF MATHEMATICS
UNIVERSITY OF TARTU
TARTU, EE2400 ESTONIA
etiit@ma.ut.ee