

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHII (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

SOI: [1.1/TAS](#) DOI: [10.15863/TAS](#)

International Scientific Journal Theoretical & Applied Science

p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online)

Year: 2017 Issue: 12 Volume: 56

Published: 30.12.2017 <http://T-Science.org>

S.U. Zhanatauov

candidate of physics and mathematical sciences,
Department «Automation and information
technologies», Associate professor,
Noncommercial joint-stock company
"Kazakh national agrarian university" Kazakhstan
sapagtu@mail.ru

SECTION 2. Applied mathematics.
Mathematical modeling.

BLOCK-DIAGONAL CORRELATION MATRICES OF Λ -SAMPLES

Abstract: The article describes the process of modeling block-diagonal correlation matrices and corresponding multidimensional Λ -samples using no more than three methods. One of the methods used is new. Its application was carried out with the use of the new Optimization Problem No. 4 (problems No. 1,2,3 were solved in [10,11]). The dimension n of the correlation matrices is $n=n_1+2k+1$, where $n_1=3,4, \dots, 10$, $k=1, \dots, 10$. In the case $n_1=0$, the cases $n=2k+1$ and $n=2k$ are considered. An example of calculations for $n=6$ is given. Multidimensional Λ -samples can be applied in multidimensional classification problems and cluster analysis problems.

Key words: inverse model of the principal component analysis, Λ -sample

Language: Russian

Citation: Zhanatauov SU (2017) BLOCK-DIAGONAL CORRELATION MATRICES OF Λ -SAMPLES. ISJ Theoretical & Applied Science, 12 (56): 101-111.

Soi: <http://s-o-i.org/1.1/TAS-12-56-17> **Doi:**  <https://dx.doi.org/10.15863/TAS.2017.12.56.17>

БЛОЧНО-ДИАГОНАЛЬНЫЕ КОРРЕЛЯЦИОННЫЕ МАТРИЦЫ Λ -ВЫБОРОК

Аннотация: В статье описан процесс моделирования блочно-диагональных корреляционных матриц и им соответствующих многомерных Λ -выборок с применением не более трех способов. Один из применяемых способов - новый. Его реализация проведена с применением новой Оптимизационной задачи №4 (задачи №1,2,3 решены в [10,11]). Размерность n корреляционных матриц имеет вид: $n=n_1+2k+1$, где $n_1=3,4, \dots, 10$, $k=1, \dots, 10$. В случае $n_1=0$ рассмотрены случаи $n=2k+1$ и $n=2k$. Приведен пример расчетов при $n=6$. Многомерные Λ -выборки могут быть применены в задачах многомерной классификации и в задачах кластерного анализа.

Ключевые слова: блочно-диагональная корреляционная матрица, Λ -выборка

Введение.

Matrix Depot-это расширяемая коллекция тестовых матриц. Сайт [1] предоставляет разнообразную коллекцию тестовых матриц, в том числе параметризованных матриц и реальных матриц размерности n -на- n . Но в задачах кластерного анализа, классификации многомерных объектов [1-3] возможны симметричные блочно-диагональные матрицы с блоками, отличающимися от блоков Жордана (Jordan block) или других видов.

В статье рассматриваются квадратные симметричные блочно-диагональные n -на- n -матрицы, состоящие из блоков 3 видов: из заданного числа n_1 -на- n_1 -блоков ($n_1 > 2$), из 2-на-2-блоков $[j:(j+1)]$, из 1-на-1-блоков («половины блока» $[(j-1):j]$ длины 1, Таблица 2). Элементы

этих блоков могут быть выбраны случайным образом, например, быть случайными числами с равномерным законом распределения $P[0,1]$. Для блоков корреляционной матрицы, у которой значения элементов не превышают 1 по абсолютной величине, а диагональные элементы равны 1. Умножением корреляционной матрицы R_{nn} слева и справа на диагональную матрицу $S_{nn}=\text{diag}(s_1, \dots, s_n)$: $W_{nn}=S_{nn}R_{nn}S_{nn}$ имеем общую симметричную блочно-диагональную матрицу.

Известные тест-матрицы [1,7] нас не удовлетворяют. Например, матрица Наура имеет только 2 неравных 1 собственных числа, остальные равны 1, параметр $f_5(W_{nn})=6/[n(n+1)(5-2n)] < 0$ для $n \geq 3$. Тест-матрица Хансена удобна тем, что имеет небольшое значение параметра $f_3(W_{nn})$, слабо зависит от n , проста для



Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

вычисления ее элементов, точно известны ее набор собственных чисел. Недостатком ее являются постоянство ее параметра $f_5(W_{nn})=1$ и невозможность существенного изменения значения $f_3(W_{nn})$ при фиксированном n . В отличие от матриц специального назначения [1] у наших тест-матриц $R^{(\ell)}_{mn}, \ell=1, \dots, \infty$, имеется несколько преимуществ, полезных для тестирования процедур кластерного анализа [3-5]. Спектры Λ_{nn} корреляционных матриц R_{nn} наших и им соответствующих выборок $Z_{mn}, m>n$, могут иметь значения f -параметров f_1, f_2, \dots, f_6 из допустимого множества их значений. Если имеем матрицу R_{nn} или ее спектр Λ_{nn} , то значения их вычисленных f -параметров попадают всегда внутрь своих интервалов изменения [2,6,7]: $f_1(\Lambda_{nn})/n \leq f_2(\Lambda_{nn}) \leq f_1^2(\Lambda_{nn}), f_1^2(\Lambda_{nn}) \geq n \times f^{(2/n)}_2(\Lambda_{nn}), f_3(\Lambda_{nn}) \times f_5(\Lambda_{nn}) \leq f^{(n/2)}_2(\Lambda_{nn}), f_6(\Lambda_{nn}) \geq (n-1) \times f_3(\Lambda_{nn})^{1/(k-1)}, 0 \leq f_4(\Lambda_{nn}) \leq 1$. Задачи задания доминирующих значений или вычисления либо недоминирующих либо всех элементов спектра Λ_{nn} корреляционных матриц $R^{(\ell)}_{mn}, \ell=1, \dots, \infty$, с заданными значениями наборов f -параметров видов $(f_3, f_6), (f_1, f_5), (f_3, f_4, f_6), (f_1, f_2, f_5), (f_1, f_2, f_4), (f_1, f_3, f_4, f_6), (c \times f_1, c^2 \times f_2, c^n \times f_5), c > 1, (f_1, f_2, f_4, f_6), (f_1, f_2, f_3, f_4, f_5, f_6)$ рассмотрены и решены в работах [2,8-11]. Для каждой корреляционной матрицы $R^{(\ell)}_{mn}, \ell=1, \dots, \infty$, получены модельные многомерные Λ -выборки $Y^{(\ell)}_{mn} \in N_s(0, \Lambda)$, многомерные R -выборки $Z^{(\ell, t)}_{mn} \in N_s(0, R^{(\ell)})$, $t=1, \dots, k_t, \ell=1, \dots, k_\ell$. Случайная n -мерная выборка объема $m>n$ $U^{(\ell)}_{mn} \in N_s(0, I_{nn})$, номер $t=1, \dots, k_t < \infty$, с выборочной корреляционной матрицей I_{nn} и с неизвестным законом распределения с помощью случайных линейных преобразований превращается в многомерные выборки $Y^{(\ell)}_{mn} \in N_s(0, \Lambda), Z^{(\ell, t)}_{mn} \in N_s(0, R^{(\ell)}), t=1, \dots, k_t, \ell=1, \dots, k_\ell$. При всех наборах f -параметров спектра Λ показаны гистограммная и модельная адекватности (R) -, (C, Λ) -, (C, Λ, Y) -выборок реальным выборкам разного объема и числа n переменных [12-14].

Модели и задачи

Модельная блочно-диагональная корреляционная матрица R_{nn} и ей соответствующая многомерная Λ -выборка Z_{mn} по построению удовлетворяют всем соотношениям прямой модели главных компонент (ПМ ГК[2,15]): $Z_{mn} \Rightarrow (R_{nn}, C_{nn}, \Lambda_{nn}, Y_{mn})$. Это не означает применимость ПМ ГК для извлечения цифровых знаний из Λ -выборки Z_{mn} . Модельная Λ -выборка Z_{mn} необходима для тестирования программ решения задач многомерной классификации и задач кластерного анализа. Присутствующие здесь ее известные 2 матрицы n -мерных выборок и 3 матрицы корреляций, дисперсий будут охарактеризованы в процессе изложения. В основе ПМ ГК лежит прямая задача анализа

главных компонент (ПЗ АГК [2-4]), решенная Г.Хотеллингом [15]. В ПЗ АГК решается прямая спектральная задача (ПСЗ) диагонализации симметрической матрицы $[1,2]: R_{nn} \Rightarrow (C_{nn}, \Lambda_{nn})$. С применением ПМ ГК проводятся анализ, интерпретация, формулируются выводы для единственной многомерной стандартизованной выборки Z_{mn} [16-18]. При решении ПЗ АГК матрицы $R_{nn}, C_{nn}, \Lambda_{nn}, Y_{mn}$ вычисляются для имеющейся реальной стандартизованной выборки-матрицы Z_{mn} . Поэтому всегда считаем, что для выборки Z_{mn} всегда существуют и вычислены (или не вычислены) матрицы $R_{nn}, C_{nn}, \Lambda_{nn}, Y_{mn}$ всегда реализована последовательность вычислений $Z_{mn} \rightarrow R_{nn} \rightarrow C_{nn}, \Lambda_{nn} \rightarrow Y_{mn}$. Теорема существования и единственности решения Y_{mn} ПЗ АГК доказана в [15]. Но наша многомерная выборка, моделируемая ниже комбинированным методом, должна быть случайной и с неопределенными значениями.

