TARTU RIIKLIKU ÜLIKOOLI
TOIMETISED

УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS

685

STATISTICAL MODELLING AND MULTIVARIATE ANALYSIS

Matemaatika- ja mehhaanikaalaseid töid
Труды по математике и механике

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS
ALUSTATUD 1893.a. VIHIK 685 ВЫПУСК ОСНОВАНЫ В 1893.г.

STATISTICAL MODELLING AND MULTIVARIATE ANALYSIS

Matemaatika- ja mehhaanikaalaseid töid
Труды по математике и механике

ТАРТУ 1984

Ученые записки Тартуского государственного университета.
Выпуск 685.
СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И МНОГОМЕРНЫЙ АНАЛИЗ.
Труды по математике и механике.
На английском языке.
Резюме на русском языке.
Тартуский государственный университет.
ЭССР, 202400, г.Тарту, ул.Юликооли, 18.
Ответственный редактор Э. Тийт.
Корректор М. Тамм.
Подписано к печати 25.07.1984.
МВ 08846.
Формат 60х90/16.
Бумага писчая.
Машинопись. Ротапринт.
Учетно-издательских листов 4,29.
Печатных листов 5,5.
Тираж 400.
Заказ № 842.
Цена 65 коп.
Типография ТГУ, ЭССР, 202400, г.Тарту, ул.Пялсона, 14.
2 - 2

ASYMPTOTIC DISTRIBUTIONS OF EIGENPROJECTORS OF COVARIANCE AND CORRELATION MATRICES FOR TESTING HYPOTHESES

T. Kollo

In this article the asymptotic distributions of the eigenprojectors of sample covariance and correlation matrices have been derived. Obtained results extend Tyler's results [1] to the correlation matrix case and enlarge the admissible class of distributions of population compared with [1]. The results have been used for testing statistical hypotheses.

1. Set-up of a problem. Preliminaries

Let M be symmetric $p\times p$-matrix. In the following treatment M will be considered as a covariance matrix Σ or correlation matrix P of population p-vector X. We denote eigenvalues of M by $\mu_1 \geqslant \ldots \geqslant \mu_p$. Let A be real $p\times r$-matrix with rank r. Following [1], let us assume that $\mu_{i-1} \neq \mu_i$; $\mu_{i+m-1} = \mu_{i+m}$. Consider the following null hypothesis

H_o: for $r \leqslant m$ the columns of A lie in the subspace generated by the set of eigenvectors of M associated with the eigenvalues $\mu_i, \ldots, \mu_{i+m-1}$.

The eigenspace of M associated with μ is

$$V(\mu) = \{ x: Mx = \mu x,\ x \in R^p \}.$$

The dimension of $V(\mu)$ is the multiplicity of μ. If λ and μ are two distinct eigenvalues of M, $V(\lambda)$ and $V(\mu)$ are orthogonal subspaces of R^p. Let $\lambda_1 > \ldots > \lambda_k$ be distinct eigenvalues of M with multiplicities $m(\lambda_1), \ldots, m(\lambda_k)$. Then

$$R^p = \sum_{i=1}^{k} V(\lambda_i) ,$$

from where for every $x \in R^p$ we get

$$x = \sum_{i=1}^{k} x_i ,$$

where $x_i \in V(\lambda_i)$. The eigenprojector of M associated with λ_i, denoted P_{λ_i}, is the projection operator onto $V(\lambda)$ with respect to the decomposition of R^p:

$$P_{\lambda_i} : \; R^p \rightarrow V(\lambda_i) ,$$

that is for every $x \in R^p$

$$P_{\lambda_i} x = x_i .$$

If v is any subset of $\{\lambda_1, \dots, \lambda_k\}$, then the eigenprojector P_v of matrix M associated with eigenvalues $\lambda_i \in v$ has the form

$$P_v = \sum_{\lambda_i \in v} P_{\lambda_i} .$$

Using eigenvectors $x_i \in V(\lambda_j)$, eigenprojector P_{λ_j} has the representation

$$P_{\lambda_j} = \sum_{i=1}^{m_j} x_i x_i' .$$

The spectral decomposition of M is

$$M = \sum_{i=1}^{k} \lambda_i P_{\lambda_i} .$$

Matrix M^+ is called the Moore-Penrose generalized inverse matrix of M, if

$$M^+ = \sum_{\substack{i=1 \\ \lambda_i \neq 0}}^{k} \frac{1}{\lambda_i} P_{\lambda_i} .$$

Let us denote

$$w = \{ \lambda_k : \; \lambda_k = \mu_l , \quad i \leq l \leq i+m-1 \} .$$

The null hypothesis can thus be rephrased as

$$H_o : \; P_w A = A . \qquad (1)$$

A more detailed review of spectral theory can be found in

Kato [3].

In the further treatment we need some special notations and notions of matrix calculus. Let A_d be a diagonal $p \times p$-matrix, obtained from $p \times p$-matrix A. If A is a $p \times q$-matrix and B is a $r \times s$-matrix, then the Kronecker product of A and B is a $pr \times qs$-block-matrix

$$A \otimes B = [a_{ij}B] \quad (i = 1,\dots,p;\ j = 1,\dots,q) .$$

The main properties of the Kronecker product can be found in Lancaster [5]. For $p \times q$-matrix $A = (a_{ij})$

$$\text{vec } A = (a_{11},\dots,a_{p1},a_{12},\dots,a_{p2},\dots,a_{1q},\dots,a_{pq})' .$$

In addition to the properties of the Kronecker product presented in [5], we need some more connections:

$$\text{vec } (ABC) = (C' \otimes A) \text{ vec } B ; \tag{2}$$

$$B \otimes A = I_{p,r}(A \otimes B)\, I_{s,q} . \tag{3}$$

The permuted identity matrix $I_{m,n}$ is a $mn \times mn$-matrix, consisting of $m \times n$-blocks, where the ji-th element is 1 and the other elements are equal to zero in the ij-th block. Connections (2) and (3) are proved in [4], for example. We denote $p \times p$-identity matrix by I_p.

If the elements of $r \times s$-matrix Y depend on the elements of $p \times q$-matrix X, matrix derivative $\frac{dY}{dX}$ is a $pq \times rs$-matrix, which we define by equality

$$\frac{dY}{dX} = \frac{d}{d \text{ vec } X} \otimes (\text{vec } Y)' ,$$

where there are partial derivatives of the coordinates of vec Y with respect to the i-th coordinate of vec X in the i-th row, and the partial derivatives of the j-th coordinate of vec Y with respect to the coordinates of vec X in the j-th column. Let us present the main properties of the matrix derivatives, assuming that the dimensions of matrices are in accordance with the used operations.

(i) $\frac{dX}{dX} = I_{p^2}$.

(ii) $\frac{dX'}{dX} = I_{p,q}$.

(iii) $\frac{dZ}{dX} = \frac{dY}{dX}\frac{dZ}{dY}$, where Z depends on Y and Y on X.

(iv) $\frac{dZ}{dX} = \frac{dY}{dX}(B \otimes A')$, if $Z = AYB$, where A,B are constant matrices.

(v) $\frac{d(ZY)}{dX} = \frac{dZ}{dX}(Y \otimes I) + \frac{dY}{dX}(I \otimes Z')$, where Z and Y depend on X.

(vi) $\frac{dX_d}{dX} = (I_{p,p})_d$.

2. Asymptotic distributions of eigenprojectors

We shall denote convergence in distribution by $\xrightarrow{\mathcal{D}}$ and convergence in probability by $\xrightarrow{P}$. Let us take the two following results as a basis.

Theorem 1 (Anderson [2], p. 108). Let $\{X_n\}$ be the sequence of random p-vectors X_i, for which

$$\sqrt{n}\,(X_n - \mu) \xrightarrow{\mathcal{D}} N(0,T)$$

when $n \to \infty$, if

$$X_n \xrightarrow{P} \mu .$$

Let g(x) be function from R^p to R^q. Then

$$\sqrt{n}\,(g(X_n) - g(\mu)) \xrightarrow{\mathcal{D}} N(0, \xi' T \xi),$$

where $p \times q$ - matrix ξ

$$\xi = \left.\frac{dg(x)}{dx}\right|_{x=\mu},$$

if g(x) has continuous partial derivatives according to all coordinates x_i at the neighbourhood of the point $x = \mu$.

Theorem 2 (Parring [6]). Let $X_1,\dots,X_n$ be the sample of size n, $EX_i = \mu$, $DX_i = \Sigma$,

$$\bar{M}_4(X_i) = E\left[(X_i - \mu)\otimes(X_i-\mu)'\otimes(X_i-\mu)\otimes(X_i-\mu)'\right]$$

Then for the sample covariance matrix

$$S(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})',$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i ,$$

the convergence

$$\sqrt{n}\ \mathrm{vec}\,(S(n) - \Sigma) \xrightarrow{\mathcal{D}} N(0, \varkappa)$$

takes place if $n \to \infty$, where

$$\varkappa = M_4(X_i) - \mathrm{vec}\Sigma(\mathrm{vec}\Sigma)'. \qquad (4)$$

Let $M(n)$ be unbiased estimate of the matrix M. We consider the cases, when $M(n)$ is the sample covariance matrix $S(n)$ or the sample correlation matrix

$$R(n) = S_d^{-\frac{1}{2}} S' S_d^{-\frac{1}{2}}. \qquad (5)$$

From the Theorems 1 and 2

$$\sqrt{n}\ \mathrm{vec}\,(M(n) - M) \xrightarrow{\mathcal{D}} N(0, \Sigma_M),$$

where Σ_M depends on the form of M.

Let us denote eigenvalues of $M(n)$ by l_i and eigenprojector of $M(n)$, associated with subset of eigenvalues $\{l_i, \ldots, l_{i+m-1}\}$ of $M(n)$, by $\tilde{P}_w$.

According to the theory of perturbations of linear operators (Kato [3]), the perturbations of the operator cause the perturbations of eigenvalues and eigenprojectors.

If for $M(n)$ expansion

$$M(n) = M + \varepsilon M_1 + \varepsilon^2 M_2 + \ldots$$

takes place, accordingly [2] eigenprojector $\tilde{P}_w$ can be presented in the form of an expansion of powers of ε. Asymptotic distribution for $\tilde{P}_w$ depends on the first two items only (including ε in the zeroth and first power). Let us denote this part of the expansion of $\tilde{P}_w$ by $\hat{P}_w$. According to Tyler [1]

$$\hat{P}_w = P_w - \sum_{\lambda_i \in w} \left[P_{\lambda_i} \varepsilon M_1 (M - \lambda_i I_p)^+ - (M - \lambda_i I_p)^+ \varepsilon M_1 P_{\lambda_i}\right].$$

For the covariance matrix we use the expansion

$$S(n) = \Sigma + \frac{1}{\sqrt{n}}(\sqrt{n}(S(n) - \Sigma)) = \Sigma + (S - \Sigma). \qquad (6)$$

Then

$$\hat{P}_w = P_w - \sum_{\lambda_i \in w} \left[P_{\lambda_i}(S(n) - \Sigma)(\Sigma - \lambda_i I_p)^+\right.$$

$$+ (\Sigma - \lambda_i I_p)^+(S(n) - \Sigma)P_{\lambda_i}]. \qquad (7)$$

Let us take the sequence $\{\operatorname{vec} S(n)\}$ in the role of sequence $\{X_n\}$ in Theorem 1. Then

$$\sqrt{n}\ \operatorname{vec}(\hat{P}_w - P_w) \xrightarrow{\mathcal{L}} N(0, \xi_w' \mathfrak{X} \xi_w),$$

where

$$\xi_w = \left.\frac{d\hat{P}_w}{dS(n)}\right|_{S(n)=\Sigma}.$$

Let us find the derivative

$$\frac{d\hat{P}_w}{dS(n)} = -\frac{d}{dS(n)}\left\{\sum_{\lambda_i \in w}\left[P_{\lambda_i}(S(n) - \Sigma)(\Sigma - \lambda_i I_p)^+ \right.\right.$$
$$\left.\left. + (\Sigma - \lambda_i I_p)^+(S(n) - \Sigma)P_{\lambda_i}\right]\right\}.$$

Using property (iv) of the matrix derivative, we obtain

$$\frac{d\hat{P}_w}{dS(n)} = -\sum_{\lambda_i \in w}\left[((\Sigma - \lambda_i I_p)^+ \otimes P_{\lambda_i}) + (P_{\lambda_i} \otimes (\Sigma - \lambda_i I_p)^+)\right].$$

Matrix $(\Sigma - \lambda_i I_p)^+$ can be presented in the following form

$$(\Sigma - \lambda_i I_p)^+ = \sum_{\substack{j \\ j \neq i}} \frac{1}{\lambda_j - \lambda_i} P_{\lambda_i}.$$

Then

$$\frac{d\hat{P}_w}{dS(n)} = -\sum_{\lambda_i \in w}\sum_{\substack{j \\ j \neq i}} \frac{1}{\lambda_j - \lambda_i}\left[(P_i \otimes P_j) + (P_j \otimes P_i)\right].$$

After reducing it we get

$$\frac{d\hat{P}_w}{dS(n)} = \sum_{\lambda_i \in w}\sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_i - \lambda_j}\left[(P_i \otimes P_j) + (P_j \otimes P_i)\right]. \qquad (8)$$

Consequently we have proved the following theorem.

Theorem 3. Let $X_1, \ldots, X_n$ be the sample of size n; $EX_i = \mu$, $DX_i = \Sigma$, $M_4(X_i) < \infty$. Then for the estimation of the eigenprojector P_w associated with roots $\lambda_i \in w$ of covariance matrix Σ the convergence

$$\sqrt{n}\ \mathrm{vec}\,(\tilde{P}_w - P_w) \xrightarrow{\mathcal{D}} N(0, \xi_w' \mathcal{X} \xi_w)$$

takes place if $n \to \infty$, where $\mathcal{X}$ is determined by equality (4) and ξ_w by (8), but $\tilde{P}_w$ is the eigenprojector of $S(n)$, associated with the subset $\{l_i, \ldots, l_{i+m-1}\}$ of roots l_i of $S(n)$.

The derivation of the asymptotic distribution of eigenprojector of correlation matrix is analogous. Let now P_w be eigenprojector of the population correlation matrix P, associated with the subset $\{\lambda_i, \ldots, \lambda_{i+m-1}\}$ of roots λ_i of P. Placing expansion (6) into the equation (5) we get

$$R(n) = P + \Sigma_d^{-\frac{1}{2}}(S(n) - \Sigma)\Sigma_d^{-\frac{1}{2}} - \tfrac{1}{2}\Big[P\Sigma_d^{-1}(S(n) - \Sigma)_d + (S(n) - \Sigma)_d\,\Sigma_d^{-1} P\Big] + \mathcal{O}\big(\tfrac{1}{\sqrt{n}}\big). \tag{9}$$

Let us denote

$$U(n) = \Sigma_d^{-\frac{1}{2}}(S(n) - \Sigma)\Sigma_d^{-\frac{1}{2}} - \tfrac{1}{2}\Big[P\,\Sigma_d^{-1}(S(n) - \Sigma)_d + (S(n) - \Sigma)_d\Sigma_d^{-1} P\Big].$$

By Kato [3] eigenprojector $\tilde{P}_w$ of $R(n)$ can be represented in the form

$$\tilde{P}_w = P_w - \sum_{\lambda_i \in w}\Big[P_{\lambda_i} U(n)(P - \lambda_i I_p)^+ + (P - \lambda_i I_p)^+ U(n) P_{\lambda_i}\Big] + \mathcal{O}\big(\tfrac{1}{\sqrt{n}}\big).$$

Asymptotic distribution of $\tilde{P}_w$ does not depend on $\mathcal{O}(\frac{1}{\sqrt{n}})$. We use the expansion

$$\hat{P}_w = P_w - \sum_{\lambda_i \in w}\Big[P_{\lambda_i} U(n)(P - \lambda_i I_p)^+ + (P - \lambda_i I_p)^+ U(n) P_{\lambda_i}\Big].$$

As in the case of covariance matrix, we get convergence

$$\sqrt{n}\ \mathrm{vec}\,(\tilde{P}_w - P_w) \xrightarrow{\mathcal{D}} N(0, \eta_w' \mathcal{X}\, \eta_w)$$

from Theorem 1 if $n \to \infty$, where

$$\eta_w = \left.\frac{d\hat{P}_w}{dS(n)}\right|_{S(n)=\Sigma} .$$

Finding of the matrix derivative η_w is analogous to the deducing of ξ_w.

$$\frac{d\hat{P}_w}{dS(n)} = -\frac{d}{dS(n)}\Big[\sum_{\lambda_i \in w} P_{\lambda_i} U(n)(P-\lambda_i I_p)^+ + (P-\lambda_i I_p)^+ U(n) P_{\lambda_i}\Big]$$

$$= -\frac{dU(n)}{dS(n)} \sum_{\lambda_i \in w} \Big[((P-\lambda_i I_p)^+ \otimes P_{\lambda_i}) + (P_{\lambda_i} \otimes (P-\lambda_i I_p)^+)\Big]$$

$$= -\frac{dU(n)}{dS(n)} \sum_{\lambda_i \in w} \sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_j - \lambda_i}\Big[(P_{\lambda_j} \otimes P_{\lambda_i}) + (P_{\lambda_i} \otimes P_{\lambda_j})\Big].$$

Using properties (iv) and (vi) of matrix derivative, we get

$$\frac{dU(n)}{dS(n)} = (\Sigma_d^{-\frac{1}{2}} \otimes \Sigma_d^{-\frac{1}{2}}) - \tfrac{1}{2}(I_{p,p})_d \Big[(I_p \otimes \Sigma_d^{-1} P) + (\Sigma_d^{-1} P \otimes I_p)\Big].$$

and

$$\frac{d\hat{P}_w}{dS(n)} = (\Sigma_d^{-\frac{1}{2}} \otimes \Sigma_d^{-\frac{1}{2}}) - \tfrac{1}{2}(I_{p,p})_d \Big[(I_p \otimes \Sigma_d^{-1} P) + (\Sigma_d^{-1} P \otimes I_p)\Big] \sum_{\lambda_i \in w} \sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_i - \lambda_j}\Big[(P_{\lambda_j} \otimes P_{\lambda_i}) + (P_{\lambda_i} \otimes P_{\lambda_j})\Big]. \quad (10)$$

Consequently we have proved the following theorem.

Theorem 4. Let $X_1, \ldots, X_n$ be the sample of size n; $EX_i = \mu$, $DX_i = \Sigma$, $M_4(X_i) < \infty$. Then for the estimate of eigenprojector P_w of the population correlation matrix P, associated with roots $\lambda_i \in w$ of P convergence

$$\sqrt{n}\,\mathrm{vec}\,(\widetilde{P}_w - P_w) \xrightarrow{\mathcal{D}} N(0, \eta_w' \varkappa \eta_w)$$

takes place if $n \to \infty$, where $\varkappa$ is determined by equality (4), η_w by (10), and P_w is the eigenprojector of $R(n)$, associated with the subset $\{l_i, \ldots, l_{i+m-1}\}$ of roots l_i of $R(n)$.

3. Testing null hypothesis

Consider the following statistic for testing H_0:

$$T_w = A - \widetilde{P}_w A = (I_p - \widetilde{P}_w)\,A\,.$$

Following Theorem 1

$$\sqrt{n}\,\mathrm{vec}\left[(I_p - \widetilde{P}_w)\,A - (I_p - P_w)\,A\right] \xrightarrow{\mathcal{D}} N(0, \xi_w' \varkappa \xi_w),$$

if $n \to \infty$, and

$$\xi_w = \left.\frac{d(I_p - \widetilde{P}_w)A}{dS(n)}\right|_{S(n)=\Sigma}.$$

Using property (iv) of matrix derivative, we get

$$\xi_w = \left.\frac{d(I_p - \widetilde{P}_w)}{dS(n)}\right|_{S(n)=\Sigma} \times (A \otimes I_p)$$

$$= -\left.\frac{d\widetilde{P}_w}{dS(n)}\right|_{S(n)=\Sigma} \cdot (A \otimes I_p)\,.$$

By means of (8) we obtain the concrete form for ξ_w, when $\widetilde{P}_w$ is an eigenprojector of $S(n)$:

$$\xi_w = \sum_{\lambda_i \in w}\ \sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_j - \lambda_i}\left[(P_{\lambda_j} \otimes P_{\lambda_i}) + (P_{\lambda_i} \otimes P_{\lambda_j})\right](A \otimes I_p)$$

$$= \sum_{\lambda_i \in w}\ \sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_j - \lambda_i}\left[(P_{\lambda_j} A \otimes P_{\lambda_i}) + (P_{\lambda_i} A \otimes P_{\lambda_j})\right].$$

Suppose that H_0 is valid, then

$$P_{\lambda_j} A = 0$$

for all $\lambda_j \bar{\in} w$. From here

$$\xi_w = \sum_{\lambda_i \in w} \sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_j - \lambda_i} (P_i A \otimes P_j) \tag{11}$$

Analogously, if $\tilde{P}_w$ is eigenprojector of R(n), from (10) we obtain

$$\xi_w = (\Sigma_d^{-\frac{1}{2}} \otimes \Sigma_d^{-\frac{1}{2}}) - \tfrac{1}{2}(I_{p,p})_d \left[I_p \otimes \Sigma^{-1} P) \right.$$

$$\left. + (\Sigma^{-1} P \otimes I_p) \right] \times \sum_{\lambda_i \in w} \sum_{\lambda_j \bar{\in} w} \frac{1}{\lambda_j - \lambda_i} (P_{\lambda_i} A P_{\lambda_j}). \tag{12}$$

For testing H_0 Tyler [1] introduced statistic

$$T_n(A) = n \left\{ \left[\mathrm{vec}\,(I_p - \tilde{P}_w)A \right]' \hat{\Sigma}_w^-(A)\, \mathrm{vec}\,(I_p - \tilde{P}_w)A \right\},$$

where $\hat{\Sigma}_w(A)$ is the sample estimate of the covariance matrix $\xi_w' \varkappa \xi_w$ of the limit distribution, and $\hat{\Sigma}_w^-(A)$ is the generalized inverse for $\hat{\Sigma}_w(A)$. In [1] it is shown that $T_n(A)$ does not depend upon the choice of the generalized inverse for $\hat{\Sigma}_w(A)$. The statistic $T_n(A)$ is asymptotically invariant under postmultiplication of A by a nonsingular $r \times r$ - matrix.

