ISSN 0494-7304 0132-053×

TARTU RIIKLIKU ÜLIKOOLI
TOIMETISED

УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS

733

ИССЛЕДОВАНИЕ СТАСТИСТИЧЕСКИХ МОДЕЛЕЙ

MULTIVARIATE STATISTICAL MODELS

Труды по математике и механике

Matemaatika- ja mehhaanikaalaseid töid

TARTU 1986

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS
ALUSTATUD 1893.a. VIHIK 733 ВЫПУСК ОСНОВАНЫ В 1893.г.

ИССЛЕДОВАНИЕ СТАСТИСТИЧЕСКИХ МОДЕЛЕЙ

MULTIVARIATE STATISTICAL MODELS

Труды по математике и механике

Matemaatika- ja mehhaanikaalaseid töid

TARTU 1986

DEFINITION OF MIXTURES WITH GIVEN MOMENTS

E.Tiit

1. Set-up of the problem.

The definition of distributions with given moments is important in several situations, for instance,

1° in data analysis, when it is necessary to identify given empirical distributions,

2° in statistical modelling, when the generation of distribution with given moments is required.

There exist several classical methods for solving the problem, for instance, using Pearson's distributions family, some other families such as Johnson family, some expansions by given distributions (Gram-Charlier', Edgeworth expansions), and mixtures of given distributions. Most studies, dealing with the mixtures, use the mixtures of normal distributions [3 - 5]. In recent time the mixtures of several other parametrical families are considered as well [1,2,7,8].

Especially for the second purpose mentioned it is convenient to use mixtures , because of the efficiency of generation of mixtures is considerable.

The aim of the paper is:

1° To demonstrate that arbitrary distribution, having moments of order k may be used as the basis for defining the mixture with given moments $\mu_1,\dots,\mu_k$.

2° To analyse the number of parameters, needed by the definition of mixtures and to demonstrate the possibility of definition of optimal mixtures (in the sense of number of defining parameters).

3° To give the analytic resolution of the problem for the case of four given moments and the arbitrary given distribution (of 4-th order).

2. The concept of optimal distribution for given moments.

Let k be natural, $k \geqslant 2$, and $\mu_1,\dots,\mu_k$ be real numbers, satisfying the condition

$$\begin{pmatrix} 1 & \mu_1 & \dots & \mu_\nu \\ \mu_1 & \mu_2 & \dots & \mu_{\nu+1} \\ \dots & \dots & \dots & \dots \\ \mu_\nu & \mu_{\nu+1} & \dots & \mu_{2\nu} \end{pmatrix} \geqslant 0 \qquad (1)$$

where $\nu = \left[\frac{k}{2}\right]$. Denote $\mu = (\mu_1, \dots, \mu_k)$ and $\mathcal{M}_k$ the set of all vectors μ fulfilling the condition (1). Let $\Pi(\mu)$ be the set of all distributions, having their first k moments equal to the corresponding components of vector μ.

Let $\mathcal{P}$ be any set of distributions, depending on parameters $\vartheta_1, \dots, \vartheta_r$. If the parameter-vector $\vartheta = (\vartheta_1, \dots, \vartheta_r)$ identifies the distribution P_ϑ in the family $\mathcal{P}$, but none of the subvectors of ϑ does it, then ϑ is the <u>defining parameter</u> of distribution P_ϑ. It is evident that in general the number of defining parameters of distributions from the set $\Pi(\mu)$ is at least k, if $\mu \in \mathcal{M}_k$. The distributions from the set $\Pi(\mu)$, having exactly k defining parameters, are said to be <u>optimal</u> and having not more than k+1 defining parameters - <u>almost optimal</u>.

Let P be arbitrary distribution of order k, $X \sim P$. Then we denote $m_h(P) = m_h(X) = EX^h$ the h-th moment of P, h=1,..., ...,k.

3. The discrete mixtures of random variables and distributions.

Let $(\Omega, \mathcal{B}, \gamma)$ be discrete (finite) probability space, $\mathfrak{X}$ - the set of random variables, having the distributions from the set $\mathcal{P}$. Then the mixture Z of random variables X_i is the transformation

$$\Omega \to \mathfrak{X},$$

given by the formula

$$Z = \sum_{i=1}^{n} \chi_{\omega_i} X_i ,$$

where ω_i are the events, $\bigcup \omega_i = \Omega$, $\omega_i \cap \omega_j = \emptyset$, $\gamma(\omega_i) = \gamma_i$ and $X_i \in \mathfrak{X}$. If $X_i \sim P_i$, then the distribution Q of mixture Z is the mixture of distributions P_i, $Q = \sum_{i=1}^{n} \gamma_i P_i$, where $\gamma_i = \gamma(\omega_i)$ is said to be the <u>mixing distribution</u>, $\gamma_i \geqslant 0$, $\sum \gamma_i = 1$.

It is well known fact that if distributions P_i are of order k, then the mixture Q has the same property and the moments of Q can be expressed through the moments of mixed distributions P_i in the following way:

$$m_h(Q) = \sum_{i=1}^{n} \gamma_i m_h(P_i), \quad h = 1,\dots,k. \tag{2}$$

4. The class of mixtures of linear parametrization.

Let P_o be the given distribution. Let us regard the set $\mathcal{P}(P_o)$ of distributions P(a,b), $(a,b \in \mathbb{R}_1)$, defined as the distributions of linear transformations of variable X_o with distribution P_o:

$$\mathcal{P}(P_o) = \{ P(a,b) : aX_o + b \sim P(a,b), \text{ if } X_o \sim P_o, \ a,b \in \mathbb{R}_1 \}.$$

Denote $\mathcal{Q}(P_o)$ the set of all discrete mixtures of distributions from $\mathcal{P}(P_o)$:

$$\mathcal{Q}(P_o) = \{ Q : Q = \sum_{i=1}^{n} \gamma_i P(a_i,b_i); \ a_i,b_i \in \mathbb{R}_1; \ \gamma \in D \},$$

where D is the set of distributions with finite support. Say $\mathcal{Q}(P_o)$ is the class of mixtures of linear parametrization of distribution P_o. Let $\mathcal{Z}(X_o)$ be the set of random variables Z, defined as

$$Z = U_a X_o + U_b, \tag{3}$$

where U_a, U_b are following random variables:

$$U_a = \sum_{i=1}^{n} \chi_{\omega_i} a_i, \quad U_b = \sum_{i=1}^{n} \chi_{\omega_i} b_i .$$

In the special case of degenerate distribution γ we have in (3)

$$Z = aX_o + b.$$

From here it follows that the set $\mathcal{Z}(X_o)$ is closed in linear transformations and finite mixtures.

Lemma 1. In every set $\mathcal{Z}(X_o)$ there exists a random variable X_*, having the following properties:

$$\begin{cases} 1^o \ X_* \text{ is standardized, } EX_* = 0, \ DX_* = 1, \\ 2^o \ X_* \text{ is symmetric: if } X_* \sim P_* \text{ then } -X_* \sim P_*. \end{cases} \tag{4}$$

Proof. Let P_o have the following moments:

$m_1(P_o) = c_1$, $m_2(P_o) = c_1^2 + c_2^2$; then $DX_o = c_2^2$.

Let us define ... oles X_1 and X_2:

$X_1 = c_2^{-1}X_o - c_1c_2^{-1}$; $X_2 = -X_1$;

$m_1(X_i) = 0$, $m_2(X_i) = 1$, $i = 1,2$, and the mixture Z of variables X_1, X_2:

$Z = \chi_\omega X_1 + \chi_{\bar\omega} X_2$, $\gamma(\omega) = 0.5$, $\bar\omega$ is the opposite event of ω.

It is clear that Z has the properties 1^o, 2^o and $Z \in \mathcal{Z}(X_o)$. Let $Z \sim Q$, then

$$Q = 0.5P(c_2^{-1}, -c_1c_2^{-1}) + 0.5P(-c_2^{-1}, c_1c_2^{-1}).$$

The lemma is proved. In future let us assume that the initial variable X_o with distribution P_o fulfills the conditions 1^o and 2^o.

Some subsets of $\mathcal{Q}(P_o)$ have been of special interest and have been studied in statistical literature steadily, for instance:

$$1^o\ \mathcal{Q}_a(P_o) = \Big\{ \sum_{i=1}^{n} \gamma_i P(a_i, 0),\ a_i \in \mathbb{R}_1 \Big\},$$

the set of scale (contamination, variance) mixtures;

$$2^o\ \mathcal{Q}_b(P_o) = \Big\{ \sum_{i=1}^{n} \gamma_i P(a, b_i),\ a, b_i \in \mathbb{R}_1 \Big\},$$

the set of shift (mean) mixtures;

$$3^o\ \mathcal{Q}_c(P_o) = \Big\{ \sum_{i=1}^{n} \gamma_i P(0, c_i) : c_i \in \mathbb{R}_1 \Big\} = D,$$

the set of discrete distributions (with finite support).

5. The defining parameters of mixtures.

Let Q be the n-component mixture of r-parameter distributions; then the number q of defining parameters of Q is, in general,

$$q = (r + 1)n - 1, \tag{5}$$

so as besides the defining parameters of initial distributions P_i the coefficients γ_i are included into the set of defining parameters of Q.

In the class of linear parametrization $\mathcal{Q}(P_o)$ all defining parameters of P_o are fixed, and every initial distribution P(a,b) is characterized with two parameters a and b only and so the n-component mixture has

$$q = 3n - 1 \tag{6}$$

defining parameters. In subclasses $\mathcal{Q}_a(P_o)$, $\mathcal{Q}_c(P_o)$ and $\mathcal{Q}_b(P_o)$ this number is correspondingly

$$q_a = q_c = 2n - 1; \tag{6'}$$

$$q_b = 2n. \tag{6''}$$

From the inequality

$$k \leq q$$

the necessary number n of components of mixture Q, having the property $Q \in \pi\mu$, follows.

6. Definition of almost optimal mixtures with given moments.

Assume k, μ from $\mathcal{M}_k$ and k-order distribution P_o, satisfying conditions (4), are given. The problem is to define the optimal or almost optimal mixture Q from the set $\mathcal{Q}(P_o) \cap \pi\mu$. The questions we must solve are the existence, uniqueness and construction of the distribution Q.

Let us write down some well-known equalities for the moments of several variables (we propose all variables to be of the order k and independent):

$$m_h(aX + bY) = \sum_{i=0}^{h} \binom{h}{i} a^i m_i(X) b^{h-i} m_{h-i}(Y); \tag{7}$$

denoting $m_h^o(X) = m_h\left(\frac{X - EX}{\sqrt{DX}}\right)$, we have

$$m_h^o = (m_2 - m_1^2)^{-h/2} \sum_{i=0}^{h} \binom{h}{i} m_i \cdot (-1)^{h-i} m_1^{h-i}, \tag{7'}$$

$$m_h = \sum_{i=0}^{h} \binom{h}{i} (m_2 - m_1^2)^{i/2} m_i^o \, m_1^{h-i}. \tag{7''}$$

We shall use the connections (7) - (7'') in following calculations, particularly in proving the following theorem.

Theorem 1. For every k ($k \geq 2$), natural, given vector μ from $\mathcal{M}_k$, satisfying (4), and given distribution P_o of order k it is possible to construct almost optimal mixture Q from the set $\mathcal{Q}(P_o) \cap \pi\mu$. The distribution Q is not unique in general, the set of distributions $\mathcal{Q}(P_o) \cap \pi\mu$ is convex.

Proof. We will define the mixture of shift-type, or what is equivalent, the variable

$$Z = aX_o + Y, \text{ where } X_o \sim P_o,\ Y \sim Q_*,\ Q_* \in D,\ Z \sim Q.$$

Let us denote $m_i(X_o) = M_i (i = 1,\dots,k)$. From the equation (7) it follows that the moments of Z have the following form:

$$m_h(Z) = \sum_{i=0}^{h} a^i M_i m_{h-i}(Y). \qquad (8)$$

From the condition $Q \in \pi_\mu$ we get the system of k linear equations with k unknowns $m_1(Y),\dots,m_k(Y)$:

$$\mu_h = \sum_{i=0}^{h} \binom{h}{i} a^i M_i m_{h-i}(Y) \quad (n = 1,\dots,k). \qquad (9)$$

The variable Y exists if and only if the vector $m(a) = (m_1(Y),\dots,m_h(Y))$ belongs to $\mathcal{M}_k$ (see [6] and [9]). Let us regard the behaviour of the vector m(a), depending on a, $a \in [0,1]$. It is evident that in the case a = 0 the solution of the system (9) satisfies the condition $m(a) \in \mathcal{M}_k$. Let $m(a_1)$ and $m(a_2)$ belong to the $\mathcal{M}_k$, $a_1 < a_2$. From the convexity of set of mixtures it follows that every $m(a)$, $a_1 \le a \le a_2$ belongs to the set $\mathcal{M}_k$. From here it follows that the set A of values a, for which $m(a) \in \mathcal{M}_k$, has the form $[0,a^*]$, $a^* \le 1$ (equality holds only when $P_o \in \pi_\mu$).

For constructing the variable Z any a, belonging to A must be found, then the system (9) solved and the moments of discrete distribution Q^* found. Then solution of the system of k nonlinear equations defines the optimal discrete distribution with k parameters. As always the optimal solution in class D exists (see [6], [9]), so the number of defining parameters of Q is, in general, k+1, q.e.d.

If P_o and Q satisfy the condition (4), then the value of the parameter a in some sense measures the similarity of the mixture to the initial distribution P_o: we have the decomposition of the variance of Z:

$DZ = a^2 DX_o + DY$, $DY = 1 - a^2$, where Y is discrete random variable. Using $a = a^*$, we get the mixture, that is the "most similar" to the initial distribution P_o.

7. Definition of optimal mixture with four given moments.

Let k = 4. Then from equations (6), (6') and (6") it follows that the optimal mixture must be of the shift type with two components, defining parameters being a, b_1, b_2 and γ. The mixture Q is the following:

$$Q = \gamma P(a,b_1) + (1-\gamma) P(a,b_2), \quad Z = \chi_\omega(aX_o + b_1) + \chi_{\bar{\omega}}(aX_o + b_2), \quad \gamma(\omega) = \gamma$$

Let us write down the equations for calculating the parameters, using the general expression for moments (2), the connection (7) and the assumption, that X_o satisfies the conditions (4) and μ - the conditions (7'). Denote $m_4(X_o) = M_4$.

$$\begin{cases} m_1(Q) = \gamma b_1+(1-\gamma)b_2, \\ m_2(Q) = \gamma(a^2+b_1^2) + (1-\gamma)(a^2+b_2^2), \\ m_3(Q) = \gamma b_1^3+(1-\gamma)b_2^3+3(\gamma a^2 b_1+(1-\gamma)a^2 b_2), \\ m_4(Q) = \gamma b_1^4+(1-\gamma)b_2^4+6a^2(\gamma b_1^2+(1-\gamma)b_2^2) \end{cases} \tag{10}$$

or after simplifications,

$$\begin{cases} \gamma b_1+(1-\gamma)b_2 = 0, \\ \gamma(a^2+b_1^2)+(1-\gamma)(a^2+b_2^2) = 1, \\ \gamma b_1^3+(1-\gamma)b_2^3 = \mu_3, \\ \gamma b_1^4+(1-\gamma)b_2^4+6a^2(1-a^2)+M_4 a^4 = \mu_4. \end{cases} \tag{10'}$$

From the first and second equations we can express b_1 and b_2 through a^2 and γ:

$$\begin{cases} b_1 = \sqrt{\dfrac{1-\gamma}{\gamma}(1-a^2)}, \\ b_2 = -\dfrac{\gamma}{1-\gamma} b_1. \end{cases} \tag{11}$$

Denoting $(1-a^2)^3 = T$, from the third equation we can get the expression of γ through T: we receive the quadratic equation

$$M_3^2(1-\gamma)\gamma = T(1-2\gamma)^2,$$

the two solutions of which are

$$\gamma = 0.5 \mp \frac{|\mu_3|}{2\sqrt{4T+\mu_3^2}}. \tag{12}$$

Let us take the smaller solution for γ, then the larger one gives $1-\gamma$ and we can add the following condition (11') to the expression (11):

$$\text{sign } b_1 = \text{sign } \mu_3. \tag{11'}$$

For finding a we express b_1 from the third and fourth equations. If we denote

$$A(a) = \mu_4 - a^4 M_4 - 6(a^2 - a^4) \tag{13}$$

and

$$B(a) = T(1-\gamma)^{-3}\gamma^{-3},$$

we get the equation

$$A(a) = \mu_3 \frac{(1-3\gamma(1-\gamma))}{1-2\gamma}(B(a))^{1/3}.$$

Using the connections $\gamma(1-\gamma) = \frac{T}{4T+\mu_3^2}$, $1-2\gamma = \frac{|\mu_3|}{4T+\mu_3^2}$ and the denotation $a^2 = y$, we get the equation of the third degree

$$G(y) = (5-M_4)y^3+(M_4-9)y^2+(\mu_4+3)y-(\mu_4-1-\mu_3^2) = 0. \quad (14)$$

If $\mu_3 \neq 0$, we have $G(1) > 0$, $G(0) \leq 0$, that means, the solution $a \in [0,1]$ always exists. If $G(0) = 0$, we get the unique solution $a = 0$, that means, $Q \in D$.

Let us regard the case $\mu_3 = 0$. Then we may choose the symmetric distribution Q. The shift-type mixture has then two defining parameters a and b, so as $b_1 = -b_2 = b$ and $\gamma = 1 - \gamma = 0.5$. Instead of the system (10') we get the system

$$\begin{cases} b^4 + 6a^2b^2 + a^4M_4 = \mu_4, \\ a^2 + b^2 = 1. \end{cases} \quad (15)$$

If we denote $x = a^2$, $1 - x = b^2$, then we get the solution from the quadratic equation

$$x = \frac{-2 \pm \sqrt{4-(M_4-5)(1-\mu_4)}}{M_4 - 5},$$

the solution in $[0,1]$ exists if and only if the condition

$$1 \leq \mu_4 \leq M_4$$

is fulfilled. That means, the shift-type mixture decreases the fourth moment of initial distribution.

If $\mu_4 > M_4$, we must use the contamination-type mixture. Then the number of defining parameters is 3, and for getting the optimal distribution we must fix one parameter, for instance, let us take $a_1^2 = 0.5$, $a_2 = ca_1$. Then we have instead of the system (10') the following system:

$$\begin{cases} M_4(\gamma+(1-\gamma)c^4) = 4\mu_4, \\ (\gamma+(1-\gamma)c^2) = 2. \end{cases} \quad (16)$$

If we use the notation $\mu_4/M_4 = S$, then we have

$\gamma = \frac{4(S-1)}{4S-3}$, $c^2 = (2-\gamma)/(1-\gamma)$. The solution γ exists, if $S > 1$, that means, the contamination-type mixture increases the fourth moment of the initial distribution P_0.

Example.

Let the vector μ of given moments be (1, 2, 3.5, 7) and the given family P be the family $\mathfrak{A}$ of uniform distributions $U(c_1, c_2)$. It is evident that the standard distribution P_o, fulfilling the conditions (4) in family $\mathfrak{P}$ is $U(-\sqrt{3}, \sqrt{3})$, $m_4(P_o) = 1.8$.

1° The first step of the solution is the calculation of corresponding standardized vector $\mu^o = (\mu_1^o, \ldots, \mu_4^o)$, using the formula (7'): We get $\mu_1^o = 0$, $\mu_2^o = 1$, $\mu_3^o = -0.5$, $\mu_4^o = 2$.

2° Now it is possible to write the equation (14), putting the values of μ_3^o and μ_4^o into the expression G(a):

$$3.2y^3 - 7.2y^2 + 5y - 0.75 = 0.$$

The solution of the equation gives $y = 0.20500471$, that means

$$a = 0.45277446.$$

From (12) we get

$$\gamma = 0.33369525, \quad 1 - \gamma = 0.66630475,$$

and from (11) and (11')

$$b_1 = -1.2599222, \quad b_2 = 0.63098761.$$

The mixture Z_o of variables from $\mathfrak{A}$ has the following form:

$$Z_o = \chi_\omega(0.45277446X_o - 1.2599222) + \chi_{\bar{\omega}}(0.45277446X_o + 0.63098761), \quad \gamma(\omega) = 0.33369525 \qquad (17)$$

and its distribution Q_o is

$$Q_o = 0.33369525P(0.45277446, -1.2599222) + 0.66630475P(0.45277446, 0.63098761).$$

Using the formula (10), it is simple to calculate the moments of Q:

$$m_1(Q_o) = 0, \quad m_2(Q_o) = 1, \quad m_3(Q_o) = -0.5, \quad m_4(Q_o) = 2$$

(with 6 decimal places).

3° Now we must find the distribution Q (random variable Z) with moments μ. For the purpose we use the linear transformation

$$Z = (\mu_2 - \mu^2)^{1/2} Z_o + \mu_1,$$

where μ_1 and μ_2 are given in μ. We get

$$Z = \chi_\omega(0.45277446X_o - 0.2599222) + \chi_{\bar{\omega}}(0.45277446X_o + 1.63098761),$$

γ is defined by (17). By formula (10) we find

$$m_1(Q) = 1, \quad m_2(Q) = 2, \quad m_3(Q) = 3.5, \quad m_4(Q) = 7 \text{ (with 6}$$

decimal digits).

4°. For using the standard generators of random numbers it is convenient to express the variable Z through $Y, Y \sim U(0;1) = P_*$. So as $X_0 = 2\sqrt{3Y} - \sqrt{3}$, we get finally

$$Z = \chi_{\omega}(1.5684567Y - 1.0441506) + \chi_{\overline{\omega}}(1.5684567Y - 0.84675923),$$

χ is defined by (7), and

$$Q = 0.3336952\, P_*(1.5684567, -1.0441506) + 0.66630475 P_*(1.5684567, 0.84675923).$$

5°. For checking the correctness of results we may use the formula (7) for calculating the moments of linear combination of variables $Z = a^* Y + W$, W having the following discrete distribution

$$\gamma(-1.0441506) = 0.33369525,$$
$$\gamma(0.84675923) = 0.66630475.$$

The moments of variables Y and W are the following

$$m_1(W) = 1/2,\ m_2(Y) = 1/3;\ m_3(Y) = 1/4,\ m_4(Y) = 1/5.$$
$$m_1(W) = 0.2157716,\ m_2(W) = 0.84155271,\ m_3(W) = 0.02465796,$$
$$m_4(W) = 0.73918664,$$

and we get

$m_1(Z)=1$; $m_2(Z)=2$, $m_3(Z)=3.5$ and $m_4(Z)=7$ with 6 decimals, q. e. d.

References

1. Everitt, B.S., Hand, D.J. Finite Mixture Distributions. Chapman and Hall, London, 1981.
2. Grim, J. On numerical evaluation of maximal likelihood estimates for finite mixtures of distributions. Kybernetika (Prague), 1982, 18, № 3, 173-190.
3. Medgessy, P. Decomposition of Superpositions of Density Functions and Discrete Distributions. Akademiai Kiado, Budapest, 1977.
4. Molenaar, W. Survey of estimation methods for a mixture of two normal distributions. Statist. Neerlandica, 1965, 19, 4, 249-265.
5. Pearson, K. Contributions to the mathematical theory of evolution 1: Dissection of frequency curves. Philos. Trans. Roy. Soc. London, Ser. A 1894, 185, 71-110.

6. Miller, A.C., Rice, T.R. Discrete approximations of probability distributions. Management Science, 1983, 29, № 3, 352-362.
7. Айвазян С.А., Бежаева В.И., Староверов О.В.Классификация многомерных наблюдений. Москва, 1974.
8. Исаенко О.К., Урбах К.Н. Разложение смесей распределений вероятностей на их составляющие. Теор. вер. и мат.стат. теор. киб. 1976, том, 13, 37-58, ВИНИТИ, Москва.
9. Тийт Э.-М. Построение дискретных многомерных распределений с заданными моментами. Дискретный аналог нормального распределения. Труды ВЦ ТГУ, 1984, 51, 142-165.

ОПРЕДЕЛЕНИЕ СМЕСЕЙ РАСПРЕДЕЛЕНИЙ С ЗАДАННЫМИ МОМЕНТАМИ

Э. Тийт

Р е з ю м е

В разных ситуациях возникает задача - определить распределение, имеющее заданные моменты $\mu_1, \ldots, \mu_k$ (удовлетворяющие, естественно, условию (I)). Оптимальным (почти оптимальным) считается распределение, имеющее k(соответственно k+1) определяющих параметров.

Одной возможностью для решения этой задачи является пользование дискретными смесями распределений. На основании любого распределения k-го порядка P_0 определяется т.н.класс смесей линейной параметризации $\mathcal{O}_l(P_0)$, состоящий из дискретных смесей распределений $\{P(a_i, b_i), a_i\ b_i \in R\}$, где $P(a, b)$ есть распределение случайной величины $aX+b$ при $x_0 \sim P$. Частными случаями смесей из класса $\mathcal{O}_l(P_0)$ являются хорошо известные смеси типа смещения и смеси типа засорения.

Доказывается теорема, что для любого комплекта заданных моментов $\mu_1, \ldots, \mu_k$ (удовлетворяющих (I)) и для любого заданного распределения (k-го порядка) P_0 возможно определить почти оптимальную смесь из класса $\mathcal{O}_l(P_0)$.

Для случая k = 4 выведены явные формулы для нахождения определяющих параметров смеси. Если моменты исходного распределения обозначаются через M_i ($i = 1, \ldots, k$), и рассматриваются стандартизированные распределения ($\mu_1 = M_1 = 0$; $\mu_2 = M_2 = 1$), то выяснилось, что в случае $\mu_4 / M_4 > 1$ необходимо пользоваться смесью типа засорения, а в случае $\mu_4 / M_4 < 1$ - смесью типа смещения.

Received September 85

RANDOM VECTORS WITH GIVEN ARBITRARY MARGINALS AND GIVEN CORRELATION MATRIX

E.Tiit

The problem of existence and definition of multivariate distribution with given marginals and some given characteristics of dependence is important in using the Monte-Carlo method in multivariate analysis, especially when the robustness of statistical methods is the subject of investigation.

The partial solutions of the problem raised have already been obtained more than forty years ago: in 1940 W. Hoeffding [5] described the set of all bivariate distributions with given marginals and gave the formula for the "minimal" and "maximal" ones, that have correspondingly the minimal and maximal correlations. In early fifties the properties of the minimal and maximal bivariate distributions were studied by Frechet [4]. In 1964 Kellerer demonstrated that the set of k-dimensional distributions with fixed marginals forms a convex polyhedra in the space of all k-dimensional distributions [5]. For 2 marginals in the complete separable metric space Strassen derived the necessary and sufficient conditions for existence of common distribution [12].

For the case when some dependence characteristics are fixed, too, there are rather few results. In 1976 Whitt surveyd all the known facts for the bivariate case [15]. For the $k > 2$ there exist some results when it is assumed that the distribution describes the subvector of stationary random process [8]. In the case all marginals are equal and correlation matrix consists of values of correlation function, $r_{ij} = R(i-j)$; $i,j = 1,\ldots,k$, see [1,9]. In [13] the author gave an algorithm for finding the k-dimensional distribution by given (equal) marginals and given correlation matrix with some generalizations in [14,16]. Some alternative resolutions are given in [3,8] and [10].

The aim of the paper is to generalize these results for the case of unequal marginals and arbitrary correlation

matrices.

1. Set-up of the problem.

Assume k is fixed natural, $k \geqslant 2$; $P_1,\ldots,P_k$ are given one-dimensional distributions with distribution functions $F_i(x_i)$ correspondingly. Let $R=(r_{ij})$ be $k \times k$ correlation matrix, given.

Our aim is to define a k-variate distribution

$$P = P(P_1,\ldots,P_k;R)$$

with given marginals P_i and given correlation matrix R. By the way we must solve the probleme of existence and uniqueness of distribution P, as well as the problem of efficient construction of vectors having the distribution P and the simulation of values of vectors with distribution P.

In paragraphs 4-6 we consider the special case of equal marginals (in 4 we add the restriction of symmetricity of distribution, too; in 5 we regard the case of positive correlations only). In paragraphs 8-9 we generalize the results of paragraphs 4-6 for the arbitrary marginals (of second order). Some illustrative examples are given in paragraphs 7 and 9.

2. Some auxiliary concepts.

Def. 1. Let $k \in N$. The natural-valued vector $I=(i_1,\ldots,i_s)$ is k-index, if it fulfills the following conditions (see [10]):

$$1 \leq i_j < i_{j+1} \leq k,\ j = 1,\ldots,s-1;\ s \leq k.$$

If k is fixed, we say index instead of k-index. We use the symbol I in the sense of $\{i_1,\ldots,i_s\}$, as well, for instance, $j \in I$ means $j \in \{i_1,\ldots,i_s\}$.