Стандартизованная многомерная выборка значений модельных z -переменных из разных моделей

С появлением ЭВМ и пакетов прикладных программ (ППП) внимание экспериментаторов сосредоточилось на центральном вопросе «анализа данных» - науки и искусства [2]-об источнике данных. Матрица исходных данных $X^0_{mn} = \{x^0_{ij}\}, i=1, \dots, m; j=1, \dots, n$; как правило преобразуют в стандартизованную многомерную выборку $Z_{mn} = \{z_{ij} = (x^0_{ij} - x_j^{cp})/s_j, i=1, \dots, m, j=1, \dots, n\}$, где x^0_{ij} - i -ое значение j -го признака реального объекта, x_j^{cp} - выборочное среднее, s_j - выборочное стандартное отклонение. Значение x^0_{ij} - число имеет свою единицу измерения, совпадающую с единицей измерения числа x_j^{cp} . При моделировании многомерных выборок, так или иначе адекватных реальной многомерной выборке, удобно моделировать безразмерную выборку Z_{mn} , а затем придать каждой из переменных свою размерность путем преобразования $\{x^0_{ij} = z_{ij}s_j + x_j^{cp}, i=1, \dots, m, j=1, \dots, n\}$. В выборке Z_{mn} моделируются случайность и неопределенность [2]. Это удобство присуще и нашей многомерной Λ -выборке Z_{mn} . Мы избегаем этап центрирования: центрированная x -переменная $(x_{ij} = x^0_{ij} - x_j^{cp})$ определяет значение s_j . Но мы предполагаем известное значение s_j . Тогда это число задает динамику изменения значений x -переменной - величин разностей между соседними значениями и частоту ненулевых разностей. Чем больше разнятся друг от друга эти разности и чем больше таких ненулевых разностей, тем лучше выражена динамика. Тем больше колеблемость (вариабельность) значений одной x -переменной.



Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

Далее имеем стандартизованные данные $Z_{1j}, Z_{2j}, \dots, Z_{mj}$, их среднее арифметическое равно нулю: $(1/m)(z_{1j} + z_{2j} + \dots + z_{mj}) = 0$, а дисперсия равна 1: $(1/m)(z_{1j}^2 + z_{2j}^2 + \dots + z_{mj}^2) = 1$, $j=1, \dots, n$. Каждая из этих z -переменных в этой статье может быть сгенерирована в рамках одной из трех моделей. 1-ая модель - Обратная модель главных компонент [2,7], 2-ая модель-вариант ОМ ГК при $n > 2$, описанная в статье [9] и 3-ая модель - модель из Оптимизационной задачи №4 (смотрите ниже). Вариант ОМ ГК при $n=2$ [9] реализуется для одного выбранного нами коэффициента $r^{(\ell)}$ с номером ℓ . Реализуемая модель имеет вид: $(r^{(\ell)}) = \Rightarrow R^{(\ell)}_{22} = \Rightarrow (\Lambda^{(\ell)}_{22}, C^{(\ell)}_{22}, U^{(\ell)}_{m2}, Y^{(\ell,t)}_{m2}, Z^{(\ell,t)}_{m2})$, $m=20$, $t=1$. Значения 2-х коррелированных z -переменных из матрицы $Z_{m2} = Y_{m2} C_{22}^T$, где i -ая строка матрицы Z_{m2} состоит из элементов

$$z_{i1} = (u_{i1} \times \sqrt{1+r}) \times (-\sin \alpha) + u_{i2} \times \sqrt{1-r} \times \cos \alpha,$$
$$z_{i2} = (u_{i1} \times \sqrt{1-r} \times \cos \alpha - u_{i2} \times \sqrt{1-r} \times \sin \alpha),$$

а ее столбцы удовлетворяют условиям: $\mathbf{z}_1 = (z_{11}, \dots, z_{m1})^T$, $\mathbf{z}_2 = (z_{12}, \dots, z_{m2})^T$, $(1/m) \mathbf{z}_1^T \mathbf{z}_2 = r_{12} = r_{21} = 0.5$, $(1/m) \mathbf{z}_1^T \mathbf{z}_1 = 1, (1/m) \mathbf{z}_2^T \mathbf{z}_2 = 1$. При переносе в матрицу Z_{m6} индексы (1,2) z -переменных примут новые значения (4,5).

В ОМ ГК [2,7,9] решена обратная задача: если известно среднее арифметическое, равное нулю, то существует способ моделирования чисел $\{z_{1j}, z_{2j}, \dots, z_{mj}\}$ с единичной дисперсией и в сумме равных нулю.