This attests the suitable choice of $T_n(A)$, because the hypothesis H_0 is also invariant under postmultiplication of A by a nonsingular $r \times r$ - matrix. It is possible to present $T_n(A)$ in the following form (see Tyler [1]):

$$T_n(A) = n\,(\mathrm{vec}\,A)'\, \hat{\Sigma}_w^+(A)\, \mathrm{vec}\,A. \tag{13}$$

Testing of hypothesis is carried out by means of asymptotic χ^2-distribution: assuming $P_w A = A$

$$\sqrt{n}\, T_n(A) \xrightarrow{\mathcal{D}} \chi^2_{r(p-m)}, \tag{14}$$

if $n \to \infty$.

All Tyler's results [1] concerning properties of $T_n(A)$ are also valid in our assumptions.

References

1. Tyler, D.E., Asymptotic influence for eigenvectors. The Annals of Statistics, 1981, 9, № 4, 725-736.
2. Андерсон Т., Введение в многомерный статистический анализ. М., 1963.
3. Като Т., Теория возмущений линейных операторов. М., 1972.

4. Колло Т., Киннар Т., Матричная производная с применением для блок-матриц. Труды ВЦ ТГУ, Тарту, вып. 51 (в печати).
5. Ланкастер П., Теория матриц. М., 1978.
6. Парринг А.-М., Вычисление асимптотических характеристик функций выборки. Уч. зап. Тартуск. ун-та, 1979, 492, 86-90.

АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ СОБСТВЕННЫХ ПРОЕКТОРОВ КОВАРИАЦИОННОЙ И КОРРЕЛЯЦИОННОЙ МАТРИЦ ДЛЯ ПРОВЕРКИ ГИПОТЕЗ

Т.Колло

Р е з ю м е

В статье выведены предельные распределения для собственных проекторов выборочной ковариационной матрицы (теорема 3) и выборочной корреляционной матрицы (теорема 4). Теорема 3 обобщает результаты [6] на случай более широкого класса распределения генеральной совокупности, предельное распределение собственных проекторов корреляционной матрицы в литературе не встречалось. Исходя из асимптотического нормального распределения собственных проекторов построена статистика $T_k(A)$ (равенство (13)), имеющая асимптотическое χ^2- распределение согласно (14). Статистика $T_k(A)$ введена для проверки нулевой гипотезы H_0 о том, содержит ли подпространство, построенное на собственные векторы, определяющие рассматриваемый проектор, заранее фиксированное подпространство p-мерного пространства.

Received May 84

ASYMPTOTIC NORMAL DISTRIBUTION OF THE SAMPLE ROOTS FOR A NONNORMAL POPULATION

I. Traat

Summary

For the latent roots of the sample covariance matrix the multivariate normal distribution is derived, which tends to the limiting distribution of these roots. Some special cases are considered.

1. Introduction

Let x be a p-dimensional random vector. Let Σ be the population covariance matrix with latent roots $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p > 0$. Suppose that $x_1, \ldots, x_N$ is a sample of $N = n+1$ independent observations of x. Then the usual unbiased estimate of Σ is

$$S = n^{-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})',$$

where $\bar{x}$ is the sample mean vector. S has the latent roots $l_1 \geqslant l_2 \geqslant \ldots \geqslant l_p$ which estimate the corresponding population roots λ_r, $r = 1, \ldots, p$.

The asymptotic distribution of the sample roots l_r has been studied by many authors. It has been obtained that the limiting joint distribution of the variates

$$z_r = \sqrt{n}\,(l_r - \lambda_r)$$

is normal with mean 0 and covariance matrix $\Omega = (\omega_{qr})$, where the expression of ω_{qr} depends on the parent population. For normal population with simple roots Girshick (1939) got

$$\omega_{qr} = \begin{cases} 2\lambda_r^2 & \text{if } q = r \\ 0 & \text{if } q \neq r. \end{cases}$$

Anderson (1963) considered λ_r with multiplicity p_r, for which he got the maximum likelihood estimate $\bar{l}_r$ as the arithmetical mean of p_r sample roots corresponding to λ_r and the asymptotical variance of $\bar{z}_r = \sqrt{n}(\bar{l}_r - \lambda_r)$ as

$$\omega_{rr} = 2\lambda_r^2 / p_r .$$

Waternaux (1976) got for nonnormal populations with finite fourth cumulants and simple roots

$$\omega_{qr} = \begin{cases} 2\lambda_r^2 + k_4^r & \text{if } q = r \\ k_{22}^{qr} & \text{if } q \neq r . \end{cases}$$

Kollo (1977) derived similar results using matrix technique.

The Edgeworth expansions of the marginal and joint distribution functions of the variates z_r have also been obtained. For nonnormal population Fujikoshi (1980) got the expansions for the case when the population roots are simple. For multiple roots Fang and Krishnaiah (1982) derived very general results, which include as special cases the Edgeworth expansions for the distribution functions of $\bar{z}_r$.

The purpose of the present paper is to find the multivariate normal distribution with parameters estimating the mean vector and covariance matrix of z_r in the case of finite n and converging to the limiting joint distribution of z_r as $n \to \infty$. It is assumed that parent distribution has finite cumulants up to order 6 and the population covariance matrix has simple roots.

2. Preliminaries

As the latent roots do not depend on the orthogonal transformations of x, then for the study of the distribution of the sample roots, it is assumed without loss of generality that the population covariance matrix is diagonal $\Sigma = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$. If the population root λ_r is simple, then for z_r we may use expansion got by Lawley (1956)

$$z_r = v_{rr} + n^{-1/2} \sum_{i \neq r} \lambda_{ri} v_{ri}^2 + n^{-1} \Big[\sum_{i \neq r} \lambda_{ri}^2 v_{ii} (v_{ri})^2 - \sum_{i \neq r} \lambda_{ri}^2 v_{rr} (v_{ri})^2 + \sum_{i \neq j \neq r} \sum \lambda_{ri} \lambda_{rj} v_{ri} v_{ij} v_{jr} \Big] + O_p(n^{-3/2}), \qquad (1)$$

where $V = (v_{ij}) = \sqrt{n}(S - \Lambda)$, $\lambda_{ri} = (\lambda_r - \lambda_i)^{-1}$.

The formulae for Ez_r, var z_r and $\operatorname{cov}(z_q, z_r)$ are derived from the expansion (1) with the help of moments of v_{qr}. The expressions of these moments (5) - (7) in terms of cumulants of the parent population are given by Cook (1951) and Kaplan (1952).

We shall denote the cumulants of $x = (x_1, \dots, x_p)'$ by

$$k_{r\dots s}(x_i, \dots, x_j) = k^{i\dots j}_{r\dots s}$$

or $k_{i\dots i\dots j\dots j}$, where i and j are repeated respectively r and s times. As the covariance matrix of x is diagonal then

$$\begin{cases} k^r_2 = \lambda_r , \\ k^{qr}_{11} = 0 . \end{cases} \tag{2}$$

If we use the notations

$$m(ij,kl) = k_{ijkl} + k_{ik}k_{jl} + k_{il}k_{jk} , \tag{3}$$

$$m(ij,kl,st) = k_{ijklst} + \sum^{12} k_{ijks}k_{lt} + \sum^{4} k_{iks}k_{jlt} + \sum^{8} k_{ik}k_{js}k_{lt} , \tag{4}$$

where the summations occur over all ways of grouping the subscripts, then

$$E(v_{ij}) = 0 ,$$

$$E(v_{ij}v_{kl}) = m(ij,kl) - n^{-1}k_{ijkl} + O(n^{-2}) , \tag{5}$$

$$E(v_{ij}v_{kl}v_{st}) = n^{-1/2}m(ij,kl,st) + O(n^{-3/2}) , \tag{6}$$

$$E(v_{ij}v_{kl}v_{st}v_{qr}) = m(ij,kl)m(st,qr) + m(ij,st)m(kl,qr) + m(ij,qr)m(kl,st) + O(n^{-1}) . \tag{7}$$

3. Parameters of the normal distribution

Let for finite n $\mu^n = (\mu^n_1, \dots, \mu^n_p)'$ be the mean vector and $\Omega^n = (\omega^n_{qr})$ the covariance matrix of $z_1, \dots, z_p$. Then

$$N(\mu^n, \Omega^n) \to N(0, \Omega) , \quad n \to \infty ,$$

where $N(0, \Omega)$ is the limiting distribution of the variates z_r. The expectation $E\, z_r = \mu^n_r$ with accuracy $n^{-1/2}$ has been received by Waternaux (1976)

$$\mu^n_r = n^{-1/2} \sum_{i \neq r} \lambda_{ri}(k^{ir}_{22} + \lambda_i \lambda_r) + O(n^{-3/2}) . \tag{8}$$

The expression of $\operatorname{var} z_r = \omega^n_{rr}$ with accuracy n^{-1} has been received by Fujikoshi (1980)

$$\omega^n_{rr} = k^r_4 + 2\lambda^2_r + 2n^{-1}b_1 + O(n^{-2}), \tag{9}$$

where

$$b_1 = -\tfrac{1}{2}k^r_4 + \sum_{i\neq r}\lambda_{ri}\Big[k^{ri}_{42} + \lambda_i k^r_4 + 5\lambda_r k^{ri}_{22} + 2k^r_3 k^{ri}_{12} + 2(k^{ri}_{21})^2 + 2\lambda^2_r\lambda_i\Big] - \sum_{i\neq r}\lambda^2_{ri}\Big[(k^{ri}_{22} + \lambda_r\lambda_i) \cdot (k^r_4 + 2\lambda^2_r - 2k^{ri}_{22} - \lambda_r\lambda_i) + 2(k^{ri}_{31})^2 - 2k^{ri}_{31}k^{ri}_{13}\Big] + \sum\sum_{i\neq j\neq r}\lambda_{ri}\lambda_{rj}\Big[k^{ri}_{31}k^{rij}_{112} + k^{rj}_{31}k^{rij}_{121} + 2(k^{rij}_{211})^2\Big].$$

We shall find the expression of $\operatorname{cov}(z_q, z_r) = \omega^n_{qr} = E(z_q z_r) - Ez_q Ez_r$, $q\neq r$ with accuracy n^{-1}. From the expansion (1) we get

$$\omega^n_{qr} = E(v_{rr}v_{qq}) + n^{-1/2}E\Big(\sum_{i\neq r}\lambda_{ri}v_{qq}v^2_{ri} + \sum_{i\neq q}\lambda_{qi}v_{rr}v^2_{qi}\Big) + n^{-1}E\Big(\sum_{i\neq r}\sum_{j\neq q}\lambda_{ri}\lambda_{qj}v^2_{ri}v^2_{qj} - \sum_{i\neq r}\lambda^2_{ri}v_{qq}v_{rr}v^2_{ri} - \sum_{i\neq q}\lambda^2_{qi}v_{rr}v_{qq}v^2_{qi} + \sum_{i\neq r}\sum_{j\neq r}\lambda_{ri}\lambda_{rj}v_{qq}v_{ri}v_{ij}v_{jr} + \sum_{i\neq q}\sum_{j\neq q}\lambda_{qi}\lambda_{qj}v_{rr}v_{qi}v_{ij}v_{jq} - \sum_{i\neq r}\sum_{j\neq q}\lambda_{ri}\lambda_{qj}Ev^2_{ri}Ev^2_{qj}\Big) + O(n^{-2}).$$

With the help of (2)-(7) we evaluate the expectations in the above expression. We must consider separately the cases when some subscripts i,j,q,r are equal to each other. For example

$$\sum_{i\neq r}\sum_{j\neq q}\lambda_{ri}\lambda_{qj}\Big[E(v^2_{ri}v^2_{qj}) - Ev^2_{ri}Ev^2_{qj}\Big] = 2\sum\sum_{i\neq j\neq r\neq q}\lambda_{ri}\lambda_{qj}k^2_{ijqr} + 2\sum_{\substack{j\neq r\neq q\\ i=q}}\lambda_{rq}\lambda_{qj}(k^{jqr}_{121})^2$$

$$
+\, 2 \sum_{\substack{i \neq r \neq q \\ j=r}} \lambda_{ri}\, \lambda_{qr} (k^{iqr}_{112})^2 + 2 \sum_{\substack{i \neq r \neq q \\ i=j}} \lambda_{ri}\, \lambda_{qi} (k^{iqr}_{211})^2
$$

$$
-\, 2 \lambda^2_{rq} (k^{rq}_{22} + \lambda_r \lambda_q)^2 .
$$

The last term differs from others and is obtained when $i = q$, $j = r$.

After calculating all expectations we have

$$
\omega^n_{qr} = k^{qr}_{22} + n^{-1} b_2 + O(n^{-2}) , \tag{10}
$$

where

$$
\begin{aligned}
b_2 = {} & \lambda_{rq} \Big\{ \Big[k^{qr}_{42} - k^{qr}_{24} + 6 k^{qr}_{22} (\lambda_q - \lambda_r) \\
& + \lambda_r k^q_4 - \lambda_q k^r_4 + 2 (k^{qr}_{12} k^q_3 - k^{qr}_{21} k^r_3) \\
& + 2 \left[(k^{qr}_{21})^2 - (k^{qr}_{12})^2 \right] + 2 \lambda_q\, \lambda_r (\lambda_q - \lambda_r) \Big\} \\
& + \lambda^2_{rq} \Big[(k^{qr}_{22} + \lambda_q \lambda_r)(k^r_4 + k^q_4 + 2 \lambda^2_r + 2 \lambda^2_q - 4 k^{qr}_{22} \\
& - 2 \lambda_q\, \lambda_r) + 2 (k^{qr}_{31} - k^{qr}_{13})^2 \Big] + \sum_{i \neq q \neq r} \Big\{ \lambda_{ri} \Big[k^{iqr}_{222} \\
& + \lambda_i k^{qr}_{22} + \lambda_r k^{iq}_{22} + 2 k^{qr}_{12} k^{iq}_{21} + 2 (k^{iqr}_{111})^2 \Big] \\
& + \lambda_{qi} \left[k^{iqr}_{222} + \lambda_i k^{qr}_{22} + \lambda_q k^{ir}_{22} + 2 k^{qr}_{21} k^{ir}_{21} + 2 (k^{iqr}_{111})^2 \right] \\
& + \lambda^2_{ri} \left[2 k^{iqr}_{121} (k^{ir}_{31} - k^{ir}_{13}) + (k^{ir}_{22} + \lambda_i \lambda_r)(k^{iq}_{22} - k^{qr}_{22}) \right] \\
& + \lambda^2_{qi} \left[2 k^{iqr}_{112} (k^{iq}_{31} - k^{iq}_{13}) + (k^{iq}_{22} + \lambda_i \lambda_q)(k^{ir}_{22} - k^{qr}_{22}) \right] \\
& + 2 \lambda_{rq} \lambda_{ri} \left[k^{qr}_{31} k^{iqr}_{211} + k^{iq}_{13} k^{iqr}_{112} + (k^{iqr}_{121})^2 - (k^{iqr}_{112})^2 \right] \\
& - 2 \lambda_{rq} \lambda_{qi} \left[k^{qr}_{13} k^{iqr}_{211} + k^{ir}_{13} k^{iqr}_{121} + (k^{iqr}_{112})^2 - (k^{iqr}_{121})^2 \right] \\
& + 2 \lambda_{ri}\, \lambda_{qi} (k^{iqr}_{211})^2 \Big\} + \sum_{i \neq j \neq q \neq r} \sum \Big[\lambda_{ri}\, \lambda_{rj} (k^{iqr}_{121} k^{ijr}_{121} \\
& + k^{ijq}_{112} k^{ijr}_{112} + k^{jqr}_{121} k^{ijr}_{211}) + 2 \lambda_{ri}\, \lambda_{qj} (k_{ijqr})^2 \\
& + \lambda_{qi}\, \lambda_{qj} (k^{iqr}_{112} k^{ijq}_{121} + k^{ijr}_{112} k^{ijq}_{112} + k^{iqr}_{112} k^{ijq}_{211}) \Big] .
\end{aligned}
$$

We see, that in general case the terms of order n^{-1} in (9) and (10) are very complicated. But there exist some special cases, which essentially simplify the expressions.

In the normal population case (9) and (10) give us Lawley's result (1956)

$$\omega_{qr}^{n} = \begin{cases} 2\lambda_r^2\left[1 - \frac{1}{n}\sum_{i\neq r}\left(\frac{\lambda_i}{\lambda_r - \lambda_i}\right)^2\right] + O(n^{-2}) & \text{if } q = r \\ \frac{2}{n}\left(\frac{\lambda_q\lambda_r}{\lambda_q - \lambda_r}\right)^2 + O(n^{-2}) & \text{if } q \neq r \end{cases}$$

In the case of independent random vector x we get

$$\omega_{qr}^{n} = \begin{cases} k_4^r + 2\lambda_r^2 - \frac{1}{n}\left[k_4^r + 2(k_4^r + \lambda_r^2) \cdot \sum_{i\neq r}\left(\frac{\lambda_i}{\lambda_r - \lambda_i}\right)^2\right] + O(n^{-2}) & \text{if } q = r \\ \frac{1}{n}\frac{\lambda_r^2 k_4^q + \lambda_q^2 k_4^r + 2\lambda_q^2\lambda_r^2}{(\lambda_r - \lambda_q)^2} + O(n^{-2}) & \text{if } q \neq r. \end{cases}$$

In the case of bivariate random vector x in the expressions (9), (10) remain the cumulants of x_r and x_q only.

The limiting values of (8), (9), (10) as $n \to \infty$ give us the results of Girshick in normal case and Waternaux in nonnormal case.

References

1. Anderson, T.W., Asymptotic theory for principal components. Ann. Math. Statist., 1963, 34, 122-148.
2. Cook, M.B., Bivariate k-statistics and cumulants of their joint sampling distribution. Biometrika, 1951, 38, 179-195.
3. Fang, C., Krishnaiah, P.R., Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations. J. Multivar. Anal., 1982, 12, 39-63.
4. Fujikoshi, Y., Asymptotic expansions for the distributions of the sample roots under nonnormality. Biometrika, 1980, 67, 45-51.
5. Girshick, M.A., On the sampling theory of the roots of determinantal equations. Ann. Math. Statist., 1939, 10, 203-224.

6. Kaplan, E.L., Tensor notation and the sampling cumulants of k-statistics. Biometrika, 1952, 39, 319-323.
7. Lawley, D.N., Tests of significance for the latent roots of the covariance and correlation matrix. Biometrika, 1956, 43, 128-136.
8. Waternaux, C., Asymptotic distribution of the sample roots for a non-normal population. Biometrika, 1976, 63, 639-645.
9. Кендалл М., Стьюарт А., Теория распределений. М., 1966.
10. Колло Т., Некоторые понятия матричного исчисления с применением в математической статистике. Труды ВЦ ТГУ, 1977, 40, 30-51.

АСИМПТОТИЧЕСКИ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНЫХ СОБСТВЁННЫХ ЗНАЧЕНИЙ ПРИ НЕНОРМАЛЬНОМ ИСХОДНОМ РАСПРЕДЕЛЕНИИ

И. Траат

Р е з ю м е

Для случайного вектора $z = \sqrt{n}(l-\lambda)$, где l и λ однократные собственные значения выборочной и теоретической ковариационной матрицы, выводится распределение $N(\mu^n, \Omega^n)$, которое сходится к предельному распределению вектора z, если $n \to \infty$. Параметры μ^n и $\Omega^n = (\omega_{ij}^n)$ оценивают моменты вектора z с точностью $n^{-1/2}$ и n^{-1} соответственно. Выражения μ^n и ω_{ii}^n в терминах кумулянтов генеральной совокупности были найдены в работах [4] и [8]. В настоящей статье выводится выражение ω_{ij}^n при $i \neq j$. Рассматривается выражение ω_{ij}^n при некоторых частных случаях.

Распределение генеральной совокупности должно иметь кумулянты до шестого порядка.

Received May 84

DEFINITION OF RANDOM VECTORS WITH GIVEN MARGINAL DISTRIBUTIONS AND GIVEN CORRELATION MATRIX

E.-M. Tiit

1. Introduction

The problem of the definition of random vectors with given marginal distributions and given characteristics of dependence (for instance - correlation matrices) arises by studying the robustness and other properties of multivariate statistical procedures by means of statistical modelling.