Def. 2. Let $k \in N$. The natural-valued vector $L=(l_1,\ldots,l_k)$ is (s,k)-indicator, if it fulfills the following conditions:

1° $1 \leq l_i \leq s$; $i=1,\ldots,k$, $s \leq k$;

2° if $l_i = h$, $h > 1$, then there exists $j < i$ with $l_j = h-1$;

From Def. 2 it follows, that the (s,k)-indicator defines the partition of set $\{1,\ldots,k\}$ into s distinct subsets; let us denote the index of the h-th subset by $I_h=(i_1^h,\ldots,i_{v_h}^h)$, $h=1,\ldots,s$, $\sum_{h=1}^{s} v_h = k$, and we have the connection:

$l_j = h \leftrightarrow j \in I_h$.

Every index I ($1 \in I$) defines a (2,k)-indicator, where $I_1 = I$, $I_2 = I^c$.

Example 1. Let k = 9, s = 4 and L = (1,2,1,1,3,2,4,3,2). Then the indices of subsets are: $I_1 = (1,3,4)$; $I_2 = (2,6,9)$, $I_3 = (5,8)$, $I_4 = (7)$.

Def. 3. Let $k \in N$. The integer-valued vector $T = (t_1,\dots,t_k)$ is (s,k)-pseudoindicator, if it fulfills the following conditions:

1° $T' = (|t_1|,\dots,|t_k|)$ is (s,k)-indicator; I_h, h=1,...,s;

2° For every h there exists index $J_h = (j_1^h,\dots,j_{q_h}^h)$, $J_h \subset I_h$, $j_1^h = i_1^h$; if $G_h = I_h \setminus J_h$, $G_h = (g_1^h,\dots,g_{p_h}^h)$, then the following connections hold, h = 1,...,s:

$$t_i = \begin{cases} h, & \text{if } i \in J_h, \\ -h, & \text{if } i \in G_h, \end{cases} \quad i=1,\dots,k;\ h=1,\dots,s.$$

From Def. 3 it follows that the (s,k)-pseudo-indicator defines the partition of the set $\{1,\dots,k\}$ into s subsets, while every subset could be partitioned into two subsets: one, corresponding to the positive and the other, corresponding to the negative termes of pseudoindicator T. The first element of every subset corresponds to the positive t.

Example 2. Let k = 9, s = 4 and T=(1,2,-1,1,3,-2,4,3,-2) Then T' = L, and the indices J_h and G_h are following: $J_1 = (1,4)$; $G_1 = (3)$; $J_2 = (2)$; $G_2 = (6,9)$; $J_3 = (5,8)$; $G_3 = \emptyset$; $J_4 = (7)$; $G_4 = \emptyset$.

Lemma 1.

Let K(s,k) be the number of different (s,k)-indicators (that is, the number of different partitions of set $\{1,\dots,k\}$ into subsets). Then

$$K(s,k) = \sum_{p=s}^{k} K(s-1,p-1)\,s^{k-p}. \qquad (1)$$

Proof. Let us regard the set of all such (s,k)-indicators that have the first value "s" on the p-th position (p=s,... ...,k:

$$(l_1,\dots,l_{p-1};s;l_{p+1},\dots,l_k).$$

Denote the set of these indicators A_p and its power $\mathfrak{a}(A_p) = a_p$. The first p-1 components of the indicator have values from the set $\{1,\dots,s-1\}$ and they fulfill the conditions

$1^0, 2^0$. That means, they are $(s-1,p-1)$-indicators and their number equals to $K(s-1,p-1)$. The values of l_j $(p+1 \le j \le k)$ belong to set $\{1,\dots,s\}$ and have no restrictions, the number of their different combinations is s^{k-p}. So we have $a_p = K(s-1,p-1)s^{k-p}$, and so as $K(s,k) = \sum_{p=s}^{k} a_p$, the lemma is proved.

Corollary 1. The number $K(k)$ of all possible partitions of set $\{1,\dots,k\}$ is the following

$$K(k) = \sum_{s=1}^{k} K(s,k). \tag{2}$$

For practical calculations it is convenient to use another form of the formula (1), namely the following formula

$$K(s,k) = \sum_{p=0}^{s-1} \frac{(-1)^p (s-p)^k}{p!(s-p)!}. \tag{1'}$$

Lemma 2. Let $Q(s,k)$ be the number of different (s,k)-pseudoindicators. Then

$$Q(s,k) = K(s,k)2^{k-s}.$$

Indeed, from every (s,k)-indicator L it is possible to form 2^{k-s} (s,k)-pseudoindicators, substituting the values of l_j^h $(j > 1)$ by their opposite values.

Corollary 2. The number $Q(k)$ of all pseudoindicators for fixed k is following:

$$Q(k) = \sum_{s=1}^{k} Q(s,k) = \sum_{s=1}^{k} K(s,k)2^{k-s}. \tag{3}$$

3. Linear decomposition of correlation matrix. The mixture of random vectors.

Let $R = (r_{ij})$ be given $k \times k$-correlation matrix (that means, R is symmetric and non-negatively defined, $|r_{ij}| \le 1$). Let $\mathfrak{R}$ be an arbitrary set of $k \times k$ correlation matrices.

Def. 4. If there exists such a set of coefficients γ_i, fulfilling the conditions

$$\begin{cases} 1^0 \ \gamma_i \ge 0, \\ 2^0 \ \sum \gamma_i = 1, \end{cases} \tag{4}$$

that the equation

$$R = \sum \gamma_i R_i, \quad R_i \in \mathfrak{R} \tag{5}$$

holds, then the linear decomposition of R by the class $\mathfrak{R}$ exists.

Denote $\mathcal{G}(\mathfrak{R})$ the convex hull of set $\mathfrak{R}$. It is evident that the linear decomposition (5) of R exists if and only if

$R \in \mathcal{G}(\mathcal{R})$.

Then a linear decomposition, containing not more than M non-zero terms, exists, where

$$M = (k\cdot(k-1))/2 + 1 \qquad (6)$$

and

$$R = \sum_{i=1}^{M} \gamma_{j_i} R_{j_i}. \qquad (5')$$

The existence of linear decomposition of several correlation matrices is considered in [2,11]; the problem is of some interest from the point of view of data analysis.

Def. 5. Let $\mathcal{Y}$ be the set of all k-dimensional random vectors Y, $Y: (\Omega_Y, C_Y, P) \to R^k$ and (Ω, C, γ) - discrete probability space. Then the transformation $Z: \Omega \to \mathcal{Y}$ is the (discrete) mixture of vectors Y:

$$Z = \sum \chi_{\omega_i} Y_i,$$

where χ_{ω_i} is the indicator-function of event ω_i, $\bigcup \omega_i = \Omega$, $\omega_i \cap \omega_j = \emptyset$, $Y_i \in \mathcal{Y}$ (see [13]). If $Y_i \sim Q_i$ and $\gamma(\omega_i) = \gamma_i$, then the distribution Q of mixture Z equals to the mixture of distributions Q_i:

$$Q = \sum \gamma_i Q_i.$$

Some properties of mixtures are regarded in [13].

4. The case of equal symmetric distributions.

Assume k is fixed and given one-dimensional distributions P_i (see paragraph 1) fulfill the conditions

$$1^\circ \quad P_1 = P_2 = \ldots = P_k, \qquad (7')$$

denote $P_i = P_o$;

$$2^\circ \quad X \sim P_o \Leftrightarrow -X \sim P_o. \qquad (7'')$$

For simplicity sake let us denote the k-variate distribution $P(P_1,\ldots,P_k;R)$ with fixed marginals P_o and fixed correlation matrix R, now $P(\hat{P}_o;R)$.

Def. 6. Let I be an arbitrary k-index,

$$1 \in I, \qquad (8)$$

and $\hat{P}_o$ an one-dimensional distribution. Then the random vector $V=(V_1,\ldots,V_k)$,

$$V_i = \begin{cases} X, & \text{if } i \in I, \\ -X, & \text{if } i \in I^c, \end{cases} \quad X \sim \hat{P}_o$$

is the simple vector with marginals $\hat{P}_o$.

Def. 7. Let V be a simple vector, defined with the help of variable $\varkappa_0, \varkappa_0 \sim \hat{P}_0$ and by k-index I. Then the k-variate distribution of vector V is simple with marginals P_0.

Let us denote it $P_I(P_0)$ and the set of all k-variate simple distributions with fixed P_0 by $\mathfrak{P}_0(P_0)$; when the marginal distribution is fixed, we omit it and write simply $\mathfrak{P}_0$.

Corollary 3. From Lemma 1 it follows that the number H(k) of different simple vectors for fixed $\varkappa_0$ (and, consequently, the number of different simple distributions $\varkappa(\mathfrak{P}_0)$ for fixed P_0 as well) is the following:

$$H(k) = 2^{k-1}. \tag{9}$$

Def.8. The correlation matrix $R^0 = (r_{ij})$ is simple, if $|r^0_{ij}| = 1, i,j = 1,\ldots,k$.

Let us denote the set of all simple correlation matrices $\mathfrak{R}_0$.

Lemma 3. The correlation matrix of a simple vector (simple distribution) is simple. Indeed, from the definition 6 it follows that if V is a simple vector then

$$r_{ij} = r(V_i,V_j) = \begin{cases} 1, & \text{if } i \in I, j \in I \text{ or } i \in I^c, j \in I^c, \\ -1, & \text{if } i \in I, j \in I^c \text{ or } i \in I^c, j \in I. \end{cases}$$

Lemma 4. Let k be fixed. Then every simple matrix R^0 and one-dimensional symmetric distribution $\hat{P}_0$ define uniquely a k-variate simple distribution $P_I(\hat{P}_0)$.

Indeed, let us define k-index I, $I = (i_1,\ldots,i_s)$ by R^0 in the following way:

$$\begin{cases} 1 \in I, \\ j \in I, \text{ if } r_{1j} = 1, \\ j \in I^c, \text{ if } r_{1j} = -1. \end{cases} \tag{10}$$

By definitions 7 and 8 $\hat{P}_0$ and I define simple vector V and simple distribution $P_I(\hat{P}_0)$. So as I, calculated by (10), is the only k-index, satisfying (8) and having the property $R(V) = R^0$ by lemma 3, then the definition is unique.

Corollary 4. There exists one-to-one correspondence between the sets $\mathfrak{R}_0$ and $\mathfrak{P}_0$ (in the case of fixed k and $\hat{P}_0$).

From Lemma 4 it may be concluded that every simple matrix R^0 defines uniquely a k-index I, satisfying (8); let us denote then $R^0 = R(I)$ and write

$$\varphi(R(I)) = P_I, \quad \varphi^{-1}(P_I) = R(I). \tag{11}$$

for all different k-indices I, satisfying (8).

Let us order the set of all different indices, satisfying (8), in arbitrary wise. Then the ordering induces the ordering in sets $\mathcal{R}_o = \{ R_1^o, \ldots, R_H^o \}$ and $\mathcal{P}_o = \{ P_o^1, \ldots, P_o^H \}$, P_o^j being the simple distribution $P_{I_j}(P_o)$; the ordering saves the correspondence (11).

Let $\mathcal{L}(\mathcal{R}_o)$ be the convex hull of set $\mathcal{R}_o$. The corresponding convex hull of the set $\mathcal{P}_o$, $\mathcal{L}(\mathcal{P}_o)$ is the set of the finite mixtures of simple distributions, see Def. 5.

Let us extend the correspondence φ to the sets $\mathcal{L}(\mathcal{R}_o)$ and $\mathcal{L}(\mathcal{P}_o)$ with the help of the connection

$$R = \sum \gamma_i R_i^o \iff Q = \sum \gamma_i P_o^i, \tag{12}$$

where γ_i satisfy the conditions (4).

Theorem 1. Let $k \geqslant 2$, P_o - arbitrary symmetrical one-dimensional distribution and $k \times k$-correlation matrix $R \in \mathcal{L}(\mathcal{R}_o)$. Then there exists k-variate distribution $P(\hat{P}_o; R)$ that is defined as a mixture of M simple distributions from the set $\mathcal{P}_o(P_o)$, $M = k(k-1)/2+1$.

Proof. From the assumption $R \in \mathcal{L}(\mathcal{R}_o)$ it follows that there exists a linear decomposition (5) of R with coefficients $\gamma_1, \ldots, \gamma_M$. Let us define the distribution Q with the help of correspondence (12), that is

$$Q = \varphi(R) = \sum_{i=1}^{n} \gamma_i P_o^i .$$

As the mixture of distributions, having equal corresponding marginal distributions, has the same marginal distributions, see [13], so the distribution Q has marginals $\hat{P}_o$, and we have $P(\hat{P}_o; R) = Q$, q. e. d.

As there exist, in general, several decompositions of R by different sets of simple matrices $R_1^o, \ldots, R_N^o$, $M \leq N \leq H$, so the distribution $P(\hat{P}_o; R)$ is not unique in general. From the definition of decomposition it can be concluded that the set of decompositions of a given matrix by simple matrices and hence the set of distributions, defined in Theorem 1, is convex.

For the efficient construction of the distribution $P(\hat{P}_o)$ the first problem that raises is that of finding the decomposition (5) of given matrix R, that means, the solution of the system of M equations with H unknowns:

$$A \cdot g = B,$$

where M × H matrix A consists of columns

$$\begin{pmatrix} \underline{\text{vecl } R_o^h} \\ 1 \end{pmatrix}, \quad h = 1,\dots,H,$$

(vecl denotes the column-vector, constructed of columns of lower triangle of the symmetrical matrix without diagonal),

$$B = \begin{pmatrix} \underline{\text{vecl } R} \\ 1 \end{pmatrix}$$

and $g = (g_1,\dots,g_H)'$ is the vector of unknown coefficients, that must satisfy the condition

$$g_i \geqslant 0. \tag{13}$$

For the solution of the problem some methods of linear programming are usable. A very simple method, based on the idea of (random) choice of M columns from matrix A, is realized by the student of mathematical faculty K.Saarestik for the personal computer APPLE II and computer ES 1060, see [17].

Let us notice that the method described is most efficient for generating the large samples of values of random vector with distribution $P(\hat{P}_o;R)$, so as for generating each value of vector Z only two random numbers must be generated: one with mixing distribution γ and the other with distribution $\hat{P}_o$.

5. The case of equal marginals and positive correlations.

Assume $k \in N$ is fixed, $k \geqslant 2$, the given distributions P_i fulfill the condition (7') (denote $P_i = P_o$) and, besides, the given correlation matrix R has only nonnegative correlations:

$$r_{ij} \geqslant 0. \tag{14}$$

Our aim is to define the distribution $P = P(P_o;R^+)$. Partly the problem is solved in [13], where one possible construction is described. Here the result will be generalised.

Def. 9. Let L be arbitrary (s,k)-indicator ($1 \leq s \leq k$) and P_o an one-dimensional distribution. Then the random vector $W = (W_1,\dots,W_k)$, satisfying the conditions

1° $W_i = X_h$, if $l_i = h$, $i = 1,\dots,k$; $h = 1,\dots,s$,

2° $X_h \sim P_o$, $h = 1,\dots,s$,

3° X_h and X_g are independent, if $h \neq g$, $h,g = 1,\dots,s$.

is quasi-simple with marginals P_o.

Def. 10. Let W be a quasi-simple vector with marginals P_o, defined by (s,k)-indicator L. Then the k-variate distribution of vector W is quasi-simple.

Let us denote the quasi-simple distribution, defined by P_o and L, $P_L(P_o)$ and the set of all quasi-simple distributions, defined by P_o and different (s,k)-indicators L ($s = 1,\ldots,k$) $\mathcal{P}_t(P_o)$ or $\mathcal{P}_t$, when P_o is fixed.

Def. 11. The correlation matrix consisting of all correlations equal to 1 or 0 is a quasi-simple matrix.

Corollary 5. The correlation matrix of quasi-simple vector is quasi-simple. Indeed, let W be a quasi-simple vector. Then its correlation matrix R(W) has the following elements:

$$r_{ij} = \begin{cases} 1, & \text{if } l_i = l_j, \\ 0, & \text{if } l_i \neq l_j. \end{cases}$$

Similarly to the case of simple matrices and simple distributions (see Lemma 4 and Corollary 4) the one-to-one correspondence φ between the quasi-simple matrices R_i^+ ($i = 1,\ldots,K$), see (2) and quasi-simple distributions (with fixed P_o) P_i^+ can be defined,

$$\varphi(R_i^+) = P_i^+ = P_{L_i}(P_o), \quad i = 1,\ldots,K.$$

Denote $\mathcal{S}(\mathcal{R}_+)$ and $\mathcal{S}(\mathcal{P}_+)$ the convex hulls of sets R_+ and P_+. The correspondence φ can be extended for the elements of convex hulls $\mathcal{S}(\mathcal{R}_+)$ and $\mathcal{S}(\mathcal{P}_+)$ as well. From here immediately follows the proof of the following

Theorem 2. Let $k \geqslant 2$, P_o - arbitrary one-dimensional distribution and $k \times k$-correlation matrix $R^+ \in \mathcal{S}(\mathcal{R}_+)$. Then there exists k-variate distribution $P(P_o;R^+)$ that is defined as a mixture of M quasi-simple distributions from the set $\mathcal{P}_+(P_o)$.

Proof is similar to that of theorem 1: we find the linear decomposition of R^+ by quasi-simple matrices,

$$R^+ = \sum_{i=1}^{M} \gamma_i R_i^+, \tag{15}$$

and define the mixture as the corresponding element in set $\mathcal{S}(P_+)$:

$$Q = \varphi(R^+) = \sum_{i=1}^{n} \gamma_i P_i^+ = P(P_o;R^+).$$

The solution is not unique, and the set of all distributions $P(P_o;R^+)$, defined as mixtures of quasi-simple distributions, form a convex set.

By the construction of the distribution $P(P_o;R^+)$ key problem is the linear decomposition of R^+. The algorithm, described in paragraph 4 is usable for the case, too, but as the system of linear equations contains K unknows now $K \gg H$ (see table 1), the method of random search is not efficient for the case. Another method for finding linear decomposition of matrix R^+, using the idea, described in [13], was realized for computers APPLE II and ES 10 60 by the student of mathematical faculty K.Floren, see [17].

Algorithm 1. Let $R^h = R^{h-1} - \gamma_h R_h^+$ be the h-th residual matrix, where γ_h is the h-th coefficient and R_h^+ - the h-th quasi-simple matrix in the linear decomposition (15); $R^o = R$ the given correlation matrix.

Let $\|A\|^2 = \sum_{\substack{i<j \\ i,j=1}}^{k} a^2_{ij}$.

On every step h,h=1,...,M-1 the coefficient γ_h and the matrix R_h^+ are defined from the conditions

$$\|R^{h-1} - \gamma_h R_h^+\| = \min_{\substack{\gamma \in \{r;h-1\} \\ R^+ \in \mathcal{R}^+}} \|R^{h-1} - \gamma R^+\|, \qquad (16)$$

where $\{r;h-1\} = \{r_{12},\dots,r_{k-1,k}\}$ is the set of out-of-diagonal elements of R^{h-1}.

After M-1 steps always $\|R^{M-1}\| = 0$ and there are two possible outcomes:

1° $r_{ii}^{M-1} \geqslant 0$; then take $R_M^+ = I$ (unity matrix) and $\gamma_M = r_{ii}^{M-1}$; the desired decomposition (15) is found.

2° $r_{ii}^{M-1} < 0$; the algorithm did not give the decomposition (15). Unfortunately that does not mean in general that the linear decomposition does not exist at all. K.Floren constructed the modification of the algorithm that allows besides the values γ_h and R_h, minimizing the (16), consider the other ones, too. In such a way it is possible to get the solution in practically all the cases, but sometimes the time of solution may be rather long.

When the decomposition (15) is found, the algorithm of generation of values of the random vector Z is very efficient (but less than in case of symmetrical P_o, described in paragraph 4), so as for getting every realization of vector Z it is necessary to generate s+1 random numbers, in general

6. Another method of defining $P(\hat{P}_o;R)$ in the case of symmetrical equal marginals $\hat{P}_o$.

The algorithm, described in paragraph 5 is usable for solving the problem in the case of nonrestricted correlations but symmetrical marginals, as well. But for dealing with this case we have to define some new, more general concepts.

Assume $k \in N$ is fixed, $k \geq 2$ and the given distributions fulfill the conditions (7') and (7'').

Def. 12. Let T be arbitrary (s,k)-pseudo-indicator ($1 \leq s \leq k$) and P_o an one-dimensional symmetrical distribution. Then the random vector $U = (U_1,\dots,U_k)$, satisfying the conditions

1^o $W_i = X_h$, if $i \in J_h$,

2^o $W_i = -X_h$, if $i \in G_h$, $i=1,\dots,k$; $h=1,\dots,s$,

3^o $X_h \sim \hat{P}_o$, $h=1,\dots,s$,

4^o X_h and X_f are independent, if $h \neq f$, $h,f=1,\dots,s$,

is pseudosimple vector with marginals $\hat{P}_o$.

Def. 13. Let U be a pseudo-simple vector with marginals $\hat{P}_o$, defined by (s,k)-pseudoindicator T. Then the k-variate distribution of vector U is pseudo-simple.

Let us denote the pseudo-simple distribution $P_T(\hat{P}_o)$. and the set of all pseudo-simple distributions (for $s=1,\dots,k$) by $\mathcal{P}(\hat{P}_o) = \mathcal{P}_-$.

Def. 14. The correlation matrix, consisting of elements equal to 0, 1 and -1 is pseudo-simple.

Corollary 6. The correlation matrix of pseudo-simple vector is pseudo-simple: R(U) is defined by the following formulae:

$$r_{ij} = \begin{cases} 1, & \text{if } t_i = t_j, \\ -1, & \text{if } t_i = -t_j, \\ 0, & \text{if } |t_i| \neq |t_j|. \end{cases}$$

Corollary 7. For every pseudo-simple matrix and symmetrical onedimensional distribution there exists one uniquely defined pseudo-simple distribution with given equal marginals.

Denote the set of all pseudo-simple matrices (k fixed) by $\mathcal{R}_-$ and let its convex hull be $\mathcal{S}(\mathcal{R}_-)$, the set of all pseudo-simple distributions (k and $\hat{P}_o$ fixed) by $\mathcal{P}_-$ and its convex hull by $\mathcal{S}(\mathcal{P}_-)$. Then similarly to theorems 1 and 2 it is possible to prove the following

Theorem 3. Let k be natural, $k \geqslant 2$; $\hat{P}_0$ arbitrary one-dimensional distribution and $k \times k$ matrix $R^- \in \mathcal{S}(\mathcal{R}_-)$. Then there exists k-variate distribution $P(\hat{P}_0; R^-)$ that is defined as the mixture of M pseudo-simple distributions from the set $\mathcal{R}_-(\hat{P}_0)$.

Corollary 7'. The sets $\mathcal{S}(\mathcal{R}_0)$ and $\mathcal{S}(\mathcal{R}_-)$ coincide for every fixed k, $k \in N$ and symmetrical $\hat{F}_0$.

From the definitions of $\mathcal{R}_0$ and $\mathcal{R}_-$ follows the relation $\mathcal{R}_0 \subset \mathcal{R}_-$, consequently $\mathcal{S}(\mathcal{R}_0) \subset \mathcal{S}(\mathcal{R}_-)$ holds, too. It is easy to demonstrate, that every pseudo-simple matrix belongs to $\mathcal{S}(\mathcal{R}_0)$, consequently $\mathcal{S}(\mathcal{R}_-) \subset \mathcal{S}(\mathcal{R}_0)$, q.e.d.

The algorithm, presented in paragraph 5, can be used for the given case after a rather slight modification, proposed by K.Floren: instead of r_{ij}^h its absolute value defines the coefficients γ_i.

The answer to the problem - if there exists a linear decomposition (either (5), (5') or (15)) for every correlation matrix R - is negative. Indeed, the set $\mathcal{S}(\mathcal{R}^0)$ is convex polyhedra in the space R^{M-1}, and by the definition $\mathcal{R} \supset \mathcal{R}^0$ where $\mathcal{R}$ is the set of all correlation matrices (of order k). So as set $\mathcal{R}$ is surrounded (because of the condition of nonnegative definitedness) with surface of the k-th degree, the sets $\mathcal{R}$ and $\mathcal{S}(\mathcal{R}^0)$ cannot coincide. In general the only common points of the sets of boundary points $\overline{\mathcal{R}}$ and $\overline{\mathcal{S}(\mathcal{R}^0)}$ of sets $\mathcal{R}$ and $\mathcal{S}(\mathcal{R}^0)$ are the simple matrices $R_1^0, \ldots, R_H^0$.

The illustration of the sets $\mathcal{S}(\mathcal{R}^0)$, $\mathcal{S}(\mathcal{R}^+)$, $\mathcal{S}(\mathcal{R}^-)$ for the case k = 3 are given in figures 1-3.

7. Example 3.

Let k = 3. Then M=4. H=4, K=5 and Q=11. Let us write down the sets $\mathcal{R}_0$, $\mathcal{V}$, $\mathcal{R}_+$, $\mathcal{W}$, $\mathcal{U}$ and $\mathcal{R}_-$.

1°. Simple vectors and simple matrices.

$V_1 = (X,X,X)$; $V_2 = (X,X,-X)$; $V_3 = (X,-X,X)$; $V_4 = (X,-X,-X)$

$$R_1^0 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}; \quad R_2^0 = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}; \quad R_3^0 = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}; \quad R_4^0 = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

2°. Quasi-simple vectors and quasi-simple matrices.

$W_1 = (X_1, X_1, X_1)$; $W_2 = (X_1, X_1, X_2)$; $W_3 = (X_1, X_2, X_1)$; $W_4 = (X_1, X_2, X_2)$; $W_5 = (X_1, X_2, X_3)$

$$R_1^+=\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix};\quad R_2^+=\begin{pmatrix}1&1&0\\1&1&0\\0&0&1\end{pmatrix};\quad R_3^+=\begin{pmatrix}1&0&1\\0&1&0\\1&0&1\end{pmatrix};\quad R_4^+=\begin{pmatrix}1&0&0\\0&1&1\\0&1&1\end{pmatrix};$$

$$R_5^+=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}.$$

3^{o}. Pseudo-simple vectors and pseudo-simple matrices. $U_i=V_i$, $R_i^-=R_i^o$, $i=1,\dots,4$; $U_5=W_2$, $R_5^-=R_2^+$, $U_7=W_3$, $R_7^-=R_3^+$, $U_9=W_4$, $R_9^-=R_4^+$, $U_{11}=W_5$, $R_{11}^-=R_5^+$ and $U_6=(X_1,-X_1,X_2)$, $U_8=(X_1,X_2,-X_1)$. $U_{10}=(X_1,X_2,-X_2)$,

$$R_6^-\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&1\end{pmatrix};\quad R_8^-=\begin{pmatrix}1&0&-1\\0&1&0\\-1&0&1\end{pmatrix};\quad R_{10}^-=\begin{pmatrix}1&0&0\\0&1&-1\\0&-1&1\end{pmatrix}.$$

Let correlation matrices R_1, R_2, R_3 be given,

$$R_1=\begin{pmatrix}1&0.7&0.3\\0.7&1&-0.1\\0.3&-0.1&1\end{pmatrix}\quad R_2=\begin{pmatrix}1&0.5&0.3\\0.5&1&-0.1\\0.3&0.5&1\end{pmatrix}\quad R_3=\begin{pmatrix}1&0.7&0.3\\0.7&1&0.1\\0.3&0.1&1\end{pmatrix}$$

The system of linear equations for finding the linear decomposition by simple matrices is the following (with three columns of absolute terms correspondingly):

	R_1	R_2	R_3
$\gamma_1+\gamma_2-\gamma_3-\gamma_4=$	0.7	0.5	0.7
$\gamma_1-\gamma_2+\gamma_3-\gamma_4=$	0.3	0.3	0.3
$\gamma_1-\gamma_2-\gamma_3+\gamma_4=$	-0.1	-0.1	0.1
$\gamma_1+\gamma_2+\gamma_3+\gamma_4=$	1	1	1

So as for the case k=3 we have H = M = 4, the solution of the system gives the answer about the existence of linear decomposition at once. Let us write down the solutions: for R_1 we have γ = (0.475,0.375,0.175,-0.025), that means, the correlation matrix R_1 (being positively determined) has no linear decomposition. For R_2 we have γ=(0.425,0.325,0.225,0.025) and for R_3 γ=(0.525,0.325,0.125,0.025). For R_3 it is possible to find the linear decomposition by quasi-simple matrices, too.Using the algorithm, described in paragraph 5, we get the following coefficients: γ^+=(0.1,0.6,0.2,0,0.1). Analogically, the decomposition by pseudo-simple matrices for R_2 is following:

γ^-=(0,0,0,0,0.5,0,0.3,0,0,0.1,0.1).

Numbers of different simple (H), quasi-simple (K) and pseudo-simple matrices, depending on the dimensionality k.

k	M	H	K(s,k) s=1	2	3	4	5	6	7	8	9	10	K	Q(s,k) s=1	2	3	4	5	6	7	8	9	10	Q
1	1	1	1										1	1										1
2	2	2	1	1									2	2	1									3
3	4	4	1	3	1								5	4	6	1								11
4	7	8	1	7	6	1							15	8	28	12	1							49
5	11	16	1	15	25	10	1						52	16	120	100	20	1						257
6	16	32	1	31	90	65	15	1					203	32	496	720	260	30	1					1539
7	22	64	1	63	301	350	140	21	1				877	64	2016	4816	2800	560	42	1				10299
8	29	128	1		966		1050		28				4332	128		30912		8400		56				78977
				126		1793		266		1					8128		30288		1064		1			
9	37	256	1		3025		6951		462		1		21147	256		193600		111216		1848		1		609441
				255		7770		2646		36					32640		248640		21168		72			
10	46	512	1		9330		42525		5880		45		116743	512		1194240		1360800		47040		90		5333603
				511		34873		22827		750					130816		2231872		365232		3000		1	

8. The case of unequal marginals.

Let $P_1,\dots,P_k$ be arbitrary one-dimensional distributions, F_i being the distribution function of P_i, $i=1,\dots,k$.