Ниже описывается решение задачи равенства суммы произведений (умноженных на постоянное число $1/m$) $z_{ki} \times z_{kj}$, $k=1, \dots, m$, заданному числу-значению коэффициента корреляции $r_{ij} = r_{ji}$ между i -ой и j -ой z -переменной. В оптимизационной задаче №4 решается задача моделирования неизвестных значений j -ой z -переменной z_{kj} , из произведений вида $z_{ki} \times z_{kj}$, $i=1, \dots, n$, при заданных значениях i -ых z -переменных z_{ki} , $k=1, \dots, m$, $i=1, \dots, n$, $i \neq j$, и заданных значениях коэффициентов корреляции $r_{1j}, r_{2j}, \dots, r_{mj}$. При $n=2$ эта задача решена в [9], при $n=1$ - ниже в оптимизационной задаче №4.

Колеблемость (вариабельность) значений одной нашей z -переменной, измеряется дисперсией, равной 1, причем числа $z_{1j}, z_{2j}, \dots, z_{mj}$, случайные в том смысле, что их поведение статистически *неопределенно*, следовательно, ее возможным (и вычисленным) значениям мы можем приписать (назначить субъективно) вероятности $\{p_i\}$, $i=1, \dots, m$. Для оценки вероятностей мы, как обычно, поступим стандартно: так как значений вероятностей возможным (и состоявшимся) значениям мы незнаем, то назначим одинаковые вероятности $\{1/m\}$. Этот подход представлен П. Лапласом в 1795г. Основанием для такого подхода является «принцип недостаточной причины»: если мы не можем определить, который из результатов

$Z_{1j}, Z_{2j}, \dots, Z_{mj}$ является более вероятным, то необходимо назначить им равную вероятность. Из разложения измеренного значения x^0 на компоненты $x^0 = x_{cp} + z s$ видно, что вариабельность зависит от единицы «шкалы измерения вариабельности», равного стандартному отклонению s (оценке для σ). Число таких единиц измерения нецелое и равно $z = (x^0 - x_{cp})/s$. Число отклонений $(x^0 - x_{cp})$ длины s равно значению z . Отметим, что z -значение является оценкой известной случайной величины $\eta = (\xi - \mu)/\sigma$, которое чаще всего находится в интервале $(\mu - 3, \mu + 3)$, если распределение гауссовское, иначе z -значение изменяется в другом интервале. Поэтому, значение z чаще всего находится в интервале $(0 - 3, 0 + 3)$. Вхождение всех z -значений в интервал $(-3, +3)$ указывает на гауссовость распределения случайной величины η , которую мы поставили в соответствие с нашей z -переменной, равной $z = (x^0 - x_{cp})/s$. Если наша z -переменная соответствует случайной величине η , то небольшое количество z -значений, больших 3, являюся «выбросами», «аномальными ми» и подлежат удалению из наших данных. Анализируются только большинство оставшихся z -значений, x^0 - значений. В оптимизационной задаче №4 Шаги к большей реальности в структуре сделаем через моделирование воздействий разных факторов неопределенности в данных, объединенных в таблицу (из m значений n свойств для m объектов, где случайность порождается с помощью компьютера (в датчиках случайных чисел) и случайных отклонений: декоррелирующих (t)-отклонений, случайных «координатных» (ℓ)-отклонений, случайных «совместных» (t, ℓ)-отклонений. Ниже рассмотрим корреляционные матрицы блочно-диагональной структуры и им соответствующие многомерные выборки значений z -переменных. При разных значениях равномерности n . Моделирование как блочно-диагональных корреляционных матриц и многомерных выборок ниже проведено тремя разными способами.

Наличие статистической *неопределенности* в значениях моделируемых переменных (в датчиках базовых распределений) не давало в ранее известных методах моделирования точного равенства значений тех или иных функций от выборки заданным значениям, но удовлетворяло критериям случайности и критериям асимптотической близости к заданным: типу распределения, значению параметра теоретической функции распределения. Но без точного равенства значений тех или иных функций от выборки заданным значениям. Главным в ранее известных моделях было получение выборки с заданным типом закона распределения. Прежде всего – закон, а точности



Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

оценок параметров функции плотности распределения вероятностей обеспечивались за счет объема выборки. Здесь работала теория асимптотической близости значений статистик к параметрам (постоянным) теоретической совокупности. Имитация вероятностных явлений при измерениях здесь полезна и используется как полезный инструмент компьютерного моделирования. Случайность порождается с помощью компьютера. Это повлияло на генерирование случайных многомерных Λ -выборок с заданными алгебраическими свойствами. Ниже мы моделируем каждую i -ую строку $\mathbf{z}_i = (z_{i1}, \dots, z_{in}), i=1, \dots, m, m > n$, тремя разными способами и объединяем одну матрицу $Z^{(t,\ell)}_{m \times 6} = [Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2} | Z^{(L)}_{m \times 1}]$, где номера t, ℓ, L изменяются независимо друг от друга: $t=1, \dots, k_t < \infty, \ell=1, \dots, k_\ell < \infty, L=1, \dots, k_L < \infty$. Например, матрицы $Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2}$ образованы: первая – с применением процедуры ИМРСЗ [19] из обратной модели главных компонент (вариант №3), вторая – вычислены программой-таблицей ([9], Таблица 2.3). Матрица $Z^{(t,\ell)}_{m \times 3}$ использует выборку $U^{(t,\ell)}_{m,10} \in N_s(0, I_{10,10})$, а выборка $Z^{(t+1,\ell)}_{m \times 2}$ использует другую выборку $U^{(t+1,\ell)}_{m,10} \in N_s(0, I_{10,10})$, что позволяет нам утверждать об одном механизме порождения неопределенности в множестве z -переменных с номерами (1,2,3,4,5) из матриц $Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2}$.

В ОЗ АГК [7, п⁰2.5.2] при декоррелирующем преобразовании случайной выборки V_{mn} с помощью случайного преобразования $C_V L_V^{(-1/2)}$, состоящего из случайного линейного C_V и случайного нелинейного преобразования $L_V^{(-1/2)}$, «теряется» закон распределения [2,7]. Здесь случайность в выборке V_{mn} порождается с помощью компьютера, выборка V_{mn} изменяется в ходе случайных преобразований $C_V L_V^{(-1/2)}$. Результат оценивается алгебраически: фиксированные функции от преобразованной выборки имеют заданные значения средних ($=0$), дисперсий ($=1$). Это был первый способ.

Далее используем два других способа, где для выборок используются два разных преобразования, превращающие их в выборки $[Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2}]$ значений z -переменных. Использование 10-мерных (можно использовать и другие размерности, большие 3) выборок $U^{(t,\ell)}_{m,10}, U^{(t+1,\ell)}_{m,10} \in N_s(0, I_{10,10})$ с разными номерами t и $t+1$ дает нам нужный эффект, порождает ненулевые значения коэффициентов корреляции $r_{14}, r_{15}, r_{24}, r_{25}, r_{34}, r_{35}$ (Таблица 1). Эту корреляционную матрицу размера 5-на-5 имеет 5-мерная выборка $[Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2}]$ значений z -переменных. Причем выборочная корреляционная матрица в точности равна матрице размера 5-на-5 $(1/m)[Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2}]^T [Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2}]$. Механизм неопределенности закона

распределения 1-мерных случайной величины $\eta = (\xi - \mu) / \sigma$ которую мы поставили в соответствие с нашей z -переменной, равной $z = (x^0 - x_{cp}) / s$, подробно описан в [7, стр.61-70]. Механизм неопределенности значений z -переменной №6 формируется при решении оптимизационной задачи №4 (смотрите ниже). При фиксированных значениях как известных значений $r^{(L)}_{1j}, r^{(L)}_{2j}, r^{(L)}_{3j}, \dots, r^{(L)}_{ij}, \dots, r^{(L)}_{nj}, n=6$ коэффициентов корреляции, так и при фиксированных значениях $m \times (n-1)$ значений $n-1$ z -переменных существует бесконечное множество $(z^{(L)}_{1j}, z^{(L)}_{2j}, \dots, z^{(L)}_{mj})^T, L=1, \dots, k_L < \infty$, значений n -ой z -переменной, каждый m -набор которого функционально связан уравнениями системы вида (*). Модельная 6-мерная выборка $Z^{(t,\ell)}_{m \times 6} = [Z^{(t,\ell)}_{m \times 3} | Z^{(t+1,\ell)}_{m \times 2} | Z^{(L)}_{m \times 1}]$, $m=20, t=1, \ell=1, L=1$, имеющая блочно-диагональную корреляционную матрицу R_{66} (Таблица 2) вида приведена в Таблице 3.

Результирующую матрицу R_{mn} необходимо диагонализировать, т. е. решить прямую спектральную задачу (ПСЗ): $R_{mn} = \Rightarrow (C_{mn}, \Lambda_{mn})$. Чтобы узнать ее f -параметры. Спектр $\Lambda_{mn} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ является решением ПСЗ (однородной спектральной задачи), которая решается методом Холецкого для симметрической матрицы $R_{mn} = R_{mn}^T$ и имеет единственное решение Λ_{mn} . В отличие от задач из [9-11] в нашей задаче нет необходимости решать обратную спектральную задачу (ОСЗ): $\Lambda_{mn} = \Rightarrow (C^{(t)}_{nn}, R^{(t)}_{nn})$ [11]. Это – отличие нашей задачи от ранее рассмотренных задач, связанных с Λ -выборками.

Рассмотрим разложение целого числа n на слагаемые вида $n = n_1 + 2k + 1$, где $n_1 = 3, 4, \dots, 10, k = 1, \dots, 10$. В случае $n_1 = 0$ рассмотрим случаи $n = 2k + 1$ или $n = 2k, k = 1, \dots, 10$.