Let

$$X = (X_1, \ldots, X_k)' \tag{1}$$

be random vector with independent components,

$$X_i \perp X_j \quad (i,j = 1,\ldots,k;\ i \neq j),$$

having given marginal distribution (of the second order) P:

$$X_i \sim P \quad (i = 1,\ldots,k). \tag{2}$$

Let $R(Z) = (r_{ij})$ denote the correlation matrix of random vector Z, $Z \in \mathcal{X}_2^k$, where $\mathcal{X}_2^k$ is the set of all k-dimensional second-order random vectors. Let T be a transformation of random vector (1) (an operator $\mathcal{X}_2^k \rightarrow \mathcal{X}_2^l$),

$$Y = TX, \quad Y = (Y_1, \ldots, Y_l)'. \tag{3}$$

The problem is to define a transformation $T = T(B,P)$ for given B, $B = (b_{ij})$, $B \in \mathcal{L}^k$ ($\mathcal{L}^k$ is the set of all non-negative symmetrical $k \times k$ matrices with $|b_{ij}| \leq 1$) and given P, $P \in \mathcal{P}$ ($\mathcal{P}$ is the set of all distributions) in such a way that following conditions are satisfied:

$$Y_i \sim P \quad (i = 1,\ldots,l) \tag{4}$$

$$R(Y) = B\,. \tag{5}$$

A well-known solution for the problem proposed is the use of linear transformations L in (3). The matrix L with the property $LL' = B$, satisfying (5), exists for every B and l, $l \leq k'$, $k' = \mathrm{rank}(B)$. But the class $\mathcal{P}'$ of distributions

P, satisfying (4), consists of stable distributions only [2].

The purpose of the paper is to give another method for the definition and construction of the transformation (3), generating the vector Y satisfying (4) and (5). This method is based on the concept of the mixture of random vectors [1].

It will be demonstrated that the transformation T(B,P) exists when one of the following complexes of conditions is fulfilled.

(A) 1^o P is symmetrical

2^o $B \in \mathcal{C}^k$, $\mathcal{C}^k$ is the class of 'constructable' $k \times k$ matrices, $\mathcal{C}^k \subset \mathcal{L}^k$, k finite.

(B) 1^o P is arbitrary

2^o $B \in \mathcal{C}^{k+}$, $\mathcal{C}^{k+} = \{B: B \in \mathcal{C}^k, b_{ij} \geq 0\}$, k finite.

It will be proved that the sets $\mathcal{C}^{k+}$ and $\mathcal{C}^k$ are not empty and enclose nontrivial matrices, and that for $k > 2$ $\mathcal{C}^k$ is the proper subset of $\mathcal{L}^k$.

The concept of correlation is generalized for the class of distributions that have no second nor the first moments and so the results are valid with no restrictions for distributions.

The algorithm of the construction of vector Y (3) satisfying (4) and (5) is given for the case if the complex of conditions (B) is fulfilled (for arbitrary k and k').

All the results are illustrated with computational examples.

2. Mixtures of random vectors

Let $(\Omega, \mathcal{A}, \gamma)$ be a probability space, where Ω is a set of elementary events, $\mathcal{A}$ is σ-algebra and γ - probability measure.

Mixture Z is a $\mathcal{A}$-measurable mapping from probability space Ω into $\mathcal{X}^k$, the space of k-dimensional random vectors X, $X \sim P$. That means, Z is also k-dimensional random vector with distribution Q. Measure γ is said to be the mixing measure. In this paper we consider the finite mixing measures only, then $\gamma = (\gamma_1, \dots, \gamma_n)$, $\sum \gamma_i = 1$ and

$$Z = \sum_{h=1}^{n} \chi_{A_n} \cdot X^h , \tag{6}$$

where χ_A is the indicator of event A, $\{A_1,\dots,A_n\}$ forms a full set of exclusive events ($A_i \in \mathcal{A}$) and $X^h \in \mathcal{X}^k$, X^h are mixture components or mixed vectors. Then the mixture's distribution Q is γ-expectation of X's distributions:

$$Q = E_{\gamma} P . \tag{7}$$

Some very simple, but useful conclusions from the definition above can be made.

Conclusion 1. If $X^h \sim P$, $h = 1,\dots,n$, then $Q = P$.

That means, the mixture of equally distributed vectors gives as result the vector with the same distribution.

Conclusion 2. If $X^h \in \mathcal{X}^k_\alpha$, then the moments of mixtures equal to γ - expectations of moments of corresponding components:

$$EZ^\alpha = E_\gamma EX^\alpha .$$

Conclusion 3. If all the components of the mixture X^h, $X^h = (X_1^h,\dots,X_k^h)'$, have equal marginal distributions,

$$X_i^h \sim P, \quad i = 1,\dots,k; \; h = 1,\dots,n,$$

then the mixture Z, $Z = (Z_1,\dots,Z_k)$, has the same marginal distributions:

$$Z_i \sim P, \quad i = 1,\dots,k .$$

Conclusion 4. If all the components X^h of the mixture have equal marginal distributions and $X^h \in \mathcal{X}_2^k$, then

$$R(Z) = E_\gamma R(X) = \sum_{h=1}^{n} \gamma_h R(X^h) . \tag{8}$$

3. Some arbitrary concepts: index-vector, index-system, index-system's sequence

The aim of the given paragraph is the elaboration of convenient tools for constructing mixtures of random vectors (with desired properties). For defining the subvectors and submatrices of given vectors and matrices the concept of the index-vector will be defined; with the help of index-systems, formed on the basis of index-vectors, some special sets of subvectors, useful in the following discussion, will be defined. Let k be the initial dimension, k fixed.

Definition 1. The vector I with natural components

(indices)

$$I = (i_1,\dots,i_m)' \qquad (1 \le i_j < i_{j+1} \le k;\ \ m \le k)$$

is said to be index-vector [3]. Let us denote $I_o = (1,\dots,k)$. For the sake of simplicity we regard I at the same time as set $\{i_1,\dots,i_m\}$. So $i \in I$ means $i \in \{i_1,\dots,i_m\}$ (or $i \in \{I\}$), $\varkappa(I)$ is the power (number of elements) of I; I^c, $I_1 \cup I_2$ and $I_1 \cap I_2$ are index-vectors that consist of elements of sets $\{I_o \setminus I\}$, $\{I_1 \cup I_2\}$ and $\{I_1 \cap I_2\}$ correspondingly; instead of $\{i\}$ and (i) we use i.

The set $H = (I_1,\dots,I_h)$ is said to be full, if

$$1^o\ \ I_i \cap I_j = \emptyset \qquad (i,j = 1,\dots,h;\ i \ne j),$$

$$2^o\ \ \bigcup I_i = I_o .$$

Definition 2. Full set H of index-vectors is said to be the index-system, if

$$3^o\ \ \varkappa(I_i) \geq \varkappa(I_{i+1}), \qquad i = 1,\dots,h-1,$$

$$4^o\ \text{if}\ \ \varkappa(I_i) = \dots = \varkappa(I_{i+j}),\ \text{then}$$

$$(I_i)_1 < (I_{i+1})_1 < \dots < (I_{i+j})_1 ,$$

where $(I)_\nu$ denotes the ν-th element of index-vector I, $\nu \le \varkappa(I)$; $\varkappa(H) = h$ is said to be the power of index-system H.

So every partition of set $\{1,\dots,k\}$ generates an index-system.

Let us regard $\mathcal{H}^k$ - the set of all possible index-systems, generated for the case of initial dimension k. We define the partial ordering in the set $\mathcal{H}^k$ as follows:

Let $H_l = (I_1^l,\dots,I_{h_l}^l)$, $l = 1,\dots,L$, $L = \varkappa(\mathcal{H}^k)$ - finite.

Assume

$$H_l \prec H_f ,$$

if there exists such index s $(1 \le s)$ that

$$I_i^l = I_i^f, \qquad i = 1,\dots,s-1,$$

$$I_s^l \supset I_s^f, \qquad I_s^l \ne I_s^f .$$

Definition 3. The sequence (H_l) $(l = 1,\dots,L)$ is said to be decreasing, if the condition

$$H_l \prec H_f \Rightarrow l < f$$

is fulfilled, $l,f = 1,\ldots,L$.

Example 1. Let us consider the case $k = 4$. Then all the possible index-systems H_l are the following:

$$H_1 = ((1,2,3,4)) \qquad H_6 = ((1,2),(3,4)) \qquad H_{11} = ((1,4),2,3)$$
$$H_2 = ((1,2,3),4) \qquad H_7 = ((1,3),(2,4)) \qquad H_{12} = ((2,3),1,4)$$
$$H_3 = ((1,2,4),3) \qquad H_8 = ((1,4),(2,3)) \qquad H_{13} = ((2,4),1,3)$$
$$H_4 = ((1,3,4),2) \qquad H_9 = ((1,2),3,4) \qquad H_{14} = ((3,4),1,2)$$
$$H_5 = ((2,3,4),1) \qquad H_{10} = ((1,3),2,4) \qquad H_{15} = (1,2,3,4)$$

Here we have the following relations:

$$H_1 < H_i\,,\quad i = 2,\ldots,15 \qquad H_5 < H_{12},H_{13},H_{14},H_{15}$$
$$H_2 < H_6,H_7,H_9,H_{10},H_{12},H_{15} \qquad H_6 < H_9,H_{15}$$
$$H_3 < H_6,H_8,H_9,H_{11},H_{13},H_{15} \qquad H_7 < H_{10},H_{15}$$
$$H_4 < H_7,H_8,H_{10},H_{11},H_{14},H_{15} \qquad H_8 < H_{11},H_{15}$$
$$H_9,H_{10},H_{11},H_{12},H_{13},H_{14} < H_{15}$$

The given sequence is decreasing, so as, for instance, the sequence

$$H_1,H_5,H_4,H_3,H_2,H_8,H_7,H_6,H_{14},H_{13},H_{12},H_{11},H_{10},H_9,H_{15}$$

too.

Let X be an arbitrary k-dimensional vector, then index-vector I defines its subvector:

$$X(I) = (X_{i_1},\ldots,X_{i_m})'\ ;$$

let B be an arbitrary $k \times k$ symmetrical matrix, then $B(I)$ is its symmetrical submatrix,

$$B(I) = \begin{pmatrix} b_{i_1 i_1} & \cdots & b_{i_1 i_m} \\ \cdot & \cdots & \cdot \\ b_{i_m i_1} & \cdots & b_{i_m i_m} \end{pmatrix}.$$

We shall use the simplified notation:

$$X_i \in X(I) \Leftrightarrow i \in I$$
$$b_{ij} \in B(I) \Leftrightarrow i,j \in I\,.$$

With the help of index-system H it is possible to define the generalized block-diagonal matrix $B(H)$ of the given matrix B that consists of submatrices $B(I_1),\dots,B(I_h)$ only (cf. Example 4).

4. Bundles and bundle-systems

Definition 4. Let J be index-vector (initial dimension k), then k-dimensional vector W, satisfying conditions

$$\begin{cases} W(J) = (X_1,\dots,X_1), \\ W(J^c) = (-X_1,\dots,-X_1) \end{cases} \tag{9}$$

is said to be the bundle (defined by J).

If $W \in \mathcal{X}_2^k$, then the correlation matrix R(W) consists of 1's and -1's only:

$$r_{ij} = \begin{cases} 1, & \text{if } (i,j \in J) \vee (i,j \in J^c) \\ -1 & \text{otherwise.} \end{cases}$$

Example 2. Let $k=5$, $J=(1,3,4)$. Then bundle W is the following:

$$W = (X_1,-X_1,X_1,X_1,-X_1)$$

and

$$R(W) = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 \end{pmatrix}.$$

Definition 5. If $J=I_0$, the bundle W is said to be simple. The correlation matrix of a simple bundle consists of 1's only.

Definition 6. Let $H = (I_1,\dots,I_h)$ be an index-system, and let J_l be such an index-vector, that

$$J_l \subset I_l, \quad l = 1,\dots,h;$$

we shall denote $I_l \setminus J_l = G_l$.

Vector $W=(W_1,\dots,W_k)$ is said to be the bundle-system, if the following conditions are fulfilled:

1° $W(I_l)$ is a bundle (defined by J_l), $l=1,\dots,h$;

2° Bundles $W(I_l)$ and $W(I_f)$ are independent if $l \neq f$, $l,f=1,\dots,h$.

On the basis of the given vector X the bundle-system $W = X(H)$ will be defined as follows:

$$W_i = \begin{cases} X_l, & \text{if } i \in J_l, \\ -X_l, & \text{if } i \in G_l, \quad l = 1,\dots,h; \; i = 1,\dots,k. \end{cases} \tag{10}$$

If $W \in \mathfrak{X}_2^k$, then correlation matrix $R(W)$ consists of 1's, -1's and 0's only:

$$r_{ij} = \begin{cases} 1, \text{ if } (i,j \in J_l) \vee (i,j \in G_l), \quad l = 1,\dots,h \\ -1, \text{ if } ((i \in J_l) \wedge (j \in G_l)) \vee ((i \in G_l) \wedge (j \in J_l)), \\ \qquad l = 1,\dots,h \\ 0 \quad \text{otherwise}. \end{cases} \tag{11}$$

<u>Definition 7.</u> If $J_l = I_l$, $l = 1,\dots,h$, then the bundle-system W is said to be simple. The correlation matrix of the simple bundle-system consists of 1's and 0's only.

<u>Example 3.</u> Let $k = 6$,

$I_1 = (1,3,4)$, $I_2 = (2,6)$, $I_3 = 5$,

$J_1 = (3,4)$, $\quad J_2 = 2$, $\quad J_3 = 5$.

Then the bundle-system W consists of 3 bundles:

$W(I_1) = (-X_1, X_1, X_1)$; $\quad W(I_2) = (X_2, -X_2)$; $\quad W(I_3) = X_3$,

$W = (-X_1, X_2, X_1, X_1, X_3, -X_2)$

and the correlation matrix $R(W)$ is the following:

$$R(W) = \begin{pmatrix} 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

<u>Example 4.</u> Let matrix B be given,

$$B = \begin{pmatrix} 1 & 0.3 & 0.4 & 0.7 & -0.2 & 0.5 \\ 0.3 & 1 & 0.5 & 0.6 & 0.3 & -0.4 \\ 0.4 & 0.5 & 1 & 0.7 & 0.2 & 0.3 \\ 0.7 & 0.6 & 0.7 & 1 & 0.1 & 0.5 \\ -0.2 & 0.3 & 0.2 & 0.1 & 1 & 0.1 \\ 0.5 & -0.4 & 0.3 & 0.5 & 0.1 & 1 \end{pmatrix},$$

index-system H consists of three index-vectors,

$$H = ((1,3,4),(2,6),5),$$

then the generalized block-diagonal matrix $B(H)$ is the following:

$$\begin{pmatrix} 1 & * & 0.4 & 0.7 & * & * \\ * & 1 & * & * & * & -0.4 \\ 0.4 & * & 1 & 0.7 & * & * \\ 0.7 & * & 0.7 & 1 & * & * \\ * & * & * & * & 1 & * \\ * & -0.4 & * & * & * & 1 \end{pmatrix}$$

where the elements, denoted by stars (*) are not defined.

5. Generalization of the concept of correlation

Let us consider random variables with arbitrary symmetrical distribution without second moments and define for them generalized correlation coefficient $\tilde{r}$:

Definition 8.

$$\tilde{r}(X,Y) = \begin{cases} 1, & \text{if } Y = X \\ -1, & \text{if } Y = -X \\ 0, & \text{if } X \quad Y . \end{cases} \tag{12}$$

Let W be the bundle-system, $X_i \sim P$, where P is arbitrary symmetrical distribution not of the second order.

If we use the generalized correlation (12) between components X_i and X_j instead of their usual correlation, we can define the correlation matrix $R(W)$ for W by formula (11). Accounting this we do not restrict ourselves with the vectors of second order in the further discussion.

6. Generation of the correlated pair of random variables

Let $X = (X_1,X_2)$, $X_i \sim P$, $X_1 \perp X_2$, and b given.

The problem is to generate $Y = (Y_1,Y_2)'$ so that the conditions (4) and

$$r(Y_1,Y_2) = b \tag{5'}$$

are fulfilled.

We shall prove that the problem is solvable if

(A) 1° P is arbitrary

2° $b \in [0,1]$ (that is, $B \in \mathcal{R}^{+}$)

or

(B) 1° P is symmetrical

2° $b \in [-1,1]$ (that is, $B \in \mathcal{R}$).

Proof of case (A). Let us define two bundle-systems W^1 and W^2:

$$W^1 = (X_1, X_1)\ ; \quad W^2 = (X_1, X_2)$$

and mixing measure: $\gamma_1 = b$; $\gamma_2 = 1 - b$. Then the mixture

$$Z = E_{\gamma} W = \chi_{A(b)} \cdot W^1 + \chi_{A(1-b)} \cdot W^2$$

(where $A(b)$ is an event with γ-measure b: $\gamma(A(b)) = b$) has correlation matrix $R(Z)$, computable by (8) and (10):

$$R(Z) = \gamma_1 R(W^1) + \gamma_2 R(W^2)$$

$$= b \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + (1-b) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix}.$$

Proof of case (B). If $b \geqslant 0$ then the proof coincides with that of in case (A). Let b be negative. Then let us define $W^1 = (X_1, -X_1)$, $W^2 = (X_1, X_2)$, $\gamma_1 = |b|$, $\gamma_2 = 1 - |b|$ and

$$R(Z) = |b| \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + (1 - |b|) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -|b| \\ -|b| & 1 \end{pmatrix}$$

It also follows that in case $k = 2$ for symmetrical distributions $\mathcal{C}^2 = \mathcal{R}^2$, for arbitrary distributions $\mathcal{C}^{2+} = \mathcal{R}^{2+}$.

7. Definition of random vector with given correlation matrix B and given (equal) marginal distributions P

Assume we have n bundle-systems W^l $(l = 1,\dots,n)$ with $X_i \sim P$, $W^l \in \mathcal{X}^k$ and given matrix B, $B \in \mathcal{R}$.

Our task is to define the mixing measure γ so that the condition (5) is fulfilled.

Using (8) we have equations

$$\sum_{l=1}^{n} \gamma_l r_{ij}^l = b_{ij}\,, \qquad i,j = 1,\dots,k\,, \tag{13}$$

where r_{ij}^l denotes the element of matrix $R(W^l)$.

The system (13) has

$$K = \frac{k(k-1)}{2} \tag{14}$$

different equations and n unknown probabilities γ_l. So as the number of all the possible different bundle-systems N_k is larger than K, the system of equations (13), as a rule, has no unique solution.

For simplifying the equation (13) let us use the notations (the index-vectors corresponding to initial dimension n)

$$I(i,j) = \{l:\ (i,j \in J_l) \vee (i,j \in G_l),\ l = 1,\dots,n\}$$

$$L(i,j) = \{l:\ ((i \in J_l) \wedge (j \in G_l)) \vee ((i \in G_l) \wedge (j \in J_l)),\ l = 1,\dots,n\}$$

$$K(i,j) = I_o \setminus (I(i,j) \cup L(i,j)) .$$

That means, I(i,j) consists of numbers l of these bundle-systems W_l, where $r^l_{ij} = 1$; L(i,j) - of numbers l of these bundle-systems W_l, where $r^l_{ij} = -1$ and K(i,j) - of numbers of these bundle-systems, where $r^l_{ij} = 0$.

Taking the definition of index-vectors I(i,j), L(i,j) and K(i,j) into account, we can rewrite the equations (13):

$$b_{ij} = \sum_{l \in I(i,j)} \gamma_l - \sum_{l \in L(i,j)} \gamma_l \quad (i,j = 1,\dots,k) \tag{13'}$$

For the case $B \in \mathcal{R}^+$ we can use only simple bundle-systems and $L(i,j) = \emptyset$. Then the system (13) takes the form

$$b_{ij} = \sum_{l \in I(i,j)} \gamma_l \quad (i,j = 1,\dots,k) \tag{13"}$$

From the equations (13') and (13") it follows that for defining the vector with the given correlation matrix B via mixtures of bundle-systems it is needed to solve the system of linear equations (13') (or, in the more simple case, (13")).

The system consists of K equations (see (14)). The solution $\gamma = (\gamma_1,\dots,\gamma_n)'$ of the system must fulfill the additional conditions:

$$\begin{cases} \sum_{i=1}^{n} \gamma_i = 1, \\ \gamma_i \in [0,1], \quad i = 1,\dots,n . \end{cases} \tag{15}$$

By given conditions the usual methods of the solution of the system of linear equations do not work, but in the next paragraph we shall demonstrate a simple algorithm for defining γ for case $B \in \mathcal{R}^+$.

8. Algorithm of the construction of random vector Z with the given correlation matrix B ($B \in \mathcal{R}^+$)

Let k be an initial dimension, B - given matrix ($B \in \mathcal{R}^{+k}$), $X = (X_1,\dots,X_k)'$ - initial vector, $X_i \sim P$, P given (arbitrary) distribution, $X_i \perp X_j$, $i \neq j$, $i,j = 1,\dots,k$.

Let (H_l) be a decreasing sequence of index-systems, $l = 1,\dots,L$, $H_l = (I_1^l,\dots,I_{h_l}^l)$.

Step 1 Take $l := 1$ (number of index-system)

$f := 0$

$B_1 := B$

Step 2 Take $f := f+1$ (number of mixture component)

Step 3 Define $G_f := B_f(H_l)$

Find $g_l = \min_{i,j} g_{ij}^l$

Step 4 If $g_l = 0$, take $l := l+1$, go to step 3

Step 5 Define $\gamma_f := g_l$

$W^f := X(H_l)$ (see (10))

$B_{f+1} := B_f - \gamma_f(R(W^f) - I_k)$

Step 6 If $\sum_{j=1}^{f} \gamma_j > 1$ then end: the construction is impossible with the given sequence (H_l)

Step 7 If $\max_{i \neq j} b_{ij}^{f+1} = 0$, then

$$\gamma_o := 1 - \sum_{j=1}^{f} \gamma_j$$

$W^o := (X_1,\dots,X_k)$

End

Step 8 Take $l := l+1$, go to step 2

During the construction from the coefficients b_{ij} of the given correlation matrix B the values of γ_l ($l = 1, \ldots, K$) (K is defined in (14)) are subtracted by the following rule:

if and only if the indices pair i,j is included in any index-vector I_s^l of index-system H_l, then from b_{ij} the value γ_l is subtracted.