For a pair of one-dimensional distributions P_1 and P_2 Hoeffding [5] defined the maximal and minimal common distributions P^* and P_* by their distribution functions $F^*(x,y)$ and $F_*(x,y)$ in the following way:

$$F^*(x,y) = \min(F_1(x),F_2(y)),$$
$$F_*(x,y) = \max(0,F_1(x)+F_2(y)-1)$$

and showed that if the distributions are of second order, then P^* and P_* have correspondingly the maximal and minimal correlations r^* and r_* in the class of all bivariate distributions P with marginals P_1 and P_2. Frechet [4] showed that these concepts may be generalized for the case $k > 2$.

We shall define some sets of k-variate distributions with marginals $P_1,\dots,P_k$, that generalize the simple, quasi-simple and pseudo-simple distributions, defined in paragraphs 4-6.

Let us denote $\min(F_{i_1}(x_{i_1}),\dots,F_{i_s}(x_{i_s})) = \min F_I(x_I)$, if $I=(i_1,\dots,i_s)$, and let the i-th marginal of F be $(F)_i$ and the ij-th bivariate marginal of F be $(F)_{ij}$.

Def. 15. Let $k \geqslant 2$, I - k-index, $1 \in I$ and $P_1,\dots,P_k$ arbitrary one-dimensional distributions of second order. The k-variate distribution $P_I^* = P_I^*(P_1,\dots,P_k)$, defined with the help of its distribution function $F_I^*(x_1,\dots,x_k)$,

$$F_I^*(x_1,\dots,x_k) = \max(0,\min F_I(x_I) + \min F_{Ic}(x_{Ic}) - 1)$$

is simple distribution with marginals $P_1,\dots,P_k$.

Lemma5. Simple distribution has the following properties:

1° The marginals of simple distribution equal to given one-dimensional distributions $P_1,\dots,P_k$.

2° The correlation coefficients of P_I have the following values:

$$r_{ij} = \begin{cases} r^*_{ij}, & \text{if } i \in I,\ j \in I \text{ or } i \in I^c,\ j \in I^c, \\ r_{*ij}, & \text{if } i \in I,\ j \in I^c \text{ or } i \in I^c,\ j \in I. \end{cases}$$

Proof. Let us calculate the marginals of $\min F_I(x_I)$:

$$(\min F_I(x_I))_i = \lim_{x_t \to \infty,\ t \neq i} F_I(x_I) = \begin{cases} 1, & \text{if } i \in I^c \\ F_i(x_i), & \text{if } i \in I. \end{cases}$$

Now we are able to calculate the marginals of simple distribution P_I:

$$(F_I^*)_i = \begin{cases} \max(0,F_i(x_i)+1 - 1, & \text{if } i \in I, \\ \max(0,1+F_i(x_i)-1, & \text{if } i \in I^c, \end{cases}$$

that means, always $(F_I)_i = F_i(x_i)$.

$$(F_I^*)_{ij} = \begin{cases} \max(0, \min I_i(x_i), F_j(x_j)) + 1 - 1 = (F^*)_{ij}, \\ \qquad \text{when } i,j \in I \text{ or } i,j \in I^c, \\ \max(0, F_i(x_i) + F_j(x_j) - 1) = (F_*)_{ij}, \\ \qquad \text{when } i \in I,\ j \in I^c \text{ or } i \in I^c, \\ \qquad j \in I, \end{cases}$$

q.e.d.

Corollary 8.

From the definition 15 and lemma 5 it follows that the simple distribution with equal symmetrical marginals, defined in Def. 6, is the special case of simple distribution. defined in 15.

Indeed, in the case $(F)^*_{ij}(x,y) = F(z)$, $(F_*)_{ij}(x,y) = 2F(z) - 1$, where $z = \min(x,y)$ and using the definition of simple vector and non-restricting assumption $X = 0$, $DX = 1$, we have:

$$r^*_{ij} = \int_{-\infty}^{\infty} xy\,dF^* = \int_{-\infty}^{\infty} x^2\,dF = 1,$$

$$r_{*ij} = \int_{-\infty}^{\infty} xy\,dF_* = -2\int_{0}^{\infty} x^2\,dF = -1.$$

Def. 16. The correlation matrix R^*, defined in the following way,

$$r_{ij} = \begin{cases} r^*_{ij}, \text{ if } i \in I,\ j \in I \text{ or } i \in I^c,\ j \in I^c, \\ r_{*ij}, \text{if } i \in I,\ j \in I^c \text{ or } i \in I^c,\ j \in I, \end{cases}$$

where I is any k-indicator, $i \in I$, is said to be simple correlation matrix for marginals $P_1,\ldots,P_k$.

It is evident that the number of all simple correlation matrices for every fixed set of marginals $P_1,\ldots,P_k$ is H.Denote the set of all simple correlation matrices $\{R^*_1,\ldots,R^*_H\} = \mathfrak{R}_*(P_1,\ldots,P_k) = \mathfrak{R}_*$ and the convex hull of the set $\mathcal{G}(\mathfrak{R}_*)$ Denote $\mathfrak{P}_*(P_1,\ldots,P_k) = \mathfrak{P}_* = \{P^*_1,\ldots,P^*_H\}$ the set of all k-variate simple distributions with marginals $P_1,\ldots,P_k$, and let $\mathcal{G}(\mathfrak{P}_*)$ be the convex hull of the set $\mathfrak{P}_*$.

Def. 17. Let k be natural, $k \geqslant 2$, L arbitrary (s,k)-indicator, defining the indeces $I_1,\ldots,I_s$, and $P_1,\ldots,P_k$ - given one-dimensional distributions of second order. The k-variate distribution $P^+_L = P^+_L(P_1,\ldots,P_k)$ is quasi-simple with marginals $P_1,\ldots,P_k$, if its distribution function is defined in the following way

$$F^+_L(x_1,\ldots,x_k) = (\min F_{I_1}(x_{I_1})) \times \ldots \times (\min F_{I_s}(x_{I_s})).$$

Def. 18. Let the assumptions of Def. 17 be fulfilled. The correlation matrix R^{+}_{*}, defined in the following way

$$r^{+}_{ij} = \begin{cases} r^{*}_{ij}, & \text{if } i,j \in I_h,\ h = 1,\dots,s, \\ 0, & \text{if } i \in I_h,\ j \in I_f,\ h \neq f,\ h,f = 1,\dots,s. \end{cases}$$

is said to be quasi-simple correlation matrix for marginals $P_1,\dots,P_k$.

The number of all quasi-simple distributions and quasi-simple matrices is K, (see [2]). Let us denote the sets $\mathcal{R}^{*}_{+} = \{R^{*+}_1,\dots,R^{*+}_K\}$ and $\mathcal{P}^{*}_{+} = \{P^{*+}_1,\dots,P^{*+}_K\}$ correspondingly and their convex hulls - $\mathcal{S}(\mathcal{R}^{*}_{+})$ and $\mathcal{S}(\mathcal{P}^{*}_{+})$.

Def. 19. Let k be natural, $k \geqslant 2$, T arbitrary (s,k)-pseudo-indicator defining the indices J_h, G_h, $h = 1,\dots,s$ and $P_1,\dots,P_k$ - given one-dimensional distributions of second order. The k-variate distribution $P^{-}_{T} = P^{-}_{T}(P_1,\dots,P_k)$ is pseudo-simple with marginals $P_1,\dots,P_k$, if its distribution function $F^{-}_{T}(x_1,\dots,x_k)$ is defined in the following way:

$$F^{-}_{T}(x_1,\dots,x_k) = (\max(0,\min(F_{J_1}(x_{J_1}))+\min(F_{G_1}))-1)\times\dots$$
$$\dots\times(\max(0,\min(F_{J_s}(x_{J_s}))+\min(F_{G_s}(x_{G_s}))-1).$$

Def. 20. Let the assumptions of Def. 19 be fulfilled. The correlation matrix R^{-}, defined in the following way,

$$r^{-}_{ij} = \begin{cases} r^{*}_{ij}, & \text{if } i,j \in J_h \text{ or } i,j \in G_h,\ h = 1,\dots,s, \\ r_{*ij}, & \text{if } i \in J_h,\ j \in G_h \text{ or } i \in G_h,\ j \in J_h, \\ & \qquad h=1,\dots,s, \\ 0, & \text{if } i \in I_h,\ j \in I_f,\ h \neq f,\ h,f = 1,\dots,s, \end{cases}$$

is said to be pseudo-simple correlation matrix for given marginals $P_1,\dots,P_k$.

The number of all different pseudo-simple distributions and pseudo-simple matrices (by fixed distributions $P_1,\dots,P_k$) is Q (see (3)). Let us denote the sets of pseudo-simple matrices and distributions and their convex hulls $\mathcal{R}^{-}_{*}$, $\mathcal{P}^{-}_{*}$, $\mathcal{S}(\mathcal{R}^{-}_{*})$ and $\mathcal{S}(\mathcal{P}^{-}_{*})$ correspondingly.

Let φ be one-to one correspondence between sets $\mathcal{R}_{*}$ and $\mathcal{P}_{*}$, $\mathcal{R}^{+}_{*}$ and $\mathcal{P}^{+}_{*}$, $\mathcal{R}^{-}_{*}$ and $\mathcal{P}^{-}_{*}$ extendeable to their convex hulls as well.

Similarly to theorems 1-3 it is possible to prove the following

Theorem 4. Let k be natural, $k \geqslant 2$, $P_1,\dots,P_k$ given one-dimensional distributions of second order, R - given correlation matrix, fulfilling at least one of the following conditions:

$$R \in \mathcal{J}(\mathcal{R}_*) \tag{17}$$

$$R \in \mathcal{J}(\mathcal{R}_*^+) \tag{17'}$$

$$R \in \mathcal{J}(\mathcal{R}_*^-), \tag{17''}$$

then there exists a k-variate distribution $P(P_1,\ldots,P_k;R)$, defined as the mixture of M distributions from the set $\mathcal{P}_*$, $\mathcal{P}_*^+$ or $\mathcal{P}_*^-$ correspondingly.

The conditions (17), (17') and (17") are not independent. So as by the definitions 16, 18 and 20 $\mathcal{R}_* \subset \mathcal{R}_*^-$ and $\mathcal{R}_*^+ \subset \mathcal{R}_*^-$, from the condition (17) or (17') follows (17"). In general, the sets $\mathcal{J}(\mathcal{R}_*)$ and $\mathcal{J}(\mathcal{R}_*^-)$ are not equivalent (as it was in the case of equal and symmetrical marginals, see corollary 7'). The fact will be proved in the example 5, see paragraph 9.

The main problem in the case of unequal marginals is, as in the case of equal marginals, too, the finding of linear decomposition of matrix R either in form (18), (18') or (18"):

$$R = \sum_{i=1}^{M} \gamma_i R_i^*, \tag{18}$$

$$R = \sum_{i=1}^{M} \gamma_i R_i^{+*}, \tag{18'}$$

$$R = \sum_{i=1}^{M} \gamma_i R_i^{-*}, \tag{18"}$$

where R_i^*, R_i^{+*} and R_i^{-*} are simple, quasi-simple and pseudo-simple correlation matrices, defined by marginals $P_1,\ldots,P_k$.

For finding these linear decompositions the methods, described in paragraphs 4-6 are of use. The calculation of correlations r_{ij}^* and r_{*ij} is only a technical problem, and in constructing general simple, quasi-simple and pseudo-simple matrices all rules, useful for a special case, are usable, only instead of 1 and -1 the corresponding values of r_{ij}^* and r_{*ij} must be placed.

The following is a necessary, but evidently not a sufficient, condition for the existence of linear decomposition of R:

$$r_{*ij} \leq r_{ij} \leq r_{ij}^*,$$

where r_{ij} are the coefficients of given correlation matrix, r_{ij}^* and r_{*ij} - the minimal and maximal correlations for given marginals P_i, P_j.

The linear decompositions (18) and (18') are useful because of the fact that the sets $\mathcal{R}_*$ and $\mathcal{R}_*^+$ have much less power than $\mathcal{R}_*^-$ and so the finding of (18) and (18') is more

efficient than the finding of (18"). Let us notice that the necessary condition for the existence of decomposition (18') is

$$r_{ij} \geqslant 0.$$

From the theorem 4 it is easy to conclude the corollary, dealing with the definition of distribution $P(P_0;R)$ in the case of equal, but asymmetrical marginals and correlation matrix R, not fulfilling the conditions (14):

Corollary 9. Let k be natural, the distributions $P_1,\dots,P_k$ fulfill the conditions (7'), P_0 being in general asymetrical.

Let r_* be the minimal correlation coefficient for equal marginals P_0, defined by its bivariate distribution function F_*:

$$F_*(x,y) = \max(0,F(x)+F(y)-1).$$

Then every correlation matrix, defined with the help of pseudo-indicator T in the following way:

$$r_{ij} = \begin{cases} 1, & \text{if } t_i=t_j, \\ r_*, & \text{if } t_i=-t_j, \\ 0, & \text{if } |t_i| \neq |t_j| \end{cases} \qquad (19)$$

is pseudo-simple for equal, asymmetrical marginals P_0, and the distribution $P(P_0;R)$ exists, when $R \in \mathcal{P}(\mathcal{R}_*^-)$, $\mathcal{R}_*^-$ being the set of pseudo-simple correlation matrices, defined by (19).

9. Example 4.

Let given distributions be the following: $EX_i = 0$, $DX_i = 1$.

P_1:

x	$-\sqrt{\frac{1}{3}}$	$\sqrt{3}$
p	0.75	0.25

P_2:

x	$-\sqrt{3}$	$\sqrt{\frac{1}{3}}$
p	0.25	0.75

P_3:

x	-1	1
p	0.5	0.5

The maximal distributions and their correlation coefficients:

P^*_{12}	$-\sqrt{1/3}$	$\sqrt{3}$
$-\sqrt{3}$	0.25	0
$\sqrt{1/3}$	0.5	0.25

P^*_{13}	$-\sqrt{1/3}$	$\sqrt{3}$
-1	0.5	0
1	0.25	0.25

P^*_{23}	$-\sqrt{3}$	$\sqrt{1/3}$
-1	0.25	0.25
1	0	0.5

$r^*_{12} = 0.33333$ $r^*_{13} = 0.57735$ $r^*_{23} = 0.57735$

The minimal distributions and their correlation coefficients:

P_{*12}	$-\sqrt{1/3}$	$\sqrt{3}$
$-\sqrt{3}$	0	0.25
$\sqrt{1/3}$	0.75	0

P_{*13}	$-\sqrt{1/3}$	$\sqrt{3}$
-1	0.25	0.25
1	0.5	0

P_{*23}	$-\sqrt{3}$	$\sqrt{1/3}$
-1	0	0.5
1	0.25	0.25

$r_{*12}=-1$ $r_{*13}=-0.57735$ $r_{*23}=-0.57735$

Simple distributions and simple matrices:

$I_1=(1,1,1)$; P_1^o

		$x_3=-1$		$x_3=1$	
	x_1	$-\sqrt{1/3}$	$\sqrt{3}$	$-\sqrt{1/3}$	$\sqrt{3}$
x_2	$-\sqrt{3}$	0.25	0	0	0
	$\sqrt{1/3}$	0.25	0	0.25	0.25

R_1^{o*}

$$\begin{pmatrix} 1 & 0.3333 & 0.5744 \\ 0.3333 & 1 & 0.5744 \\ 0.5774 & 0.5774 & 1 \end{pmatrix}$$

$I_2=(1,1,2)$; P_2^o

		$x_3=-1$		$x_3=1$	
x_2	$-\sqrt{3}$	0	0	0.25	0
	$\sqrt{1/3}$	0.25	0.25	0.25	0

R_2^{o*}

$$\begin{pmatrix} 1 & 0.3333 & -0.5774 \\ 0.3333 & 1 & -0.5774 \\ -0.5774 & -0.5774 & 1 \end{pmatrix}$$

$I_3=(1,2,1)$; P_3^o

		$x_3=-1$		$x_3=1$	
x_2	$-\sqrt{3}$	0	0	0	0.25
	$\sqrt{1/3}$	0.5	0	0.25	0

R_3^{o*}

$$\begin{pmatrix} 1 & -1 & 0.5774 \\ -1 & 1 & -0.5774 \\ 0.5774 & -0.5774 & 1 \end{pmatrix}$$

$I_4=(1,2,2)$; P_4^o

		$x_3=-1$		$x_3=1$	
x_2	$-\sqrt{3}$	0	0.25	0	0
	$\sqrt{1/3}$	0.25	0	0.5	0

R_4^{o*}

$$\begin{pmatrix} 1 & -1 & -0.5774 \\ -1 & 1 & 0.5774 \\ -0. & 0.5774 & 1 \end{pmatrix}$$

Quasi-simple distributions and quasi-simple matrices

$L_1 = I_1$, $P_1^+ = P_1^o$; $R_1^+ = R_1^o$

$L_2=(1,1,2)$; P_2^+

		$x_3=-1$		$x_3=1$	
x_2	$\sqrt{3}$	0.125	0	0.125	0
	$\sqrt{1/3}$	0.25	0.125	0.25	0.125

R_2^{+*}

$$\begin{pmatrix} 1 & 0.3333 & 0 \\ 0.3333 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$L_3=(1,2,1)$; P_3^+

		$x_3=-1$		$x_3=1$	
x_2	$\sqrt{3}$	0.125	0	0.0625	0.0625
	$\sqrt{1/3}$	0.375	0	0.1875	0.1875

R_3^{+*}

$$\begin{pmatrix} 1 & 0 & 0.5774 \\ 0 & & 0 \\ 0.5774 & 0 & \end{pmatrix}$$

$L_4=(1,2,2)$; P_4^+

		$x_3=-1$		x_3	
x_2	$-\sqrt{3}$	0.1875	0.0625	0	
	$\sqrt{1/3}$	0.1875	0.0625	0.375	0.125

R_4^{+*}

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5774 \\ 0 & 0.5774 & 1 \end{pmatrix}$$

$L_5 = (1,2,3); P_5^+$

	$x_3 = -1$		$x_3 = 1$	
x_2 \ x_1	$\sqrt{1/3}$	$\sqrt{3}$	$\sqrt{1/3}$	$\sqrt{3}$
$\sqrt{3}$	0.09375	0.09125	0.09375	0.09125
$\sqrt{1/3}$	0.28125	0.09375	0.28125	0.09375

$$R_5^{+*} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Pseudo-simple distributions and pseudo-simple matrices:

$T_1 = I_1;\ P_1^- = P_1,\ R_1^- = R_1 \quad T_5 = L_2;\ P_2^- = P_2^+;\ R_2^- = R_2^+$
$T_2 = I_2;\ P_2^- = P_2^0,\ R_2^- = R_2^0 \quad T_7 = L_3 \quad P_3^- = P_3^+;\ R_3^- = R_3^+$
$T_3 = I_3;\ P_3^- = P_3^0,\ R_3^- = R_3^0 \quad T_9 = L_4 \quad P_4^- = P_4^+;\ R_4^- = R_4^+$
$T_4 = I_4;\ P_4^- = P_4^0,\ R_4^- = R_4^0 \quad T_{11} = L_5 \quad P_{11}^- = P_5^+;\ R_{11}^- = R_5^+$

T_6, T_8, T_{10} are new and the corresponding distributions and matrices:

$T_6 = (1,-1,2); P_6^-$

	$x_3 = -1$		$x_3 = 1$	
x_2 \ x_1	$\sqrt{1/3}$	$\sqrt{3}$	$\sqrt{1/3}$	$\sqrt{3}$
$\sqrt{3}$	0	0.125	0	0.125
$\sqrt{1/3}$	0.375	0	0.375	

$$R_6^{-*} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$T_8 = (1,2,-1); P_8^-$

	$x_3 = -1$		$x_3 = 1$	
x_2 \ x_1	$\sqrt{1/3}$	$\sqrt{3}$	$\sqrt{1/3}$	$\sqrt{3}$
$\sqrt{3}$	0.0625	0.0625	0.125	0
$\sqrt{1/3}$	0.1875	0.1875	0.125	0

$$R_8^{-*} = \begin{pmatrix} 1 & 0 & -0.5774 \\ 0 & 1 & 0 \\ -0.5774 & 0 & 1 \end{pmatrix}$$

$T_{10} = (1,2,-2); P_{10}^-$

	$x_3 = -1$		$x_3 = 1$	
x_2 \ x_1	$\sqrt{1/3}$	$\sqrt{3}$	$\sqrt{1/3}$	$\sqrt{3}$
$-\sqrt{3}$	0	0	0.1875	0.0625
$\sqrt{1/3}$	0.375	0.125	0.1875	0.0625

$$R_{10}^{-*} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.5774 \\ 0 & -0.5774 & 1 \end{pmatrix}$$

Let the given correlation matrix be the following:

$$R = \begin{pmatrix} 1 & 0.2 & 0.4 \\ 0.2 & 1 & 0.3 \\ 0.4 & 0.3 & 1 \end{pmatrix}$$

I. For finding the linear decomposition by simple matrices we have the system of equations:

$$\begin{cases} 0.33333\gamma_1 + 0.3333\gamma_2 - \gamma_3 - \gamma_4 = 0.2 & (A) \\ 0.57735\gamma_1 - 0.57735\gamma_2 + 0.57735\gamma_3 - 0.57735\gamma_4 = 0.4 & (B) \\ 0.57735\gamma_1 - 0.57735\gamma_2 - 0.57735\gamma_3 + 0.57735\gamma_4 = 0.3 & (C) \\ \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1 & (D) \end{cases}$$

The solution of the system is: $\gamma_1 = 0.7531$; $\gamma_2 = 0.1469$, $\gamma_3 = 0.0933$; $\gamma_4 = 0.0067$. That means, one of the desired distributions P is the following:

$$0.7531\ P_1^0 + 0.1469\ P_2^0 + 0.0933\ P_3^0 + 0.0067\ P_4^0 = P$$

P:

x_2 \ x_1	$x_3 = -1$		$x_3 = 1$	
	$-\sqrt{1/3}$	$\sqrt{3}$	$-\sqrt{1/3}$	$\sqrt{3}$
$-\sqrt{3}$	0.1883	0.0017	0.0367	0.0233
$\sqrt{1/3}$	0.2733	0.0367	0.2517	0.1883

For checking, if P has the correlation matrix R, let us compute the bivariate marginals of R and their correlation coefficients:

P_{12}	$-\sqrt{1/3}$	$\sqrt{3}$
$-\sqrt{3}$	0.225	0.025
$\sqrt{1/3}$	0.525	0.225

$r_{12} = 0.2$

P_{13}	$-\sqrt{1/3}$	$\sqrt{3}$
-1	0.4616	0.0384
1	0.2884	0.2116

$r_{13} = 0.4$

P_{23}	$-\sqrt{3}$	$\sqrt{1/3}$
-1	0.19	0.31
1	0.06	0.44

$r_{23} = 0.3$

q.e.d.

II. So as all correlations of the given matrix R are positive, it is possible to find the linear decomposition by quasi-simple matrices, too. Using the algorithm, described in paragraph 5, we have:

$$\gamma_1^+ = 0.3 : 0.57735 = 0.5196.$$

The first residual matrix:

$$R_+^1 = \begin{pmatrix} 0.4804 & 0.0268 & 0.1 \\ 0.0268 & 0.4804 & 0 \\ 0.1 & 0 & 0.4804 \end{pmatrix}$$

$$\gamma_2^+ = 0.0268 : 0.3333 = 0.0804.$$

The second residual matrix

$$R_+^2 = \begin{pmatrix} 0.4 & 0 & 0.1 \\ 0 & 0.4 & 0 \\ 0.1 & 0 & 0.4 \end{pmatrix}$$

$$\gamma_3^+ = 0.1 : 0.57735 = 0.1732.$$

The third residual matrix

$$R_+^3 = \begin{pmatrix} 0.2268 & 0 & 0 \\ 0 & 0.2268 & 0 \\ 0 & 0 & 0.2268 \end{pmatrix}$$

$$\gamma_4^+ = 0; \quad \gamma_5^+ = 0.2268,$$

and we have: $P = 0.5196 \cdot P_1^+ + 0.0804 \cdot P_2^+ + 0.1732 P_3^+ + 0.2268 P_5^+$.

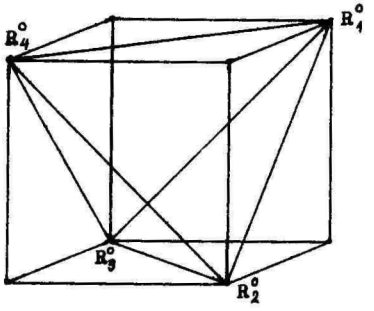

Fig. 1

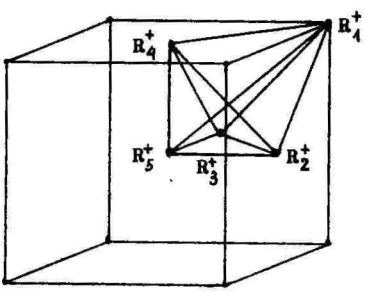

Fig. 2

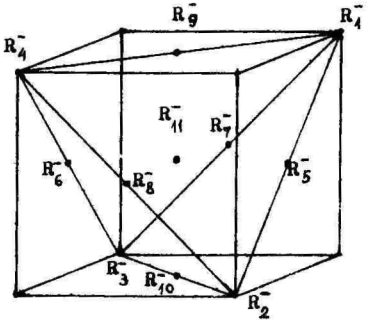

Fig. 3

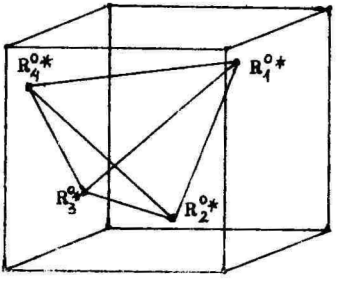

Fig. 4

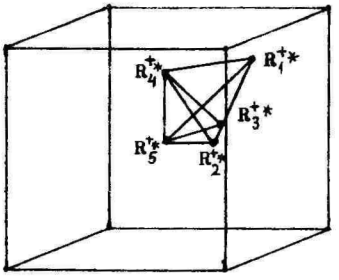

Fig. 5

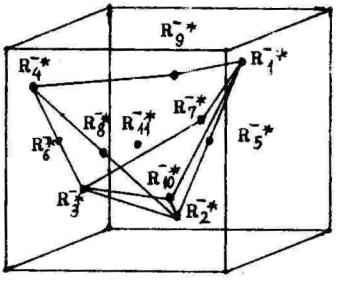

Fig. 6

$x_3 = -1$			$x_3 = 1$	
x_1 / x_2	$-\sqrt{1/3}$	$\sqrt{3}$	$-\sqrt{1/3}$	$\sqrt{3}$
$-\sqrt{3}$	0,1829	0.0071	0.0421	0.0179
$\sqrt{1/3}$	0.2787	0.0313	0.2463	0.1937

The solutions I and II do not coincide, it follows from here that every convex combination $\alpha P(I) + (1-\alpha)P(II)$ being the mixture of solutions, is the solution again, $\alpha \in [0,1]$.

Example 5. Let us take k = 3, the distributions P_1, P_2, P_3 equal to these of example 4,

$$R = R_3^- = \begin{pmatrix} 1 & 0.57735 & 0 \\ 0.57735 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and let us write down the system for finding the linear decomposition of R by simple matrices

$$\begin{cases} 0.3333\gamma_1 + 0.3333\gamma_2 - \gamma_3 - \gamma_4 = 0 \\ 0.57735\gamma_1 - 0.57735\gamma_2 + 0.57735\gamma_3 - 0.57735\gamma_4 = 0.57735 \\ 0.57735\gamma_1 - 0.57735\gamma_2 - 0.57735\gamma_3 + 0.57735\gamma_4 = 0 \\ \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1. \end{cases}$$

The solution $\gamma_1 = 0.125$, $\gamma_2 = 0.625$, $\gamma_3 = -0.125$, $\gamma_4 = 0.375$ does not satisfy the condition (13) and hence $R_8^- \notin \mathcal{P}(\mathcal{R}_0)$.

From here it can be concluded that in the case of unequal marginals the sets $\mathcal{P}(\mathcal{R}_*)$ and $\mathcal{P}(\mathcal{R}_*^-)$ do not coincide in general.

The sets $\mathcal{P}(\mathcal{R}_*)$, $\mathcal{P}(\mathcal{R}_*^+)$ and $\mathcal{P}(\mathcal{R}_*^-)$ for distributions, fixed in example 4, are given in figures 4-6.