Случай $n = 2k$ в прикладных задачах классификации многомерных объектов встречается редко. Реализация модели, соответствующей этому случаю описана в [9]. Проводится k -кратное моделирование пар u -переменных, y -переменных и z -переменных. Недиagonalный элемент $r_{12} = r_{21} = r$ корреляционной матрицы R_{22} , образующую квадратный блок длины 2: блок $[j: (j+2)]$ с началом в столбце (строке) № j и концом в столбце (строке) № $(j+2)$ явно присутствует во всех формулах: элементов ее спектра Λ_{22} , элементов ее матрицы собственных векторов $C^{(L)}_{22}$, элементов выборок $Y^{(t)}_{m,2}, Z^{(L)}_{m,2}, t=1, \dots, k_t < \infty, L=1, 2$. Схематичное изображение этой модели ($n=2$) при бесконечном числе значений элемента $r^{(t)}_{12} = r^{(t)}_{21} = r^{(t)}: R^{(t)}_{22} = \Rightarrow (\Lambda^{(t)}_{22}, C^{(t,L)}_{22}, Y^{(t)}_{m,2}, Z^{(t,L)}_{m,2}), \ell=1, \dots, k_\ell < \infty, t=1, \dots, k_t < \infty$, при любых значениях $k=1, \dots, 10$.

Случай $n = 2k + 1$ в прикладных задачах классификации многомерных объектов можно рассматривать как случай $n = n_1 + n_2 + n_3 + \dots + n_L$, где

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

$n_i=2k+1 \geq 3$, $i=1, \dots, 10$. Здесь некоторые значения размерности n_i могут принимать значения вида $n_i=2k$.

Для случая $n=2k+1$ после реализации модели ($r < 0$) $\Rightarrow (\Lambda_{22}, C_{22}^{(L)}, Y_{m2}^{(t)}, Z_{m2}^{(L,t)})$, $t=1, \dots, k_t < \infty$, $L=1, 2$, $m=20$ или модели ($r > 0$) $\Rightarrow (\Lambda_{22}, C_{22}^{(L)}, Y_{m2}^{(t)}, Z_{m2}^{(L,t)})$, $t=1, L=1, 2$, $m=20$, необходимо реализовать ниже изложенную оптимизационную задачу №4. Эта необходимость обусловлена невозможностью использования модели при $n=2$. Нужно разработать модель при $n=1$. Суть модели ($r > 0, z_1, \dots, z_n$) $\Rightarrow (U_{m2}^{(t)}, Y_{m2}^{(t)}, Z_{m2}^{(L,t)})$, $t=1, L=1, 2$, $m=20$ состоит в наличии так называемой «половины блока» $[j:(j+1)]$ длины 1. Смысл термина «половина блока» $[j:(j+1)]$ наглядно иллюстрируется в Таблице 2. Излагаемый ниже алгоритм решения оптимизационной задачи №4 не применяется для блоков длины 2, а применяется для блоков длины 1.

Модельную матрицу R_{nn} можно построить, применяя модели из работ [2,7]. В случае $n=n_1+2k$ используем 2 способа.

Способ 1. ОМ ГК (при размерности выборки n_1). Для выбора элементов n_1 -на- n_1 матрицы R можно использовать элементы другой корреляционной матрицы R_{nn} большей размерности. Если в новую матрицу попали подходящие элементы. Элементы n_1 -на- n_1 матрицы можно моделировать по заданному спектру Λ размерности n_1 -на- n_1 . Его элементы моделируются, если заданы значения f -параметров из их множества $\{f_1, f_2, f_3, f_4, f_5, f_6\}$. Процедуры моделирования спектра Λ апробированы в [9-11]. Для построения k штук 2-блоков используем модель из работы [9]: $R_{22}^{(t)} \Rightarrow (\Lambda_{22}^{(t)}, C_{22}^{(L,t)}, Y_{m2}^{(t)}, Z_{m2}^{(L,t)})$, $\ell=1, \dots, k_t < \infty$, $L=1, 2$, $t=1, \dots, k_t < \infty$, при любых значениях коэффициента корреляции $r^{(t)}$ (при моделировании 2-блоков). При этом сперва k раз решаются ПСЗ при $n=2$ [9]: $\{r^{(t)}\} \Rightarrow R_{22}^{(t)} \Rightarrow (\Lambda_{22}^{(t)}, C_{22}^{(L,t)})$, $L=1, 2$, $\ell=1, \dots, k = k_t < \infty$. Здесь важен выбор индексы $(1,j), (2,j), (3,j), \dots, (j,j), \dots, (n,j)$, соответствующие известным элементам корреляционной матрицы $r_{1j}, r_{2j}, r_{3j}, \dots, r_{jj}, \dots, r_{nj}$.