In every step the least non-zero correlation coefficient, indices of which fulfill the above condition, is changed to zero. In such a way after K steps all correlation coefficients of the initial matrix B are made equal to zeroes, and the vector γ, $\gamma = (\gamma_1, \ldots, \gamma_K)$ is chosen in such a way that all initial correlation coefficients b_{ij} are equal to sums of γ's, γ_l being nonnegative.

For proving that γ is a solution of the system of equations (13") let us define for every pair i,j the index-vector I(i,j) with the initial dimension K (see Definition 1).

Let us take

$$I(i,j) = \left\{ l: i,j \in I_s^l \ (s = 1, \ldots, h_l) \right\} \quad (i,j = 1, \ldots, k).$$

From the construction it follows that the equation

$$b_{ij} = \sum_{l \in I(i,j)} \gamma_l$$

is true for every pair of indices i,j and consequently γ, $\gamma = (\gamma_1, \ldots, \gamma_K)$ is a solution of system (13").

Corresponding mixture components W^l are defined on each step as bundle-systems $X(H^l)$.

From the definition of γ_o (step 7) it can be concluded that the conditions (15) are fulfilled, that means, the mixture Z,

$$Z = E_{\gamma} W,$$

where γ and W are defined with the help of the given algorithm, fulfills the conditions (4) and (5).

Note that the construction is not unique, but depends on the choice of the sequence H^l.

It is possible that the constructions based on some sequences H^l give the solution, but others do not give. But as the number of different sequences is finite, it is always possible to check if any of the sequences gives the solution to the problem proposed.

<u>Example 5.</u> Let $k = 4$, and given

$$B = \begin{pmatrix} 1 & 0.5 & 0.4 & 0.4 \\ 0.5 & 1 & 0.3 & 0.2 \\ 0.4 & 0.3 & 1 & 0.1 \\ 0.4 & 0.2 & 0.1 & 1 \end{pmatrix}.$$

Let us use the index-sequence from Example 1.

$f = 1$; $H_1 = I_1^1 = ((1,2,3,4))$; $G_1 = B_1 = B$;

$g_1 = 0.1$; $\gamma_1 = 0.1$; $W^1 = (X_1, X_1, X_1, X_1)$;

$$B_2 = \begin{pmatrix} 1 & 0.4 & 0.3 & 0.3 \\ 0.4 & 1 & 0.2 & 0.1 \\ 0.3 & 0.2 & 1 & 0 \\ 0.3 & 0.1 & 0 & 1 \end{pmatrix};$$

$f = 2$; $H_2 = ((1,2,3),4)$;

$$G_2 = \begin{pmatrix} 1 & 0.4 & 0.3 & * \\ 0.4 & 1 & 0.2 & * \\ 0.3 & 0.2 & 1 & * \\ * & * & * & 1 \end{pmatrix};$$

$g_2 = 0.2$; $\gamma_2 = 0.2$; $W^2 = (X_1, X_1, X_1, X_2)$;

$$B_3 = \begin{pmatrix} 1 & 0.2 & 0.1 & 0.3 \\ 0.2 & 1 & 0 & 0.1 \\ 0.1 & 0 & 1 & 0 \\ 0.3 & 0.1 & 0 & 1 \end{pmatrix};$$

$f = 3$; $H_3 = ((1,2,4),3)$;

$g_3 = 0.1$; $\gamma_3 = 0.1$; $W^3 = (X_1, X_1, X_2, X_1)$;

$$B_4 = \begin{pmatrix} 1 & 0.1 & 0.1 & 0.2 \\ 0.1 & 1 & 0 & 0 \\ 0.1 & 0 & 1 & 0 \\ 0.2 & 0 & 0 & 1 \end{pmatrix};$$

$f = 4$; $H_9 = ((1,2),3,4)$;

$g_9 = 0.1$; $\gamma_4 = 0.1$; $W^4 = (X_1, X_1, X_2, X_3)$;

$$B_5 = \begin{pmatrix} 1 & 0 & 0.1 & 0.2 \\ 0 & 1 & 0 & 0 \\ 0.1 & 0 & 1 & 0 \\ 0.2 & 0 & 0 & 1 \end{pmatrix};$$

$$f = 5;\ H_{10} = ((1,3),2,4);$$
$$g_{10} = 0.1;\ \gamma_5 = 0.1;\ W^5 = (X_1,X_2,X_1,X_3);$$
$$f = 6;\ H_{11} = ((1,4),2,3);$$
$$g_{11} = 0.2;\ \gamma_6 = 0.2;\ W^6 = (X_1,X_2,X_3,X_1)$$

and the construction is finished.

So as

$$\sum_{i=1}^{6} \gamma_i = 0.1 + 0.2 + 0.1 + 0.1 + 0.1 + 0.2 = 0.8,$$

we need to define

$$\gamma_o = 1 - 0.8 = 0.2$$
$$W^o = (X_1,X_2,X_3,X_4)$$

and so

$$Z = E_\gamma W = \chi_{A(0.1)} \cdot W^1 + \chi_{A(0.2)} \cdot W^2 + \dots + \chi_{A(0.2)} \cdot W^o.$$

9. Construction of the constant-correlation matrix

Let us prove that for every k sets $\mathcal{C}^k$ and $\mathcal{C}^{+k}$ include an infinite set of different matrices.

Definition 9. A correlation matrix B is said to be a constant-correlation matrix if the following condition is fulfilled:

$$b_{ij} = b \qquad (i,j = 1,\dots,k;\ i \neq j). \tag{16}$$

Let us denote the constant-correlation matrix (16) B(b).

It is well-known that $B(b) \in \mathcal{L}^k$ for every k and for every b,

$$b \in \left[-\frac{1}{k-1}, 1\right].$$

This result is an immediate consequence from the values of eigenvalues of B(b): $\lambda_1 = 1 + (k-1)b$, $\lambda_2 = \dots = \lambda_k = 1 - b$.

We shall demonstrate that $B(b) \in \mathcal{C}^{k^2}$ for every k and for every b,

$$b \in \left[-\frac{1}{K}, 1\right].$$

(A) Let $b \geqslant 0$. Then the construction of mixture Z is quite analogous to that of the case $k = 2$ (in paragraph 6): Take $\gamma_1 = b$; $\gamma_2 = 1 - b$; $W_1 = (X_1,\dots,X_1)$; $W_2 = (X_1,X_2,\dots,X_k)$,

$$Z = E_\gamma W,$$

$$R(Z) = b \begin{pmatrix} 1 & 1 & \dots & 1 \\ . & . & . & . & . \\ 1 & 1 & \dots & 1 \end{pmatrix} + (1-b)I_k = \begin{pmatrix} 1 & b & \dots & b \\ . & . & . & . & . \\ b & b & & 1 \end{pmatrix}$$

q.e.d.

(B) Let $b < 0$.

Let us define bundles $W(i,j) = (W_1,\dots,W_k)$ as follows:

$$W_l = \begin{cases} -X_i, & \text{if } l = j \\ X_l & \text{otherwise}, \quad l = 1,\dots,k; \ i = 1,\dots,j-1; \\ & \qquad j = 2,\dots,k \end{cases}$$

$$W_o = (X_1,\dots,X_k).$$

The mixture measure will be:

$$\gamma(i,j) = -b, \quad i = 1,\dots,j-1; \quad j = 2,\dots,k,$$
$$\gamma_o = 1 - \sum \gamma(i,j);$$

then

$$Z = E_\gamma W,$$

$$R(Z) = \sum_{i,j} \gamma(i,j)\, R(W(i,j)) + \gamma_o R(W_o) = \begin{pmatrix} 1 & b & \dots & b \\ b & 1 & \dots & b \\ . & . & . & . & . \\ b & b & \dots & 1 \end{pmatrix},$$

so as $R(W(i,j))$ has only two non-zero out-of-diagonal elements: $r_{ij} = r_{ji} = -1$.

Now let us prove that $\mathcal{C}$ is the proper subset of $\mathcal{R}$, that means, there exists such a matrix $B \in \mathcal{R}$ that it is impossible to construct mixture Z so that $R(Z) = B$.

Take $k = 3$, $b = -0.5$.

For $k = 3$ there exist 5 different index-systems: $H_1 = ((1,2,3))$, $H_2 = ((1,2),3)$, $H_3 = ((1,3),2)$, $H_4 = ((2,3),1)$, $H_5 = (1,2,3)$.

The equation system (13) is the following:

$$\begin{cases} \gamma_1 r^1_{12} + \gamma_2 r^2_{12} & = -0.5 \\ \gamma_1 r^1_{13} + \gamma_3 r^3_{13} & = -0.5 \\ \gamma_1 r^1_{23} + \gamma_4 r^4_{23} & = -0.5 \\ \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 \leq 1 & (\gamma_1 \geq 0) \end{cases} \tag{17}$$

For the bundle W^1 there are 8 different possibilities of choosing components X_1 and $-X_1$, but as a result the number of different correlation matrices $R(W^1)$ is only four:

(1) $r^1_{12} = r^1_{13} = r^1_{23} = 1$,

(2) $r^1_{12} = 1$, $r^1_{13} = r^1_{23} = -1$,

(3) $r^1_{12} = -1$, $r^1_{13} = 1$, $r^1_{23} = -1$,

(4) $r^1_{12} = r^1_{13} = -1$, $r^1_{23} = 1$.

For the bundles W^2, W^3 and W^4 there are two possibilities for each: $r^2_{12} = 1$ or $r^2_{12} = -1$; $r^3_{13} = 1$ or $r^3_{13} = -1$ and $r^4_{23} = 1$ or $r^4_{23} = -1$. So as system (17) is symmetrical for all pairs of indices, only 10 different cases are of interest. Let us denote them schematically in the following wise:

```
(a) ++     (b) ++     (c) ++     (d) +-     (e) ++
    + +        + +        + -        + -        - +
    +  +       +  -       +  -       +  -       -  +

(f) ++     (g) +-     (h) ++     (i) +-     (j) +-
    - +        - +        - -        - +        - -
    -  -       -  +       -  -       -  -       -  -
```

So as the solution γ must consist of positive components only, it is evident that cases (a)-(c), (e), (f), (h) can not give any solution.

For investigating all other cases let us sum all four equations of (17). We get the following results:

(d) $4\gamma_1 \leq -0.5$; (g) $2(\gamma_3 + \gamma_4) \leq -0.5$;

(i) $2\gamma_3 \leq -0.5$; (j) $0 \leq -0.5$,

that means, no solution in domain $\gamma_i \in [0,1]$ exists. So as for all the possibilities the right part of the system is less than the left one, it is evident that no combination of bundle-systems can give the solution of (17) in domain $\gamma_i \in [0,1]$. Consequently, matrix B with $b_{ij} = -0.5$ does not belong to set $\mathcal{C}$ of matrices, constructable by means of mixtures of bundle-systems.

References

1. Tukey J.W. A survey of sampling from contaminated distributions. In: Contributions to Prob. and Statist. (Ed. Olkin I. et al.) Stanford Univ. Press, 1960.
2. Золотарев В.М. Одномерные устойчивые распределения. Москва, 1980.

3. Тийт, Э. Обобщенное понятие признака-вектора. Труды ВЦ ТГУ. № 32, Тарту, 1975, 3–30.

ОПРЕДЕЛЕНИЕ СЛУЧАЙНЫХ ВЕКТОРОВ С ДАННЫМИ МАРГИНАЛЬНЫМИ РАСПРЕДЕЛЕНИЯМИ И С ДАННОЙ КОРРЕЛЯЦИОННОЙ МАТРИЦЕЙ

Э.Тийт

Резюме

При исследованиях, применяющих метод статистического моделирования в многомерном статистическом анализе, возникает задача определения и генерирования случайных векторов с заданными маргинальными распределениями и заданной матрицей ковариации.

Так как для вектора X с независимыми компонентами $X = (X_1,\dots,X_k)'$ проблема решается просто, то задачей является определить такое преобразование T,

$$Y = TX, \quad Y = (Y_1,\dots,Y_\ell)', \tag{3}$$

что выполняются следующие условия

$$Y_i \sim P \tag{4}$$

где P заданное распределение,

$$R(Y) = B \tag{5}$$

где $R(Y)$ – корреляционная матрица вектора Y, а B – заданная матрица.

Если X имеет устойчивое распределение, то T может быть линейным преобразованием, притом задача решается для всех $B \in \mathcal{R}^k$, где $\mathcal{R}^k$ – множество симметричных неотрицательно определенных (к х к)-матриц.

Для других типов распределения возможно определение T в виде некоторой смеси к-мерных векторов, притом множество всевозможных (к х к)-матриц B таких, что выполняются условия (4) и (5), обозначаемое символом C^k, является подмножеством множества $\mathcal{R}^k$

В статье рассматриваются конечные смеси векторов, определяемые как измеримые отображения из вероятностного пространства в пространстве случайных векторов. Распределение смеси определяется как математическое ожидание

$E_\gamma P$,

где γ – мешающая мера, P – распределения взмешенных векто-

ров.

Выводятся некоторые простые свойства смесей, например

$$R(Z) = E_\gamma R(X), \tag{8}$$

где $R(Z)$ - корреляционная матрица смеси, γ - мешающая мера и $R(X)$ - корреляционные матрицы взмешенных векторов.

Далее определяются некоторые вспомогательные понятия, такие как индекс-вектор, индекс-система и связка случайных величин. Связка является вектором, состоящим из компонентов X и $-X$, где X имеет симметрическое распределение. Корреляционная матрица связки состоит из 1 и -1, а если случайный вектор Z состоит из нескольких независимых связок, то корреляционная матрица $R(Z)$ содержит элементы $1, -1$ и 0.

На основании равенства (8) можно выписать систему уравнений для определения смеси, решение которого (если такое существует) и дает преобразование T для заданных P и B.

В статье излагается и алгоритм для решения поставленной задачи в случае, когда $t_{ij} \geq 0$, $i,j = 1,\ldots,k$.

Received Juny 84

EXPERIMENTAL DESIGNING FOR MONTE-CARLO STUDY IN MULTIVARIATE STATISTICS

E.-M. Tiit, I. Traat

1. Introduction

Let us sketch a general scheme of the use of the statistical modelling (Monte-Carlo method) in the multivariate statistical analysis.

Let $X = (X_1, \ldots, X_n)'$ be the initial (theoretical) random vector. Usually the general task of the multivariate statistical analysis is the building of model

$$Y = F(\mathcal{X}, \vartheta), \tag{1}$$

where $Y = (Y_1, \ldots, Y\ )'$ is a random vector, either measured (in cases of regression, variance or discriminant analyses), or unmeasured, 'latent' (in case of the component, factor and cluster analyses); $\vartheta = (\vartheta_1, \ldots, \vartheta_r)'$ is an unknown parameter vector and $\mathcal{X} = (x_{ij})$ $(i = 1, \ldots, m;\ j = 1, \ldots, n)$ is the sample. On the basis of the sample one of the following typical problems of mathematical statistics will be solved:

S1 the estimation of the parameter vector ϑ (the estimate is $\tilde{\vartheta}$);

S2 the estimation of the estimation error (e.g.: $D\tilde{\vartheta}$);

S3 testing the suitability and exactness of the model (1).

The problems S2 and S3 are analytically solvable only by existence of sufficient prior information (e.g., X having some 'good' distribution). Otherwise the only possibility is to use the experiment, that is - statistical modelling.

For effective planning of modelling experiments it is useful to exploit some results of the experimental designing theory. Thereby, as a rule, in the role of the criterion variable some distance $d(\tilde{\vartheta}, \vartheta)$ between the estimation $\tilde{\vartheta}$ and the true value of parameter vector ϑ appears.

The complex of factors, describing the experimental points (that is: samples $\mathcal{X}$) depends on the purpose of the investigation. As a rule, always some factors from the following list appear:

F1: sample size n;

F2: initial vector's dimension m;

F3: type of the parent distribution P;

F4: characteristics of X (for instance, moments m^1, m^2, correlation matrix $R(X)$ etc.);

F5: characteristics of model F (e.g., dimension r of vector ϑ);

F6: percentage of outliers and missing observations;

F7: higher moments (m^3, m^4) or cumulants of X.

The factors F6 and F7 are used when special studies, connected with the robustness of methods, considered in data analysis, are of interest. The factors F7 are needed in studies about convergence rate in the asymptotical theory. The factors F1-F5 are needed in the most of statistical problems.

There arise several problems with the experimental-designing approach to statistical modelling studies in multivariate analysis. Some of the most disturbing ones are following:

P1: High dimensionality of several factors (F4, F7).

P2: The structure of the set of all admissible points of the experiment is unknown.

P3: The exact value of the parameter-vector ϑ, and followingly, the distance $d(\vartheta, \bar{\vartheta})$ is unknown.

P4: As a rule, every experiment (cycle of modelling) has a considerably high cost, so the number of experiments must not be too large.

2. Optimal parametrization of the initial vector

One of the possibilities dealing with the problems described is the definition of optimal parametrization of the initial vector X. The parametrization is said to be optimal, if it fulfills the following conditions:

C1 (the universality condition). The number and the meaning of the parameters do not depend on the type of the parent distribution P, and on the dimension m of the initial vector X.

C2 (the minimality condition). The number of independent parameters describing all essential characteristics of X (e.g., F4 and F7) is minimal.

C3 (the computability condition). From the known values of parameters the exact value of ϑ is simply computable.

The connection between the parameter ϑ of model and the parameters of vector X is sometimes more useful in the inverse sense:

C'3. From the given values ϑ_0 of ϑ the values of parameters of vector X can be computed and a concrete realization (sample) of X can be generated.

The purpose of the given paper is to demonstrate some possibilities of such parametrization and give some examples of the practical usage of described principles in the investigation of asymptotical behaviour of the characteristics of the regression and component analyses.

3. The random vectors with constant-correlation matrix

A very simple family of random vectors, depending on only two parameters and fulfilling the conditions C1-C3 (or C'3) of optimality can be defined with the help of constant-correlation matrix.

The correlation matrix R with elements r_{ij} $(i,j = 1,\dots,m)$ is said to be constant and denoted $R(\alpha)$, if

$$r_{ij} = \alpha, \quad i \neq j, \ i,j = 1,\dots,m. \tag{2}$$

Let $EX_i = 0$, $DX_i = 1$ and all the marginal distributions of X are equal. Then the random vector X is characterized by two parameters of joint distribution - vector's dimension m $(m < 1)$ and correlation coefficient α, $\alpha \in [-\frac{1}{m-1}, 1]$.

In [1] it is proved that for the arbitrary symmetrical distribution P and dimension m it is possible to define $R(\alpha)$ for $\alpha \in [-\frac{2}{m(m-1)}, 1]$ and to generate X with the correlation matrix $R(\alpha)$ and with all marginals equal to P. That means, the pair of parameters (m, α) fulfills the universality condition C1.

So as we have taken all first moments equal to zero and all second moments equal either 1 or α, the minimality condition C2 is fulfilled too.

Let us prove that for some known statistical models the condition C3 is fulfilled too.

From the well-known fact that the eigenvalues of the constant-correlation matrix $R(\alpha)$ have only two different values,

$$\begin{cases} \lambda_1 = 1 + (m-1)\alpha, \ \lambda_2 = \ldots = \lambda_m = 1 - \alpha \ \text{if} \ \alpha \geqslant 0, \\ \lambda_1 = \ldots = \lambda_{m-1} = 1 - \alpha, \ \lambda_m = 1 + (m-1)\alpha \ \text{if} \ \alpha < 0, \end{cases}$$

it follows that only one principal component (first or last) is defined, all others define the $(m-1)$-dimensional subspace. In the case of the classical factor analysis (on the basis of the so-called reduced correlation matrix $R - D$) for $\alpha > 0$ only one factor exists (its variance equals $m\alpha$) and the variances of the uniquenesses of all variables X_1 to X_m equal to $1 - \alpha$.

The parameters of the regression analysis are easy to compute too.

From the values of λ_i it follows that $\det R(\alpha) = (1 + (m-1)\alpha)(1 - \alpha)^{m-1}$, so it is easy to test the definiteness of the correlation matrix of regressors.

Choosing arbitrary X_i for the regressand and taking a complex $X_{j_1}, \ldots, X_{j_k}$ $(X_{j_h} \neq X_i, \ k \leqslant m-1)$ for regressors, we get the regression equation (see [4])

$$X_i = \beta \sum_{h=1}^{k} X_{j_h},$$

where $\beta = \dfrac{\alpha}{1 + (k-1)\alpha}$, $R^2 = \dfrac{k\alpha^2}{1 + (k-1)\alpha}$.

In such a way the condition C3 is fulfilled for the component, factor and regression analyses (in particular, for the step-regression procedures, too).

With some additional parameters it is also possible to find the parameter vector ϑ of the other linear multivariate models. For discriminant analysis the distance d between the group centres is needed, for canonical analysis - the partition of initial vector X into two subvectors etc.

4. The random vector with the given eigenvalues of correlation matrix

Let $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ be given diagonal matrix of eigenvalues, $\sum \lambda_i = m$, $\lambda_i > \lambda_{i+1} \geqslant 0$.