References

1. Andĕl J. Dependent random variables with a given marginal distribution. - Acta Univ. Carol. Math. et phys., 1983, v. 24, No. 1, p. 3-12.
2. Anderson T.W. Asymptotically efficient estimation of covariance matrices with linear structure. - Ann. Statist., 1973, v. 1, No.1,p. 135-141.
3. Benzécri, J.P. Sur une généralisation du probleme de l'ajustment d'une mesura a desmarges. - Les Cahiers de l'Analyse des Données, 1985. Vol. VIII, N°3, 359-370.
4. Fréchet M. Sur les tableaux de correlation dont les marges sont données. - Ann. Univ. Lyon. Sect. A, 1951, fasc. 14, p. 53-77.
5. Hoeffding W. Masstabinvariante Korrelationstheorie. - Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin,

1940, B. 5, S. 179-233.

6. Kellerer H.G. Masstheoretische Marginalprobleme. - Math. Ann., 1964, B. 153, H. 3, S. 168-198.

7. Kieffer J.C. Generators with prescribed marginals for nonergodic automorphisms. - In: Trans. 9th Prag. Conf. Inf. Theory, Statist. Decis. Funct., Random Process., Prague, June 28 - July 2, 1982. Vol. B., Prague, 1983, p. 17-21.

8. Kobliha F. Náhodná čísla s roveměrným rozdělením a danou korelací. Ekonomicko-Matematicky Obzor., 1984, v. 20, 1, 82 - 87.

9. Liu Bede, Munson D.C., Jr. Generation of a random sequence having a jointly specified marginal distribution and autovariance. - IEEE Trans. Acoust., Speech, and Signal Process., 1982, v. 30, No. 6, p. 973-983.

10. Mei-ling Ting Lee. Dependence by total positivity. - The Annals of Probability, 1985, Vol. 13, № 2, 572-582.

11. Shapiro A. Asymptotic distribution theory in the analysis of covariance structures. - South African Statist. J., 1983, v. 17, No. 1, p. 33-81.

12. Strassen V. The existence of probability measures with given marginals. - Ann. Math. Statist., 1965, v. 36, No 2, p. 423-439.

13. Tiit E.-M. Definition of random vectors with given marginal distributions and given correlation matrix. - Tartu Riikliku Ülikooli Toimetised, 1984, No 685, p. 21-36.

14. Tiit E.-M. The multivariate distributions with given marginals. Fourth International Vilnius Conference on Probability theory and mathematical statistics. Abstracts of Communications I-IV, 305-307. Vilnius, 1985.

15. Whitt W. Bivariate distributions with given marginals. - Ann. Statist., 1976, v. 4, No 6, p. 1280-1289.

I6. Тийт Э.-М. Определение векторов по заданным корреляционным матрицам и маргинальным распределениям или моментам. "Теоретические и прикладные вопросы математики", тезисы докладов т. II, I42-I44. Тарту, I985.

I7. Тийт Э.-М.А., Саарестик К.К., Флорен К.П. Алгоритм для моделирования многомерных случайных векторов с заданкорреляционной матрицей и заданными маргинальными распределениями. - Применение многомерного статистиче-

ского анализа в экономике и оценке качества продукции. Тезисы II, 219–221, Тарту, 1985.

СЛУЧАЙНЫЕ ВЕКТОРЫ С ЗАДАННЫМИ ПРОИЗВОЛЬНЫМИ МАРГИНАЛЬНЫМИ РАСПРЕДЕЛЕНИЯМИ И ЗАДАННОЙ КОРРЕЛЯЦИОННОЙ МАТРИЦЕЙ

Э. Тийт

Р е з ю м е

Проблема существования и определения многомерного распределения с заданными маргинальными распределениями и заданной корреляционной матрицей является существенной при применений метода Монте-Карло в статистике. Частичные решения данной проблеме предложены в [1,3–10,12,13].

В настоящей статье решается следующая проблема: заданы одномерные распределения 2-го порядка $P_1,\ldots,P_k$ и $k \times k$ корреляционная матрица R. Требуется определить k-мерное распределение $P = P(P_1,\ldots,P_k; R)$, обладающее заданными маргинальными распределениями и заданной корреляционной матрицей.

В пункте 4 задача решается для случая, когда все маргинальные распределения равны друг другу и симметричны: $P_i = P_0$, $i = 1,\ldots,k$. Находится линейное разложение (5) матрицы R (см. [2,11]) по т.н. простым корреляционным матрицам R_i^0 (это такие матрицы, элементами которых являются только 1 и -1). При фиксированном P_0 каждой простой корреляционной матрице соответствует простое k-мерное распределение (т.е. распределение, при котором все компоненты случайного вектора равны друг другу по абсолютной величине). Искомым распределением P является конечная смесь простых распределений, причем смешивающее распределение определяется коэффициентами линейного разложения. В теореме 1 доказывается, что распределение P существует, если $R \in \mathcal{Y}(R^0)$, где $\mathcal{Y}$ – выпуклая оболочка множества R_0 простых корреляционных матриц. Решение не единственное.

В пунктах 5 и 6 предлагаются альтернативные решения поставленной задачи, основывающие на разложении матрицы R по корреляционным матрицам, содержащим элементы $\{-1,0,1\}$. Излагается алгоритм нахождения линейного разложения матрицы R.

В пункте 8 результаты обобщаются для случая неравных и неисмметричных маргинальных распределений, пользуясь притом минимальными и максимальными совместными распределениями Хоффдинга (см. [6]). Полученные результаты иллюстрируются при помощи 3 пример и 6 рисунков.

Received August 1985

EXACT SAMPLES OF MULTIVARIATE DISTRIBUTIONS AND THEIR EXPLOITATION IN STATISTICAL ALGORITHM'S TESTING

E. Tiit

§1. The set-up of the problem

1. The possible ways of testing of algorithms in multivariate statistics.

In the situation, when many different statistical packages and special programs are available for numerous types of computers, exploited by great number of users (most of them non-mathematicians), the problem of testing, comparing and estimation of algorithms and programs is of increasing interest, see [1,6,7].

Usually for testing of algorithms of multivariate statistical analysis the following two methods are used:

1° The algorithms are tested with the old "classical" examples, used by many authors and described in published papers.

2° The algorithms are tested with the help of Monte-Carlo method.

Both methods have their shortages. The "old" examples are in some sense random, they are not suitable for the systematic study of the algorithms. Besides, in general, the accuracy of the results of old examples is sometimes questionable. For using Monte-Carlo method several random samples must be generated, but it is also connected with some problems. At first, there are no generators for arbitrary multivariate distributions. The second problem is that the Monte-Carlo results consist random errors, and for attaining the significant estimations the large series of repeated trials are needed. From here it follows that the Monte-Carlo study needs many resources of computer (time, memory, completing software) and therefore cannot be used in the case of micro or personal computer for the systematical analysis of multivariate methods of mathematical statistics.

In the paper the alternative approach that is grounded on the concept of exact sample is suggested. The method is useable for testing the procedures of classical multivariate statistics, sufficient statistics for them being the first and second moments of the parent distribution. The construction of some samples for random vectors with dimensionality 20 - 30 and the calculation of the exact values of the parameters of multivariate models for them is possible with help of arbitrary computers (including the most modest personal ones).

The methodology of the constructions is based on the definition of multivariate distributions with given correlation matrices and given marginals, described in [4].

2. The concept of exact sample.

Let k be fixed $k \geqslant 2$, and $\mathcal{P}_k$ the set of all k-variate distributions. Let $\vartheta = (\vartheta_1, \ldots, \vartheta_r)$ be any parameter-vector of k-variate distribution, $\vartheta \in \Theta$. For every fixed value ϑ_0 of ϑ we have the set π_{ϑ_0} of all distributions with parameter $\vartheta = \vartheta_0$, $\pi_{\vartheta_0} \subset \mathcal{P}_k$.

Say

$$M(X, \beta) = 0, \tag{1}$$

where X is k-variate random vector, is the statistical model, identified by the parameter-vector β, $\beta = (\beta_1, \ldots, \beta_s)$, $\beta \in B$. The aim of the statistical procedure is to estimate the model parameter β on the basis of some sample $\mathbb{X}$, $\mathbb{X}$ being representative for population X with any distribution $P = P_\vartheta$ from $\mathcal{P}_k$. Let us assume that the model parameter β is a function of distribution parameter ϑ,

$$\beta = \beta(\vartheta).$$

In the case we have the theoretical (true) value β_* of model parameter,

$$\beta_* = \beta(\vartheta_0),$$

equal for all distributions from the set π_{ϑ_0} for every fixed $\vartheta_0 \in \Theta$.

Let $\mathbb{X}$ be $k \times n$ matrix, $\mathbb{X} = (x_{ij})$, $i = 1, \ldots, k$; $j = 1, \ldots, n$. We say $\mathbb{X}$ is sample and regard every row $x_{(i)}$ as sample of the i-th component X_i of the random vector X and every column $x^{(j)}$ as one object (the realization or complex of measurements of the object). The sample $\mathbb{X}$ defines uniquely the empirical distribution $P(\mathbb{X})$, where every object of the

sample has the probability 1/n.

If the empirical distribution P(X) belongs to the class of distributions π_{ϑ_0}, then X is said to be the exact sample for all distributions from π_{ϑ_0}. The exact sample with minimal size is said to be the optimal exact sample for the class π_{ϑ_0} (or for $P, P \in \pi_{\vartheta_0}$).

It is evident that if X is the exact sample for a class π_{ϑ}, then for every natural m the sample $X^{:m}$, defined as block-matrix

$$X^{:m} = \underbrace{(X:X:\ldots:X)}_{m \text{ times}}$$

is also the exact sample for π_{ϑ} with sample size mn, when X has size n.

3. The existence of the exact sample.

For us it is of interest to find such a subset Θ_f of set Θ, that for each set $\pi_{\vartheta'}$ of distributions there exists an exact sample, if $\vartheta' \in \Theta_f$.

It is evident that in the case in a set of distributions $\pi_{\vartheta'}$ any distribution P_0, fulfilling following conditions, must exist:

$$\begin{cases} 1^\circ\ P_0 \text{ has finite support.} \\ 2^\circ\ \text{All probabilities of } P_0 \text{ are rational.} \end{cases} \qquad (2)$$

Of course, that does not mean that all distributions in $\pi_{\vartheta'}$ fulfil the conditions $1^\circ - 2^\circ$; on the contrary, in $\pi_{\vartheta'}$ continuous distributions P may exist, as well. Then every distribution P_0 is one of discrete analogues of continuous distribution P (see, for instance, [10]).

So as our aim is the study of classical multivariate procedures, and for most parameters of these models are the first and second empirical moments sufficient statistics, then in the future we are concerned with the case when ϑ is the vector of first and second moments. Then its dimension r depends on the dimension of parent distribution k in the following way:

$$r = 0.5\,(k^2 + 3k). \qquad (3)$$

4. The error of algorithm and the admissibility region of algorithm.

Let A be any algorithm for estimation β by sample (empirical distribution). Let X be an exact sample (of size

n) for the set π_ϑ, ϑ fixed. Let $\beta(\vartheta, n, A)$ be the value of β found by X with help of the algorithm A. Let us express the difference between the calculated and theoretical values of parameter β of the model as the sum of two terms:

$$\beta(\vartheta, n, A) - \beta^* = b(n) + \varepsilon(\vartheta, A), \qquad (4)$$

where the first term b(n) depends on sample size and expresses the expected bias in the estimation of β due to finiteness of sample (assuming that the population is infinite). Then for consistent estimations we have

$$\lim_{n\to\infty} b(n) = 0.$$

Another term of the left side of the expression (4) is the error of the algorithm A. The source of the error may be insufficient precision of calculation, some peculiar value of ϑ or any mistake in algorithm or program. In algorithm-testing problems we assume that there is given some positive constant η in such a way that if

$$\varepsilon(\vartheta, A) < \eta, \qquad (5)$$

then the algorithm A is admissible in the case of the given value of ϑ, and if the condition (5) is not fulfilled, the algorithm is not admissible. For every A it is of interest to find the set $\Theta(A)$ of parameters ϑ fulfilling the condition (5) - it is the admissible region of algorithm A. In the admissible region it is possible to neglect the error of algorithm $\varepsilon(\vartheta, A)$.

5. The study of algorithm errors with the help of experimental design.

It is natural to find for any algorithm A its admissible region $\Theta(A)$. As a rule, a great amount of trials is needed and it seems that the results of experimental design theory must be used for increasing the efficiency of study.

The general scheme of research is standard for statistical investigation: Let us choose a sample of points $\vartheta^1, \ldots, \vartheta^N$ (the plan), in the space Θ, construct any exact sample for every point ϑ^j, $j = 1, \ldots, N$ and calculate the parameter $\beta(\vartheta^j) = \beta^j$ of model (1) by every exact sample. Then calculate by formula (4) the error $\varepsilon(\vartheta^j, A)$, dependent on ϑ^j. Now the problem of fin-

ding the model $G(\vartheta)$, describing the dependence of error on ϑ:

$$\varepsilon(\vartheta, A) = G(\vartheta), \qquad (6)$$

is solvable with the help of regression or variance analysis. On the basis of model (6) the admissible region $\Theta(A)$, see (5), can be described as well.

By realizing the described idea in practice there arise several problems:

1° The number r of different parameters ϑ_i (components of distribution parameter ϑ) is too large even in the case, when we restrict us with the first and second moments only (see (3)). The parameters are not independent (for instance, the correlation coefficients are connected with the condition of non-negative definiteness of correlation matrix).

2° There do not exist many standard methods of
- defining the multivariate distributions with given parameters,
- construction of exact samples for them.

Let us notice that when using the random samples instead of exact samples the trials must be repeated multiple times to estimate the error.

3° For calculating the value of algorithm error we must know the exact value of model parameter $\beta = (\beta^j)$, calculated independently from the algorithm A, and the bias b(n) as well.

In the paper some possible ways of solving the problems mentioned are given.

6. Decreasing of the number r of distribution's parameters ϑ

One possible solution of the problems 1° and 3° is the suitable parametrization of the parameter-vector ϑ

Let $\psi = (\psi_1, \ldots, \psi_t)$, $t \leq r$, ψ_i does not depend functionally on ψ_j $(j = 1, \ldots, t,\ j \neq i)$, $\psi \in \Psi$, and there exist analytical expressions

$$\begin{cases} \vartheta_i = \vartheta_i(\psi), & i = 1, \ldots, r; \\ \beta_j = \bar{\beta}_j(\psi) = (\beta_j(\vartheta)), & j = 1, \ldots, s. \end{cases} \qquad (7)$$

Then we have the parametrization rules for distribution parameter vector ϑ and its function - model parametervector β. If for every $\vartheta \in \Theta_0$, where Θ_0 is some subset of Θ, and

$\delta > 0$, there exists such $\psi, \psi \in \Psi$, that the inequality

$$d(\vartheta, \vartheta(\psi)) < \delta \qquad (8)$$

is fulfilled, where d is some metrics in space $\mathcal{M}$, then the reparametrization (7) is exact for $\mathcal{M}_0$. For our aim $\mathcal{M}_0$ is defined as some "critical" region of $\mathcal{M}$, where the algorithm error is of special interest.

In §2 of the paper some practical examples of optimal parametrization are given. With the help of the described methods some packages of programs of applied multivariate statistics are tested.

In the following part of §1 we will deal with possible solution of the problem 2^0.

7. Construction of exact samples for given first and second moments of k-dimensional distribution.

Let us prove some lemmas.

Lemma 1. The optimal exact sample for the set $\mathcal{T}(0,1)$ of 1-variate distributions, fulfilling the conditions

$$EX = 0, \quad DX = 1 \qquad (9)$$

is following matrix X_0 (with sample size 2):

$$X_0 = (1,-1).$$

The proof is evident. Notice that the sample X_0 has all even moments equal to 1, all odd moments - 0.

Lemma 2. Let k be fixed, $k \geqslant 2$. There exists an exact sample for the set $\mathcal{T}_\vartheta$ of k-variate distributions, marginals fulfilling conditions (9) and

$$r(X_i, X_j) = 0, \quad i,j = 1,\ldots,k, \quad i \neq j. \qquad (10)$$

The exact sample has the size

$$n = 2^h, \quad h = \left[\log_2 k\right] + 1. \qquad (11)$$

Proof gives the construction of the exact sample by its rows $x_{(i)} = (x_{i1},\ldots,x_{in})$, $i = 1, \ldots, k$. The construction is similar to some construction of ortogonal contrasts in experimental design theory.

At first we shall define an auxiliary $h \times n$ matrix Y by its rows $y_{(i)} = (y_{i1},\ldots,y_{in})$, $i = 1,\ldots,h$ in the following way:

$$y_{ij} = \begin{cases} 1, \text{ if } j = 2l\cdot 2^{h-i} + 1, \ldots, (2l+1)2^{h-i}; \\ -1, \text{ if } j = (2l+1)2^{h-i} + 1, \ldots, (2l+2)2^{h-i}; \\ l = 0, \ldots, 2^{i-1}-1,\ i = 1, \ldots, h;\ j = 1, \ldots, n. \end{cases}$$

The rows of matrix $\mathbb{X}$ will be defined with the help of the elementwise (star) product of vectors:

if $U_i = (u_{i1}, \ldots, u_{in})$, $i = 1,2,3$, then $U_1 = U_2 * U_3$ is defined with equations $u_{1j} = u_{2j}u_{3j}$, $j = 1, \ldots, n$.

For defining matrix $\mathbb{X}$ we use the following iterative algorithm:

$$x_{(1)} = y_{(1)};$$

When $x_{(1)}, \ldots, x_{(2^l-1)}$ are already defined, then the following rows will be defined with help of following connections:

$$x_{(2^l)} = y_{(l+1)},$$

$$x_{(2^l+i)} = y_{(l+1)} * x_{(i)},\ i = 1, \ldots, s,$$

$$s = \begin{cases} 2^l - 1, \text{ if } l < h-1, \\ k - 2^{h-1}, \text{ if } l = h-1. \end{cases}$$

From the definition it follows that $\mathbb{X}$ is $\mathbb{X} \times n$ matrix, $\mathbb{X} = 2^{h-1} + k - 2^{h-1} = k$, and $\sum_{j=1}^{n} x_{ij} = 0$, $\sum_{j=1}^{n} x_{ij}x_{gj} = \begin{cases} n, \text{if } i=g, \\ 0, \text{if } i \neq g, \end{cases}$ so as the rows $x_{(i)}$ of matrix $\mathbb{X}$ are defined as orthogonal contrasts, q. e. d. (See, for instance, the example 1, where $\mathbb{X}_2$ is defined by the rule, given here).

Lemma 3. Let P be simple k-variate distribution (see [4] defined with the help of index-vector $I = (i_1, \ldots, i_s)$, $(1 \leq s \leq k)$, $i_1 = 1$. Let the marginals of P fulfill the condition (4). Then there exists an exact sample $\mathbb{X} = (x_{ij})$ for distribution P, with sample size 2, defined with help of following equations:

$$x_{1j} = \begin{cases} 1, \text{ if } j \in I, \\ -1, \text{ if } j \in I^0; \end{cases} \qquad x_{2j} = \begin{cases} -1, \text{ if } j \in I, \\ 1, \text{ if } j \in I^c. \end{cases}$$

The proof of the lemma 3 is evident.

Lemma 4. Let P be quasi-simple k-variate distribution, see [4], defined with help of (s,k)-indicator L and let the marginals of P fulfill the conditions (9). Then there exists the exact sample $\mathbb{X} = (x_{ij})$ for distribution P with sample size n,

$$n = 2^h, \text{ where } h = [\log_2 s] + 1. \qquad (11')$$

Proof follows from the construction. Let us define the exact sample $\mathbb{X}_0$ for s-variate distribution with independent marginals with the help of lemma 3, $\mathbb{X}_0 = (x^0_{ij})$, $i = 1,\ldots,s$, $j = 1,\ldots,n$. Now let us define $\mathbb{X} = (x_{ij})$, $i = 1,\ldots,k$, $j = 1,\ldots,n$ in the following way:

$$x_{ij} = x^0_{hj}, \text{ if } l_i = h,\ i = 1,\ldots,k,\ h = 1,\ldots,s,\ j = 1,\ldots,n.$$

From the construction it follows that $\sum_{j=1}^{n} x_{ij} = 0$,

$$\sum_{j=1}^{n} x_{ij}x_{gj} = \begin{cases} n, & \text{if } l_i = l_g, \\ 0, & \text{if } l_i \neq l_g, \end{cases}$$

q . e . d.

Corollary 1. Let P be pseudo-simple k-variate distribution defined with the help of (s,k)-pseudo-indicator H. Let the marginals of P fulfill the condition (9). Then there exists an exact sample $\mathbb{X}$ with sample size (11).

Lemma 5. If there exists an exact sample $\mathbb{X}_0$ of size n for some k-variate distribution with marginals, fulfilling the conditions (9) and having correlation matrix R, then there exists an exact sample of the same size for the distribution P with the same correlation matrix and arbitrary marginal means and variances.

Proof. If $EX = a$, $DX = b^2$, then evidently for the one-dimensional distribution from $\pi_{(a,b)}$ the optimal exact sample $\mathbb{X}$ is following:

$$\mathbb{X} = (a-b,\ a+b).$$

Let the correlation matrix R and the constants a_i, b_i of the distribution P be given,

$$EX_i = a_i,\ DX_i = b_i^2,$$

and $\mathbb{X}^0$ be the exact sample for the distribution with correlation matrix R and marginals, fulfilling the conditions (9). Then the exact sample $\mathbb{X} = (x_{ij})$ for the distribution P may be constructed in the following way:

$$x_{ij} = x^0_{ij}b_i + a_i,\ \ i = 1,\ldots,k,\ j = 1,\ldots,n.$$

Lemma 6. Let the distribution Q have the form of finite mixture, $Q = \sum_{i=1}^{n} \gamma_i P_i$, all γ_i being rational, $\gamma_i = v_i/u$. Let $\mathbb{X}_i$ be the exact sample for the distribution P_i with the

sample size n_i, $i = 1,\ldots,m$. Then there exists an exact sample $\mathbb{X}$ for the distribution Q with sample size n, constructed as the sum of exact samples $s_i\mathbb{X}_i$,

$$(\mathbb{X}_1^{:s_1}:\ldots:\mathbb{X}_m^{:s_m}),\ n = \sum_{i=1}^{m} s_i m_i.$$

Proof consists of the calculation of the terms s_i and n: $n_i s_i = n\nu_i/u$, where n is any such natural that all s_i are natural. One possibility is to choose $n = \hat{n}$, $\hat{n} = u\prod_{i=1}^{m} n_i$, but the size n of the optimal exact sample is, as a rule, much less than $\hat{n}$.

From all the lemmas proved, immediately follows

Theorem 1. Let ϑ be the vector of the first and second moments of a k-dimensional distribution, satisfying the following condtions:

1° all parameters ν_i are rational;

2° correlation matrix R has a linear decomposition by simple, quasi-simple or pseudo-simple matrices, see [4].

Then there exists an exact sample $\mathbb{X}$ for the set of distributions π_ϑ.

Proof. From the assumption 2° it follows that the distributions from π_ϑ can be expressed through the finite mixture of simple, quasi-simple or pseudo-simple distributions. As the weights of mixture are rational functions of correlation coefficients, so from the assumption 1° follows that the mixing distribution has rational probabilities. From lemmas 3 - 6 follows the possibility of construction of ecaxt sample.

8. The almost exact samples with fixed higher marginal moments.

At first let us consider the one-dimensional case, k=1, when all the fixed parameters are the moments, $\nu_i = \mu_i$, $i = 1,\ldots,r$ $r \geqslant 2$. From the paper [3] it follows that if the values of ν_i fulfill the condition of moments (see (1) in [3]), then there exists a discrete distribution with support having not more than $\left[\frac{r}{2}\right]+1$ points. But, so as the probabilities are the solutions of the system of equations of the r-th order, the probabilities are not rational, in general. That means, the conditions of existence of exact sample (2) are not fulfilled.

Say, sample $\mathbb{X}$ is α-almost exact for the set of k-dimensional distributions π_ϑ (or for any distribution P from π_ϑ), k natural, if the empirical values of parameters ϑ_e, calculated by $\mathbb{X}$, fulfill the condition

$$d(\vartheta_e, \vartheta) < \alpha . \qquad (12)$$

It is evident that replacing the exact values of parameter ϑ of the discrete distribution $P_0, P_0 \in \pi_\vartheta$ (calculated by given parameters ϑ) with its rational approximation $\bar{\vartheta}$, it is possible to generate the exact sample $\mathbb{X}$ for $\bar{\vartheta}$ if $\bar{\vartheta}$ fulfils the condition (12); then sample $\mathbb{X}$ is by the definition α-almost exact for ϑ.

The calculation of k-variate exact samples with fixed correlation matrix and fixed (equal or unequal) marginals can be performed (in the case when R has linear decomposition by simple, quasi-simple or pseudo-simple matrices for given marginals) with the help of the method, described in theorem 1 and the lemmas 1 - 6.

§2. Some concrete families of exact samples for multivariate distributions with given parameters

9. The family of constant correlation. In the part of the paper we shall be concerned about the finding of convenient parametrization of the first and second moments of k-variate distributions. For simplicity sake we assume that all marginals are standardized, fulfilling the condition (9). When defining the correlation matrix $R_k(\rho) = R(\rho)$ with the help of equations

$$r_{ij} = \rho , \quad i,j = 1,2,\ldots,k, \quad i \neq j, \qquad (13)$$

we get an one-parameter family of distributions (the new parameter $\psi = \rho$, $t = 1$) for every fixed k, named the family of constant correlation. Let us denote the family of k-variate distributions with standardized marginals and constant correlation by $\pi_{(\rho,k)}$.

10. The construction of the exact sample for the set $\pi_{(\rho,k)}$.

Every constant correlation matrix $R(\rho)$ with $\rho \in [0,1]$ has the following linear decomposition by quasi-simple matrices

$$R(\rho) = \rho R_1 + (1 - \rho)R_2,$$

where $R_1 = (r_{ij})$, $r_{ij} = 1$, $i,j = 1,\ldots,k$ and $R_2 = I_k$ (unit matrix).

Then by theorem 2 of [4] there exists a mixture Q of quasi-simple distributions P_1 and P_2,

$$Q = \rho P_1 + (1 - \rho)P_2$$

with marginals, fulfilling the condition (9), and the correlation matrix of Q equals to given $R(\rho)$. The k-variate distribution P_1 has always all components equal, and P_2 — all components independent.

Now we can use the lemmas 1 and 2; from them it follows that there exist exact samples $\mathbb{X}_1$ and $\mathbb{X}_2$ of the distributions P_1 and P_2 with sample sizes $n_1 = 2$ and $n_2 = 2^h$ correspondingly, $h = [\log_2 k] + 1$ (see (11)). By lemma 6 for all rational values of ρ there exists the exact sample $\mathbb{X}$ for $T(\rho,k)$, defined with the help of the equation

$$\mathbb{X} = (X_1^{:s_1} : X_2^{:s_2}),$$

where s_1 and s_2 are naturals, fulfilling the condition

$$s_1/s_2 = (2^{h-1}\rho)/(1 - \rho). \qquad (14)$$

Then the sample size n of the exact sample $\mathbb{X}$ can be calculated in the following way:

$$n = 2s_1 + 2^h s_2,$$

and the exact sample is optimal in the case when s_1 and s_2 are the minimal naturals, satisfying the condition (14).

Our next task is the calculation of expressions (7) of the parameters of model β by the new parameters ρ and k. Here we can use the fact that in most classical models the parameters depend on the inverses and eigenvectors and eigenvalues of correlation matrix $R(\rho)$ and its submatrices. In our case all diagonal submatrices (blocks) of matrix $R(\rho)$ have the same structure as $R(\rho)$, out-of diagonal blocks being constant. We have the following simple connections:

1° The inverse matrix of $l \times l$ diagonal block of $R(\rho)$ is $R^{-} = (r^{ij})$,

$$r^{ij} = \begin{cases} (1 + (l - 2)\rho)/((1 + (l - 1)\rho)(1 - \rho)), & i = j, \\ -\rho/((1 + (l - 1)\rho)(1 - \rho)), & i \neq j, \end{cases}$$

$l = 2,\ldots,k.$

2° The eigenvalues of the same 1×1 diagonal block are:

$$\lambda_1 = 1 + (1 - 1)\rho, \quad \lambda_2 = \ldots = \lambda_1 = 1 - \rho.$$

The determinant D of the block equals

$$D = (1 + (1 - 1)\rho) \cdot (1 - \rho)^{1-1}.$$

From here it is evident that for $\rho \in [0,1)$ the constant correlation matrix is always positively defined, but it is possible to find the distributions satisfying the condition

$$D < \eta$$

for every given positive η, that means, the family is useable for the investigation of the influence of non-definiteness of parent distribution for the error of algorithm.

11. Example 1. Let us take $k = 8$, $\rho = 0.5$. From (11') we find that $h = 4$ and the samples $\mathbf{X}_1$ and $\mathbf{X}_2$ are following:

$$\mathbf{X}_1 = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \quad \mathbf{X}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

Using the value $\rho = 0.5$ we get from the equation (14) $s_1 = 8$, $s_2 = 1$ and $n = 2 \cdot 8 + 16 = 32$.