Здесь известные значения переменных с номерами $1, 2, 3, \dots, j-1, j+1, \dots, n$. Но не известны значения j -ой переменной: $z_{1j}, z_{2j}, \dots, z_{mj}$. Каждое из этих неизвестных значений, например z_{kj} , умножается на известное значение z_{k1} : $z_{k1} \times z_{kj}$. Сумма таких произведений при изменении индекса $k=1, \dots, m$, умноженная на постоянную величину $(1/m)$ равна коэффициенту корреляции r_{1j} : $(1/m) (z_{11} \times z_{1j} + z_{21} \times z_{2j} + \dots + z_{k1} \times z_{kj} + \dots + z_{m1} \times z_{mj})$. Аналогичный вид имеют формулы для коэффициентов корреляции $r_{2j}, r_{13j}, \dots, r_{jj}, \dots, r_{nj}$.

Оптимизационная задача №4

Для известных значений $r_{1j}, r_{2j}, r_{3j}, \dots, r_{jj}, \dots, r_{nj}$ n коэффициентов корреляции удовлетворяющих $n-1$ линейным уравнениям и 1 нелинейному уравнению:

$$\begin{aligned} (1/m) \times (z_{11} \times z_{1j} + z_{21} \times z_{2j} + \dots + z_{k1} \times z_{kj} + \dots + z_{m1} \times z_{mj}) &= r_{1j} \\ (1/m) \times (z_{21} \times z_{1j} + z_{22} \times z_{2j} + \dots + z_{k2} \times z_{kj} + \dots + z_{m2} \times z_{mj}) &= r_{2j} \\ \dots & \dots \\ (1/m) \times (z_{1j} \times z_{1j} + z_{2j} \times z_{2j} + \dots + z_{kj} \times z_{kj} + \dots + z_{mj} \times z_{mj}) &= r_{jj} = 1 \\ \dots & \dots \\ (1/m) \times (z_{1n} \times z_{1j} + z_{2n} \times z_{2j} + \dots + z_{kn} \times z_{kj} + \dots + z_{mn} \times z_{mj}) &= r_{nj} (*) \end{aligned}$$

В нашей задаче требуется найти m -мерное решение - вектор $(z_{1j}, z_{2j}, \dots, z_{mj})$, удовлетворяющее системе уравнений (*).

Указанные номера $1, 2, 3, \dots, j-1, j+1, \dots, n$ расположены в скобках слева: $(1,j), (2,j), (3,j), \dots, (j, j-1), (j, j+1), \dots, (n,j)$. В правой части системы уравнений (*) нижние индексы известных коэффициентов корреляции имеют вид: $(1,j), (2,j), (3,j), \dots, (j,j-1), (j,j+1), \dots, (n,j)$. Одна пара индексов (j,j) является индексом дисперсии стандартизованной j -ой z -переменной, равной 1: $r_{jj}=1$. При $j=n$ известна выборка Z_1 размерности $m \times (n-1)$ с известной корреляционной матрицей $R_{11} = (1/m) Z_1^T Z_1$, размерности $(n-1) \times (n-1)$. Стандартизованная n -ая z -переменная с неизвестными значениями $z_{1n}, z_{2n}, \dots, z_{mn}$ расположена в n -ом столбце матрицы $Z_{mn} = [Z_1 | Z_2]$, $Z_2 = (z_{1n}, z_{2n}, \dots, z_{mn})^T$. Матрица $R_{12} = (1/m) Z_1^T Z_2$, является вектором длины $(n-1)$ и ее элементы $r_{1j}, r_{2j}, r_{3j}, \dots, r_{nj}$ в рамках оптимизационной задачи №4 задаются случайно, а определяемый им вектор $\beta = R_{11}^{-1} R_{12}$ является вектором регрессионных коэффициентов. Задачу моделирования выборки $Z_{mn} = [Z_1 | Z_2]$, корреляционных матриц R_{11}, R_{12} для заданного вектора β , обратную модель многомерной регрессии, рассмотрим в другой статье.

Система (*) из n уравнений с $m > n > 2$ неизвестными $z_{1j}, z_{2j}, \dots, z_{mj}$ при фиксированном номере j z -переменной, где $j \leq n$, является недоопределенной системой: имеются $n(m-1)$ известных значений z -переменных, n известных значений $r_{1j}, r_{2j}, r_{3j}, \dots, r_{jj}, \dots, r_{nj}$, $m > n$ неизвестных $z_{1j}, z_{2j}, \dots, z_{mj}$, имеет бесконечное множество решений $\{z_{1j}, z_{2j}, \dots, z_{mj}\}$. Решение Оптимизационной задачи №4 проведем с применением процедуры Solver. Программа-таблица приведена в Таблице 1, где приведены начальные значения и полученные значения $z_{1j}, z_{2j}, \dots, z_{mj}$ и формулы целевой функции, формул 5 функций ограничений из (*) программы-таблицы для процедуры Solver при $m=20$. Окно процедуры «Поиск решения» в ЭТ Excel с введенными формулами целевой функции оптимизационной задачи №4 ($n=6, m=20$) и 5 функций ограничений из системы (*) приведены на Рисунке 2, а настроенные параметры процедуры «Поиск решения» в ЭТ Excel для



Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

решения оптимизационной задачи №4 – на Рисунке 1.