Then it is possible to fix an arbitrary matrix of eigenvectors, $\Gamma = (\gamma_1 \vdots \ldots \vdots \gamma_m)$, $\gamma_j = (\gamma_{1j}, \ldots, \gamma_{mj})'$, fulfilling the following condition

$$\Gamma'\Gamma = I_m \qquad (4)$$

and to calculate the correlation matrix R.

$$R = \Gamma\Lambda\Gamma' \qquad (5)$$

Then Λ and Γ are eigenvalues' and eigenvectors' matrices for R.

This procedure is especially convenient for the case $m = 2^{\ell}$. Then it is possible to define matrix Γ in this way that in addition to (4) the following conditions are fulfilled:

$$\begin{cases} \Gamma = \Gamma' \\ |\gamma_{ij}| = 1/\sqrt{m} \quad (i,j = 1,\dots,m), \\ \gamma_{1j} \geqslant 0, \quad \gamma_{j1} \geqslant 0, \; j = 1,\dots,m \end{cases} \qquad (6)$$

Then R has only $m - 1$ different non-diagonal elements $r_1,\dots,r_{m-1}$, all being simple linear combinations of λ's, and R has the structure of latin square (in every column and every row every element r_i occurs exactly* once).

So as

$$R^{-1} = \Gamma\Lambda^{-1}\Gamma',$$

all the calculations connected with regression analysis may be simply analytically carried out. For instance, if R is the covariation matrix of regressors, and all covariations between the regressand and the regressors equal to β, then $R^2 = \frac{m\beta^2}{\lambda_1}$ (see [4]), where R^2 is the multiple correlation coefficient.

The given parametrization is suitable for the investigation of the behaviour of the statistics of the component and factor analysis. From (6) it follows that all the loadings of every principal component have equal absolute values:

$$|f_{ij}| = \frac{\lambda_i}{\sqrt{m}}, \; j = 1,\dots,m; \; i = 1,\dots,m,$$

where f_{ij} are the loadings of the i-th component.

---

* In the particular case when some of the elements r_i and r_j are equal to each other, of course, in every column and row there are some equal elements.

For defining the optimal parametrization in the sense of paragraph 2 it is necessary to express all the eigenvalues with the help of one parameter. For instance, let us define an arithmetical progression with m nonnegative elements and sum m:

$$\begin{cases} \lambda_i = a + (i - 1)b, \\ \text{where} \\ b = \frac{2(1 - a)}{m - 1}, \quad a \in [0, 2]. \end{cases} \tag{7}$$

If we take again $EX_i = 0$, $DX_i = 1$ and define correlation matrix with the help of equations (5)-(7), then we get the optimal parametrization of vector X, parameters being (m,a)or (b,a).

In the case described it is possible to define and generate the random vector X with the correlation matrix, given by (5)-(7) and with arbitrary equal and symmetrical marginal distributions. Vector X will be constructed as a mixture of m bundle systems, where every bundle system consists from independent pairs of variables. The mixing measure consists of values $r_1, \ldots, r_{m-1}$ and $1 - \sum r_i$;

So as from (5) and (6) the following equation follows

$$\sum_{i=1}^{m-1} r_i + 1 = \lambda_1,$$

and from (7) it can be concluded that $\lambda_1 \leq 2$, it is evident that the given construction is possible.

5. Definition of random variables and vectors with the given fourth moments

Let X have the arbitrary symmetrical distribution of the fourth order.

$$EX^2 = D, \quad EX^4 = K.$$

A. Let M be arbitrary, $M > 1$. Then it is always possible to define such a mixture Z of two random variables hX and HX ($h < 1$, $H > 1$) with the probabilities γ and $1 - \gamma$ respondingly that the following conditions are fulfilled

$$\begin{cases} EZ^i = EX^i, \ i = 1,2,3, \\ EZ^4 = MEX^4 = MK. \end{cases} \tag{8}$$

From the equation for mixture's moments (see [1], (C2)) we get the system of equation:

$$\begin{cases} \gamma h^2 + (1-\gamma)H^2 = 1, \\ \gamma h^4 + (1-\gamma)H^4 = M, \end{cases} \tag{9}$$

that is evidently solvable for $\gamma, h \in [0,1)$, $H, M \in (1,\infty)$, where from 3 unknowns h, H and γ one may be arbitrarity chosen.

If X is random vector with equal marginal distributions, then the mixture Z of hX and HX has the following property,

$$EZ_iZ_jZ_kZ_l = MEX_iX_jX_kX_l, \tag{10}$$

that means, all the fourth moments of vector Z are proportional to these of vector X.

B. For $M < 1$ it is possible to define the mixture Z of "shifted" variables $h(X - m)$, $h(X + m)$(where $\gamma = 1 - \gamma = 0.5$) in such a way that conditions (8) are fulfilled. The solution exists if and only if the Cauchy-Bunjakovski inequality

$$(EZ^2)^2 \leq EZ^4$$

or

$$D^2 \leq MK$$

is fulfilled.

As a rule, the property (10) is not fulfilled in this case.

<u>Example 1</u>[*]. The investigation of the behaviour of the regression parameters.

In this study the influence of the sample size and the value of fourth moments on the convergence rate of regression parameters is investigated.

The parent distributions are.

(i) the 8-dimensional standardized normal distribution with constant correlation matrix $R(\alpha)$, $\alpha = 0.5$.

(ii) the mixture of two 8-dimensional normal distributions with all moments up to the third order equal to these of the case (i), proportionality coefficient M for fourth moments being 2.

---

[*] The numerical results in the examples 1, 2 are obtained on the computer EC-1060 with fortran-programs written by two students of mathematical faculty E.Sillat and K.Karpender. For the generation of pseudorandom numbers the subroutines RANDU and GAUSS were used.

The sample sizes were 8, 32, 128 and 512, the process was repeated 10 times.

In every cycle the regression equation with 7 regressors and multiple correlation coefficient were found. The theoretical confidence intervals for the regression parameters were found with the help of the results of Parring [3].

For every cycle the mean of the mean of 7 (theoretically equal) sample regression coefficients and the mean of sample multiple correlation coefficients were found. The empirical frequencies of belonging parameters to their confidence intervals were counted, too. All experiments were repeated 2 times.

In the Table 1 the distances between empirical and theoretical parameters are presented.

The distances between the empirical and theoretical regression parameters

TABLE 1

n	Normal I exp.		Normal II exp.		Mixture I exp.		Mixture II exp.	
	b	R^2	b	R^2	b	R^2	b	R^2
8	do not exist	do not exist	0.052	0.5617	0.0028	0.4683	0.521	0.2045
32	0	0.0731	0.0001	0.1209	0.0265	0.2138	0.0447	0.0318
128	0.0149	0.0694	0.0031	0.0577	0.1177	0.4189	0.0356	0.1583
512	0.0003	0.0069	0.0018	0.0094	0.0007	0.0043	0.0042	0.0229

For characterizing the convergence rate let us use the maximum distance for sample sizes equal or greater than given n and compute the mean of two experiments.

In Table 2 we present these mean maximum distances $\bar{d}$ and the percentage of belonging to confidence intervals for characteristics studied:

TABLE 2

n	Normal b		Mixture b		Normal R^2		Mixture R^2	
	$\bar{d}$	%	$\bar{d}$	%	$\bar{d}$	%	$\bar{d}$	%
8	-	51	-	61	0.562	70	0.336	90
32	0.009	57	0.070	83	0.097	90	0.280	70
128	0.009	84	0.076	74	0.084	90	0.289	70
512	0.001	94	0.002	91	0.008	100	0.014	100

As a result we can conclude that the tendency of convergence is more rapid in the case of normal distribution. On the basis of frequencies in Table 2 it was not possible to find any considerable difference between the two types of distribution.

<u>Example 2</u>. Monte-Carlo study of the distribution of the eigenvalues of sample correlation matrix

A Monte-Carlo experiment was carried out, first to compare the sampling mean and covariance matrix of the eigenvalues of sample correlation matrix with the asymptotic ones, and secondly to test the univariate and joint normality of the eigenvalues. The effect of parent distribution and sample size n was investigated.

The 4-dimensional random vector Z with given eigenvalues (an arithmetical progression)

$$\Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \mathrm{diag}\,(1.9,\ 1.3,\ 0.7,\ 0.1)$$

and given moments

$$\begin{aligned} &EZ = 0; \\ &DZ_i = 1,\ i = 1,\dots,4; \\ &EZZ' = P; \\ &EZ \otimes Z' \otimes Z = 0 \end{aligned} \qquad (11)$$

was observed. The correlation matrix

$$P = \begin{pmatrix} 1 & -0.3 & 0 & 0.6 \\ & 1 & 0.6 & 0 \\ & & 1 & -0.3 \\ & & & 1 \end{pmatrix} \qquad (12)$$

was determinated by Λ with the equations (5), (6). The fourth moment

$$M_4(Z) = EZ \otimes Z' \otimes Z \otimes Z'$$

is different for each observed distribution of Z:

(i) a multivariate uniform distribution (U);

(ii) a multivariate normal (N);

(iii) a mixture of multivariate normals (M), defined by (8), (9), where

$\gamma = 0.8$, $h = 0.5$, $H = 3$, $M = 2$.

Sample sizes n = 20, 80, 360, 500 were used.

First the independent vector X with moments (11), where $EXX' = I$, and with distributions (i), (ii) or (iii) was generated. With the help of linear transformation matrix

$$W = \Gamma\Lambda^{1/2}$$

the vector

$$Z = WX$$

with the given theoretical correlation matrix P was obtained. The eigenvalues $l = (l_1,l_2,l_3,l_4)'$ of the correlation matrix R based on the generated values of Z were computed. The process was repeated k = 100 times. The sample mean $\bar{l} = (\bar{l}_1,\bar{l}_2,\bar{l}_3,\bar{l}_4)'$ of the eigenvalues and the sample covariance matrix $\bar{\beta} = (\bar{\beta}_{ij})$ of the random vector $\sqrt{n}(l-\lambda)$ were calculated. Their distances from asymptotical ones were observed. The univariate normality of the eigenvalues was tested with the help of Kolmogorov-Smirnov criterion. The ellipsoids

$$U_{1-\alpha} = (l - \bar{l})'\, n\bar{\beta}^{-1}(l - \bar{l}) \leq C_\alpha$$

of equal density of the distribution $N(\bar{l}, \frac{1}{n}\bar{\beta})$ were used to estimate the multivariate normality of the eigenvalues. The percentile points C_α of the distribution $X^2_{\alpha,3}$ were taken by the values $\alpha = 0.5, 0.3, 0.1$.

The asymptotic covariance matrix β of $\sqrt{n}\,(l - \lambda)$ is given in [2]:

$$\beta = \xi'\varkappa\xi,$$

where

$$\varkappa = M_4(Z) - \mathrm{vec}\, P(\mathrm{vec}\, P)'$$

and ξ is a matrix function of Δ and Γ.

The fourth moment $M_4(Z)$ is expressed through the fourth moment of X [5]:

$$M_4(Z) = (W \otimes W)\, M_4(X)(W \otimes W)'.$$

For the distributions (i), (ii), (iii) the fourth moments are:

$$M_4(Z_U) = (W\otimes W)\big[\, I_{m^2} + I_{m,m} + \mathrm{vec}\, I_m(\mathrm{vec}\, I_m)' - 1\cdot 2\,(I_{m,m})\mathrm{diag}\big](W\otimes W)';$$

$$M_4(Z_N) = (W\otimes W)\,\big[I_{m^2} + I_{m,m} + \mathrm{vec}\, I_m(\mathrm{vec}\, I_m)'\big]\,(W\otimes W)';$$

$$M_4(Z_M) = 2\cdot M_4(Z_N) \quad \text{(by the equation (10))},$$

where $I_{m,m}$ is the permuted identity matrix, $m' = 4$.

The asymptotic covariance matrices β for the observed distributions and for the correlation matrix (12) are presented in Table 3.

The asymptotic covariance matrices β

TABLE 3

Uniform distribution	1.784	-1.422	-0.367	0.005
		1.951	-0.497	-0.032
			0.860	0.004
				0.024
Normal distribution	2.978	-2.161	-0.765	-0.052
		2.408	-0.250	0.003
			0.992	0.023
				0.026
Mixture of normal distributions	5.957	-4.323	-1.529	-0.105
		4.817	-0.500	0.007
			1.985	0.046
				0.053

The results of the described experimental study are presented in Tables 4-8. The influence of the distribution type and the influence of the sample size is measured by the distance between empirical and asymptotic values of the observed statistics.

The sampling and asymptotic theoretical means of the eigenvalues. The distances are $d = \max_{1 \le i \le 4} |\bar{l}_i - \lambda_i|$

TABLE 4

Distribution	$\bar{l}_i$ \ n	20	80	320	500	Asymptotic λ_i
1	2	3	4	5	6	7
Uniform	$\bar{l}_1$	1.949	1.923	1.900	1.899	1.9
	$\bar{l}_2$	1.319	1.301	1.301	1.304	1.3
	$\bar{l}_3$	0.648	0.681	0.700	0.697	0.7
	$\bar{l}_4$	0.084	0.095	0.099	0.099	0.1
	d_U	0.052	0.023	0.001	0.004	$\emptyset$

1	2	3	4	5	6	7
Normal	$\bar{l}_1$	2.020	1.921	1.885	1.895	1.9
	$\bar{l}_2$	1.219	1.288	1.313	1.300	1.3
	$\bar{l}_3$	0.673	0.694	0.702	0.705	0.7
	$\bar{l}_4$	0.088	0.097	0.100	0.099	0.1
	d_N	0.120	0.021	0.015	0.005	Ø
Mixture	$\bar{l}_1$	2.128	1.960	1.901	1.894	1.9
	$\bar{l}_2$	1.191	1.273	1.297	1.302	1.3
	$\bar{l}_3$	0.596	0.672	0.703	0.704	0.7
	$\bar{l}_4$	0.084	0.095	0.099	0.100	0.1
	d_M	0.228	0.060	0.003	0.006	Ø

The sampling and asymptotic variances of $\sqrt{n}\,(l_i - \lambda_i)$.
The distances are $d = \max_{1 \le i \le 4} |\bar{\beta}_{ii} - \beta_{ii}|$.

TABLE 5

Distribution	n / $\bar{\beta}_{ii}$	20	80	320	500	Asymptotic β_{ii}
Uniform	$\bar{\beta}_{11}$	0.659	0.858	1.115	0.902	1.784
	$\bar{\beta}_{22}$	0.613	0.841	1.086	0.705	1.951
	$\bar{\beta}_{33}$	0.393	0.360	0.339	0.356	0.860
	$\bar{\beta}_{44}$	0.010	0.012	0.010	0.009	0.024
	d_U	1.338	1.110	0.865	1.146	Ø
Normal	$\bar{\beta}_{11}$	1.422	2.496	2.282	2.158	2.978
	$\bar{\beta}_{22}$	0.840	1.642	1.843	2.324	2.408
	$\bar{\beta}_{33}$	0.669	0.908	0.832	0.879	0.992
	$\bar{\beta}_{44}$	0.025	0.031	0.031	0.026	0.026
	d_N	1.568	0.766	0.696	0.820	Ø
Mixture	$\bar{\beta}_{11}$	2.136	3.849	4.246	5.150	5.957
	$\bar{\beta}_{22}$	1.073	2.875	4.438	4.227	4.817
	$\bar{\beta}_{33}$	0.936	1.675	2.003	2.186	1.985
	$\bar{\beta}_{44}$	0.018	0.039	0.039	0.045	0.053
	d_M	3.821	2.108	1.711	0.807	Ø

The distances of sampling covariance matrices of $\sqrt{n}\ (l - \lambda)$ from asymptotic ones: $d = \max_{1 \le i,j \le 4} |\bar{\beta}_{ij} - \beta_{ij}|$

TABLE 6

Distribution	d \ n	20	80	320	500
Uniform	d_U	1.338	1.110	0.865	1.146
Normal	d_N	1.568	0.766	0.696	0.820
Mixture	d_M	3.821	2.108	1.711	0.837
	$d_{M/N}$	1.335	0.871	2.030	2.172

The distance $d_{M/N}$ is calculated between the sample covarince matrix $\bar{\beta}$ based on the mixture-type distribution and asymptotic β based on the normal distribution.

The Kolmogorov-Smirnov distances d_i between the empirical distribution function of the eigenvalue l_i and $N(\bar{l}_i, \frac{1}{n}\bar{\beta}_{ii})$

TABLE 7

Distribution	n	20	80	320	500
Uniform	d_1	0.237	0.189	0.142	0.160
	d_2	0.191	0.166	0.123	0.187
	d_3	0.205	0.202	0.138	0.153
	d_4	0.292	0.209	0.145	0.176
Normal	d_1	0.225	0.069	0.124	0.080
	d_2	0.211	0.083	0.119	0.051
	d_3	0.123	0.063	0.060	0.075
	d_4	0.192	0.179	0.104	0.087
Mixture	d_1	0.285	0.159	0.099	0.141
	d_2	0.215	0.135	0.141	0.114
	d_3	0.239	0.175	0.134	0.139
	d_4	0.270	0.185	0.131	0.129

The critical value for testing hypothesis H_1: the distribution of $l_i \neq N(\bar{l}_i, \frac{1}{n}\bar{\beta}_{ii})$, is 0.134; $\alpha = 0.05$, $k=100$.

Frequencies of sampling vectors l belonging to the confidence ellipsoids $U_{0.5}$, $U_{0.7}$, $U_{0.9}$

TABLE 8

Distribution	n	20	80	320	500	theoretical frequence
Uniform	$U_{0.5}$	46	52	52	58	50
	$U_{0.7}$	69	77	74	77	70
	$U_{0.9}$	92	90	95	96	90
Normal	$U_{0.5}$	50	50	57	55	50
	$U_{0.7}$	71	68	70	71	70
	$U_{0.9}$	92	96	87	91	90
Mixture	$U_{0.5}$	49	47	53	47	50
	$U_{0.7}$	68	68	73	72	70
	$U_{0.9}$	91	93	91	91	90

Table 3 shows how the fourth moment of parent distribution effects to the asymptotic covariance matrix β of $\sqrt{n}$ $(l - \lambda)$. To the increasing M_4 $(M_4(Z_U) < M_4(Z_N) < M_4(Z_M))$ corresponds increasing β. As a rule the variances of greater eigenvalues are greater than the variances of smaller eigenvalues.

From Table 4 it follows that the sample mean $\bar{l}_i$ is converging to λ_i, when n is increasing. The rate of convergence is greater when the fourth moment M_4 is smaller.

The sampling variances in Table 5 differ significantly from asymptotic values when n is small. Hence, in the case of small samples, it is not admissible to find the confidence intervals of the eigenvalues with the help of asymptotic variances. It is interesting that by small n, the sampling variances are smaller (not greater) than the asymptotic values. When n is increasing, the sampling variances are also increasing converging to its asymptotic values. The fourth moment of parent distribution has a remarkable influence on the variances of the eigenvalues. The behavior of the sample covariance matrices $\bar{\beta}$ is analogical to that of the sample variances.

From Tables 5,6 it follows that in the case of small n it is better to estimate the variances and covariance matrix of $\sqrt{n}$ $(l - \lambda)$ with asymptotic β which corresponds to the distribution with the smaller fourth moment than the parent distribution has. For instance, by n = 20, 80 and mixture type parent distribution, it is better to use β corresponding to normal parent distribution.

From Table 7 it appears that the parent distribution influences the convergence rate of marginal distributions of eigenvalues l_i to normal distribution $N(\bar{l}_i, \frac{1}{n}\bar{\beta}_{ii})$. By the normal population the convergence is quicker. For instance, by n = 320 we can take the hypothesis H_0 for all l_i. If the parent distribution is different from the normal, there exists some l_i the distribution of which is not $N(\bar{l}_i, \frac{1}{n}\bar{\beta}_{ii})$, even when n = 500.

The frequencies in Table 8 do not sharply differ from theoretical values. Thus, the multivariate distribution of eigenvalues is quite close to $N(\bar{l}, \frac{1}{n}\bar{\beta})$. The effects of parent distribution and sample size are not noticeable in these frequencies.

References

1. Tiit, E.-M., Definition of random vector with given marginal distribution and given correlation matrix.

2. Колло Т., О применении асимптотических распределений в модели главных компонент. Тезисы докладов II всесоюзной научно-технической конференции, Тарту 1981, 282-283.
3. Парринг А.-М., Оценка функции линейной регрессии и ее асимптотическое поведение. Уч. зап. Тартуск. ун-та, 1979, 492, 91-99.
4. Тийт Э.А., Об априорном выборе регрессоров в линейной регрессионной модели. Заводская лаборатория, 1983, № 7, 44-48.
5. Траат И., Моменты выборочной ковариационной матрицы. Труды ВЦ ТГУ, 1984, 51

ПРИМЕНЕНИЕ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА ПРИ ИССЛЕДОВАНИЯХ МЕТОДОМ МОНТЕ-КАРЛО В МНОГОМЕРНОМ СТАТИСТИЧЕСКОМ АНАЛИЗЕ

Э.Тийт, И.Траат

Р е з ю м е

Метод Монте-Карло является весьма перспективным при ис-

следовании поведения разных статистик в многомерном анализе. Для повышения эффективности таких исследований целесообразно пользоваться результатами теории планирования эксперимента. Тогда "точками плана" являются совокупности выборок, генерированные заданными параметрами, а откликом – разность между истинным значением исследуемого параметра и значением, его оценки (статистики), найденной по данной совокупности выборок. Факторы – объем выборки, размерность исходного вектора и его теоретическое распределение, разные заранее фиксируемые параметры исходного вектора и исследуемой модели.