12. The expressions of parameters of classical multivariate procedures in the case of constant correlation matrix $R_k(\rho)$.

1° Correlation analysis. The partial correlations of the 1-th order (that is, the correlations between X_i and X_j when the linear influence of $X_{g_1}, \ldots, X_{g_l}$ is eliminated) depends only on the order l, but does not depend on concrete indices $i, j, g_1, \ldots, g_l$, and is expressed in the following way:

$$r(X_i, X_j / X_{g_1}, \ldots, X_{g_l}) = \rho / (1 + l\rho).$$

Example 2. The partial correlations, calculated on the basis of the exact sample, generated in Example 1, are given in Table 1:

Table 1

Order of partial correlation	Number of different coefficients	The value of correlation coefficient
0	28	0.5
1	168	0.333333
2	420	0.25
3	560	0.2
4	420	0.166667
5	168	0.142857
6	28	0.125

2° Regression analysis. The linear model of the l-th order for prognosing the regressand X_i by the regressors $X_{g_1},\dots,X_{g_l}$,

$$X_i = \sum_{j=1}^{l} b_j^l X_{g_j} + b_o^l \qquad (15)$$

depends only on the number of regressors, but does not depend on concrete indices $i,g_1,\dots,g_l$. All regression coefficients in the model (15) are equal:

$$b_j^l = \rho/(1 + (l - 1)\rho),$$
$$b_o^l = 0, \qquad j = 1,\dots,l,\ l = 1,\dots,k-1.$$

In this case the coefficients of standardized regression (the socalled betha-regression) are equal to b_j's. The multiple correlation coefficient of the l-term model (15) (see [5,9]) is

$$R = \rho\sqrt{1/(1 + (l - 1)\rho)}.$$

The estimated variance s^2 of regressand:

$$s^2 = \frac{(1 - \rho)(1 + l\rho)\, n}{(1 + (l - 1)\rho)(n - l - 1)},$$

and the value of F in the ANOVA table for testing the significance of the whole l-term model (15):

$$F = \frac{\rho^2 (n - l - 1)}{(1 - \rho)(1 + l\rho)}.$$

The variance s_b^2 of every regression coefficient b_j^l in model (15):

$$s_b^2 = \frac{(1 + (l - 2)\rho)(1 + l\rho)}{(1 + (l - 1)\rho)^2 (n - l - 1)}$$

and the value of F for testing the significance of every single regression coefficient b_j^1:

$$F(b) = \frac{\rho^2(n - 1 - 1)}{(1 + (1 - 2)\rho)(1 + 1\rho)} .$$

Residuals of the linear prognosis

$$(x_{if} - \sum_{j=1}^{1} b_j^1 x_{g_{jf}})\sqrt{\frac{n - 1}{n}} , \quad f = 1,\ldots,n.$$

Example 3. The parameters of all possible linear models, calculated by the exact sample, generated in example 1, are given in Table 2.

Table 2

1	b_j	R	s^2	F	s_b^2	F(b)	Residuals of the first object
1	0.5	0.5	0.8	10	0.025	10	0.49213
2	0.33333	0.57735	0.73563	7.25	0.03065	3.625	0.328083
3	0.25	0.61237	0.71428	5.6	0.03348	1.867	0.246062
4	0.2	0.63246	0.71111	4.5	0.03555	1.125	0.1968502
5	0.16667	0.64550	0.71795	3.71429	0.03739	0.743	0.16404183
6	0.14286	0.65465	0.73143	3.125	0.03918	0.521	0.14060728
7	0.125	0.66144	0.75	2.66667	0.041014	0.381	0.123031

3° The factor analysis. Principal components without iterations. All loadings of the first factor are equal,

$$f_{1i} = \sqrt{(1 + (k - 1)\rho)/k}, \quad i = 1,\ldots,k.$$

If the stopping rule $\lambda_i \leq 1$ is used, then only one factor is computed. All other factors are not defined uniquely, but they have all equal "length":

$$\sum_{i=1}^{k} f_{ji}^2 = 1 - \rho , \quad j = 2,\ldots,k.$$

The individual factor scores

$$l_{1i} = 1/\sqrt{(1 + (k - 1)\rho)k}, \quad i = 1,\ldots,k.$$

The estimated values of the individual factor scores of the first factor

$$\sqrt{\frac{n - 1}{n}}\Big/\sqrt{(1 + (k - 1)\rho)\cdot k}\cdot\sum_{i=1}^{k} x_{ij}, \quad j = 1,\ldots,n.$$

The classical factor analysis (principal components with iterations).

The eigenvalues of the reduced correlation matrix are:

$\lambda_1 = k\rho$, $\lambda_i = 0$, $i = 2,\ldots,k$.

The diagonal matrix of the variances of the so-called uniquenesses:

$\mathrm{diag}(1-\rho,\ldots,1-\rho)$.

Consequently, only one factor exists, having equal loadings:

$f_{1i} = \sqrt{\rho}$, $i = 1,\ldots,k$

and, hence the coefficients of individual factor scores of the first factor:

$l_{1i} = \sqrt{\rho}\,/(1 + (k-1)\rho)$, $i = 1,\ldots,k$.

Example 4. The parameters of factor analysis calculated on the basis of the exact sample, created in the example 1, are given in Table 3.

Table 3

Model	Eigenvalues	f_{1i}	$f^2_{ji}, j\geq 2$	uniquenesses	l_{1i}	Estimated factor of the 1obj
Principal factors	4.5,0.5,...,0.5	0.75	0.5	0	0.1666667	1.31233
Classical factor analysis	4.0,0,...,0	0.707107	0	0.5	0.157135	1.23728

4° Canonical analysis. Assume the number of variables in the first group is q and in the second group p, $p+q \leq k$. Then matrix $R_{12}R_{22}^{-1}R_{21}R_{11}^{-1}$ is constant with all elements equal

$$\frac{q\rho^2}{(1+(p-1)\rho)(1+(q-1)\rho)}$$

and has only one non-zero eigenvalue:

$$\lambda_1 = \frac{pq\rho^2}{(1+(p-1)\rho)(1+(q-1)\rho)}, \quad \lambda_2 = \ldots = \lambda_s = 0,$$

$s = \min(q,p)$.

Consequently, the canonical correlations are:

$$\rho_1=\rho\sqrt{\frac{pq}{(1+(p-1)\rho)(1+(q-1)\rho)}}\,,\quad \rho_2=\ldots=\rho_s=0$$

and the coefficients of the canonical variables of the first and second group are

$$\sqrt{\frac{1}{q(1+(q-1)\rho)}},\quad \sqrt{\frac{1}{p(1+(p-1)\rho)}}\,.$$

Example 5. The parameters of canonical analysis are computed on the basis of the exact sample, created in the example 1. The results are given in Table 4.

Table 4

Number of variables		Canonical	Coefficients of the 1-st var.	
1-st group	2-nd group	correlations	1-st group	2-nd group
4	4	0.8, 0, 0, 0	0.31623 i=1,...,4	0.31623 i=5,...,8
5	3	0.79057,0, 0	0.25820 i=1,...,5	0.40825 i=6,7,8
6	2	0.75593,0	0.21822 i=1,...,6	0.57735 i=7,8

5° Discriminant analysis. Let us assume the first variable X_1 to be the grouping one. Then we have the following parameters of groups

Number of group	1	2
Mean of x_1	-1	1
Means of x_i, $i = 2,\ldots,k$	$-\rho$	ρ
Estimated standard deviations of x_i, $i=2,\ldots,k$	$\sqrt{(1-\rho^2)\frac{n}{n-2}}$	$\sqrt{(1-\rho^2)\frac{n}{n-2}}$

The correlation coefficients $r(x_i,x_j)$ $(i \neq j,\ j=2,\ldots,k)$ in pooled within group correlation matrix are

$$r_{ij} = \frac{\rho}{1+\rho} \qquad i,j=2,\ldots,k,\ i\neq j.$$

In the step-wise procedure the values of statistics "F to enter" on the s'th step (when already s arguments are included) depend only on the number of step,

$$F = \frac{\rho^2(n-2-s)}{(1+(s-1)\rho^2)(s+1)}$$

with 1 and n-2-s degrees of freedom, $s = 0,1,\ldots,k-2$.

The two different values of F, characterising the difference between the groups (F-matrix and the "approximate F") on the (s+1)'st step are in the case equal,

$$F = \frac{\rho^2(n-2-s)}{1 + (s-1)\rho^2}$$ with n - 2 - s and s+1 degrees of freedom, s = 0,1,...,k-2.

On the first step we have the precent of correctly classified objects (in both groups) equal

$$0.5(1 + \rho),$$

and the coefficients of the discriminant functions (for the first and the second group) are

$$-\frac{\rho(n-2)}{(1-\rho^2)n}, \quad \frac{\rho(n-2)}{(1-\rho^2)n}.$$

Example 6. The parameters of discriminant analysis calculated on the basis of the exact sample, created in the example 1, are given in the table 5

Table 5

The characteristics of groups				The characteristics of step-wise procedures				
	1'st group	2'nd group	Equal for both groups					
			St.Dev.	Step N°	F to enter	F	degrees of freed	
Means x_1	-1	1	0.89443	0	10	-	-	
$x_{j,j=2,...,k}$	-0.5	0.5	Within group correl.	1	3.625	10.00	1	30
				2	1.8666	7.25	2	29
Coeff. of discr. fn.	-0.625	0.625	0.33333	3	1.125	5.6	3	28
				4	0.748	4.5	4	27
% of correctly classific.	75%	75%		5	0.5208	3.7143	5	26
				6	0.3809	3.125	6	25
				7	-	2.666	7	24

On the basis of the given methods and three testing examples ($\rho = 0.2$, $\rho = 0.5$ (as in examples 1-5) and $\rho = 0.8$) the programs of four packages of multivariate statistics were tested. These programs were two original packages, created in Tartu State University, BMDP'79, and SAISI (constructed from some models of SPSS). Practically all the results of programs coincided with the theoretical values, calcula-

ted analytically with the help of personal computer. The testing was carried with the help of workers of the Laboratory of applied mathematics of Tartu State University S.Koskel, M.Vähi and E.Ehasalu. The testing programs made M.Somelar. The author is very thankful to them.

13. The family with fixed eigenvalues of correlation matrix.

Assuming that all marginals of the parent distribution fulfill the condition (9) (that is not restrictive for the case, when only the 1-st and 2-nd moments are fixed, as it was demonstrated in Lemma 5), then the task of convenient parametrization of the parameter-vector ϑ can be reduced to the task of parametrization of the correlation matrix R only. One possibility is to use the eigenvalues of the correlation matrix as the new parameters ψ (then the number of new parameters is $\leq k$), see [2,5,8]. As in general the matrix R,

$$R = H \Lambda H', \tag{16}$$

where H is the matrix of eigenvectors, $H = (h_{(1)};\ldots;h_{(k)})$, $h(i) = (h_{1i},\ldots,h_{ki})'$, and

$$\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_k), \tag{17}$$

in the case of given eigenvalues depends on the matrix H, so we have to fix the matrix H in any suitable form. The simplest structure for H is possible in the special case, when

$$k = 2^m. \tag{18}$$

Then it is evident that there exists biorthogonal $k \times k$ matrix H, fulfilling the following conditions:

$$H' = H,\ HH = I_k,\ |h_{ij}| = 1/\sqrt{k},\ i,j = 1,\ldots,k. \tag{19}$$

In future we assume that the conditions (18) and (19) are fulfilled.

Lemma 7. Let H be matrix, fulfilling the conditions (19) by k, expressed in (18), and Λ the diagonal matrix (17), fulfilling the conditions

$$\lambda_i \geq \lambda_{i+1},\ i = 1,\ldots,k-1,\ \sum_{i=1}^{k}\lambda_i = k,\ \lambda_i \geq 0.$$

Then the correlation matrix R, calculated by (16), has the structure of Latin square, that means, there exist not more than k-1 different correlation coefficients r_{ij}, every one of them lies (at least) once in every row and every column.

Proof. From the conditions (19) it follows that every row (column) of the matrix H has the properties of normal contrast. From here it can be concluded that the following equation holds for any pair of indices i and j:

$$h_{(i)} * h_{(j)} = (1/\sqrt{k}) \cdot h_{(1)},$$ where $h_{(1)}$ is a row of H.

That means, we receive only k-1 different expressions of r_{ij}:

$$r_{ij} = \sum_{f=1}^{k} h_{if} \lambda_f h_{jf} = (1/\sqrt{k}) \sum_{f=1}^{k} h_{1f} \lambda_f; \quad i \neq j, \ i,j=1,\ldots,k$$

and for different values of i, when j is fixed, the expressions of $r_{i_1 j}$ and $r_{i_2 j}$ through $h_{(1)}$ are different, although in some case some concrete values of $r_{i_1 j}$ and $r_{i_2 j}$ may coincide.

The proof of the lemma is completed.

The definition of the random vector, having given marginals and correlation matrix R, may always be performed with the help of the construction given in [4], and when only the first and second moments of marginals are fixed, the exact sample may be constructed with the help of the method given in Theorem 1. It is evident, that the condition 1° must be replaced with the condition that all eigenvalues are rational. Otherwise only the almost exact sample may be constructed. So as the correlation matrix in the case considered includes only k-1 different correlations, the linear decomposition of it has not more than k terms, and, consequently, the exact sample will be the sum of not more than k exact samples of simple distributions.

The number t of parameters ψ, defining the parameter-vector ϑ and parameter of model β, may be decreased with the help of the special construction of matrix Λ. For instance, the eigenvalues may form arithmetical or geometrical progression (then t = 1 or 2)

Let us notice that for the construction of the matrix H the algorithm of construction X, given in Lemma 2, may be used with slight modification:

$$\begin{cases} h_{(1)} = 1/\sqrt{k} \cdot (1,\ldots,1), \\ h_{(i)} = 1/\sqrt{k} \cdot x_{(i-1)}, \ i = 2,\ldots,2^m-1. \end{cases} \qquad (20)$$

The distributions with given eigenvalues form a special tool for testing several properties of factor analysis, since

the given parameters λ_i are in fact the parameters of factor analysis and the expressions of model parameters β through new parameters ψ (formulae (7)) are very simple:

$$f_{ij} = \sqrt{\lambda_i}\, h_{ij};\ i,j = 1,\ldots,k.$$

Example 7.

Let us have k = 8 (m = 3) and the matrix Λ is the following:

Λ = diag(4,2,1,0.5,0.25,0,0). The matrix H, constructed by (19), is the following:

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \cdot 8^{-1/2}.$$

The correlation coefficients, calculated by the formulae (16), are the following:

$r_{12}=r_{34}=r_{56}=r_{78}=(4+2+1+0.5-0.25-0.25)/8=0.875$

$r_{13}=r_{24}=r_{57}=r_{68}=(4+2-1-0.5+0.25+0.25)/8=0.625$

$r_{14}=r_{23}=r_{58}=r_{67}=(4+2-1-0.5-0.25-0.25)/8=0.5$

$r_{15}=r_{26}=r_{37}=r_{48}=(4-2+1-0.5+0.25-0.25)/8=0.3125$

$r_{16}=r_{25}=r_{38}=r_{47}=0.3125$

$r_{17}=r_{28}=r_{35}=r_{46}=(4-2-1+0.5-0.25+0.25)/8=0.1875$

$r_{18}=r_{27}=r_{36}=r_{45}=0.1875.$

With the help of the algorithm, described in [4], the decomposition by the quasi-simple matrices of the correlation matrix defined is found. In the case in the decomposition only 5 quasi-simple matrices were:

$$R_1=\begin{pmatrix}1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1\end{pmatrix}\quad R_2=\begin{pmatrix}1&1&1&1&0&0&0&0\\1&1&1&1&0&0&0&0\\1&1&1&1&0&0&0&0\\1&1&1&1&0&0&0&0\\0&0&0&0&1&1&1&1\\0&0&0&0&1&1&1&1\\0&0&0&0&1&1&1&1\\0&0&0&0&1&1&1&1\end{pmatrix}\quad R_3=\begin{pmatrix}1&1&0&0&1&1&0&0\\1&1&0&0&1&1&0&0\\0&0&1&1&0&0&1&1\\0&0&1&1&0&0&1&1\\1&1&0&0&1&1&0&0\\1&1&0&0&1&1&0&0\\0&0&1&1&0&0&1&1\\0&0&1&1&0&0&1&1\end{pmatrix}$$

$$R_4=\begin{pmatrix}1&1&0&0&0&0&0&0\\1&1&0&0&0&0&0&0\\0&0&1&1&0&0&0&0\\0&0&1&1&0&0&0&0\\0&0&0&0&1&1&0&0\\0&0&0&0&1&1&0&0\\0&0&0&0&0&0&1&1\\0&0&0&0&0&0&1&1\end{pmatrix}\quad R_5=\begin{pmatrix}1&0&1&0&0&0&0&0\\0&1&0&1&0&0&0&0\\1&0&1&0&0&0&0&0\\0&1&0&1&0&0&0&0\\0&0&0&0&1&0&1&0\\0&0&0&0&0&1&0&1\\0&0&0&0&1&0&1&0\\0&0&0&0&0&1&0&1\end{pmatrix}.$$

The linear decomposition is

$$R=0.1875R_1 + 0.3125R_2 + 0.125R_3 + 0.25R_4 + 0.125R_5.$$

Corresponding quasi-simple vectors are

$$W_1=(X_1,X_1,X_1,X_1,X_1,X_1,X_1,X_1),$$
$$W_2=(X_1,X_1,X_1,X_1,X_2,X_2,X_2,X_2),$$
$$W_3=(X_1,X_1,X_2,X_2,X_1,X_1,X_2,X_2),$$
$$W_4=(X_1,X_1,X_2,X_2,X_3,X_3,X_4,X_4),$$
$$W_5=(X_1,X_2,X_1,X_2,X_3,X_4,X_3,X_4).$$

The following step of our construction is the finding of the exact samples for all quasi-simple vectors.

$$\mathfrak{X}_1=\begin{pmatrix}1&-1\\1&-1\\1&-1\\1&-1\\1&-1\\1&-1\\1&-1\\1&-1\end{pmatrix}\quad \mathfrak{X}_2=\begin{pmatrix}1&1&-1&-1\\1&1&-1&-1\\1&1&-1&-1\\1&1&-1&-1\\1&-1&1&-1\\1&-1&1&-1\\1&-1&1&-1\\1&-1&1&-1\end{pmatrix}\quad \mathfrak{X}_3=\begin{pmatrix}1&1&-1&-1\\1&1&-1&-1\\1&-1&1&-1\\1&-1&1&-1\\1&1&-1&-1\\1&1&-1&-1\\1&-1&1&-1\\1&-1&1&-1\end{pmatrix}\quad \mathfrak{X}_4=\begin{pmatrix}1&1&1&1&-1&-1&-1&-1\\1&1&1&1&-1&-1&-1&-1\\1&1&-1&-1&1&1&-1&-1\\1&1&-1&-1&1&1&-1&-1\\1&-1&1&-1&1&-1&1&-1\\1&-1&1&-1&1&-1&1&-1\\1&1&-1&-1&-1&-1&1&1\\1&1&-1&-1&-1&-1&1&1\end{pmatrix}$$

$$\mathfrak{X}_5=\begin{pmatrix}1&1&1&1&-1&-1&-1&-1\\1&1&-1&-1&1&1&-1&-1\\1&1&1&1&-1&-1&-1&-1\\1&1&-1&-1&1&1&-1&-1\\1&-1&1&-1&1&-1&1&-1\\1&1&-1&-1&-1&-1&1&1\\1&-1&1&-1&1&-1&1&-1\\1&1&-1&-1&-1&-1&1&1\end{pmatrix}.$$

The coefficients s_i can be calculated with the help of Lemma 6 from the equations

$$2s_1 = 0.1875n,$$
$$4s_2 = 0.3125n,$$

$2s_1 = 0.1875n,$

$4s_2 = 0.3125n,$

$4s_3 = 0.125\ n,$

$8s_4 = 0.25n,$

$8s_5 = 0.125n.$

Taking n = 64, we have $s_1 = 6$, $s_2 = 5$, $s_3 = 2$, $s_4 = 2$, $s_5 = 1$, and

$$X = (X_1^{:6}; X_2^{:5}; X_3^{:2}; X_4^{:2}; X_5), \qquad (21)$$

and the size of the exact sample is indeed $6 \cdot 2 + 5 \cdot 4 + 2 \cdot 4 + 2 \cdot 8 + 8 = 64$. The parameters of factor analysis, computed by the exact sample (21), are the following:
(only four first factors (principal components) are defined)

F_1	F_2	F_3	F_4
0.70711	0.5	0.35356	0.25
0.70711	0.5	0.35356	0.25
0.70711	0.5	-0.35356	-0.25
0.70711	0.5	-0.35356	-0.25
0.70711	-0.5	0.35356	-0.25
0.70711	-0.5	0.35356	-0.25
0.70711	-0.5	-0.35356	0.25
0.70711	-0.5	-0.35356	0.25

Communalities of 1-, 2-, 3- and 4-factor system are correspondingly 0.5, 0.75, 0.875 and 0.9375; the "length" of the 5-th and 6-th factors are equally 0.25, the 7-th and 8-th factors have zero length.

For the nonrotated factor-structure the factor score coefficients are (for the two first factors):

$l_{1i} = 0.17678$, $i = 1,\ldots,8$;

$l_{2i} = 0.25$, $i = 1,\ldots,4$, $l_{2i} = -0.25$, $i = 5,\ldots,8$.

The values of the individual factor scores for the first object of the exact sample X are

1.403125, 0.00000.

Every method of orthogonal rotation (varimax, quartimax e.c.) gives for the first pair of the factors the optimal angle 45^0.

After the rotation we have the following factor-matrix:

F_1	F_2
0.85355	0.14645
0.85355	0.14645
0.85355	0.14645
0.85355	0.14645
0.14645	0.85355
0.14645	0.85355
0.14645	0.85355
0.14645	0.85355

With the help of the method of testing, described in 1, we found that some algorithms (for instance, in package SAISI) did not rotate the factors in the model, given in Example 7. It seems that the point of rotation for angle 45° is a peculiar one and does not include in the admissible region of some algorithms of factor rotation.

References

1. Francis I., Heiberger R., Velleman P. - Criteria and Considerations in the Evaluation of Statistical Programs Package. - American Statistician, 1975, 29, № 1,52-56.
2. Lin, S.P., Bendel R.B. Generation of Population Correlation Matrices with Specified Eigenvalues. - J.R.Statist. Soc. B, 1985, 47, № 1, 67-75.
3. Tiit, E. Definition of mixtures with given moments, the same issue, 3-14.
4. Tiit, E. Random vectors with given arbitrary marginals and given correlation matrix, the same issue 14-39.
5. Tiit, E.-M. Formal Computation of Regression Parameters. Compstat 1984, 497-502.
6. Vahle H., Weber R.Matematische-statistische Programmsammlungen. - Rechnentechnik datenverarbeitung 1982, 19, № 6, 19-21.
7. Губарев В.В., Котов Ю.А., Молин А.Б. Статистические программы. Новосибирск, 1984. ВИНИТИ, 7883-84 Деп.
8. Жанатауов С.У. Метод получения выборки с заданными собственными числами ее корреляционной матрицы. - Математические вопросы анализа данных. Новосибирск, 1980, 62-76.
9. Тийт Э.А. Об априорном выборе регрессоров в линейной регрессионной модели, Заводская лаборат. 49, № 7.44-49.
10. Тийт Э.-М. Построение дискретных многомерных распределений с заданными моментами. Дискретный аналог нормально-

го распределения. ВЦ ТГУ 1984, 51, 142-165.

"ТОЧНЫЕ ВЫБОРКИ" ИЗ МНОГОМЕРНЫХ РАСПРЕДЕЛЕНИЙ И ИХ ПРИМЕНЕНИЕ ПРИ ТЕСТИРОВАНИИ СТАТИСТИЧЕСКИХ АЛГОРИТМОВ

Э. Тийт

Р е з ю м е

Проблема тестирования матобеспечения (алгоритмов и их реализаций) математической статистики является в настоящее время весьма актуальной. Стандартными методами для этого являются:

1° Использование "традиционных" примеров, опубликованных многократно.

2° Применение метода Монте-Карло.

В настоящей статье предлагается альтернативный подход, основанный на т.н. "точных выборках". Это специальным образом построенные конечные совокупности точек - реализаций случайного вектора, имеющие наперед заданные параметры эмпирического распределения. Для таких "выборок" возможно аналитически (без применения исследуемых алгоритмов) вычислить точные значения β^* параметров статистических моделей. Затем, путем сравнения β^* с результатами применения исследуемого алгоритма A_K "точной выборке" $\beta(\mathcal{T}, \mathcal{N}, A)$ по формуле (4) возможно оценить ошибку алгоритма A и выяснить его допустимую область, где имеет место неравенство (5). В пункте 7 излагаются теоретические основы для построения "точных выборок" по заданным первым и вторым моментам (притом используется методика построения многомерных распределений с заданной корреляционной матрицей, изложенная в [4]).

Во втором параграфе вводится семейство $\pi_{(\varrho,\kappa)}$ ($\varrho \in [0,1)$, рациональное, $\kappa \in \mathbb{N}$) к-мерных распределений с стандартизированными маргинальными распределениями и постоянными корреляциями $r_{ij} = \varrho$ ($i,j = 1,\dots,k$; $i \neq j$). Алгоритмы для построения точных выборок из семейства $\pi_{(\varrho,\kappa)}$ изложен в п. 10, а в п.12 излагаются точные формулы параметров многомерного статистического анализа для распределений из семейства $\pi_{(\varrho,\kappa)}$.

Для конкретного семейства π при $\varrho = 0.5$ и $\kappa = 8$ в таблицах 1-5 приведены значения основных статистически корреляционного регрессионного, компонентного, факторного, канонического и дискриминантного анализа.

Received September 1985

MATRIX CALCULUS FOR MULTIVARIATE DISTRIBUTIONS

I. Traat

1. Introduction

In the last few years the matrix calculus, especially such operations like the Kronecker product and the matrix derivative have been applied to many statistical problems (see e.g. [2-8, 12, 15]). The matrix technique seems to be very convenient in the multivariate case. We try extend the usage of matrix operations to one more field of mathematical statistics, namely to that which is connected with the representation of multivariate distributions. From this point of view the moments, cumulants and asymptotic expansions are considered. The observed material needs the application of higher-order matrix derivative which is defined here on the basis of MacRae's first-order matrix derivative [3].

The definitions which generalize the notions of moments and cumulants of a random vector are given. It is shown that the expressions for generalized moments and cumulants of any distribution can be obtained by matrix differentiation of its characteristic function. This result is applied to the multivariate normal distribution to derive the matrix expressions of its central moments up to the sixth order.

Quite often the moments of some random function are calculated from its Taylor expansion. Our definition for moment needs the matrix form of Taylor expansion. It is given here for the vector, the components of which are the functions of another vector. In mathematical statistics various functions of symmetric matrices (like covariance and correlation matrices) are of interest. As there exist some additional requirements when dealing with the derivative with respect to symmetric matrix [5], then the Taylor expansion for the symmetric argument is separately considered.

We also describe the matrix technique when deriving the multivariate Edgeworth expansion. For that some auxiliary matrix results concerning the Fourier transform and the

multivariate Hermite polynomials are given. Finally, the formal Edgeworth expansion with the accuracy $o(n^{-1})$ for the class of multivariate statistics is derived.

One can see that the matrix operations make the multivariate formulae simple and compact. They are expressible without multiple indices and in many cases the outside look of the multivariate formula coincides with that of the corresponding univariate formula. In addition, the matrix form of the formulae is convenient for computer calculations.

2. Preliminaries

Here we list the operations and their properties given in [2-8], which will be used to obtain the results of this paper. An additional notion - the power of a vector is formulated.

DEFINITION 1. The Kronecker product of $m \times n$ matrix $X = [x_{ij}]$ and $s \times t$ matrix Y is the $ms \times nt$ block matrix

$$X \otimes Y = [x_{ij} Y].$$

This product has the following properties, assuming the matrices are comfortable to the indicated operations:

$$\begin{aligned} &(cX) \otimes Y = X \otimes (cY) = c(X \otimes Y), \ c - \text{scalar}, \\ &(X \otimes Y)(Z \otimes W) = XZ \otimes YW, \\ &(X \otimes Y)' = X' \otimes Y', \\ &(X + Y) \otimes (Z + W) = X \otimes Z + X \otimes W + Y \otimes Z + Y \otimes W, \\ &X \otimes (Y \otimes Z) = (X \otimes Y) \otimes Z. \end{aligned} \tag{2.1}$$

DEFINITION 2. The star product [3] of $m \times n$ matrix X and $ms \times nt$ block matrix Z is the $s \times t$ matrix

$$X * Z = \sum_{i,j} x_{ij} Z_{ij},$$

where Z_{ij} is the ij-th $s \times t$ block of Z.

In the case of scalar c and vector t, the star product is equal to the usual product:

$$c * t = t \cdot c \tag{2.2}$$

DEFINITION 3. The permutation matrix $I_{m,n}$ (also commutation matrix and permuted identity matrix) is the $mn \times mn$ block matrix consisting of mn blocks of size $m \times n$ such that the ij-th block has a 1 in the ji-th position and zeros elsewhere.