Применению этих идей на практике препятствует весьма большое количество возможных параметров, а иногда и сложность выражения отклика через факторы и недостаточная точность первого.

С целью решения упомянутых проблем целесообразно ввести некоторую оптимальную параметризацию исходного вектора, которая минимизирует число меняющихся параметров вектора X и позволяет практически точно вычислить значения параметров модели.

В настоящей статье приводятся две такие параметризации: векторы с постоянной корреляционной матрицей $R(\alpha)$ и векторы, имеющие корреляционную матрицу, собственные значения которой образуют арифметическую прогрессию. Оба типа применимы для случайных векторов любой размерности и любого (равного и симметрического) маргинального распределения.

Приведены примеры пользоваться изложенных идей в регрессионном (пример I) и факторном (пример 2) анализах.

В примере I исследуется поведение множественного коэффициента корреляции и коэффициентов регресии в случае когда исходный вектор имеет постоянную корреляционную матрицу, либо нормальное распределение, либо смесь нормальных с четвертым моментом $2M_4$, где M_4 – четвертый момент нормального распределения.

В примере 2 изучается распределение собственных значений выборочной корреляционной матрицы методом Монте-Карло. Рассматривается 4-мерный случайный вектор Z с фиксированными собственными значениями и с фиксированными первыми моментами, притом распределение вектора Z одно из следующих:

(i) многомерное равномерное (U)

(ii) многомерное нормальное (N);

(iii) смесь многомерных нормальных (M).

Генерируются выборки вектора ξ с объемами n = 20, 80, 320, 500. При каждом n вычисляются собственные значения корреляционной матрицы к = 100 раз. Их выборочные характеристики приведены в таблицах 4–8.

Выяснилось, что распределение вектора ξ влияет на поведение моментов собственных значений. Чем больше четвертый момент вектора ξ, тем большие дисперсии в таблице 5, притом дисперсии отличаются значительно от асимптотических значений, если n маленькое. Распределение вектора ξ влияет на маргинальное распределение собственных значений (таблица 7). Скорость сходимости к нормальному закону быстрее при нормальном векторе ξ. Из частот в таблице 8 следует, что многомерное распределение собственных значений близко к $N(\bar{l}, 1/n\,\bar{\beta})$, притом объем выборки и распределение генеральной совокупности не оказывает влияние.

Received Juny 84

A MONTE-CARLO STUDY OF THE DISTRIBUTION OF SOME CLUSTERING CRITERIA

K. Pärna, M. Raus

1. Introduction

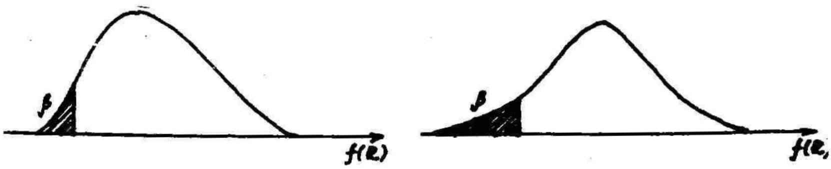
The clustering problem often rises as an optimization problem: one has to determine the partition R of a given set $A = \{a_1, a_2, \ldots, a_n\}$ into mutually exclusive classes or clusters, which minimizes (without loss of generality) some clustering criterion $f = f(R)$. It is well-known that the global extremum is effectively obtainable only in the case of special criterion such as within-cluster variance [2]. This is caused by a too large number of partitions of n elements even for relatively small values of n. For example, in the case of n = 30 elements we have approximately $8.5 \cdot 10^{23}$ distinct partitions. For detailed discussion of the problem see [4], ch.6 or [1], ch.2.

In our opinion, one of the possible ways to overcome the difficulties mentioned above, is the method of random search. Using this method, one has to generate (in some random way) partitions $R_1, R_2, \ldots, R_n$ and then find the 'best' of these in the sense of the given clustering criterion.

In order to determine the sample size N needed in random search, one must fix the portion β of 'best' partitions. Here, the 'best' partition is one of those, for which the value of clustering criterion is reasonably close to the optimum. Now, it is an easy exercise to show that for obtaining at least one 'best' partition with probability $1 - \alpha$, the sample size of

$$N(\alpha, \beta) = \left[\frac{\log \alpha}{\log (1 - \beta)} \right] + 1$$

is needed. From this formula we can conclude that the larger portion β of 'best' partitions, the smaller sample size is needed in random search. But it must be taken into account that the large value of β is justified only in the case when the corresponding 'best' partitions are good indeed i.e. are close to optimum in the sense of criterion chosen. In terms of probability distributions it means that the distribution of criterion values over all possible partitions must have the short left tail (but short right tail in the maximization problem). Figure 1(a) represents such a favorable distribution of criterion values. On the other hand, Figure 1(b) represents the case of long tailed distribution of criterion values; here the 'best' partition need not give the criterion value near to minimum.



(a) (b)

Fig. 1. Two shapes of distribution for criterion values

In the present paper the forms of the distribution of values of some clustering criteria are established. This is done by Monte-Carlo method. We shall see that the distributions criteria under study have short tails that are of interest. This provides a good reason for random search in clustering.

Note that such a favorable result does not hold for all clustering criteria. Solomon and Fortier detected the long tail distribution for 'Holzinger's coefficient of belongness' criterion values [3].

Section 2 of this paper makes the reader acquainted with the four clustering criteria under study. In sections 3 and 4 we describe our Monte-Carlo experiments and discuss the results.

2. Clustering criteria

We shall consider 4 clustering criteria, including (a) summed Hamming's distance and (b) three information-theoretical criteria. To give a precise description, we shall introduce some notions.

Let the raw data which we have to clusterize be a $(n \times m)$-matrix, the rows of which correspond to n objects and columns to m features. Without loss of generality, these features are considered as measured on nominal scale. Therefore, they merely determine m partitions $R_1, R_2, \ldots, R_m$ on the set A of n objects $A = \{a_1, a_2, \ldots, a_n\}$. Let $\pi(A)$ be the set of all possible partitions of A into mutually exclusive nonempty subsets (named as classes or clusters). The number of such partitions is known as Bell's number and will be designated by B_n. Let R be a partition of A. Upper indices will designate the classes of the partition. So, the partition R has classes $R^1, R^2, \ldots, R^k$ $(k \leq n)$ and the partition R_i has classes $R_i^1, R_i^2, \ldots, R_i^{k_i}$ $(k_i \leq n;\ i = 1,2,\ldots,m)$. Let class sizes be $n^1, n^2, \ldots, n^k$ and $n_i^1, n_i^2, \ldots, n_i^{k_i}$, accordingly. The product of partitions R and R_i is defined as the partition $RR_i \in \pi(A)$ with classes $R^r \cap R_i^s$ and class sizes n_i^{rs}. We have at most $k \times k_i$ nonempty classes for product-partition. Of course, we may write

$$\sum_{r=1}^{k} n^r = \sum_{s=1}^{k_i} n_i^s = \sum_r \sum_s n_i^{rs} = n.$$

Now we are ready to define the clustering criteria.

(a) Summed Hamming's distance.

Hamming's distance $d(R, R_i)$ between partitions R and R_i is the number of 'disconcordant' pairs (u,v) of objects of of the set A,

$$d(R, R_i) = \sum_{u,v=1}^{n} |r(u,v) - r_i(u,v)|,$$

where

$$r(u,v) = \begin{cases} 1, & \text{if objects u and v belong} \\ & \text{to the same class in R,} \\ 0, & \text{if objects u and v belong} \\ & \text{to the distinct classes in R,} \end{cases}$$

and $r_i(u,v)$ is the same for partition R_i.

The more practical formula is

$$d(R, R_i) = \sum_r (n^r)^2 + \sum_s (n_i^s)^2 - 2\sum_r \sum_s (n_i^{rs})^2,$$

(see [5], p.50).

Our first clustering criterion is the summed Hamming's distance

$$f_1(R) = \sum_{i=1}^{m} d(R,R_i)$$

which must be minimized over all partitions $R \in \mathcal{R}(A)$.

(b) Information-theoretical criteria.

Let $p^r = n^r/n$, $p_i^s = n_i^s/n$ and $p_i^{rs} = n_i^{rs}/n$ be the relative frequencies (probabilities) of the classes of partitions R, R_i and product partition RR_i. To define entropies of these partitions we follow the Shannon formula:

$$H(R) = -\sum_{r} p^r \log p^r$$

$$H(R_i) = -\sum_{s} p_i^s \log p_i^s$$

$$H(RR_i) = -\sum_{r,s} p_i^{rs} \log p_i^{rs} .$$

The quantity

$$I(R,R_i) = H(R) + H(R_i) - H(R,R_i)$$

is known as the mutual (or transmitted) information of partitions R and R_i. It indicates the strength of one-to-one correspondence between classes of R and R_i. We normalize the mutual information by $H(R)$, $H(R_i)$ or $H(RR_i)$ to produce clustering criteria:

$$f_2(R) = \sum_{i=1}^{m} I(R,R_i)/H(R),$$

$$f_3(R) = \sum_{i=1}^{m} I(R,R_i)/H(R_i)$$

and

$$f_4(R) = \sum_{i=1}^{m} I(R,R_i)/H(RR_i) .$$

(See [6] for interpretation of these criteria.)

Functionals $f_2(R)$, $f_3(R)$ and $f_4(R)$ must be maximized in the clustering process.

3. Evaluation procedure

Our Monte-Carlo experiment consists of the following steps:

(1) forming a sample of random partitions $R_1, R_2, \ldots, R_N$ to simulate uniform probability distribution on the set of all partitions $\mathcal{R}(A)$;

(2) calculate the values of criteria $f_1(R), \ldots, f_4(R)$ for obtained random partitions;

(3) construct empirical distributions for the values of each criterion.

In turn, the generator of random partitions (step (1)) consists of four parts: (1) generator of the number of clusters k, (2) generator of the random set cluster frequencies $\{n^1, n^2, \ldots, n^k\}$, (3) generator of the random permutation of objects $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$, and (4) partition of obtained permutation into the clusters in accordance with obtained frequencies. Necessary combinatorial considerations were taken into account, of course, to warrant equal chances for all partitions in $\mathcal{R}(A)$ to be chosen into the sample. Two distinct numbers of objects, $n = 10$ and $n = 30$ were used in our experiments. In both cases $N = 1000$ random partitions were generated. Still, this is a negligible number compared with Bell's numbers of all possible partitions $B_{10} \approx 1.2 \cdot 10^6$ and $B_{30} \approx 8.5 \cdot 10^{23}$. Initial partitions $R_1, \ldots, R_6$, corresponding to certain sociological attribute-variables, were used.

The experiment was carried out on ES-1022 large scale computer[1]. The time required made up some minutes in the case of $n = 10$ and about an hour in the case of $n = 30$.

4. Results

On Fig. 2 obtained empirical distributions of criteria values are represented. The left and the right columns of figures correspond to the case $n = 10$ and $n = 30$, accordingly. A thousand random values, obtained for every criterion are classified into 32 classes of the same length depending on the criterion.

Firstly, the reader can see that the left and the right columns of figures are quite similar, i.e. the form of the distribution of f(R) does not depend on n significantly. Secondly, the forms of the obtained distributions are quite close to that of the normal distribution. Thirdly, the most important disclosure is that the tails being of interest are

[1] All programs are written by the second author.

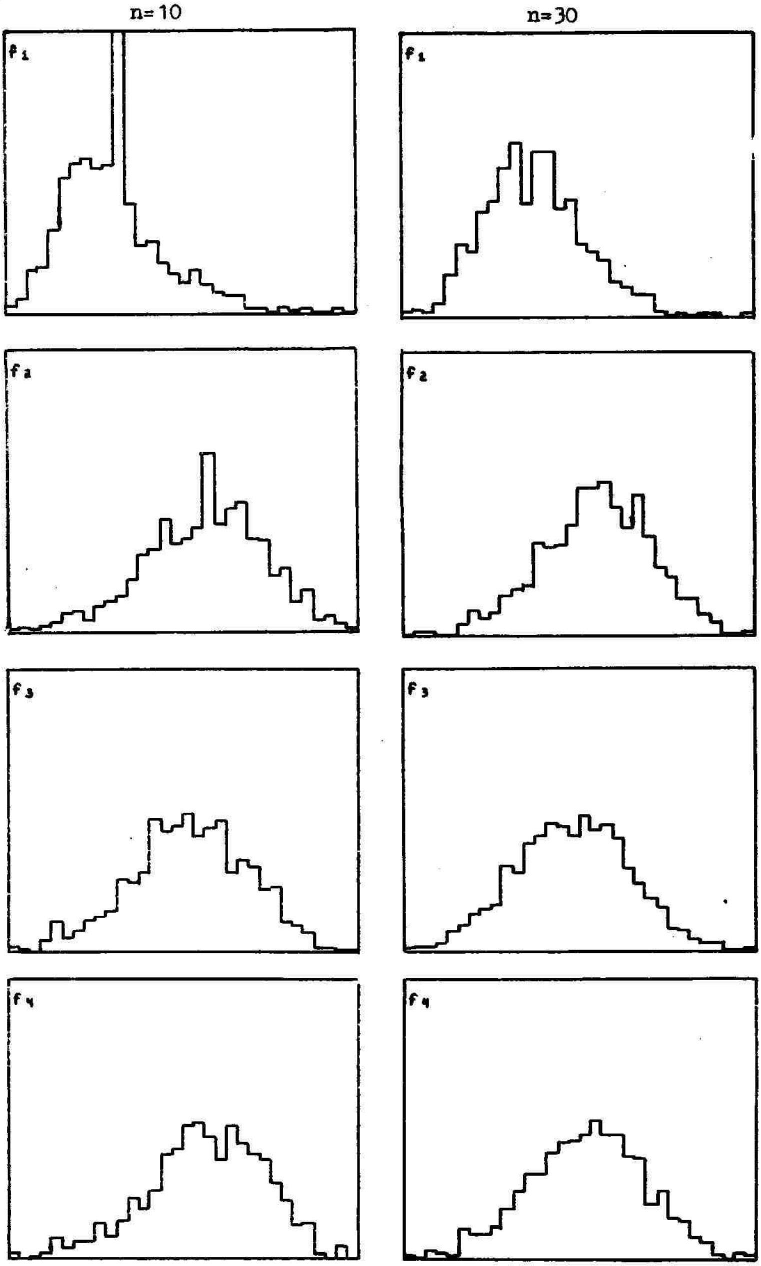


Fig. 2. Empirical distributions of the values of four clustering criteria

not long for these distributions (such tails are marked with an asterisk). Hence, the portion of the best partitions may be considered large. Now, we can state the main result of this work: random search seems to be an effective method in cluster analysis, if above-given criteria are exploited.

Still, some obscurities remain. It is not clear how the form of the distribution depends on initial data $R_1, \ldots, R_m$, if any significant dependence exists. In addition, the obtained results cannot be extended automatically to other clustering criteria. These problems require further investigation.

References

1. Duran, B.S., Odell, P.L. Cluster Analysis. A Survey. Springer-Verlag, Berlin e.a., 1974.
2. Fisher, W.D. On grouping for maximum homogeneity. Journal of the American Statistical Society, 1958, vol. 53, pp. 789-798.
3. Fortier, J.J., Solomon, B.B. Clustering Procedures. In "Multivariate Analysis I" P.R.Krishnaiah (Ed.), Academic Press, N.Y., 1966, pp. 493-506.
4. Hartigan, J.A. Clustering Algorithms. Wiley, New-York, 1975.
5. Миркин Б.Г. Анализ качественных признаков и структур. М., Статистика, 1980.
6. Пярна К.А. Метод факторного анализа для номинальных признаков. В сб. "Опыт применения прикладных методов математики и вычислительной техники в народном хозяйстве". М., Статистика, 1978, стр. 80-90.

ИЗУЧЕНИЕ РАСПРЕДЕЛЕНИЯ НЕКОТОРЫХ КРИТЕРИЕВ КЛАСТЕРИЗАЦИИ МЕТОДОМ МОНТЕ-КАРЛО

К.А.Пярна, М.С.Раус

Резюме

Рассматривается следующая задача кластерного анализа: найти разбиение R конечного множества А, при котором заданный целевой функционал достигает минимального значения. Одной возможностью решения этой задачи является метод случайного поиска. При этом нужно, чтобы распределение значе-

ний целевого функционала имело короткий левый хвост. В данной работе методом Монте-Карло изучается форма этого распределения для четырех функционалов: суммарное расстояние Хэмминга и три информационно-теоретических функционала. Эмпирические распределения, полученные на основе тысячи случайных разбиений, изображены на рис. 2.

Received Juny 84

CLUSTERING AS A PROCESS OF FINDING OF KERNELS OF MONOTONIC SYSTEM

R. Kääremaa

The solving of a classification problem is the examination of all possible partitions of the set of operational taxonomic units (OTU's) to find the partition, which gives optimum to the objective function. Usually the search of the partition corresponding to the optimum of the objective function is realized in such a manner which allows to find this partition without looking through all possible partitions. One of the ways for searching the classification is the use of monotonic systems theory [3].

In this article we deal with such a classification of OTU's, which is based on the dissimilarity (or similarity) matrix. The classification process, in which the first step is the construction of a likeness matrix on the units to be classified and the second step is the construction of a classification on these units, is usually called cluster-analysis; producing classes, which may be non-overlapping as well as overlapping, and are called clusters. In this article the theory of monotonic systems has been used with the aim to realize the cluster-analysis problem in such a manner which on the one hand gives a new way for the construction of objective functions and finding their global extremums and on the other hand it is a generalization of various kinds of cluster-methods.

1. Monotonic system, kernel

A system is defined as a pair (W, π_W), where W is a finite set of elements and π_W is a function which associates with each element α of W a non-negative real number $\pi_W(\alpha)$. The value $\pi(\alpha)$, measuring the importance (relevance) of the element α in the system, is called a weight of α and function π_W - a weighting function on W. The weighting function is defined on each subset W' of W.

It is said that there is θ - influence exerted on the subsystem $W' \subset W$ if it induces changing of weights of all elements $\alpha \in W/W'$ only in one direction - non-increasingly. It is said that there is $\oplus$-influence exerted on the subsystem $W' \subset W$ if it induces changing of weights of all elements $\alpha \in W/W'$ non-decreasingly.

The system (W, π_W) is called to be monotonic if $\ominus$-influence (or $\oplus$-influence) is defined for each subsystem $W'' \subseteq W' \subset W$, i.e. accordingly

$$\pi_{W/W''}(\alpha) \leq \pi_{W/W'}(\alpha) \leq \pi_W(\alpha)$$

(or $\pi_{W/W''}(\alpha) \geqslant \pi_{W/W'}(\alpha) \geqslant \pi_W(\alpha)$).

In case of $\ominus$-influence (or $\oplus$-influence) we denote a set of weights of the elements of the subsystem $V \subseteq W$ by $\{\pi^-_V(\alpha) \mid \alpha \in V\}$ (or $\{\pi^+_V(\alpha) \mid \alpha \in V\}$). We define a function on each subsystem $V \subseteq W$ as

$$F_-(V) = \min_{\alpha \in V} \pi^-_V(\alpha) \qquad (\text{or } F_+(V) = \max_{\alpha \in V} \pi^+_V(\alpha)).$$

The kernel is the subsystem of W on which F_- has global maximum (or F_+ has global minimum). The union of kernels is called to be the maximum kernel.

2. Searching of the kernel of the monotonic system

In this section we present an algorithm for searching of the maximum kernel of the monotonic system. The algorithm is given in the elementary form and because of that it contains much computational redundancy, but in such a form it performs the principles of searching of the kernel in a more understandable way.

1. Fix the elements of the set W, define the weighting function π_W and $\ominus$ (or $\oplus$) -influence or the instruction for the recalculation of weights. Γ_i denotes a subset of W. If $|W| = N$, then $i \in [0, N-1]$. $\Gamma_0 = W$. In case of $\ominus$-influence we denote the weight of the element α of the set Γ_i by $\pi^-_{\Gamma_i}(\alpha)$ (in case of $\oplus$-influence accordingly by $\pi^+_{\Gamma_i}(\alpha)$).
2. $i = 0$.
3. In the set Γ_i find the element μ_i so that

$$\pi^-_{\Gamma_i}(\mu_i) = \min_{\delta \in \Gamma_i} \pi^-_{\Gamma_i}(\delta)$$

$$(\text{or } \pi^+_{\Gamma_i}(\mu_i) = \max_{\delta \in \Gamma_i} \pi^+_{\Gamma_i}(\delta)).$$

(i.e. can be linked by a chain of links each less than or equal to r).

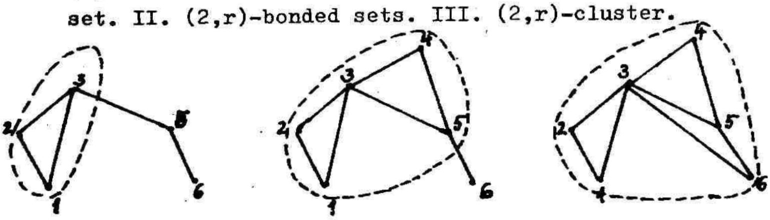
2. A nonempty subset S of the set of OTU's X is (K,r)-bonded, if for each $x \in S$, there is a K-element subset T of S for which $x \notin T$ and $d(x,t) \leq r$ for all $t \in T$. Find B_r, the maximal (K,r)-bonded set of X.
3. If $B_r \neq \emptyset$, decompose B_r into r-connected components (clusters).
4. If a new cluster is found, save all relevant information about this cluster (r, cluster size, identification of the elements).
5. Increase r by 1 and repeat steps 2 - 4 until the largest cluster X is reached.