An important property of $I_{m,n}$ is, that it reverses the order of a Kronecker product:

$$X \otimes Y = I_{s,m}(Y \otimes X)\, I_{n,t}. \qquad (2.3)$$

In addition, the following is valid:

$$I_{m,1} = I_{1,m} = I_m,$$
$$I_{m,n} = I'_{n,m},$$
$$I_{m,n}I_{n,m} = I_{mn}.$$

DEFINITION 4. The operator vec forms the vector from the columns of a matrix stacking them one under another:

$$\operatorname{vec} X = (x_{11},\ldots,x_{m1},\ldots,x_{1n},\ldots,x_{mn})'.$$

Clearly

$$\operatorname{vec}(cX) = c \operatorname{vec} X,$$
$$\operatorname{vec}(X + Y) = \operatorname{vec} X + \operatorname{vec} Y.$$

The operator vec is connected with the Kronecker product and permutation matrix by the equations:

$$\operatorname{vec} XZY = (Y' \otimes X) \operatorname{vec} Z, \qquad (2.4)$$
$$I_{m,n} \operatorname{vec} X = \operatorname{vec} X'. \qquad (2.5)$$

In the case of symmetric $p \times p$ matrix Z it is useful to form the vector from the unequal elements of Z.

DEFINITION 5. The operator vech forms the vector from the above diagonal elements of a square matrix:

$$\operatorname{vech} Z = (z_{11},z_{12},z_{22},\ldots,z_{1p},\ldots,z_{pp})'.$$

There exist the following connections between operators vec and vech when working with symmetric matrix Z:

$$\operatorname{vec} Z = G \operatorname{vech} Z, \qquad (2.6)$$

where G is unique $p^2 \times p(p+1)/2$ matrix consisting of ones and zeros only;

$$\operatorname{vech} Z = H \operatorname{vec} Z, \qquad (2.7)$$

where H is nonunique (e.g. $H = (G'G)^{-1}G'$);

$$\operatorname{vech} Z = HG \operatorname{vech} Z,$$
$$\operatorname{vec} Z = GH \operatorname{vec} Z. \qquad (2.8)$$

On the basis of the Kronecker product we give the definition which appears to be useful in considering the multivariate distributions.

DEFINITION 6. The k-th power of a vector x is

$$x^k = x \otimes x \otimes \ldots \otimes x,$$

where x is multiplied k times.

If $x = [x_i]$ is the vector of order p, then x^{k} is the vector of order p^{k}, the components of which are all possible products $x_{i_1} x_{i_2} \ldots x_{i_k}$, $i_1, i_2, \ldots, i_k \in \{1, \ldots, p\}$. So x^{k} consists of all mixed products of order k of the components of vector x. In particular, if the order p = 1, then definition 6 gives the usual power of a scalar. Obviously it holds

$$x^{k} \otimes x^{j} = x^{k+j}.$$

This part of preliminaries we finish with listing some properties about vectors x, y:

$$x \otimes y' = xy', \tag{2.9}$$

$$x \otimes y' = y' \otimes x,$$

$$\operatorname{vec}(x \otimes y') = y \otimes x. \tag{2.10}$$

Next we consider the matrix derivative which is an important notion in our further results. A good review on matrix differentation is given by Nel [6].

The notion of matrix derivative corresponds to the notion of Fréchet derivative in the finite Euclidean spaces [11]. Probably the main reason for taking into use the matrix derivative in statistics was, that it is the convenient tool for practical finding of the derivatives of one matrix with respect to another matrix (both may also be vectors or scalars).

Here we represent the MacRae's [3] definition of matrix derivative

DEFINITION 7. Let the elements of $s \times t$ matrix Y be functions of the elements of $m \times n$ matrix X and let the elements of X be functionally independent. Let $d/dx = [\partial / \partial x_{ij}]$ be a matrix of derivative operators. Then the derivative of matrix Y with respect to X is defined to be an $ms \times nt$ matrix, given by

$$\frac{dY}{dX} = Y \otimes \frac{d}{dx},$$

where multiplication of a matrix element by a derivative operator corresponds to the operation of differentiation.

By the definition, the result of matrix differentiation is a matrix, consisting of partial derivatives of all elements of Y with respect to all elements of X.

Let matrix W be a function of Y, but Y, Z and $p \times q$ matrix U be functions of X. Then following differentiation rules are valid:

$$dX/dX = \text{vec } I_m \text{ vec}' I_n,$$
$$dX'/dX = I_{m,n},$$
$$d(Y+Z)/dX = dY/dX + dZ/dX,$$
$$d(YZ)/dX = (dY/dX)(Z \otimes I_n) + (Y \otimes I_m)(dZ/dX) \qquad (2.11)$$
$$d(Y \otimes U)/dX = (Y \otimes dU/dX) + (I_{p,s} \otimes I_m)(U \otimes dY/dX)(I_{t,q} \otimes I_n), \qquad (2.12)$$
$$dW/dX = W \otimes (d/dY * dY/dX).$$

From these equalities the special cases for the p-vector t and scalar composite function w(Y) follow:

$$dt/dt = \text{vec } I_p, \qquad (2.13)$$
$$dt'/dt = I_p, \qquad (2.14)$$
$$dw/dX = dw/dY * dY/dX. \qquad (2.15)$$

The definition of matrix derivative and some of the given differentiation rules do not hold when X is symmetric, as in this case the elements of X are functionally dependent. In the case of symmetric X, the derivative has to be found with respect to independent elements of X, i.e. with respect to vech X. But that is not easy to do analytically. In [5], there are given useful formulae which connect the derivatives with respect to vech X and with respect to the whole matrix X. In the latter case the ji-th element of symmetric X is considered to differ from its ij-th element even though their scalar values are equal (i.e. $\partial x_{ij} / \partial x_{ji} = 0, \; i \neq j$). With the help of these results the above differentiation technique is applicable also to the functions of symmetric matrices

Let the scalar y and s * t matrix Y be the functions of symmetric matrix Z. Then due to [5]

$$dy/d\text{vech } Z = G' \text{vec}(dy/dZ),$$
$$dy/d\text{vech } Z = G'(dy/d\text{vec } Z), \qquad (2.16)$$
$$dY/d\text{vech } Z = (I_s \otimes G')(dY/d\text{vec } Z), \qquad (2.17)$$

where matrix G is defined by the equality (2.6) and the derivatives at the right sides of these formulae are found without considering Z to be symmetric (i.e. the elements of Z are observed as functionally independent).

3. The higher-order matrix derivative

In the literature on matrix differentiation the first-order (sometimes also the second order) matrix derivative is mainly observed. We need to use the higher-order matrix de-

rivative.

DEFINITION 8. The matrix derivative of order k is inductively defined as the matrix derivative of the matrix derivative of order k-1:

$$\frac{d^k Y}{(dX)^k} = \frac{d}{dX}\left[\frac{d^{k-1}Y}{(dX)^{k-1}}\right].$$

So at each step of differentiation we may use the rules of finding the first-order matrix derivative given by MacRae. From the definition it follows

$$\frac{d^k Y}{(dx)^k} = (\ldots((Y \otimes \frac{d}{dX}) \otimes \frac{d}{dX}) \otimes \ldots) \frac{d}{dX},$$

where the matrix of derivative operators is applied k times.

The important special case for our results is the derivative of scalar function y(t) with respect to p-vector t. From the definition it follows, that dy(t)/dt is the p-vector consisting of the first order partial derivatives of y(t), but $d^k y(t) / (dt)^k$ is the p^k-vector consisting of all k-order partial derivatives of y(t).

The next result is the generalization of MacCulloch's result (2.16).

Theorem 1. Let Z be a p × p symmetric matrix. Then for the k-order matrix derivative of a scalar function y(Z), the following is valid:

$$\frac{d^k y(Z)}{(\text{dvech } Z)^k} = G_k' \frac{d^k y(Z)}{(\text{dvec } Z)^k}, \qquad (3.1)$$

where $G_k = G \otimes G \otimes \ldots \otimes G$ is k-times Kronecker product of matrix G, defined by (2.6), and the derivative at the right side of (3.1) is found without considering Z to be symmetric.

Proof. If k = 1, then by virtue of (2.16) the equality (3.1) holds. Assume that it holds in the case of some k-1:

$$\frac{d^{k-1} y(Z)}{(\text{dvech } Z)^{k-1}} = G_{k-1}' \frac{d^{k-1} y(Z)}{(\text{dvec } Z)^{k-1}} \qquad (3.2)$$

The differentiation of both sides of (3.2) with respect to vech Z and application of the rule (2.11) lead to

$$\frac{d^k y(Z)}{(\text{dvech } Z)^k} = \frac{d}{\text{dvech } Z}\left[G_{k-1}' \frac{d^{k-1} y(Z)}{(\text{dvec } Z)^{k-1}}\right] =$$

$$= \left[G_{k-1}' \otimes I_{p(p+1)/2}\right] \frac{d}{\text{dvech } Z}\left[\frac{d^{k-1} y(Z)}{(\text{dvec } Z)^{k-1}}\right].$$

As the vector $d^{k-1}y(Z)/(\mathrm{dvec}\ Z)^{k-1}$ of dimension $p^{2(k-1)}$, is the function of symmetric matrix Z, then according to the property (2.17) we get

$$\frac{d}{\mathrm{dvech}\ Z}\left[\frac{d^{k-1}y(Z)}{(\mathrm{dvec}\ Z)^{k-1}}\right]=\left[I_{2(k-1)}\otimes G'\right]\frac{d^{k}y(Z)}{(\mathrm{dvec}\ Z)^{k}}\,.$$

Finally

$$\frac{d^{k}y(Z)}{(\mathrm{d\ vech}\ Z)^{k}}=\left[G'_{k-1}\otimes I_{p(p+1)/2}\right]\left[I_{2(k-1)}\otimes G'\right]\frac{d^{k}y(Z)}{(\mathrm{d\ vec}\ Z)^{k}}$$

which by virtue of (2.1) proves the theorem, q.e.d.

4. The matrix form of the Taylor expansion

Let the q-vector g(x) be a function of some p-vector x ($g : R^p \to R^q$). Using the notions of higher-order matrix derivative and the power of vector, we give the following form for the Taylor expansion of g(x) (the earlier matrix representations of Taylor expansion see e.g. in [9, 10]).

Theorem 2. If all the partial derivatives of g(x) of orders n + 1 and less are continuous in a neighbourhood D of point a, then

$$g(x)=\sum_{k=0}^{n}\frac{1}{k!}\left[\frac{d^{k}g'(x)}{(dx)^{k}}\right]'_{x=a}(x-a)^{k}+$$

$$+\frac{1}{(n+1)!}\left[\frac{d^{n+1}g'(x)}{(dx)^{n+1}}\right]'_{x=\xi}(x-a)^{n+1}, \qquad (4.1)$$

where $\xi\in D$[1].

Proof. Note, that the expansion (4.1) includes the expansion for each component $g_i(x)$ of vector g(x). As according to the definition $d^{k}g'(x)/(dx)^{k}$ is the matrix, the i-th column of which is $d^{k}g_i(x)/(dx)^{k}$, then from (4.1) follows for $g_i(x)$:

$$g_i(x)=\sum_{k=0}^{n}\frac{1}{k!}\left[\frac{d^{k}g_i(x)}{(dx)^{k}}\right]'_{x=a}(x-a)^{k}+$$

$$+\frac{1}{(n+1)!}\left[\frac{d^{n+1}g_i(x)}{(dx)^{n+1}}\right]'_{x=\xi}(x-a)^{n+1}. \qquad (4.2)$$

It is enough to show that this expansion is identical to the usual representation of it (given for instance in [13]). For that we only need to rewrite the k-th term of this expansion with the help of definitions of k-order derivative and the

---

[1] Note, that the power $(x-a)^{k}$ is meant in the sense of definition 6.

k-th power. According to these, the vector $d^k g_i(x)/(dx)^k$ consists of all k-order partial derivatives $\partial^k g_i(x)/\partial x_{i_1} \ldots \partial x_{i_k}$, and the vector $(x-a)^k$ consists of all products $(x_{i_1} - a_{i_1})\ldots(x_{i_k} - a_{i_k})$, where $i_1,\ldots,i_k \in \{1,\ldots,p\}$. Thus, their scalar product gives the following multiple sum for the k-th term of (4.2).

$$\frac{1}{k!}\left[\frac{d^k g_i(x)}{(dx)^k}\right]'(x-a)^k =$$

$$= \frac{1}{k!} \sum_{i_1,\ldots,i_k=1}^{p} \frac{\partial^k g_i(x)}{\partial x_{i_1} \ldots \partial x_{i_k}} (x_{i_1} - a_{i_1})\ldots(x_{i_k} - a_{i_k}).$$

q.e.d.

5. The Taylor expansion for the function of symmetric matrix

If the q-vector $g(Z)$ is a function of some symmetric $p \times p$ matrix Z, then it may be considered to be the function of vech Z. As vech Z is the vector with functionally independent elements, then theorem 2 may be used for expanding $g(Z)$ in the following way:

$$g(Z) = \sum_{k=0}^{n} \frac{1}{k!}\left[\frac{d^k g'(Z)}{(\text{dvech } Z)^k}\right]'_{Z=A} \left[\text{vech}(Z - A)\right]^k +$$

$$+ \frac{1}{(n+1)!}\left[\frac{d^{n+1} g'(Z)}{(\text{dvech } Z)^{n+1}}\right]'_{Z=A_1} \left[\text{vech}(Z-A)\right]^{n+1} \qquad (5.1)$$

$A, A_1 \in D$.

For the practical expanding of $g(Z)$ the following form of Taylor expansion is useful, in which the derivatives are found with respect to all elements of Z.

Theorem 3. If the function of symmetric matrix - $g(Z)$ has all continuous partial derivatives of orders n+1 and less in a neighbourhood D of matrix A, then the vector $g(Z)$ may be expanded in the following way:

$$g(Z) = \sum_{k=0}^{n} \frac{1}{k!}\left[\frac{d^k g'(Z)}{(\text{dvec } Z)^k}\right]'_{Z=A} \left[\text{vec}(Z-A)\right]^k +$$

$$+ \frac{1}{(n+1)!}\left[\frac{d^{n+1} g'(Z)}{(\text{dvec } Z)^{n+1}}\right]'_{\substack{Z=A_1,\\ A_1 \in D}} \left[\text{vec}(Z-A)\right]^{n+1} \qquad (5.2)$$

where the derivatives are found without considering Z to be

symmetric.

Proof. We need to prove that the expressions (5.1) and (5.2) are identical.

Let us observe the expansion of the component $g_i(Z)$, obtained from (5.1). The k-th term of this expansion has the form

$$d_k = \frac{1}{k!}\left[\frac{d^k g_i(Z)}{(dveoh\ Z)^k}\right]' \left[vech(Z - A)\right]^k.$$

Application of theorem 1 and equality (2.7) leads to

$$d_k = \frac{1}{k!}\left[\frac{d^k g_i(Z)}{(dvec\ Z)^k}\right]' G_k\left[H\ vec(Z - A)\right]^k.$$

Further transformations with repeated application of the property (2.8) give

$$G_k\left[Hvec(Z - A)\right]^k = (G \otimes \ldots \otimes G)\left[H\ vec(Z-A) \otimes \ldots \ldots \otimes Hvec(Z-A)\right] = GH\ vec(Z-A) \otimes \ldots \otimes GH\ vec(Z-A) = \left[GH\ vec(Z-A)\right]^k = \left[vec(Z-A)\right]^k.$$

Finally

$$d_k = \frac{1}{k!}\left[\frac{d^k g_i(Z)}{(dvec\ Z)^k}\right]'\left[vec(Z-A)\right]^k,$$

which is the desired k-th term of the expansion of $g_i(Z)$ given by (5.2). Q.E.D.

6. The moment of a random vector

The purpose is to define the k-moment of some random vector x in the form of another vector, which includes all the k-order mixed moments of vector x. First this idea in a little different form was used in [12].

DEFINITION 7. The k-moment μ_k of some random vector x is the expectation of its k-th power:

$$\mu_k = E\ x^k. \tag{6.1}$$

If x is the p-vector, then due to the definition of the power of vector, μ_k is the p^k-vector consisting of all k-order mixed moments of x. In particular, if p = 1, then (6.1) defines the usual moment of a random variable.

We say that the moment μ_k exists, if all its elements exist.

DEFINITION 8. The central k-moment $\bar{\mu}_k$ of vector x is the

expectation of the k-th power of vector $x - \mu_1$:

$$\bar{\mu}_k = E(x - \mu_1)^k. \tag{6.2}$$

For example, application of (2.10), (2.9) gives for the second central moment

$$\bar{\mu}_2 = E(x - \mu_1)^2 = E[(x - \mu_1) \otimes (x - \mu_1)] = E \operatorname{vec}[(x - \mu_1)(x - \mu_1)'],$$

from which follows

$$\bar{\mu}_2 = \operatorname{vec} \Sigma, \tag{6.3}$$

where Σ is the covariance matrix of x.

The expressions of some first moments for the vector of sample covariance matrix vec S and for its vector-function g(S) are given by the author in [16]. The technique of deriving the moments of vec S is described in [15].

Next we show that the matrix differentation is a powerful tool in finding the moments μ_k, $\bar{\mu}_k$ of random vector x, if characteristic function

$$\psi(t) = E\, e^{it'x}$$

is known.

Theorem 4. If the moment μ_k of random vector x exists, then

$$\frac{d^k \psi(t)}{(dt)^k} \Big|_{t=0} = i^k \mu_k. \tag{6.4}$$

Proof. If k = 1, then we have to find the first order matrix derivative of composite function $e^{it'x}$ with respect to the vector t. Application of the properties (2.15),(2.2) (2.11), (2.14) gives

$$\frac{d\,\psi(t)}{dt} = E\,\frac{de^{it'x}}{dt} = E\left[e^{it'x} * \frac{d(it'x)}{dt}\right] =$$

$$= E\left[\frac{d(it'x)}{dt}\, e^{it'x}\right] = E\left[i\,\frac{dt'}{dt}\, x\, e^{it'x}\right] = E\left[ixe^{it'x}\right]$$

and consequently

$$\frac{d\psi(t)}{dt} \Big|_{t=0} = iEx = i\,\mu_1.$$

Let us assume that in the case of some k-1 the following

$$\frac{d^{k-1}\psi(t)}{(dt)^{k-1}} = E\,\frac{d^{k-1}e^{it'x}}{(dt)^{k-1}} = E\,(i^{k-1}\, x^{k-1}\, e^{it'x})$$

is valid.

Differentiating this with the application of rule (2.11), we find

$$\frac{d^k\psi(t)}{(dt)^k} = i^{k-1}\, E\,\frac{d(x^{k-1}e^{it'x})}{dt} = i^{k-1}E\left[(x^{k-1}\otimes I_p)\frac{de^{it'x}}{dt}\right] =$$

$$= i^k\, E\left[(x^{k-1}\otimes I_p)\, x\, e^{it'x}\right].$$

As according to (2.1)

$$(x^{k-1}\otimes I_p)x = (x^{k-1}\otimes I_p)(1\otimes x) = x^{k-1}\otimes x = x^k,$$

then

$$\frac{d^k\psi(t)}{(dt)^k}\Big|_{t=0} = i^k\, E\left[x^k e^{it'x}\right]_{t=0} = i^k Ex^k = i^k\mu_k.$$

Q.E.D.

An obvious conclusion of theorem 4 is following.

Corollary 4. If $\psi_y(t)$ is the characteristic function of vector $y = x - \mu_1$ ($\mu_1 = Ex$), then the matrix differentiation of $\psi_y(t)$ gives the central moments of x:

$$\frac{d^k\psi_y(t)}{(dt)^k}\Big|_{t=0} = i^k\,\bar{\mu}_k. \qquad (6.5)$$

Note, that the expressions (6.4), (6.5) define the usual moments of random variable, if $x,t \in R^1$.

7. The cumulant of a random vector

Let $f(t) = \ln\psi(t)$ be the logarithmic characteristic function of a random vector x ($x,t \in R^p$).

DEFINITION 9. The k-cumulant of a random p-vector x is the p^k-vector $\varkappa_k$ defined by the equality

$$\frac{d^k\ln\psi(t)}{(dt)^k}_{t=0} = i^k\,\varkappa_k. \qquad (7.1)$$

As the k-order matrix derivative of the scalar function $\ln\psi(t)$ is the vector consisting of all k-order partial derivatives of $\ln\psi(t)$, then $\varkappa_k$ consists of all k-order mixed cumulants of x. In particular, if the dimension p = 1, then (7.1) gives the usual cumulant of a random variable.

The connections between cumulants and moments can be obtained when finding the matrix derivatives of both sides of the equality

$$f(t) = \ln\psi(t).$$

For three first cumulants the following is valid:

$$\varkappa_1 = \mu_1,\ \varkappa_2 = \bar{\mu}_2 = \mathrm{vec}\,\Sigma\ ,\ \varkappa_3 = \bar{\mu}_3.$$

8. The Taylor expansion for characteristic functions

Theorem 5. The characteristic function $\psi(t)$ and the logarithmic characteristic function $\ln \psi(t)$ of a random vector x can be formally expanded as follows:

$$\psi(t) = \sum_{k=0}^{\infty} \mu_k' \frac{(it)^k}{k!}, \tag{8.1}$$

$$\ln \psi(t) = \sum_{k=1}^{\infty} \varkappa_k' \frac{(it)^k}{k!}. \tag{8.2}$$

Proof. Applying theorem 2 for the functions $\psi(t)$ and $\ln \psi(t)$ in the point $t = 0$ and using the expressions (6.4), (7.1) of their matrix derivatives, the expansions (8,1) and (8.2) follow easily. Q.E.D.

We note that using the new notions ($\mu_k, \varkappa_k$, the power of a vector t^k), the expansion of multivariate characteristic function can be put down in the same simple form as in the univariate case.

9. Central moments and cumulants of the normal random vector

The usual form of the characteristic function of the distribution $N_p(0, \Sigma)$ is

$$\varphi(t) = e^{-\frac{1}{2} t'\Sigma t}. \tag{9.1}$$

Applying the operator vec to the scalar $t'\Sigma t$ and using the connections (2.4), (6.3) we get

$$t'\Sigma t = \operatorname{vec} t'\Sigma t = (t \otimes t)' \operatorname{vec} \Sigma = (t \otimes t)' \bar{\mu}_2,$$

from which follows the other form of (9.1):

$$\varphi(t) = e^{-\frac{1}{2} (t \otimes t)' \bar{\mu}_2} \tag{9.2}$$

Further, in this paper, we also use the notation $N(\mu_1, \bar{\mu}_2)$ for the distribution $N(\mu_1, \Sigma)$.

Theorem 6. The central moments $\bar{\mu}_4$, $\bar{\mu}_6$ of the multivariate normal distribution $N_p(\mu_1, \bar{\mu}_2)$ are expressed as the functions of moment $\bar{\mu}_2$ in the form:

$$\bar{\mu}_4 = J_1 \bar{\mu}_2^2, \tag{9.3}$$

$$\bar{\mu}_6 = J_2 J_3 \bar{\mu}_2^3, \tag{9.4}$$

where

$$J_1 = I_{p4} + I_{p,p^2} \otimes I_p + I_{p,p^3},$$

$$J_2 = I_{p^6} + I_{p,p^2} \otimes I_{p^3} + I_{p,p^3} \otimes I_{p^2} + I_{p,p^4} \otimes I_p + I_{p,p^5},$$

$$J_3 = I_{p^6} + I_{p^2} \otimes I_{p,p^2} \otimes I_p + I_{p^2} \otimes I_{p,p^3} .$$

Proof. We shall derive the expressions for $\bar{\mu}_4$ and $\bar{\mu}_6$ with the help of matrix differentiation of characteristic function (9.2). Step by step we shall find the derivatives of higher orders and find their values in the point $t = 0$.

The following auxiliary equalities are needed:

$$\frac{d(t' \otimes I_{p^k})}{dt} = I_{p,p^k} \tag{9.5}$$

$$\left[I_{p,p^k}(\bar{\mu}_2 \otimes I_{p^{k-1}})\right] \otimes I_{p^\ell} = (I_{p,p^k} \otimes I_{p^\ell})(\bar{\mu}_2 \otimes I_{p^{k-1+\ell}}). \tag{9.6}$$

The equality (9.5) follows straightforwardly when applying the differentiation rule (2.12). The equality (9.6) is obtained by substituting $I_{p^\ell} = I_{p^\ell} \cdot I_{p^\ell}$ and applying the property (2.1).

1) Using the rules (2.15), (2.2), (2.11) we get for the derivative of first order

$$\frac{d\varphi(t)}{dt} = -\frac{1}{2}\frac{d[(t \otimes t)'\bar{\mu}_2)]}{dt}\varphi(t) = -\frac{1}{2}\frac{d(t \otimes t)'}{dt}\bar{\mu}_2(t).$$

The differentiation of the Kronecker product according to (2.12) yields

$$\frac{d\varphi(t)}{dt} = -\frac{1}{2}[t' \otimes I_p + I_p(t' \otimes I_p)\, I_{p,p}]\bar{\mu}_2\varphi(t) =$$
$$= -\frac{1}{2}(t' \otimes I_p)(I_{p^2} + I_{p,p})\bar{\mu}_2\varphi(t).$$

As by virtue of (6.3), (2.5)

$$I_{p,p}\bar{\mu}_2 = \bar{\mu}_2, \tag{9.7}$$

then

$$\frac{d\varphi(t)}{dt} = -(t' \otimes I_p)\bar{\mu}_2\varphi(t), \tag{9.8}$$

from which

$$\frac{d\varphi(t)}{dt}\Big|_{t=0} = 0. \tag{9.9}$$

2) The differentiation of (9.8) according to rules (2.11), (9.5) gives

$$\frac{d^2\varphi(t)}{(dt)^2} = -\frac{d(t' \otimes I_p)}{dt}\bar{\mu}_2\varphi(t) - (t' \otimes I_p \otimes I_p)\frac{d(\bar{\mu}_2\varphi(t))}{dt} =$$
$$= -I_{p,p}\bar{\mu}_2\varphi(t) - (t' \otimes I_{p^2})(\bar{\mu}_2 \otimes I_p)\frac{d\varphi(t)}{dt}.$$

Finally, by means of (9.7)

$$\frac{d^2\varphi(t)}{(dt)^2} = -\bar{\mu}_2\varphi(t) - (t' \otimes I_{p^2})(\bar{\mu}_2 \otimes I_p)\frac{d\varphi(t)}{dt}, \tag{9.10}$$

from which for t = 0 follows

$$\frac{d^2\varphi(t)}{(dt)^2}\Big|_{t=0} = -\bar{\mu}_2. \qquad (9.11)$$

3) Applying the properties (2.11), (9.5), we find the derivative of (9.10):

$$\frac{d^3\varphi(t)}{(dt)^3} = -(\bar{\mu}_2 \otimes I_p)\frac{d\varphi(t)}{dt} - I_{p,p^2}(\bar{\mu}_2 \otimes I_p)\frac{d\varphi(t)}{dt} - (t' \otimes I_{p^3})(\bar{\mu}_2 \otimes I_{p^2})\frac{d^2\varphi(t)}{(dt)^2}. \qquad (9.12)$$

According to (9.9)

$$\frac{d^3\varphi(t)}{(dt)^3}\Big|_{t=0} = 0.$$

4) The differentiation of (9.12) in an analogical way yields:

$$\frac{d^4\varphi(t)}{(dt)^4} = -(\bar{\mu}_2 \otimes I_{p^2})\frac{d^2\varphi(t)}{(dt)^2} - \left[I_{p,p^2}(\bar{\mu}_2 \otimes I_p) \otimes I_p\right]\frac{d^2\varphi(t)}{(dt)^2} - I_{p,p^3}(\bar{\mu}_2 \otimes I_{p^2})\frac{d^3\varphi(t)}{(dt)^3} - (t' \otimes I_{p^4})(\bar{\mu}_2 \otimes I_{p^3})\frac{d^3\varphi(t)}{(dt)^3}.$$

The application of (9.9), (9.10) gives

$$\frac{d^4\varphi(t)}{(dt)^4}\Big|_{t=0} = (\bar{\mu}_2 \otimes I_{p^2})\bar{\mu}_2 + \left[I_{p,p^2}(\bar{\mu}_2 \otimes I_p) \otimes I_p\right]\bar{\mu}_2 + I_{p,p^3}(\bar{\mu}_2 \otimes I_{p^2})\bar{\mu}_2.$$

For simplification we use (9.6) and the equality

$$(\bar{\mu}_i \otimes I_{p^j})\bar{\mu}_j = \bar{\mu}_i \otimes \bar{\mu}_j,$$

which is easy to show by substituting $\bar{\mu}_j = 1 \otimes \bar{\mu}_j$ and applying the property (2.1). It leads to

$$\frac{d^4\varphi(t)}{(dt)^4}\Big|_{t=0} = \bar{\mu}_2^2 + (I_{p,p^2} \otimes I_p)\bar{\mu}_2^2 + I_{p,p^3}\bar{\mu}_2^2,$$

where by our definition $\bar{\mu}_2^2 = \bar{\mu}_2 \otimes \bar{\mu}_2$. Taking the vector $\bar{\mu}_2^2$ behind the brackets yields the expression (9.3) of theorem 6. The expression (9.4) is obtained by continuing the process of differentiation in an analogical way. Q.E.D.

In particular, if p = 1, then

$$I_{1,1} = I_1 = 1$$

and from theorem 6 follow the expressions of moments for univariate normal distribution

$$\bar{\mu}_4 = 3\bar{\mu}_2^2, \quad \bar{\mu}_6 = 15\,\bar{\mu}_2^3,$$

where $\bar{\mu}_2$ is the variance.

Theorem 7. The cumulants of multivariate normal distribution $N_p(\mu_1, \bar{\mu}_2)$ are

$$\varkappa_1 = \mu_1, \ \varkappa_2 = \bar{\mu}_2, \ \varkappa_i = 0, \ i > 2.$$

Proof. The logarithmic characteristic function of $N_p(\mu_1, \bar{\mu}_2)$ is

$$f(t) = it'\mu_1 - \frac{1}{2}(t \otimes t)'\bar{\mu}_2.$$

Differentiating this we get

$$\frac{df(t)}{dt} = i\frac{dt'}{dt}\mu_1 - \frac{1}{2}\left[t' \otimes \frac{dt'}{dt} + (t' \otimes \frac{dt'}{dt})I_{p,p}\right]\bar{\mu}_2,$$

from which by virtue of (2.14) follows

$$\frac{df(t)}{dt} = i\mu_1 - (t' \otimes I_p)\bar{\mu}_2. \tag{9.13}$$

As from this

$$\frac{df(t)}{dt}\Big|_{t=0} = i\mu_1,$$

then $\varkappa_1 = \mu_1$.

The differentiation of (9.13) yields

$$\frac{d^2f(t)}{(dt)^2} = -I_{p,p}\bar{\mu}_2 = -\bar{\mu}_2 = i^2\bar{\mu}_2, \tag{9.14}$$

from which

$$\varkappa_2 = \bar{\mu}_2.$$

As the second derivative (9.14) is constant then higher order derivatives are zeros. Q.E.D.

10. Multivariate Hermite polynomials

Let us see the density function

$$\Phi(x) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}x'\Sigma^{-1}x} \tag{10.1}$$

of $N_p(0,\Sigma)$. It is known that the multivariate k-order Hermite polynomial is obtained by finding the k-order partial derivative of $\Phi(x)$. With the help of matrix differentiation, it is possible to find all k-order Hermite polynomials at the same time, as the components of a vector h_k.

Theorem 8. The vector h_k of multivariate Hermite polynomials, defined by the equation

$$\frac{d^k\Phi(x)}{(dx)^k} = (-1)^k h_k \Phi(x), \tag{10.2}$$

has for k = 1,2,3 the following forms:

$$h_1 = \Sigma^{-1} x, \tag{10.3}$$

$$h_2 = (\Sigma^{-1} \otimes \Sigma^{-1})\, x^2 - \operatorname{vec} \Sigma^{-1}, \tag{10.4}$$

$$h_3 = (\Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1})\, x^3 - (\operatorname{vec} \Sigma^{-1} \otimes \Sigma^{-1}) x - $$
$$- (\Sigma^{-1} \otimes \operatorname{vec} \Sigma^{-1}) x - \operatorname{vec}(\Sigma^{-1} x \otimes \Sigma^{-1}). \tag{10.5}$$

Proof. According to the definition (10.2), we have to find matrix derivatives of $\Phi(x)$ up to the third order.

1) For the first derivative, the application of properties (2.15), (2.2), (2.11), yields

$$\frac{d\Phi(x)}{dx} = \frac{d(-\frac{1}{2} x' \Sigma^{-1} x)}{dx} \Phi(x) =$$
$$= -\frac{1}{2}\left[\frac{dx'}{dx} \Sigma^{-1} x + (x' \otimes I_p) \frac{d(\Sigma^{-1} x)}{dx}\right] \Phi(x).$$

As

$$\frac{d(\Sigma^{-1} x)}{dx} = (\Sigma^{-1} \otimes I_p) \operatorname{vec} I_p,$$

from which by virtue of (2.4) follows

$$\frac{d(\Sigma^{-1} x)}{dx} = \operatorname{vec} \Sigma^{-1}, \tag{10.6}$$

then application of this and (2.14) gives

$$\frac{d\Phi(x)}{dx} = -\frac{1}{2}\left[\Sigma^{-1} x + (x' \otimes I_p) \operatorname{vec} \Sigma^{-1}\right] \Phi(x).$$

Finally, using of (2.4) yields

$$\frac{d\Phi(x)}{dx} = -\Sigma^{-1} x \Phi(x), \tag{10.7}$$

from which follows the expression (10.3) of h_1.

2) The differentiation of (10.7) according to (2.11) leads to

$$\frac{d^2\Phi(x)}{(dx)^2} = -\frac{d(\Sigma^{-1} x)}{dx} \Phi(x) + (\Sigma^{-1} x \otimes I_p) \Sigma^{-1} x \Phi(x).$$

As $\Sigma^{-1} x$ is a vector, then according to (2.4) and (2.10)

$$(\Sigma^{-1} x \otimes I_p) \Sigma^{-1} x = \operatorname{vec}(\Sigma^{-1} x x' \Sigma^{-1}) =$$
$$= (\Sigma^{-1} \otimes \Sigma^{-1}) \operatorname{vec} x x' = (\Sigma^{-1} \otimes \Sigma^{-1}) x^2.$$

The application of this and of (10.6) yields

$$\frac{d^2\Phi(x)}{(dx)^2} = \left[-\operatorname{vec} \Sigma^{-1} + (\Sigma^{-1} \otimes \Sigma^{-1}) x^2\right] \Phi(x), \tag{10.8}$$

from which follows the expression (10.4) of h_2.

3) The differentiation of (10.8) gives the expression (10.5) of h_3. Q.E.D.

11. The Fourier transform connected with the multivariate normal distribution

The characteristic function of $N_p(0,\Sigma)$ is defined by the following Fourier transform of $\Phi(x)$:

$$\varphi(t) = \int_{R^p} e^{it'x}\Phi(x)dx, \qquad (11.1)$$

where $dx = dx_1 \ldots dx_p$ and $\Phi(x)$ is the density function of $N_p(0,\Sigma)$. It is also possible to find the Fourier transforms of several partial derivatives of $\Phi(x)$. We give the appropriate result in the matrix form.

Theorem 9. For the functions $\Phi(x)$ and $\varphi(t)$ defined by (10.1) and (11.1), the following result is valid:

$$(it)^k\varphi(t) = (-1)^k \int_{R^p} e^{it'x} \frac{d^k\Phi(x)}{(dx)^k}\,dx, \qquad (11.2)$$

where the integral is taken separately from each component of vector $e^{it'x}d^k\Phi(x)/(dx)^k$, $(x,t \in R^p)$.

Proof. To prove the equality of vectors, we have to prove the equality for the corresponding components of these vectors. Using the definitions of the power of vector and higher-order matrix derivative of a scalar function, the following equality for corresponding components is obtained from (11.2):

$$i^k t_{i_1}\cdots t_{i_k}\varphi(t) = (-1)^k \int_{R^p} e^{it'x} \frac{\partial^k\Phi(x)}{\partial x_{i_1}\cdots\partial x_{i_k}}\,dx, \qquad (11.3)$$

$i_1,\ldots,i_k \in \{1,\ldots,p\}$.

This equality can be proved easily with the help of mathematical induction by k, using integration by parts. Q.E.D.

The important conclusion of this theorem for deriving the multivariate Edgeworth expansion is following.

Corollary 9. If c is a constant p^k-vector, then the following expression is valid:

$$(-1)^k c' \frac{d^k\Phi(x)}{(dx)^k} = (2\pi)^{-p} \int_{R^p} c'(it)^k\varphi(t)e^{-it'x}dt,$$

i.e. left side of this equality is the inverse Fourier transform of the scalar product $c'(it)^k\varphi(t)$.

Proof. Theorem 9 says that the vector $(it)^k\varphi(t)$ is the Fourier transform of the vector $(-1)^k d^k\Phi(x)/(dx)^k$, which therefare must be expressed with the help of inverse Fourier transform:

$$(-1)^k \frac{d^k \Phi(x)}{(dx)^k} = (2\pi)^{-p} \int_{R^p} (it)^k \varphi(t) e^{-it'x} dt.$$

Multiplying the both sides of this equality on the left by the vector c transposed and applying the additivity of integral, the corollary is proved. Q.E.D.

12. The matrix form of the multivariate Edgeworth expansion

The Edgeworth expansion is an expansion of the density function of a statistic (sample function), basing on the density function of a normal distribution and its derivatives. The expansion is ordered corresponding to the powers of sample size n.

Let z be a p dimensional statistic with
cumulants - $\varkappa_i$,
density function - f(x),
characteristic function - $\psi(t)$.

Deriving the Edgeworth expansion of f(x) we need to know, how the cumulants $\varkappa_i$ of statistic z depend on the sample size n. Let us assume

$$\varkappa_1 = n^{-\frac{1}{2}} \gamma_1 + o(n^{-1}), \tag{12.1}$$

$$\varkappa_2 = k_2 + n^{-1} \gamma_2 + o(n^{-1}), \tag{12.2}$$

$$\varkappa_3 = n^{-\frac{1}{2}} \gamma_3 + o(n^{-1}), \tag{12.3}$$

$$\varkappa_4 = n^{-1} \gamma_4 + o(n^{-1}), \tag{12.4}$$

$$\varkappa_j = o(n^{-1}), \quad j \geq 5, \tag{12.5}$$

where k_2, γ_1, ..., γ_4 are the matrix expressions, depending on the population cumulants, but not on n. For example, the cumulants of statistics $\sqrt{n}\ \mathrm{vec}(S - \Sigma)$, $\sqrt{n}[g(S) - g(\Sigma)]$, where S and Σ are the sample and population covariance matrices, g is some function of them, are represented in the forms (12.1) - (12.5).

Theorem 10. The formal Edgeworth expansion for the p dimensional statistic z with cumulants (12.1)-(12.5), is given up to the order n^{-1} in the following way:

$$f(x) = \Phi(x) - n^{-\frac{1}{2}}[\gamma_1' d_1 + \frac{1}{6} \gamma_3' d_3] +$$
$$+ n^{-1}[\tfrac{1}{2}(\gamma_2 + \gamma_1^2)' d_2 + \tfrac{1}{24}(\gamma_4 + 2\gamma_1 \otimes \gamma_3 + 2\gamma_3 \otimes \gamma_1)' d_4 + \tfrac{1}{72}(\gamma_3^2)' d_6] +$$

$$+ o(n^{-1}),$$

where

$$d_k = \frac{d^k \Phi(x)}{(dx)^k}, \tag{12.6}$$

$\Phi(x)$ is the density function of $N_p(0,k_2)$, k_2 is the limiting ($n \to \infty$) second cumulant[1] of statistic z, γ_1, γ_2, γ_3, γ_4 are the terms from the expressions of cumulants determined by (12.1) - (12.5).

Proof. Inserting the expansion of $\ln\psi(t)$ given by (8.2) into identity

$$\psi(t) = \exp(\ln\psi(t)),$$

we get

$$\psi(t) = \exp\Big[\sum_{k=1}^{\infty} \varkappa'_k \frac{(it)^k}{k!}\Big]. \tag{12.7}$$

The substitution of cumulants (12.1)-(12.5) into (12.7) yields

$$\psi(t) = \exp\Big[n^{-\frac{1}{2}} \gamma'_1 (it) + k'_2 \frac{(it)^2}{2} + n^{-1}\gamma'_2 \frac{(it)^2}{2} + n^{-\frac{1}{2}}\gamma'_3 \frac{(it)^3}{3!} + n^{-1}\gamma'_4 \frac{(it)^4}{4!} + o(n^{-1})\Big].$$

Separating here the multiplier

$$\varphi(t) = \exp\left(k'_2 \frac{(it)^2}{2}\right),$$

which is the characteristic function of $N_p(0,k_2)$, we obtain

$$\psi(t) = \exp\Big[n^{-\frac{1}{2}}\Big(\gamma_1' it + \gamma'_3 \frac{(it)^3}{3!}\Big) + n^{-1}\Big(\gamma_2' \frac{(it)^2}{2} + \gamma_4' \frac{(it)^4}{4!} + o(n^{-1})\Big]\varphi(t).$$

Expanding of the exponental function above to the power series, gives

$$\psi(t) = \Big\{1 + n^{-\frac{1}{2}}\Big(\gamma_1' it + \gamma'_3 \frac{(it)^3}{3!}\Big) + n^{-1}\Big(\gamma_2' \frac{(it)^2}{2} + \gamma_4' \frac{(it)^4}{4!}\Big) + \frac{1}{2} n^{-1}\Big[\Big(\gamma_1' it\Big)^2 + \Big(\gamma_1' it\Big)\Big(\gamma_3' \frac{(it)^3}{3!}\Big) + \Big(\gamma_3' \frac{(it)^3}{3!}\Big)\Big(\gamma_1' it\Big) - \Big(\gamma_3' \frac{(it)^3}{3!}\Big)^2\Big] + o(n^{-1})\Big\}\varphi(t) \tag{12.8}$$

The summands in the square brackets are transformed with the help of (2.1). For example

---

[1] Note, that according to (6.3), k_2 is the limiting covariance matrix of z in the vextor form.

$$(\gamma_1'it)^2 = (\gamma_1'it)\otimes(\gamma_1'it)=(\gamma_1\otimes\gamma_1)'(it\otimes it)=(\gamma_1^2)'(it)^2,$$

$$(\gamma_1'it)(\gamma_3'\frac{(it)^3}{3!}) = (\gamma_1'it)\otimes(\gamma_3'\frac{(it)^3}{3!}) = (\gamma_1\otimes\gamma_3)'\frac{(it)^4}{3!},$$

Thus, (12.8) will take the form

$$\psi(t) = \{1 + n^{-\frac{1}{2}}[\gamma_1'it + \tfrac{1}{6}\gamma_3'(it)^3] + n^{-1}[\tfrac{1}{2}\gamma_2'(it)^2 + \tfrac{1}{2}(\gamma_1^2)'(it)^2 + \tfrac{1}{24}\gamma_4'(it)^4 + \tfrac{1}{12}(\gamma_1\otimes\gamma_3)'(it)^4 + \tfrac{1}{72}(\gamma_3^2)'(it)^6] + o(n^{-1})\}\varphi(t). \qquad (12.9)$$

Carring out the inverse Fourier transform of (12.9), $\psi(t)$ will be replaced by the density function $f(x)$, first summand $\varphi(t)$ will be replaced by the density function $\Phi(x)$ of the distribution $N(0,k_2)$ and the vectors $(it)^k\varphi(t)$ will be replaced according to corollary 9 by the vectors $(-1)^k d_k$, where d_k is the matrix derivative (12.6). Q.E.D.

The Edgeworth expansion for a concrete statistic follows from theorem 10 if we only know the expressions γ_1, k_2 of that statistic. In [16] the author has given the Edgeworth expansion up to the order $n^{-1/2}$ for the multivariate statistic $\sqrt{n}\,[g(S) - g(\Sigma)]$ without assuming the normality of population. Here S, Σ are the sample and population covariance matrices, g is some vectorfunction of them.

Theorem 10 gives the formal Edgeworth expansion for the class of statistics. This expansion will approximate the real distribution of the statistic with given accuracy, if some assumptions about the population distribution and about that statistic are made. These assumptions are given in [1] for the class of statistics which involves all appropriately smooth functions of sample moments.

References

1. Bhattacharya, R.N., Ghosh, J.K., On the validity of the formal Edgeworth expansion. Ann. Statist., 1978, 6, 434-451.
2. Henderson, H.V., Searle, S.R., Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics. Can. J. Statist., 1979, 7, 65-81.
3. MacRae, E.C., Matrix derivatives with an application to an adaptive linear decision problem. Ann. Statist., 1974, 2, 337-346.

4. Magnus, J.R., Neudecker, H., The commutation matrix: some properties and applications. Ann. Statist.,1979, 7, 381-394.
5. McCulloch, C.E. Symmetric matrix derivatives with applications. J. Amer. Statist. Assoc., 1982, 77, 679-682.
6. Nel, D.G. On matrix differentiation in statistics.South African Statist. J., 1980, 14, 137-193.
7. Neudecker, H. Some theorems on matrix differentiation with specisl reference to Kronecker matrix products. J. Amer. Statist. Assoc., 1969, 64, 953-963.
8. Neudecker, H., Wansbeek, T. Some results on commutation matrices with statistical applications. Can.J.Statist. 1983, 11, 221-231.
9. Turnbull, H.W. A matrix form of Taylor's theorem. Proc. Edinburgh Math. Soc.: Series 2, 1929, 2, 33-54.
10. Vetter, W.J. Matrix calculus operations and Taylor expansions. SIAM Review, 1973, 15, 352-368.
11. Картан А., Дифференциальное исчисление. Дифференциальные формы. М.: Мир, 1971.
12. Колло Т. Некоторые понятия матричного исчисления с применением в математической статистике. Тр. Вычисл. центра Тартуск. ун-та, 1977, 40, 30-51.
13. Корн Т., Корн Т., Справочник по математике: для научных работников и инженеров. М.: Наука, 1968.
14. Крамер Г. Математические методы статистики. М.: Мир, 1975.
15. Траат И. Моменты выборочной ковариационной матрицы. Тр. Вычисл. центра Тартуск.ун-та, 1984, 51, 108-125.
16. Траат И. Многомерное распределение Эджворта для функции выборочной ковариационной матрицы. Тезисы III Всесоюзной научнотехн.конференции "Применение многомерного статистического анализа в экономике и оценые качества продукции" Тарту, 1985, 65-68.

МАТРИЧНОЕ ИСЧИСЛЕНИЕ ДЛЯ МНОГОМЕРНЫХ РАСПРЕДЕЛЕНИЙ

И. Траат

Р е з ю м е

Операции кронекеровского произведения и матричной производной позволяют удобным образом представить (а также вывести) многие результаты для многомерных распределений. Матричный вид получаемых результатов является легко реализуемым на ЭВМ.

В настоящей статье на основе указанных операций дано новое определение моментов и кумулянтов случайного вектора. Показывается, что выражения таким образом определенных моментов и кумулянтов получаются путем матричного дифференцирования характеристической и логарифмической характеристической функций. Приводится матричный вид разложения Тэйлора, который может быть использован при выведении выражений моментов вектор-функции. Разложение Тэйлора приводится и в случае, когда аргументом рассматриваемой функции является симметричная матрица.

При помощи матричного дифференцирования выводятся выражения центральных моментов до шестого порядка многомерного нормального распределения, а также матричные выражения многомерных полиномов Эрмита до третьего порядка.

Приводится матричная техника выведения разложения Эджворта для многомерной статистики. Для класса многомерных статистик представляется явная формула разложения до порядка n^{-1} (включительно).

Received August 1985

STRONG CONSISTENCY OF k-MEANS CLUSTERING CRITERION IN SEPARABLE METRIC SPACES

K. Pärna

1. Summary

A random sample from the population in the separable metric space is partitioned into the k clusters that minimise the given clustering criterion. Conditions are found that ensure the almost sure convergence of the sample minimum of the criterion to that of the population. Analogous result for Euclidean spaces has been proved by Pollard (1981).

2. Preliminaries and notations

At first we describe, in a simple manner, the k-means procedure.

Let $x_1, x_2, \ldots, x_n$ be n points in the metric space T. The k-means clustering procedure prescribes the following criterion for partition n points into k groups: first choose cluster centres $a_1, a_2, \ldots, a_k$ in T to minimize

$$W^n = \frac{1}{n} \sum_{i=1}^{n} \min_j d^2(x_i, a_j),$$

where $d(\cdot, \cdot)$ is the metrics in T, then assign each x_i to its nearest cluster. In this way, each centre a_j acquires a subset of the x's as its cluster. In the case when T is Euclidean space R^k and d is common Euclidean distance it is not difficult to see that optimal a_j equal to the means of their own clusters - hence the term "k-means".

It is important to say that in metric spaces the minimum of the criterion is possibly not attainable on any set of centres. Thereby, throughout this paper we are working with the infimum value of the criterion and with "ε-optimal" sets of centres, but not with optimal sets.

In what follows T is a separable metric space and $x_1, x_2, \ldots, x_n$ is assumed to be a random sample of n independent observations on some probability distribution P on T.

Conditions are given that ensure the almost sure convergence of the sample minimum of W^n to the population minimum. Our result remains valid for more general clustering criteria. To be precise, for each probability measure Q on T and each finite subset A of T define the sum of dispersions within the clusters

$$W(A,Q) = \int \min_{a \in A} \varphi(d(x,a))Q(dx).^{1)}$$

Here φ is nonnegative and nondecreasing function with real values (other requirements on φ will be given in Section 3). Further, let

$$W_j(Q) = \inf\{W(A,Q): A \text{ contains } j \text{ points}\}, \qquad (1)$$

from which the simple inequalities

$$W_1(Q) \geqslant W_2(Q) \geqslant \ldots \geqslant W_k(Q)$$

follow (addition of a point to the given set A never increases the criterion value).

In these notations we are minimizing $W(A,P_n)$, where P_n is the empirical measure obtained from the sample by placing mass 1/n at each of $x_1,x_2,\ldots,x_n$; the sample minimum of W^n is $W_k(P_n)$ now. By the strong law of large numbers (SLLN) for each fixed A $W(A,P_n)$ converges to $W(A,P)$ almost surely (a.s.). It may be expected therefore that under some conditions the convergence

$$W_k(P_n) \longrightarrow W_k(P) \quad \text{a.s.}$$

holds too. This is the result to be proved. The corresponding theorem will be formulated in Section 3 and proved in Sections 4 and 5.

Our result generalizes one of the assertions of Pollard (1981), who proved the convergence of the sample minimum (and convergence of the optimal sample centres themselves) in the case of the Euclidean spaces R^s. Some related pioneering works have been written by MacQueen (1967) and Hartigan (1978). In Thygeson-Sverdrup (1981) the case k = 1 for the compact metric spaces is considered.

Our method of proof uses some ideas from Pollard (1981). Say a set A, containing k points, is ε-optimal with respect to the measure Q, if

---

1) The domain of integration is T everywhere if not specified.

$$W(A,Q) < W_k(Q) + \varepsilon \ , \quad \varepsilon > 0.$$

Let A_{ε_n} be ε_n - optimal with respect to P_n. First it is shown that for appropriate sequence ε_n, which converges to zero, A_{ε_n} eventually lie in some ball $B \subset T$. The proof of this takes an inductive form, starting from the simple 1-means case. The second stage of the proof involves showing that, almost surely, $W(A,P_n) - W(A,P)$ converges to zero uniformly over those subsets of B containing k points. The proof of this uniform SLLN is based on a uniform convergence theorem given by Ranga Rao (1962) and is deferred to Section 5. At the 3-rd step the asymptotical equivalence of minimizing $W(\cdot,P_n)$ and $W(\cdot,P)$ - the desired result-is shown.

3. The convergence theorem

Introduce some regularity conditions on the function φ. We shall need φ defined on the interval $[0,\infty)$, being continuous, convex and nondecreasing, with $\varphi(0) = 0$ and $\varphi(r) \to \infty$ as $r \to \infty$. In order to control the growth of φ in the tails, assume that there exists a constant λ such that $\varphi(2r) \leq \lambda \cdot \varphi(r)$ for every $r > 0$. For example, any function of the form $\varphi(r) = r^s$ with positive s will go. As long as $\int \varphi(d(x,z))P(dx)$ is finite for some $z \in T$, these conditions ensure that $W(A,P)$ is finite for each A: for each $a \in T$

$$\int \varphi(d(x,a))P(dx) \leq \int \varphi(d(x,z) + d(z,a))P(dx) \leq$$
$$\leq \frac{1}{2} \int [\varphi(2d(x,z)) + \varphi(2d(z,a))]P(dx) \leq$$
$$\leq \frac{\lambda}{2} \int \varphi(d(x,z))P(dx) + \frac{\lambda}{2} \varphi(d(z,a)) < \infty \ .$$

Now we are ready to formulate our main result.

Theorem. Suppose that $\int \varphi(d(x,z))P(dx) < \infty$ for some $z \in T$ and inequalities

$$W_1(P) > W_2(P) > \ldots > W_k(P) \qquad (2)$$

hold. Then $W_k(P_n) \to W_k(P)$ almost surely.

It will be shown in Appendix (Lemma 2) that relations (2) are satisfied whenever there exist subsets A(j) of T, containing exactly j distinct points (j = 1,2,...,k-1), which minimize clustering criterion, i.e. $W(A(j),P) = W_j(P)$.

4. Proof of the theorem

The proof consists of three stages.

1) Let A_{ε_n} be an ε_n-optimal set of k centres for the sample $x_1, x_2, \ldots, x_n$, i.e. $W(A_{\varepsilon_n}, P_n) - W_k(P_n) < \varepsilon_n$, with $\varepsilon_n > 0$. We show that there exist a sequence $\varepsilon_n \to 0$ (as $n \to \infty$) and a closed ball $B(5M, x_0)$ with the centre x_0 and radius 5M so that A_{ε_n} will be contained in $B(5M, x_0)$, for n large enough.

The proof of that assertion has an inductive form. First we take an arbitrary $x_0 \in T$ and find an M so large that at least one point of A_{ε_n}, for n large enough, will be included in the closed ball $K = B(M, x_0)$. To do this, find $r > 0$ so that closed ball $B(r, x_0)$ has positive P-measure, $P(K) > 0$. Now take an M ($M > r$) large enough to satisfy the inequality

$$\varphi(M - r)P(K) > W(x_0, P), \tag{3}$$

where $W(x_0, P) = \int \varphi(d(x, x_0))P(dx) < \infty$.[1] The existence of such M is ensured by $\lim \varphi(r) = \infty$, as $r \to \infty$.

Further, let

$$\Delta_n = W(x_0, P_n) - W_1(P_n) \geqslant 0, \tag{4}$$

$$\varepsilon_n = \begin{cases} \frac{1}{n}, & \text{if } \Delta_n = 0 \text{ or } \Delta_n > \frac{1}{n}, \\ \Delta_n, & \text{if } 0 < \Delta_n \leqslant \frac{1}{n}. \end{cases} \tag{5}$$

So we have the sequence $\{\varepsilon_n\}$, satisfying $0 < \varepsilon_n \leqslant \frac{1}{n}$. For ε_n-optimality of A_{ε_n} and for (1), (2) and (5) it is easy to check that the relations

$$W(A_{\varepsilon_n}, P_n) < W_k(P_n) + \varepsilon_n < W_1(P_n) + \varepsilon_n \leqslant W(x_0, P_n) + \frac{1}{n}$$

hold.

Sequentially,

$$\limsup_n W(A_{\varepsilon_n}, P_n) \leqslant \limsup_n W(x_0, P_n) = W(x_0, P) \text{ a.s.}, \tag{6}$$

where the equality holds for the ordinary SLLN (remember that $W(x_0, P) < \infty$).

Now suppose (controversially) that for infinitely many values of n the ball $B(M, x_0)$ does not contain any point of A_{ε_n}. Then we have along some subsequence $\{n'\}$ that $d(x, a) >$

---

[1] For the sake of simplicity we use notation $W(x_0, P)$ rather than $W(\{x_0\}, P)$.

$> M-r$ for each $x \in K$ and $a \in A_{\varepsilon_{n'}}$, which implies $\min_{a \in A_{\varepsilon_{n'}}} \varphi(d(x,a)) > \varphi(M-r)$, $x \in K$ (since φ is monotone). To attain a contradiction let us write

$$\limsup_{n} W(A_{\varepsilon_n}, P_n) \geqslant \limsup_{n'} W(A_{\varepsilon_{n'}}, P_{n'}) \equiv$$
$$\equiv \limsup_{n'} \int \min_{a \in A_{\varepsilon_{n'}}} \varphi(d(x,a)) P_{n'}(dx) \geqslant \limsup_{n'} \int_K (M-r) P_{n'}(dx) =$$
$$= \varphi(M-r) P(K) \text{ a.s.}$$

by the SLLN. Together with (3) this would make $\limsup_{n} W(A_{\varepsilon_n}, P_n) > W(x_0, P)$ a.s., which contradicts (6). Without loss of generality we therefore may assume that almost surely each A_{ε_n} contains at least one point (say a_n^1) of $B(M, x_0)$.

If $k = 1$ the remaining part of 1) may be skipped, steps (2) and (3) will complete the proof of the theorem in this case. Thus we have the induction basis: theorem is valid in the case $k = 1$.

If $k > 1$, then we have to show that, for n large enough, the closed ball $B(5M, x_0)$, of radius $5M$ and centred at x_0, contains all the points of A_{ε_n}. To do this, assume that the theorem is valid for each number of centres from 1 to k-1 (inductive hypothesis). Now we need M large enough to ensure

$$\lambda \int_{\{x: d(x,x_0) \geqslant 2M\}} \varphi(d(x,x_0)) P(dx) < \frac{\varepsilon}{2}, \qquad (7)$$

where ε is chosen to satisfy

$$0 < \varepsilon < W_{k-1}(P) - W_k(P). \qquad (8)$$

According to (2) the latter discrepancy is positive as it is necessary for the choice of positive ε. Condition (7) is the second requirement placed on M.