Now we present an illustrative example [2] for finding the (2,r)-clustering using the graph representation. On a graph the vertices represent the OTU's and the edges join namely those pairs of vertices which represent OTU's with dissimilarity $\leq$ r.

Matrix of ranks

	2	3	4	5	6
1	1	5	15	12	9
2		2	13	10	11
3			6	3	8
4				7	14
5					4

(2,r)-clustering of six OTU's. I. Maximal r-connected set. II. (2,r)-bonded sets. III. (2,r)-cluster.



r = 5	r = 7	r = 8
I. {1,2,3,5,6}	I. {1,2,3,4,5,6}	I. {1,2,3,4,5,6}
II. {1,2,3}	II. {1,2,3}, {3,4,5}, {1,2,3,4,5}	II. {1,2,3}, {3,4,5}, {3,5,6}, {3,4,5,6}, {1,2,3,4,5}, {1,2,3,4,5,6}
III. {1,2,3}	III. {1,2,3,4,5}	III. {1,2,3,4,5,6}

4. Fix the level $\varepsilon_i = \pi^-_{\Gamma_i}(\mu_i)$ (or $\varepsilon_i = \pi^+_{\Gamma_i}(\mu_i)$). Place μ_i into the set W_i as the first element of this set.
5. In the set Γ_i/W_i find element ξ, satisfying the condition

$$\pi^-_{\Gamma_i/W_i}(\xi) \leq \varepsilon_i \quad (\text{or } \pi^+_{\Gamma_i/W_i}(\xi) \geq \varepsilon_i).$$

6. If in the set Γ_i/W_i is such an element ξ, place it into the set W_i and repeat step 5. Otherwise, go to step 7.
7. If the remaining set Γ_i/W_i is not empty, go to step 8. Otherwise, the kernel $\Gamma = \Gamma_i = W_i$ is found on level ε_i.
8. Increase i by 1, define set $\Gamma_i = \Gamma_{i-1}/W_{i-1}$ and go to step 3.

The kernel is a subsystem, the elements of which are very sensitive to exerting of $\ominus$- or $\oplus$-influence. Removing of such a subsystem changes the whole nature of the system. If the kernel Γ of the system (W, π_W) is found, it is possible to eliminate the influence of the elements of the kernel on the other elements completely or partially, and in the received system it is possible to find the kernel again. If such a process is repeated, a sequence of kernels, accordingly non-intersect (non-overlapping) or intersect (overlapping) will be received.

3. Searching of clusters by some well-known methods

It is possible to present various kinds of cluster-methods using the monotonic systems theory so that searching of clusters is successive searching of kernels of monotonic system. In this section we refer to the two well-known algorithms (in the elementary form) for the construction of clusters which we shall present in the light of monotonic systems theory in the next section.

3.1. Clustering algorithm proposed by Ling [2].

1. Convert matrix of dissimilarities D, computed on n OTU's, to the corresponding matrix of ranks Δ as follows: the rank of the element of d_{ij} of matrix D is given by the position of d_{ij} in the resulting vector, when all elements of D are sorted into non-descending order of magnitude. Let K be given, $K \in [1, n-1]$, and $r = \binom{K+1}{2}$, the smallest rank that can give rise to a (K,r)-cluster. The (K,r)-cluster has the property that each of its elements is within a distance r of at least K other elements of the same cluster and the entire set can be r-connected

3.2. Clustering algorithm proposed by Jardine and Sibson [1]. It is an algorithm for applying the method B_k, $k \in [1, n-2]$, for clustering n OTU's.

1. For the set X of n OTU's a dissimilarity coefficient d can take a maximum of $n(n-1)/2$ distinct values. Fix the value k. List the subsets of X with exactly $k+2$ elements in an arbitrary order.
2. Consider the value of dissimilarity coefficient d taken on each pair in the first subset. If d takes a unique maximum value on a single pair of the subset reduce this value to the next value taken by d on any pair from the subset. Otherwise leave d unchanged.
3. Repeat the process on the next subset starting with the modified d. Continue until all the subsets have been considered.
4. Repeat 2 and 3 until the list can be run through without further modification of d. The resultant dissimilarity coefficient is $B_k(d)$.

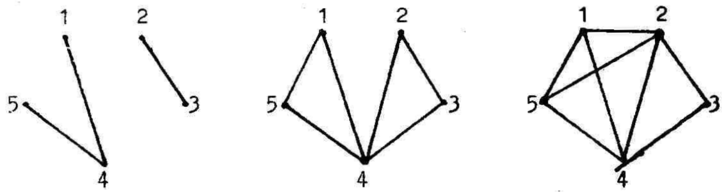
The construction of the clusters on the level h from $B_k(d)$ is graph-theoretically presented in the following way. Each OTU may be represented by a vertex on a graph, and all pairs of vertices which correspond to pairs of OTU's having a dissimilarity at largest h are connected. All maximal complete subgraphs are found and all pairs of such subgraphs that intersect in at least k vertices are further connected. The method B_k, $k > 1$, induces overlapping clusters since two intersecting maximal complete subgraphs which have less than k overlap vertices are distinguished as separate clusters. As at $k = 1$ no overlap occurs, the procedure is identical to single linkage.

We present an illustrative example [1] for carrying out method B_2.

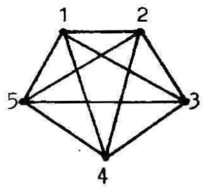
Matrix of dissimilarities.

	2	3	4	5
1	0.3	0.4	0.1	0.2
2		0.1	0.2	0.3
3			0.2	0.4
4				0.1

A graph representation of the dissimilarity coefficient.

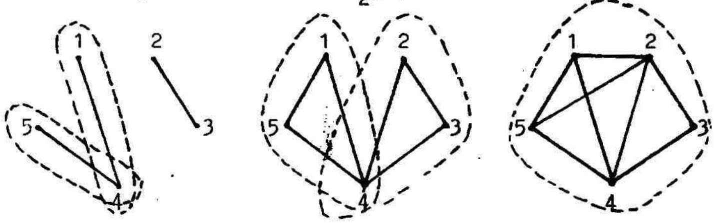


level 0.1 level 0.2 level 0.3



level 0.4

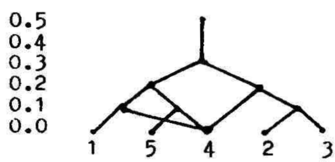
A graph representation of $B_2(d)$.



level 0.1 level 0.2 level 0.3

A numerical characterization of $B_2(d)$ and 'tree' diagram representing $B_2(d)$.

	2	3	4	5
1	0.3	0.3	0.1	0.2
2		0.1	0.2	0.3
3			0.2	0.3
4				0.1



4. Searching of clusters as searching of the kernels of the monotonic system

In this section we construct kernels which are identical to clusters formed by Ling's method and by B_k-method. We use the algorithm described above for searching the kernels of the monotonic system.

4.1. For receiving the result, corresponding to the application of Ling's algorithm, we take OTU's x_i, $i = 1,2, \dots,n$, as the elements of set W. Thus $W = \{x_1,x_2,\dots,x_n\}$, $|W| = n$.

Let the dissimilarity matrix on n OTU's, having $n(n-1)/2$ elements, and the corresponding matrix of ranks Δ be given. A rank Δ_{ij} numerically presents a tie between elements x_i and x_j.

We define the weight of the element $x_i \in W$ as follows. Fix the value $r \in [1,n(n-1)/2]$. The weight of x_i is the number of ties between x_i and all other OTU's having numerical value of tie less than or equal to r.

θ-influence exerted on element x_i is the removal of the element x_i from the set W; thus each tie between x_i and other elements of W is removed too.

We take the example of section 3.1 and demonstrate searching of the kernel of the monotonic system, defined above, by algorithm given in section 2.

$W = \{1,2,3,4,5,6\}$.

Let r = 5, then

$\pi^-_W(1) = 2,\ \pi^-_W(2) = 2,\ \pi^-_W(3) = 3,$

$\pi^-_W(4) = 0,\ \pi^-_W(5) = 1,\ \pi^-_W(6) = 1.$

If $i = 0$, then $\Gamma_0 = W$; $\varepsilon_0 = 0$; $W_0 = \{4\}$.

If $i = 1$, then $\Gamma_1 = \Gamma_0/W_0 = \{1,2,3,5,6\}$; $\varepsilon_1 = 1$; $W_1 = \{5,6\}$.

If $i = 2$, then $\Gamma_2 = \Gamma_1/W_1 = \{1,2,3\}$; $\varepsilon_2 = 2$; $W_2 = \{1,2,3\}$.

Now the set $\Gamma_2/W_2 = \emptyset$ and the maximal kernel $\Gamma = \{1,2,3\}$, got on level $\varepsilon_2 = 2$. It is identical to maximal (K,r)-bonded set in Ling's sense, where K = 2, r = 5.

Let r = 7, then

$\pi^-_W(1) = 2,\ \pi^-_W(2) = 2,\ \pi^-_W(3) = 4,$

$\pi^-_W(4) = 2,\ \pi^-_W(5) = 3,\ \pi^-_W(6) = 1.$

If $i = 0$, then $\Gamma_0 = W$; $\varepsilon_0 = 1$; $W_0 = \{6\}$.

If $i = 1$, then $\Gamma_1 = \Gamma_0/W_0 = \{1,2,3,4,5\}$; $\varepsilon_1 = 2$; $W_1 = \{1,2,3,4,5\}$.

Now the set $\Gamma_1/W_1 = \emptyset$ and the maximal kernel $\Gamma = \{1,2,3,4,5\}$, got on level $\varepsilon_1 = 2$. It is identical to maximal (K,r)-bonded set, where $K = 2$, $r = 7$.

Let $r = 8$, then

$$\pi^-_W(1) = 2,\ \pi^-_W(2) = 2,\ \pi^-_W(3) = 5,$$

$$\pi^-_W(4) = 2,\ \pi^-_W(5) = 3,\ \pi^-_W(6) = 2.$$

If $i = 0$, then $\Gamma_o = W$; $\varepsilon_o = 2$; $W_o = \{1,2,3,4,5,6\}$.

Now the set $\Gamma_o/W_o = \emptyset$ and the maximal kernel $\Gamma = \{1,2,3,4,5,6\}$, got on level $\varepsilon_o = 2$. It is identical to maximal (K,r)-bonded set, where $K = 2$, $r = 8$.

To find (K,r)-clusters Ling decomposes maximal (K,r)-bonded set into r-connected components (which correspond to clusters). Analogically it is possible to find kernels belonging to the maximal kernel.

4.2. For receiving the result, corresponding to the application of B_k-method, we take the pairwise ties (x_i,x_j), $i \neq j$, $i,j = 1,2,\ldots,n$, on n OTU's as the elements of set W. Thus $W = \{(x_1,x_2),(x_1,x_3),\ldots,(x_{n-1},x_n)\}$, $|W| = n(n-1)/2$.

Let the dissimilarity matrix on n OTU's be given. We define the weight of the element (x_i,x_j) as the value of dissimilarity between the OTU's x_i and x_j.

Fix k and find subsets of W with exactly $(k+2)(k+1)/2$ elements. Among these subsets it is possible to separate subsets including the element (x_i,x_j). Further we shall call such subsets (x_i,x_j)-subsets.

θ-influence exerted on element (x_i,x_j) is expressed as follows. Inspect all (x_i,x_j)-subsets. If in any (x_i,x_j)-subset the weight of the element (x_i,x_j) is equal to maximum over the weights of all elements of this subset, then (x_i,x_j) will get the weight equal to maximal weight of other elements of this subset. Repeat this process until the weight of (x_i,x_j) is not unique maximum over the weights of elements of any (x_i,x_j)-subset and then remove element (x_i,x_j) from the set W. If there is any subset with $(k+2)(k+1)/2$ elements from which all elements with the exception of only one element (x_s,x_t) are removed, this remaining element will obtain the weight equal to the maximum weight of the removed elements of this subset. Then this element (x_s,x_t) will be removed from W too, i.e. θ-influence is exerted on element (x_s,x_t).

We take the example of section 3.2 and demonstrate searching of the kernel of the monotonic system, defined above, by algorithm given in section 2.

$$W = \{(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)\}.$$

$\pi^-_W(1,2) = 0.3$, $\pi^-_W(1,3) = 0.4$, $\pi^-_W(1,4) = 0.1$,

$\pi^-_W(1,5) = 0.2$, $\pi^-_W(2,3) = 0.1$, $\pi^-_W(2,4) = 0.2$,

$\pi^-_W(2,5) = 0.3$, $\pi^-_W(3,4) = 0.2$, $\pi^-_W(3,5) = 0.4$,

$\pi^-_W(4,5) = 0.1$.

If $i = 0$, then $\Gamma_0 = W$; $\varepsilon_0 = 0.1$; $W_0 = \{(1,4),(2,3),(4,5)\}$.

If $i = 1$, then $\Gamma_1 = \Gamma_0/W_0 = \{(1,2),(1,3),(1,5),(2,4),(2,5),(3,4),(3,5)\}$; $\varepsilon_1 = 0.2$; $W_1 = \{(1,5),(2,4),(3,4)\}$.

If $i = 2$, then $\Gamma_2 = \Gamma_1/W_1 = \{(1,2),(1,3),(2,5),(3,5)\}$; $\varepsilon_2 = 0.3$; $W_2 = \{(1,2),(2,5),(1,3),(3,5)\}$.

Now the set $\Gamma_2/W_2 = \emptyset$ and the maximal kernel $\Gamma = \{(1,2),(2,5),(1,3),(3,5)\}$, received on level 0.3.

This kernel is the first kernel Γ' of the monotonic system W. To receive the kernel on the next level we remove the elements of the first kernel Γ' from the set W, so $W' = \{(1,4),(1,5),(2,3),(2,4),(3,4),(4,5)\}$. Applying the algorithm of searching of the kernel on the system $(W', \pi_{W'})$ we receive the second kernel $\Gamma'' = \{(1,5),(2,4),(3,4)\}$ on level 0.2, and on system $(W'', \pi_{W''})$, where $W'' = \{(1,4),(2,3),(4,5)\}$, - the third kernel $\Gamma''' = \{(1,4),(2,3),(4,5)\}$ on level 0.1.

The kernels consist of ties, because the elements of W are ties between OTU's. In order to find clusters on OTU's, it is necessary to pass on from ties to OTU's. The analogical situation arises in case of carrying out B_k - method, where it is necessary to find clusters on OTU's on the basis of the modified matrix of dissimilarities $B_k(d)$.

5. Cluster-analysis in the light of monotonic systems theory

The previous examples demonstrated the possibility to regard cluster-analysis as a process of searching the kernels of the monotonic system, which has elements either OTU's or pairwise ties on OTU's. In practice it is difficult to separate OTU's and their ties. If any OTU is similar to many other OTU's, this OTU has a relatively great weight in the system, simultaneously their ties have relatively great weights too.

It is not possible to carry out both methods given above without taking into consideration the inherent compatibility of OTU's and their ties. In case of Ling's method the ties for finding of r-connected components (clusters) in the maximal (K,r)-bonded set are needed. In case of B_k-method, going over from ties to OTU's takes place.

We set up the problem of searching of the kernels in the monotonic system (W, π_W), which has both the OTU's and their pairwise ties as the elements of W.

The question arises, whether it is possible to define the weighting function so that the values of the weights of all the elements (OTU's and ties) are located in the same interval (for example in interval $[0,1]$), and whether it is possible to define θ- or $\bar{\theta}$-influence on all the subsets of W so that the monotony of the system will be guaranteed. It appears that it is possible to present π_W and θ- or $\bar{\theta}$-influence in several ways so that these demands are satisfied.

We present one of these possibilities of defining π_W and θ-influence, which has proved to be suitable for various kinds of clustering.

Let the elements of W be n OTU's $\{x_1, x_2, \ldots, x_n\}$ and $n(n-1)/2$ pairwise ties on n OTU's $\{(x_1,x_2),(x_1,x_3),\ldots,(x_{n-1},x_n)\}$, thus

$$W = \{x_1,\ldots,x_n,(x_1,x_2),\ldots,(x_{n-1},x_n)\}.$$

Let the weight of the element (x_i,x_j) be equal to the similarity so, that $\pi_W(x_i,x_j) = 1 - \rho_{ij}$, where ρ_{ij} is the taxonomic distance between OTU's x_i and x_j. We remind that in case of m attributes, measured on each object, the taxonomic distance

$$\rho_{ij} = \sqrt{\frac{\sum_{l=1}^{m} (x_{il} - x_{jl})^2}{m}} .$$

We define the weight of OTU x_j in system W as

$$\pi_W(x_j) = \sqrt{\frac{\sum_{i=1}^{n} (1 - \rho_{ij})^2}{n}} .$$

Thus we received the system (W, π_W), where $W = \{\bar{\alpha}, \bar{\beta}\}$, $\bar{\alpha}$ is the set of OTU's, $\bar{\alpha} = \{x_1, \ldots, x_n\}$, $\bar{\beta}$ is the set of pairwise ties, $\bar{\beta} = \{(x_1, x_2), \ldots, (x_{n-1}, x_n)\}$ and weighting function π_W is defined so

$$\pi_W(\gamma_p) = \begin{cases} \sqrt{\frac{\sum_{i=1}^{n} (1 - \rho_{ij})^2}{n}} , & \text{if } \gamma_p \in \bar{\alpha},\ \gamma_p \equiv x_j \\ 1 - \rho_{ij}, & \text{if } \gamma_p \in \bar{\beta},\ \gamma_p \equiv (x_i, x_j). \end{cases}$$

Let θ-influence exerted on any element of the system be the removal of this element from the system. If θ-influence is exerted on OTU, each tie between this OTU and other OTU's is removed too. The weight of the removed element (either OTU or tie) will obtain the value 0.

It is possible to generalize the choice of elements into W. The OTU's, units to be clusterized, are usually chosen as objects (presented by attributes), but they may also be attributes (presented by objects). In general, the elements of W may be objects and pairwise ties on objects, or attributes and pairwise ties on attributes, or objects and their pairwise ties and attributes and their pairwise ties. On the set W it is necessary to define the weighting function and θ- or Θ-influence on all the subsets of W so that the monotony of the system will be guaranteed.

It is possible to present various kinds of interpretations of the cluster-analysis, modifying the elements of W and the weighting function π_W in the monotonic system (W, π_W).

References

1. Jardine, N., Sibson, R., The construction of hierarchic and non-hierarchic classifications. The Computer Journal, 1968, 11, 177-184.

2. Ling, R.F., On the theory and construction of K-clusters, The Computer Journal, 1972, 15, 326-332.
3. Муллат И.Э., Экстремальные подсистемы монотонных систем. Автоматика и телемеханика, 1976, 5, 130-139.

Кластеризация как процесс нахождения ядер монотонной системы

Р.Ээремаа

Р е з ю м е

Рассматривается задача кластер-анализа объектов (или признаков) как задача поиска ядер монотонной системы. Дается соответствующий алгоритм и примеры получения ядер, идентичных с кластерами, найденными методами Линга и Жардина-Сибсона. Представляется возможность обобщения монотонной системы для кластеризации объектов и признаков одновременно.

Received Juny 84

A SIMULATION-BASED STUDY OF TWO FINITE MARKOV CHAIN CRITERIA

T. Möls

1. Introduction and problems

In this paper we consider a finite discrete stationary stochastic process $x(t)$ (a chain) with states $a_1,\dots,a_K$ and probabilities $p_i = P\{x(t)=a_i\}$, $p_{ij} = P\{x(t)=a_i \wedge x(t+1)=a_j\}$, $p_{ijk} = P\{x(t)=a_i \wedge x(t+1)=a_j \wedge x(t+2)=a_k\}$. Two statistical problems concerning chains are treated: (1) testing the Markov property and (2) testing whether the observations on a chain's trajectory are made randomly. In both cases only one asymptotic criterion (formulae (8) and(10)) is investigated. Our aim is to get some idea of how these criteria work in small samples, where the asymptotics is not guaranteed. Due to the complexity of the problem, only binary (K=2) chains are considered, and only by using the Monte-Carlo modelling.

The Markov Property Criterion. If $x(t)$ has the Markov property, then

$$\sum_j p_{ij}p_{jk}/p_j = \sum_j p_{ijk} . \qquad (1)$$

Substituting all probabilities in (1) with corresponding frequencies (in some set of chain's trajectory fragments), one can test whether the chain has the Markov property or not. A natural idea is to exploit the statistic

$$H^2 = 2\sum_{i,j} n_{i\cdot k}\ln(n_{i\cdot k}/\tilde{n}_{i\cdot k}), \qquad (2)$$

where $n_{i\cdot k}$ and $\tilde{n}_{i\cdot k}$ are observed and expected (if proposed (1)) frequencies of the triplet $\langle a_i a_j a_k \rangle$ (after summation over the index, denoted by point). If data form a contingency table and (1) holds, then (2) is asymptotically (in $n \to \infty$) χ^2-distributed with $(K-1)\times(K-1)$ degrees (see [1],ch. 8.7) . But in chains data do not form a contingency table because of the triplets $\langle a_i a_j a_k \rangle$ are dependent and, moreover, inequalities like $n_{ij} > n_{ij\cdot}$ hold. Therefore an experimental study is needed to make clear whether in real situations (2) has a χ^2-distribution. A further problem is, how the distribution of (2) depends upon the deviation from the markovity.

We have studied by means of Monte-Carlo technique the distribution of a variant of (2) (see (8)) in binary chains (K=2), varying sample size N (the lenght of a single uninterrupted trajectory interval), two transition probabilities $q_{1/i} = P\{x(t)=a_1 \mid x(t-1)=a_i\}$ (i=1,2) and a parameter τ of nonmarkovity (defined later). The empirical significance level or power of the corresponding criterion was approximated(after a suitable transformation)with a second order polynomial (see Table 1), using three-level plan of Box and Benken as given in [3].

The Random Sampling Criterion. If data are collected from short trajectory intervals (replications),the sampling time may depend on chain's state. In ergodic case the initial state of a random replication is invariantly distributed for the chain. This fact leads to the criterion of form

$$h^2 = 2\sum_i n_i' \ln (n_i'/n' q_i) , \tag{3}$$

where q_i is the chain's invariant distribution, n_i' is the frequency of replications with initial state a_i and n' is the number of all used replications. The statistical behaviour of (3) depends greatly on various factors and may be predicted only in simulation experiments.

We have studied the distribution of a variant of (3) (see (10)) in binary chains, varying the length N of replications, the distribution on replications' starting states (p_1), the invariant distribution in chain (q_1) and the transition probability $q_{1/1}$. The method was the same as for the Markov Property Criterion.

2. Modelling technique

To generate a nonmarkov chain with transition probabilities $q_{i/j}$, we use the following way. Denote $p_{k/j}$ some conditional probability for the jump $j \to k$ and let

$$q_{k/ji} = \tau p_{k/i} + (1-\tau) p_{k/j} \tag{4}$$

where $q_{k/ji} = P\{x(t)=a_k \mid x(t-1)=a_j \wedge x(t-2)=a_i\}$ and $\tau \in [0,1]$ is a parameter of nonmarkovity (if $\tau=0$ then (4) defines a Markov chain). The formula (4) gives a convenient way to construct random trajectory with required nonmarkovity τ and transition probabilities $q_{i/j}$. But at first the probabilities $p_{k/j}$, needed in generation must be expressed in terms of required probabilities $q_{i/j}$. Inverting of (4) produces the necessary formula

$$P = Q(\tau A + (1-\tau)I)^{-} , \tag{5}$$

where $P=(p_{i/j})$, $Q=(q_{i/j})$, $A_{ij}= q_{j/i}q_i/q_j$ and B^{-} denotes the generalized inverse for B. In binary case (5) leads to the formulae

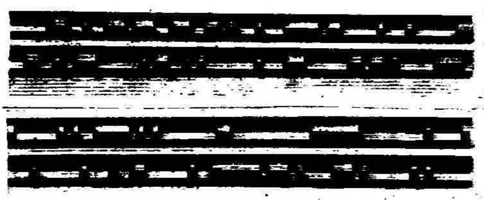
$$p_{1/1} = 1 - (1-\tau)q_{2/1}/(1-\tau(q_{1/2}+q_{2/1})) \tag{6'}$$

and

$$p_{1/2} = (1-\tau)q_{1/2}/(1 - \tau(q_{1/2}+q_{2/1})) . \tag{6''}$$

In Figure 1 two trajectories are shown, one with $\tau=0$, the other with $\tau=0.4$. The notable feature of this figure

a



b

Fig.1. Examples of trajectories: a - Markov chain, b - nonmarkov chain, generated from (4). In both cases $q_{1/1}=0.7$ and $q_{1/2}=0.5$.

is that with naked eye one cannot discriminate convincingly between Markov and nonmarkov chains.

Some notations are necessary before we shall represent the Monte-Carlo modelling results. Suppose that M independent modelling trials are made, each consisting of R trajectory intervals (replications). Denote s_{mrk} the chain's state at the moment k in the replication r within the trial m. Further let us denote the frequencies

$$N_i^m = \sum_{r,k}(s_{mrk}=a_i) , \tag{7'}$$

$$N_{ij}^m = \sum_{r,k}(s_{mrk}=a_i \wedge s_{mr(k+1)}=a_j) , \tag{7''}$$

$$N_{ijk}^m = \sum_{r,k}(s_{mrk}=a_i \wedge s_{mr(k+1)}=a_j \wedge s_{mr(k+2)}=a_k) . \tag{7'''}$$

Now the statistic (2) may be rewritten for a chain's trajectory m as follows:

$$H_m^2 = 2\sum_{i,k}^{(o)} N^m_{i\cdot k} \ln\left(N^m_{i\cdot k} \Big/ \left[\frac{N^m_{i\cdot\cdot}}{N^m_{i\cdot}} \sum_j^{(o)} \frac{N^m_{ij} N^m_{jk}}{N^m_{j\cdot}}\right]\right), \qquad (8)$$

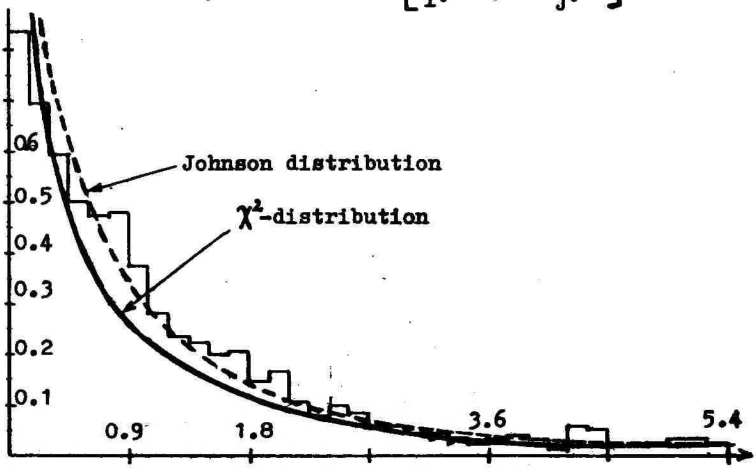


Fig.2. The empirical probability distribution of the criterion (8) in 1000 trials for binary Markov chain with $q_{1/1}$=0.7 and $q_{1/2}$=0.5.

where (o) indicates, that failing (not existing) addends are omitted. Figure 2 shows the empirical probability distribution of (8) in 1000 trials for a binary chain, each trial consisting of five 5-moment long replications. The expected χ^2-distribution with 1 degree and a Johnson distribution, fitted to the empirical data by the method of Slifker and Shapiro [2], are shown, too. We can conclude, that χ^2-distribution is not quite perfect for the (8).

Let next define the frequencies of starting states in m-th replication,

$$N_i^{m\prime} = \sum_r (s_{mr1} = a_i) ,$$

and the estimates for transition probabilities

$$q^m_{i/j} = \begin{cases} N^m_{ji}/N^m_{j\cdot} & \text{if } N^m_{j\cdot} \neq 0 \\ \delta_{ij} & \text{if } N^m_{j\cdot} = 0 \end{cases}$$

($\delta_{ij}=(i=j)$). Denote by $\tilde{q}_i^m$ the invariant probability distribution for the matrix $(\tilde{q}^m_{i/j})$, i.e. the solution of

$$\sum_j \tilde{q}^m_{i/j}\tilde{q}^m_j = \tilde{q}^m_i \ . \tag{9}$$

Now the statistic (3) may be written for the m-th trial as follows:

$$h_m^2 = 2 \sum_i^{(o)} N_i^{m\prime} \ln(N_i^{m\prime}/R\tilde{q}_i^m) \ . \tag{10}$$

Here, if (9) has many solutions, that one which minimizes (10) must be taken.

3. Results for Markov Property Criterion

In this section the empirical power of the criterion (8) is estimated. After fixing the required significance level α, the actual level was estimated in a Monte-Carlo experiment. We varied four factors, each on three levels:

$$\begin{cases} x_1 = \sqrt{N} & = 3 \text{ or } 5 \text{ or } 7 \\ x_2 = q_{1/1} & = 0.5 \text{ or } 0.7 \text{ or } 0.9 \\ x_3 = q_{1/2} & = 0.25 \text{ or } 0.5 \text{ or } 0.75 \\ x_4 = \tau & = 0 \text{ or } 0.4 \text{ or } 0.8 \end{cases} \tag{11}$$

A Box-Benken plan on cube prescribes the following 27 combinations of these levels (denoted here respectively by L, O and H) (see [3], p. 131):

x_1	H	L	H	L	0	0	0	0	L	H	L	H	0	0	0	0	0	0	0	0	L	H	L	H	0	0	0
x_2	L	L	H	H	0	0	0	0	0	0	0	0	L	H	L	H	L	H	L	H	0	0	0	0	0	0	0
x_3	0	0	0	0	H	L	H	L	0	0	0	0	L	L	H	H	0	0	0	0	L	L	H	H	0	0	0
x_4	0	0	0	0	L	L	H	H	L	L	H	H	0	0	0	0	L	L	H	H	0	0	0	0	0	0	0

100 trials were modelled for each plan's point, which produces the empirical probability distribution of the statistic (8). Now the empirical power of the criterion was determined as the relative frequency of cases, in which the statistic jumps over the α-quantile of the χ^2-distribution (we used $\alpha=0.1$, 0.05 and 0.025). Here we attempted to improve the power estimates by smoothing the empirical distribution with a Johnson curve. But this method, very useful in some situations, was not effective in our case.

For every α the 27 values of empirical power were transformed in $\arcsin\sqrt{\ }$-scale and then approximated with a second

TABLE 1

The polynomials, approximating arcsin$\sqrt{\ }$ -transformed empirical power of the Markov Property Criterion (8) in cube (11)

Argument	α=0.1	α=0.05	α=.025
1	0,89	0.67	0.90
$\sqrt{N}$	-0.27	-0.19	-0.24
$q_{1/1}$	0.26	0.52	0.14
$q_{1/2}$	1.02	0.53	0.30
τ	-2.10	-2.54	-2.30
N	0.0084	0.0026	0.0096
$\sqrt{N}q_{1/1}$	0.23	0.19	0.14
$\sqrt{N}q_{1/2}$	-0.081	-0.078	-0.051
$\sqrt{N}$	0.26	0.31	0.28
$q^2_{1/1}$	-0.54	-0.74	-0.41
$q_{1/1}q_{1/2}$	-2.38	-1.80	-1.37
$q_{1/1}$	0.85	1.15	1.10
$q^2_{1/2}$	1.32	1.32	1.17
$q_{1/2}$	-0.51	-0.33	-0.58
τ^2	1.07	0.90	0.93
χ^2(method 1)	56.0	66.5	64.7
χ^2(method 2)	55.1	66.8	64.8

order polynomial of arguments x_1, x_2, x_3, x_4. The estimated polynomial coefficients are given in Table 1. After inverse transformation, the polynomial values give power estimates.

The approximation quality in frequency data may be expressed in terms of χ^2, using two methods. The first method starts from the residual sum of squares (denote SQ) between the arcsin$\sqrt{\ }$-transformed data and polynomial. Taking into account that a single arcsin$\sqrt{\ }$-variate is approximately normal distributed with an asymptotic variance 1/4M (M = 100 trials), the statistic

$$\chi^2_{\text{arcsin}} = 4M \cdot SQ \tag{12}$$

is asymptotically χ^2-distributed with 27 - 15 = 12 degrees (if the null hypothesis is true).

The other way is to sum up the single-degree information chi-squares for all plan's points. The corresponding statistic is

$$\chi^2_{inf} = 2\sum_{i=1}^{27}(\nu_i \ln(\nu_i/\tilde{\nu}_i) + (M-\nu_i)\ln((M-\nu_i)/(M-\tilde{\nu}_i))), \quad (13)$$

where ν_i and $\tilde{\nu}_i$ are respectively the observed and calculated frequencies of rejecting the null hypothesis. Both methods, though based on different ideas, lead to similar results (see tables 1 and 3).

The main conclusion from our experiment is that a second order polynomial is not adequate to fit the power function of the test (8) in cube (11). Nevertheless, comparing the observed and prognosed values in plan points (Table 2), one can see a concordance, which is sufficient for many practical purposes.

TABLE 2

Observed power of the Markov Property Criterion (α=5%, 100 trials), compared with the power, evaluated from Table 1. The χ^2 shows a bad fit.

Plan point	Observed power%	Calculated power%	Plan point	Observed power%	Calculated power%
1	4	5.0	15	15	12.4
2	7	2.5	16	1	2.4
3	11	12.5	17	6	6.8
4	1	0.0	18	2	0.4
5	8	9.4	19	5	14.3
6	3	5.7	20	22	27.9
7	42	27.6	21	2	2.1
8	44	33.3	22	14	18.3
9	7	11.4	23	4	5.1
10	3	0.4	24	8	12.9
11	0	2.3	25	7	4.1
12	66	58.5	26	2	4.1
13	5	2.8	27	4	4.1
14	9	11.1	$\chi^2(12)$ = 66.8		

4. Results for Random Sampling Criterion

The criterion (10) was treated by the same method as the previous one. The controlled parameters were the length N of replications, the transition probability $q_{1/1}$, the invariant probability q_1 of the state a_1 and the proba-

TABLE 3

The polynomial, approximating arcsin$\sqrt{\ }$-transformed empirical power of the Random Sampling Criterion (10) (α=5%) in the cube (14).

Argument	Coefficient	Argument	Coefficient
1	-0.107	$Nq_{1/1}$	0.0076
N	0.151	p_1^2	1.48
p_1	-1.26	p_1q_1	-2.04
q_1	0.667	$p_1q_{1/1}$	0.617
$q_{1/1}$	0.691	q_1^2	-0.356
N^2	-0.0085	$q_1q_{1/1}$	1.92
Np_1	-0.0611	$q_{1/1}^2$	-1.87
Nq_1	-0.0345		
χ^2(method 1) = 11.91		χ^2(method 2) = 11.22	

TABLE 4

Observed power of the Random Sampling Criterion (α=5%, 100 trials), compared with the power, evaluated from Table 3. The χ^2 shows a good fit.

Plan point	Observed power %	Calculated power %	Plan point	Observed power %	Calculated power %
1	19	19.6	15	25	29.6
2	7	9.7	16	0	0.2
3	0	0.2	17	22	19.3
4	0	0.6	18	0	0.6
5	1	1.7	19	19	14.5
6	3	3.7	20	1	1.2
7	4	4.9	21	0	0.2
8	0	0.4	22	1	2.0
9	1	0.6	23	4	1.8
10	2	1.4	24	3	2.1
11	0	1.4	25	4	4.3
12	1	0.3	26	3	4.3
13	11	10.2	27	6	4.3
14	7	4.0	$\chi^2(12)$ = 11.7		

bility p_1 of the state a_1, used in the initializing of the replications. The parameters were set on following levels:

$$\begin{cases} x_1 = N = 3 \text{ or } 6 \text{ or } 9 \\ x_2 = p_1 = 0.25 \text{ or } 0.5 \text{ or } 0.75 \\ x_3 = q_1 = 0.4 \text{ or } 0.5 \text{ or } 0.6 \\ x_4 = q_{1/1} = 0.35 \text{ or } 0.5 \text{ or } 0.65 \end{cases} \qquad (14)$$

The other parameters were constant: $R=5$, $M=100$, $\tau=0$.

Results of the Monte-Carlo experiment are given in tables 3 and 4. We can see, that a second order polynomial is an excellent approximation for the power transform, but in small samples the χ^2-distribution is quite far from the true distribution of (10).

References

1. Kullback, S., Information theory and statistics.New-York, 1958.
2. Slifker, J.F., Shapiro, S.S., The Johnson system: selection and parameter estimation. Technometrics, 1980, 22, 234-246.
3. Голикова Т.И., Панченко Л.А., Фридман М.З., Каталог планов второго порядка. М., 1974.

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИМ МОДЕЛИРОВАНИЕМ ДВУХ КРИТЕРИЕВ ДЛЯ КОНЕЧНЫХ ЦЕПЕЙ МАРКОВА

Т. Мелс

Р е з ю м е

Излагаются результаты статистического моделирования критерия Марковости и критерия случайной выборки для бинарных случайных цепей. Эмпирическая мощность критерия, найденная при 27 комбинациях уровней 4-х факторов (длина траекторий, переходные вероятности, немарковость, начальное распределение) преобразуется арксин-формулой, а затем аппроксимируется полиномом 2 степени. Приводятся таблицы коэффициентов для полиномов. Для критерия случайности выборки аппроксимация хорошая.

Received July 84

ABOUT A POSSIBLE GENERALIZATION OF THE EFRON'S BOOTSTRAP

M.Unt

1. Introduction.

In this note Efron's bootstrap method [1] is considered as a method of random distributions.

Let us remind the idea of the original bootstrap method. Let $\{X_1,\ldots,X_n\}$ be a random sample of size n from the population of the unknown probability distribution P_{-1}. Putting mass 1/n at every point of the sample X we construct the sampling distribution P_o. Having specified the random variable of interest $\varphi(P)$, we approximate the distribution of $\varphi(P_{-1})$ by the distributions of $\hat{\varphi}(P_o)$ and $\hat{\varphi}(P_1)$ where P_1 denotes the bootstrap-distribution: the distribution of the random sample of size n from P_o.

2. The random distributions.

To express our basic idea of bootstrap as a method of random distributions more exactly, let us define the random distribution. Let $\mathcal{M}$ be the space of random variables, which is a linear and partially ordered space of functions taking real values (including constant functions). The distribution of a random variable may be expressed as a linear, normed monotonic and continuous functional (a distribution functional) on a suitable space $\mathcal{F}$ of functions (for example the space of bounded and continuous functions C). To generalize this notion for the case of random distributions we have to replace the numerical values of the distribution functional by random variables. In this generalization monotonity of the functional P, $P: F \to \mathcal{M}$ means that if $f_1 \leq f_2$ ($f_i \in F$, $i = 1,2$), then $Pf_1 \leq Pf_2$, linearity means that $P(\alpha f_1 + \beta f_2) = \alpha Pf_1 + \beta Pf_2$, normerity means that $P(1) = 1$, continuity means that if $f_i \downarrow 0$ (pointwise), then $Pf \downarrow 0$; these properties are assumed to be true with probability 1. Using these 4 properties we may extend this functional from domain F to some maximal domain F^*. Such a generalized functional is an operator of random distribution.

The examples of a random distribution are the sampling distribution P_o and the associated bootstrap distribution P_1.

In an ordinary bootstrap [1] the distribution of P_1 depends on the concrete value of P_o. The operator P_1 acts in the following way. Let f be a bounded continuous function and $\{x_1,\ldots,x_n\}$ an observed sample. Then

$$P_1 f = \sum_{i=1}^{n} \frac{1}{n} f(x_i^*),$$

where x_i^* $(i = 1,\ldots,n)$ are randomly chosen from $\{x_1,\ldots,x_n\}$. Of course, if one uses smoothed or symmetrized bootstrap, the operator P_1 is defined differently.

3. The generalization of the bootstrap method

Now we are ready to generalize the bootstrap idea.

Let $P_{-1}, P_o, P_1, \ldots$ denote the sequence of random distributions, where the first distribution P_{-1} is degenerated (non-random). Propose in this sequence the Markov property holds, i.e. the distribution of P_i depends on $P_o,\ldots,P_{j-1},P_j$ $(j < i)$ only through P_j.

A sequence P_{-1}, P_o, $P_1,\ldots$ arises when the bootstrap procedure is applied repeatedly. Then P_{-1} is a fixed probability distribution, P_o is an empirical distribution, P_1 is a bootstrap distribution generated on P_o, P_2 is a bootstrop distribution generated on P_1 and so on.

A sequence's trajectory is a sequence of distributions p_i, where p_i is a realization of the random distribution P_i $(i = 0,1,2,\ldots)$. The Markov property makes it easy to construct a sequence's trajectory.

Denote by φ a bounded functional of interest, assigning to every distribution p_i a number $\varphi(p_i)$, and to every random distribution P_i the mean value $E\,\varphi(P_i)$. For example, φ may be median or some moment.

Def. The sequence of random distributions P_{-1}, P_o, $P_1,\ldots$ is called Efron's sequence,*) if it has Efron property which means that for some non-random function f the following inequality holds:

$$|\varphi(P_{-1}) - E\hat{\varphi}| \le |\varphi(P_{-1}) - E f[\varphi(P_0), \varphi(P_1), \ldots]|$$

where

$$\hat{\varphi} = 2\varphi(P_0) - \varphi(P_1).$$

If the sequence P_i is Efron's one, then only the distributions P_o and P_1 are useful in estimating $\varphi(P_{-1})$. In opposite

*) with apologizements and compliments to Professor B.Efron.

case, the bootstrappings of higher rank should be used for improving the estimation of $\varphi(P_{-1})$.

Two small Monte-Carlo experiments were run to test the Efron property in estimating the mean and variance, while P_{-1} was the exponential distribution, and the rank statistics (min, max) of the uniform distribution, but there is no reason to assert our sequences not to be Efron's ones.

References

1. Efron, B. Bootstrap methods. Another look at the Jackknife. Ann. Statist., 1979, 7, 1-26.

ВОЗМОЖНОЕ ОБОБЩЕНИЕ МЕТОДА БУТСТРЭП ЭФРОНА

М.Унт

Р е з ю м е

Метод Эфрона бутстрэп [I] рассматривается методом случайных распределений. Обобщая понятия функционала распределения, получают понятия: оператор случайного распределения и последовательность случайных распределений. Приведен термин "последовательность Эфрона", где для оценивания функции параметров теоретического распределения нужны только первые члены последовательности.

Received July 84

CONTENTS