Suppose that for some value of n the set A_{ε_n} contains at least one point outside $B(5M, x_0)$. Then it can be shown that after deleting all such points the value of clustering criteria does not increase significantly (less than $\varepsilon/2$). Indeed, the worst case is when the centre a_n^1 that is known lie in $B(M, x_0)$ accepts to its own cluster all the sample points presently assigned to cluster outside $B(5M, x_0)$. These sample points must have been a distance at least 2M from x_0, otherwise they would have been closer to the cluster centre

a_n^1 than to any centre outside $B(5M,x_o)$. The increment of $W(\cdot,P_n)$ due to such deleting would therefore at most

$$\int_{\{x:d(x,x_o)\geqslant 2M\}} \varphi(d(x,a_n^1))P_n(dx) \leqslant \int_{\{x:d(x,x_o)\geqslant 2M\}} \varphi(d(x,x_o)+d(x_o,a_n^1))P_n(dx) \leqslant$$

$$\leqslant \int_{\{x:d(x,x_o)\geqslant 2M\}} \varphi(2d(x,x_o))P_n(dx) \leqslant \lambda \int_{\{x:d(x,x_o)\geqslant 2M\}} \varphi(d(x,x_o))P_n(dx),$$

which is less than $\varepsilon/2$ by the inequality (7). Thus we have

$$W(\bar{A}_{\varepsilon_n},P_n) < W(A_{\varepsilon_n},P_n) + \frac{\varepsilon}{2}, \tag{9}$$

where $\bar{A}_{\varepsilon_n}$ consists of points of A_{ε_n} inside $B(5M,x_o)$. The set $\bar{A}_{\varepsilon_n}$ has at most k-1 points and is a candidate for minimizing $W(\cdot,P_n)$ over all sets of k-1 (or fewer) points, hence the inequality

$$W_{k-1}(P_n) \leqslant W(\bar{A}_{\varepsilon_n},P_n) \tag{10}$$

holds. Further, by our inductive hypothesis

$$\lim_n W_{k-1}(P_n) = W_{k-1}(P) \text{ a.s.} \tag{11}$$

If $A_{\varepsilon_n} \not\subset B(5M,x_o)$ along some subsequence $\{n'\}$ of values of n, the relations (9)-(11) give us

$$W_{k-1}(P) \leqslant \limsup_{n'} W(A_{\varepsilon_{n'}},P_{n'}) + \frac{\varepsilon}{2} \text{ a.s.} \tag{12}$$

Now we need, in addition to (5), $\varepsilon_{n'}$ small enough to satisfy

$$\varepsilon_{n'} < W(A_{\varepsilon/2},P_{n'}) - W_k(P_{n'}), \tag{13}$$

where $A_{\varepsilon/2}$ is $\varepsilon/2$-optimal set of k centres with respect to P, i.e.

$$W(A_{\varepsilon/2},P) < W_k(P) + \frac{\varepsilon}{2}. \tag{14}$$

For the $\varepsilon_{n'}$-optimality of $A_{\varepsilon_{n'}}$ and for (13) we have

$$W(A_{\varepsilon_{n'}},P_{n'}) < W_k(P_{n'}) + \varepsilon_{n'} < W(A_{\varepsilon/2},P_{n'}). \tag{15}$$

Relating this to (12) we get

$$W_{k-1}(P) \leqslant \limsup_{n'} W(A_{\varepsilon/2},P_{n'}) + \frac{\varepsilon}{2} = W(A_{\varepsilon/2},P) + \frac{\varepsilon}{2} \text{ a.s.}$$

by the SLLN. Together with (14) this makes $W_{k-1}(P) < W_k(P) + \varepsilon$: we reached contradiction with (8).

We know now that there exists a sequence $\varepsilon_n \to 0$ such that A_{ε_n}, the ε_n-optimal sets of k centres, will be contained in the closed ball $B(5M,x_o)$, sequentially.

2) The second step of the proof of the theorem verifies the following uniform SLLN.

Lemma 1. For arbitrary closed ball $B(R,c) \subset T$ of radius R and centred at c the uniform convergence

$$\limsup_{n} \sup_{A \in \mathcal{E}(k,R,c)} | W(A,P_n) - W(A,P) | = 0 \text{ a.s.}, \tag{16}$$

where $\mathcal{E}(k,R,c) = \{A: A \subset B(R,c), A \text{ contains } k \text{ points}\}$, takes place.

The proof of this lemma is deferred to Section 5.

3) The third step completes the proof of the theorem. Partition all the values of n into the two following disjoint and exhaustive classes:

$$\{n'\} = \{n: W_k(P) - W_k(P_n) \geqslant 0\}, \tag{17}$$

$$\{n''\} = \{n: W_k(P) - W_k(P_n) < 0\}. \tag{18}$$

If one of the classes is finite, then it suffices to look at another only.

Let $\{\varepsilon_{n'}\}$ be the zero-sequence, defined by (5) and let $A_{\varepsilon_{n'}}$ be an $\varepsilon_{n'}$-optimal set of k points with respect to the empirical measure $P_{n'}$. By the definition of the ε-optimality and $W_k(p)$ the inequality

$$W_k(P) - W_k(P_{n'}) < W(A_{\varepsilon_{n'}},P) - W(A_{\varepsilon_{n'}},P_{n'}) + \varepsilon_{n'} \tag{19}$$

holds. Since $A_{\varepsilon_{n'}} \in \mathcal{E}(k,R,c)$, for n large enough, the right side of (19) converges to zero almost surely by the Lemma 1. Together with (17) it implies that

$$\lim_{n'} W_k(P_{n'}) = W_k(P) \text{ a.s.} \tag{20}$$

To obtain a similar result for the class $\{n''\}$, we need to show that there exists a zero-sequence of positive numbers $\{\delta_{n''}\}$ and a closed ball $B(M_1,x_1)$, which sequentially contains all the sets $A_{\delta_{n''}}$, defined as $\delta_{n''}$-optimal with respect to P and consisting of k points. It may be done by repeating, with some simplifications, the first step of the theorem; we will omit this part of the proof. Further, take M so large that $B(M_1,x_1) \subset B(5M,x_0)$. This is the final requirement on M. Now, from the inequality

$$W(A_{\delta_{n''}},P) < W_k(P) + \delta_{n''} \tag{21}$$

and by the definition of $W_k(P)$ we get, for n large enough, that

$$W_k(P_{n''}) - W_k(P) < W(A_{\delta_{n''}},P_{n''}) - W(A_{\delta_{n''}},P) + \delta_{n''} < \\ < \sup_{A \in \mathcal{E}(k,5M,x_0)} | W(A,P_{n''}) - W(A,P)| + \delta_{n''}, \tag{22}$$

which converges to zero almost surely, as $n'' \to \infty$. Together with (18) it yields in

$$\lim_{n''} W_k(P_{n''}) = W_k(P) \text{ a.s.} \qquad (23)$$

So, both subsequences of $W_k(P_n)$ have the same limit $W_k(P)$ a.s. This completes the proof of the theorem.

5. Proof of the Lemma 1

For the convenience denote

$$f_A(x) = \min_{a \in A} \varphi\, d(x,a)$$

and

$$\mathcal{A} = \{ f_A(x) : A \in \mathcal{E}(k,R,c) \}.$$

Clearly, $\mathcal{A}$ is a family of continuous functions on T. Now Lemma 1 can be reformulated as

$$\lim_n \sup_{f_A \in \mathcal{A}} \left| \int_T f_A(x) P_n(dx) - \int_T f_A(x) P(dx) \right| = 0 \text{ a.s.} \qquad (24)$$

The proof of this assertion is based on the following uniform convergence theorem, given in Ranga Rao (1962) (Theorem 3.2):

Let $\mathcal{A}$ be a family of continuous functions on separable metric space T satisfying the following conditions: (i) there exists a continuous function $g(x)$ on T such that $|f(x)| \leq g(x)$ for all $f \in \mathcal{A}$ and $x \in T$; (ii) $\mathcal{A}$ is equicontinuous. Suppose that μ_n, μ is a sequence of measures on T such that (a) $\mu_n \Rightarrow \mu$ (weakly) and (b) $\int g(x)\mu_n(dx) \to \int g(x)\mu(dx)$ $(< \infty)$. Then we have

$$\lim_n \sup_{f \in \mathcal{A}} \left| \int f(x)\mu_n(dx) - \int f(x)\mu(dx) \right| = 0.$$

In our case it suffices to take $\mu_n = P_n$, $\mu = P$. The assumption (a) i.e. $P_n \Rightarrow P$ is fulfilled with probability one, as shown in Varadarajan (1958). Suitable $g(x)$ is $g(x) = \varphi(d(x,c) + R)$, which is continuous on T and

$$f_A(x) = \min_{a \in A} \varphi(d(x,a)) < \min_{a \in A} \varphi(d(x,c)+d(c,a)) \leq$$

$$\leq \varphi(d(x,c) + R) = g(x)$$

for each $A \in \mathcal{E}(k,R,c)$ and $x \in T$.

To establish the equicontinuity of $\mathcal{A}$ fix an arbitrary $x_0 \in T$ and $\varepsilon > 0$. We need to show that there exists a $\delta = \delta(x_0, \varepsilon) > 0$ such that $|f_A(x_0) - f_A(y)| < \varepsilon$ for each $A \in \mathcal{E}(k,R,c)$ and $y \in B(\delta, x_0)$. To get a suitable δ, note

that φ is uniformly continuous on the closed interval

$$I = \begin{cases} [d(x_0,c)-R;\ d(x_0,c)+R], & \text{if } x_0 \notin B(R,c) \\ [0;\ d(x_0,c)+R], & \text{if } x_0 \in B(R,c). \end{cases}$$

Thereby, for the given $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon, x_0) > 0$ (not depending on A), such that

$$|\varphi(r) - \varphi(t)| < \varepsilon \tag{25}$$

for each $r \in I$ and $t > 0$ satisfying $|r-t| < \delta$. Prove now this δ is just needed.

For given x_0 and y let $a(x_0)$ and $a(y)$ be their nearest points in A. In these notations

$$\min_{a \in A} d(x_0,a) - \min_{a \in A} d(y,a) = d(x_0,a(x_0)) - d(y,a(y)) \leqslant$$
$$\leqslant d(x_0,a(y)) - d(y,a(y)) \leqslant d(x_0,y)$$

by the triangular inequality. By the same way we also get

$$\min_{a \in A} d(y,a) - \min_{a \in A} d(x_0,a) \leqslant d(x_0,y).$$

Hence, from $y \in B(\delta, x_0)$ the inequality

$$\left| \operatorname{mind}(x_0,a) - \operatorname{mind}(y,a) \right| \leqslant \delta \tag{26}$$

follows.

Now write

$$|f_A(x_0) - f_A(y)| = \left| \min_{a \in A} \varphi(d(x_0,a)) - \min_{a \in A} \varphi(d(y,a)) \right| =$$
$$= \left| \varphi\left(\min_{a \in A} d(x_0,a)\right) - \varphi\left(\min_{a \in A} d(y,a)\right) \right|,$$

which is less than ε, for each $A \in \mathcal{E}(k,R,c)$ and $y \in B(\delta, x_0)$, by taking $r = \min_{a \in A} d(x_0,a)$ and $t = \min_{a \in A} d(y,a)$ in (25). (Note that $r \in I$ and $|r-t| \leqslant \delta$ for each $y \in B(\delta, x_0)$, as follows from (26)). Hence, the family $\mathcal{A}$ is equicontinuous.

At last, it follows from the ordinary SLLN that the assumption (b) of the theorem of Ranga Rao is fulfilled with probability one. Indeed, the random variable $g(x)$ has finite expectation:

$$\int g(x)P(dx) = \int \varphi(d(x,c) + R)P(dx) =$$
$$= \int_{\{x:d(x,c) \leqslant R\}} \varphi(d(x,c)+R)P(dx) + \int_{\{x:d(x,c) > R\}} \varphi(d(x,c)+R)P(dx) \leqslant$$
$$\leqslant \int_{\{x:d(x,c) \leqslant R\}} \varphi(2R)P(dx) + \int_{\{x:d(x,c) > R\}} \varphi(2d(x,c))P(dx) \leqslant$$
$$\leqslant \varphi(2R) \cdot P(B(R,c)) + \lambda \int \varphi(d(x,c))P(dx) < \infty.$$

This completes the proof of the Lemma 1.

6. Appendix

In this section we give a sufficient condition to ensure the inequalities (2) be satisfied.

Lemma 2. Let T be a separable metric space and P a probability measure on T, not concentrated at any k points. In addition to other conditions, formulated in Section 3, let φ be strictly increasing. Suppose that for each j = 1,2,... ...,k-1 there exists an P-optimal set A(j), containing exactly j points i.e. $W(A(j),P) = W_j(P)$. Then inequalities (2) hold.

Proof. Consider the case k=2. (The proof of the general case is analogous.) We have a probability measure P,not concentrated at any single point; and a point a = A(1),satisfying $W(a,P) = W_1(P)$. It is not difficult to see that there exist another point $b \neq a$ and a $\delta > 0$ such that open ball $B(\delta,b)$ has positive P-measure and does not intersect with $B(\delta,a)$. We now show that the pair $\{a,b\}$ is strictly better in the sense of $W(\cdot,P)$ than the single point a. Indeed, write first that

$$W(\{a,b\}, P) = \int_{\{x:d(a,x) \leq d(b,x)\}} \varphi(d(a,x))P(dx) + \int_{\{x:d(a,x) > d(b,x)\}} \varphi(d(b,x))P(dx).$$

Since the domain of integration of the last integral contains P-positive ball $B(\delta,b)$, this integral is strictly less than

$$\int_{\{x:\, d(a,x) > d(b,x)\}} \varphi(d(a,x))P(dx)$$

and we get

$$W(\{a,b\},P) < \int \varphi(d(a,x))P(dx) = W(a,P) = W_1(P).$$

Hence $W_2(P) < W_1(P)$ - the needed result.

Acknowledgement. I am indebted to professor T.Arak for stimulating discussions and helpful comments.

References

1. Hartigan, J.A. Asymptotic distributions for clustering criteria. Ann. Statist., 1978, 6, 117-131.
2. MacQueen, J.B. Some methods for classification and analysis of multivariate observations. Proc. Fifth Berceley Symp. Math. Statist. Prob. (LeCam, L.L. and Neyman, J. Eds.),N.-Y., 1967, 1, 181-196.
3. Pollard, D. Strong consistency of k-means clustering.Ann. Statist., 1981, 9, 135-140.

4. Ranga Rao, R. Relations between weak and uniform convergence with applications. Ann. Math. Statist., 1962, 33, 659-680.
5. Sverdrup-Thygeson, H. Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. Ann. Statist.,1981, 9, 141-145.
6. Varadarajan, V.S. On the convergence of probability distributions. Sankhya, 1958, 19, 23-26.

КЛАСТЕРИЗАЦИЯ МЕТОДОМ к-СРЕДНИХ В СЕПАРАБЕЛЬНОМ МЕТРИЧЕСКОМ ПРОСТРАНСТВЕ: СИЛЬНАЯ СОСТОЯТЕЛЬНОСТЬ

К. Пярна

Р е з ю м е

Пусть $\{x_1, x_2, \ldots, x_n\}$ - случайная выборка из распределения P на сепарабельном метрическом пространстве T, и пусть P_n - соответствующее эмпирическое распределение. Метод к-средних, примененный к выборке, состоит в нахождении к-элементного множества $A_n \subset T$, при котором функционал $W(A, P_n) = \frac{1}{n} \sum_{i=1}^{n} \min_{a_j \in A} \varphi(d(x_i, a_j))$ достигает минимальное (по A) значение. (Функция φ удовлетворяет некоторым условиям регулярности). Пусть $A^* \subset T$ к-элементное множество, которое минимизирует функционал $W(A,P) = \int_T \min_{a_j \in A} \varphi(d(x, a_j))\, P(dx)$.

Теорема: Пусть $\int_T \min_j \varphi(d(x, a_j))\, P(dx) < \infty$ при некотором $z \in T$. Тогда с вероятностью единица $W(A_n, P_n) \to W(A^*, P)$ при $n \to \infty$. Результат обобщает одно утверждение из [3].

Received September 1985

PROOF OF A THEOREM ON Δ-REGRESSION

T.Möls

This report contains the proof of a theorem stated in /1/. We denote by $\mathbf{X}$ the (mxm)-matrix of m i.i.d. values of a vector $\mathbf{x} = (x_1,\dots,x_m)$. The rows of $\mathbf{X}$ will be called the 'data points'. Let consider the model

$$\mathbf{i} \approx \mathbf{XA}\,, \tag{1}$$

where $\mathbf{i} = (1,\dots,1)'$, $\mathbf{A}$ is a column of constant coefficients and '$\approx$' means the approximation in some sense. If $\mathbf{A}$ is somehow estimated from (1) we get the equation of a plane

$$1 = \mathbf{xA}\,, \tag{2}$$

which approximates the functional relationship between arguments ($\mathbf{x}$ is the row-vector of arguments). In special case, where plane (2) minimizes the sum of squared deviations of data points from it assumed the deviations are measured in the direction of some vector $\Delta = (\Delta_1,\dots,\Delta_m)$, the plane (2) is called the Δ-regression plane.

The Theorem 1 facilitates the calculation of $\mathbf{A}$ given the direction Δ. According to the Theorem, at first all data points are shifted in the direction of Δ, producing the new data matrix $\mathbf{X}_\Delta = \mathbf{X} + \mathbf{i}\Delta$. After that $\mathbf{A}$ is calculated in standard way as

$$\mathbf{A}_\Delta = (\mathbf{X}_\Delta'\mathbf{X}_\Delta)^{-1}\mathbf{X}_\Delta'\mathbf{i}\,. \tag{3}$$

Clearly $\mathbf{i} \approx (\mathbf{X} + \mathbf{i}\Delta)\mathbf{A}_\Delta$ or

$$\mathbf{i} \approx \mathbf{XA}_\Delta(1 - \Delta\mathbf{A}_\Delta)^{-1}\,. \tag{4}$$

It follows from the Theorem 1, that for large Δ the vector $\mathbf{A}_\Delta(1 - \Delta\mathbf{A}_\Delta)^{-1}$ is close to the vector $\mathbf{A}$ of Δ-regression coefficients in (2).

In the proof we shall use the following notations:

$$\left.\begin{array}{ll} \mathbf{X} = (\mathbf{Y}:\mathbf{Z}) \ (\mathbf{Y} \text{ is nxf}) & \alpha = (\mathbf{Y}_\Delta'\mathbf{Y}_\Delta)^{-1} \\ \beta = \mathbf{Y}_\Delta'\mathbf{Z} & \gamma = \mathbf{Z}'\mathbf{Z} \\ \mu = \mathbf{Y}_\Delta'\mathbf{i} & \nu = \mathbf{Z}'\mathbf{i} \end{array}\right\} \tag{5}$$

Theorem 1. If $\Delta = c\delta$, $\delta' \in R^m$, $c \in R$ and $c \to \infty$, then conditionally (X held fixed) for some A_δ

$$A_\Delta(1 - \Delta A_\Delta)^{-1} \to A_\delta , \tag{6}$$

where the limiting vector A_δ determines through (2) the Δ-regression plane.

Proof. Consider first the case of $\delta = (1,0,\dots,0)$. In this case the equation $1 = xA_\Delta(1 - \Delta A_\Delta)^{-1}$ can be written as

$$x_1 = a_0(\Delta) + a_2(\Delta)x_2 + \dots + a_m(\Delta)x_m , \tag{7}$$

where

$$\left.\begin{aligned} a_0(\Delta) &= (1 - cA_{\Delta,1})/A_{\Delta,1} \\ a_i(\Delta) &= -A_{\Delta,1}/A_{\Delta,1} \quad (i = 2,\dots,m). \end{aligned}\right\} \tag{8}$$

But if $\Delta \to \infty$, the uncertainty 0/0 occurs. To see what are the limiting values in (8), we shall write

$$A_\Delta = \begin{pmatrix} X_\Delta'X_\Delta & \vdots & X_\Delta'Z \\ \cdots & \vdots & \cdots \\ Z'X_\Delta & \vdots & Z'Z \end{pmatrix}^{-1} \begin{pmatrix} X_\Delta'i \\ \cdots \\ Z'i \end{pmatrix} = \begin{pmatrix} \alpha(\mu+\beta(\gamma-\beta'\alpha\beta)^{-1}(\beta'\alpha\mu-\nu)) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ -(\gamma-\beta'\alpha\beta)^{-1}(\beta'\alpha\mu-\nu) \end{pmatrix} \tag{9}$$

(see /2/ page 46). For large c

$$\alpha \approx (1 - 2\mu_0/nc)/nc^2, \qquad \beta = \beta_0 + c\nu', \quad \mu = \mu_0 + nc , \tag{10}$$

where μ_0 and β_0 correspond to $\Delta = 0$ in (5). Using (10) in (9), after tedious calculations we get the asymptotic expansions

$$\left.\begin{aligned} A_{\Delta,1} &\approx \frac{1}{nc^2}\left(nc + \nu'\left(\gamma - \tfrac{1}{n}\nu\nu'\right)^{-1}\left(\beta' - \frac{\mu_0\nu}{n}\right) - \mu_0\right) \\ A_{\Delta,i} &\approx -\frac{1}{c}\left(\gamma - \tfrac{1}{n}\nu\nu'\right)^{-1}\left(\beta' - \frac{\mu_0\nu}{n}\right) \quad (i = 2,\dots,m). \end{aligned}\right\} \tag{11}$$

Substituting these expansions into (8) and making some further asymptotic simplifications we obtain the limits for $a_i(\Delta)$ $(i = 0,2,\dots,m)$:

$$\left.\begin{aligned} a_0(\Delta) \to a_0 &= \bar{y} - \left(\tfrac{1}{n}i'Z\right)(Z'(1-U)Z)^{-1}Z'(1-U)Y \\ (a_2(\Delta),\dots,a_m(\Delta))' \to (a_2,\dots,a_m)' &= (Z'(1-U)Z)^{-1}Z'(1-U)Y, \end{aligned}\right\} \tag{12}$$

where $\bar{y}$ is the arithmetical mean of x_1-variable and $U = ii'/n$.

It follows now that convergence in (6) is true because the expression

$$A_\Delta(1 - \Delta A_\Delta)^{-1} = -\frac{1}{a_0(\Delta)}(-1, a_2(\Delta), \ldots, a_m(\Delta))'$$

converges according to (12) if $c \to \infty$. On the other hand, a_0 and $a_2, \ldots, a_m$ in (12) are the ordinary least squares estimates for the model (7) and consequently the plane (2) with $\Delta = \Delta_\delta$ minimizes data deviations in direction of δ. This completes the proof for $\delta = (1, 0, \ldots, 0)$.

The general case where $\delta = (\delta_1, \ldots, \delta_m)$ can easily be reduced to the studied one by using the orthogonal transformation $x \to x\delta'$ in x-space. If the orthogonal matrix δ is chosen from the condition $\delta\delta' = (c, 0, \ldots, 0)$, then the data translation in direction of $\delta\delta'$ (in the new coordinate system) leads to the plane, which minimizes the residual deviations in that direction. But this is just the direction of δ in initial coordinate system. The proof is thus completed.

References

1. Мелс Т. Понятие Δ-регрессии. Тезисы конференции "Теоретические и прикладные вопросы математики", Тарту, 1985, 88-90 (II часть).
2. Рао С.Р. Линейные статистические методы и их применения. Москва, 1968.

ДОКАЗАТЕЛЬСТВО ОДНОЙ ТЕОРЕМЫ О Δ-РЕГРЕССИИ

Т. Мелс

Р е з ю м е

Понятие Δ-регрессии дано в /I/. Здесь приводится с доказательством следующая Теорема. Пусть строки (nxm)-матрицы X представляют наблюдения над m-мерным случайным вектором и $\delta' \in R^m$, $c \in R$, $\Delta = c\delta$. Обозначим $X_\Delta = X + \Delta\dot{i}$, где $\dot{i} = (1, \ldots, 1)'$ и

$$A_\Delta = (X_\Delta' X_\Delta)^{-1} X_\Delta' \dot{i} .$$

Тогда вектор $A_\Delta(1 - \Delta A_\Delta)^{-1}$ сходится при $c \to \infty$ к вектору A_δ коэффициентов Δ-регрессии. . Это значит, что сумма квадратов отклонений наблюдаемых значений случайного вектора от плоскости $xA_\delta = i$ минимальна, если отклонения измеряются в направлении вектора Δ.

Received December 1985

INVALIDITY OF BOOTSTRAP IN FINITE POPULATIONS?

M. Unt

Bootstrap-method /1/ has become very popular and many practising statisticians have mastered it. In theoretical works the bootstrap has been considered for the case of infinite populations. But in practice we often have to deal with finite populations.

Hence, applying the bootstrap in this case the problem of accuracy arises. In this note we try to investigate how the bootstrap works if all we know about the population of interest is the size s of the population G(s) with unknown probability distribution F_G, and a small sample $X(n) \sim F_G$.

A natural idea how to apply the bootstrap in such situation is following. From the sample X(n) we generate a bootstrap population BG(s) with size s. When generating bootstrap sample BX(n) from BG(s) we remind that because of finiteness of G(s) the random selection must be subsampling not resampling.

Let us denote by T_G the parameter of the population G(s) and T_X is its sample estimate. The corresponding variables T_{BG}^j and T_{BX}^j denote the statistics of the jth bootstrap population and bootstrap sample respectively (j= 1,...,m assumed bootstrap populations are generated m times).

The most interesting object of investigation is the difference $D2 = T_X - T_G$. The bootstrap approximation to D2 is $D4 = T_{BX} - T_{BG}$.

To get idea about the quality of this approximation, a simulation experiment was run. The populations G(s) were generated of exponential distribution Exp(1) for s= 7,10,15, 18,20,25,30, the sample size remaining n = 5. We investigated the behaviour of coefficient of variation defined by formula

$$T_X = \sqrt{(\Sigma x_i^2 - (\Sigma x_i)^2/n)/(n-1)} \; / (\Sigma x_i/n).$$

The number of independent populations G(s) for every s was 40, from each G(s) was drawn a sample X and on its basis m = = 100 bootstrap populations BG(s) were generated. From each BG(s) a bootstrap sample BX was drawn, and so we obtained

40×100 pairs of differences D4 and D3 = $T_{BX} - T_X$. Differences D2 and D1 = $T_X - T_F$, where T_F=1 is the coefficient of variation of Exp(1), were evaluated in another simulation experiment (number of simulations was about 4000).

The results of the experiments are presented in Table 1. For each difference 90%-confidence interval is given.

Table 1

s	D2	D3	D4
7	-.0407, -.0201	-.1082, -.0925	-.0239, -.0138
10	-.0672, -.0487	-.1334, -.1180	-.0372, -.0289
15	-.0891, -.0689	-.1269, -.1157	-.0413, -.0306
18	-.0951, -.0747	-.1352, -.1200	-.0498, -.0353
20	-.0904, -.0699	-.1350. -.1238	-.0532, -.0425
25	-.1096, -.0884	-.1336, -.1182	-.0504, -.0431
30	-.1189, -.0886	-.1404, -.1231	-.0518, -.0351

D1 does not depend on the value of s, its estimated confidence interval is (-.1436, -.1291).

From the table we may do two inferences. The first: the standard bootstrap performs well (compare D3 with D1!). The second is that D4 does not mimic D2. The reason is that if $s \to \infty$, T_{BG} does not converge to T_X. Note, that if we should define T_X differently, exchanging n-1 in the denominator by n, the convergence would take place, and D2 were in good agreement with D4. But the conditions for good agreement between D2 and D4 may be more complicated in other situations.

References

1. Efron, B. Bootstrap Methods: Another look at the Jackknife. Ann. Statist. 1979, 7, 1-26.

НЕПРИМЕНИМОСТЬ МЕТОДА "БУТСТРЭП" ПРИ КОНЕЧНЫХ ГЕНЕРАЛЬНЫХ СОВОКУПНОСТЯХ?

М. Унт

Р е з ю м е

В статье изучается возможность применения метода "бутстрэп" в практически важном случае, когда генеральная совокупность конечная. Экспериментальное исследование метода "бутстрэп" при определении смещения оценки коэффициента вариации экспоненциального распределения показало его неприменимость в данном случае. В тех же экспериментах классический вариант "бутстрэпа" работал хорошо.

Received December 1985

CONTENTS

Ученые записки Тартуского государственного университета.
Выпуск 733.
ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ.
Труды по математике и механике.
На английском языке.
Резюме на русском языке.
Тартуский государственный университет.
ЭССР, 202400, г.Тарту, ул.Юликооли, 18.
Ответственный редактор Т. Мельс.
Подписано к печати 21.02.1986.
МВ 01676.
Формат 60x90/16.
Бумага писчая.
Машинопись. Ротапринт.
Учетно-издательских листов 6,02. Печатных листов 6,5.
Тираж 400.
Заказ № 210.
Цена 90 коп.
Типография ТГУ, ЭССР, 202400, г.Тарту, ул.Пялсона, 14.