Конкретные значения полученного решения $Z_{1j}, Z_{2j}, \dots, Z_{mj}$ системы уравнений (*) зависят от начальных значений, назначаемых пользователем процедуры «Поиск решения» и вводимых в соответствующее поле окна этой процедуры. В примере мы выбрали любой нормированный вектор $(z_{1j}, z_{2j}, \dots, z_{mj})$ такой, что $1 = (z_{1j}^2 + \dots + z_{mj}^2)/m$, где $n=6, m=20$. Этот вектор в паре с другими векторами из столбцов матрицы Z_{mn} не дает желаемых значений коэффициентов корреляции, но процедура «Поиск решения» легко преобразует этот нормированный вектор $(z_{1j}, z_{2j}, \dots, z_{mj})$ в вектор, являющийся решением системы (*). Размерность $n=6$ в нашем примере позволяет сравнивать результаты расчетов с данными, полученными в результате анализа данных из работ [16-18].

Пример моделирования блочно-диагональной корреляционной матрицы и многомерной Λ -выборки

Пусть требуется получить (модельно вычислить) корреляционную матрицу размерности $n=6$, состоящую из 2-х квадратных блоков размерностей [1-3], [4-5] и с элементом $r_{56}=r_{65}$ из «половины блока [5-6]» длины 1. Элемент r_{55} симметрической матрицы R_{66} входит и в блок [4-5], и в левую половину блока [5-6] (зеленый блок матрицы R_{66} на Таблице 2) с общим элементом $r_{55}=1$. Поэтому матрицу R_{66} мы представляем состоящей из 3 блоков: целые блоки [1-3], [4-5], половина блока [5-6]. Этим мы реализуем случай из общего разложения целого числа n на слагаемые $n=n_1+2k+1$, где $n_1=3, k=1$. В случае $n=n_1+2k$ или $n=2k$ излагаемый выше алгоритм решения оптимизационной задачи №4 не применяется, а k раз применяется модель $r^{(t)} \Rightarrow R^{(t)}_{22} \Rightarrow (\Lambda^{(t)}_{22}, C^{(t,L)}_{22}, Y^{(t,\ell)}_{m_2}, Z^{(t,\ell)}_{m_2})$, $\ell=1, \dots, k_t < \infty$, $L=1, 2$, $t=1, \dots, k_t < \infty$, при любых значениях коэффициента корреляции $r^{(t)}$ (при $n=2$) из работы [9]. В случае $n=n_1+2k$ применяется ОМ ГК при ее размерности n_1 и k раз применяется модель из работы [9] для построения k штук 2-блоков. Если все u -переменные из одной выборки $U_{mn} \in N_s(0, I_{mn})$, то выбор только из нее в нашу новую выборку, k штук выборок u -переменных с корреляционной матрицей имеющей I_{22} , то последующее преобразование в выборку $Z^{(t)}_{m_2}$ z -переменных. обеспечит этой корреляционной матрице «чистый» блочно-диагональный вид с k штук 2-блоками. Если все выбранные u -переменные из двух или более выборок: $U^{(t)}_{mn} \in N_s(0, I_{mn})$, то выбор двух из них в нашу одну новую выборку $[U^{(t)}_{m_2}] U^{(t+1)}_{m_2}$ имеющей корреляционную матрицу I_{44} , с последующим преобразованием в

m -на- 4 выборку z -переменных. обеспечит этой корреляционной матрицы вид отличающийся от блочно-диагонального (Таблица 2).

Если один из блоков формируемой нами корреляционной матрицы имеет размерность $n > 2$, то моделируем ее элементы, применяя программу CORMAP (CORMAT, COMA31) из ППП «Спектр» [19]. Допустим, что нам надо моделировать блок корреляционной матрицы размерности $n=3$, располагаемый в первых 3 столбцах корреляционной матрицы размерности $n=6$. По нашему выбору величины (и знака) коэффициента корреляции будем руководствоваться шкалой Чеддока (Chaddock scale). По этой шкале количественная мера тесноты связи: абсолютное значение коэффициента, принадлежащее интервалу от 0 до 0.3 - качественно интерпретируется как «очень слабая», интервалу от 0.3 до 0.5 - «слабая», умеренная», интервалу от 0.5 до 0.7 - «заметная» (moderate positive), интервалу от 0.7 до 0.9 - «высокая», интервалу от 0.9 до 1 - «очень высокая». Назначим значения ее недиагональных элементов [7]. Предположим, что эти значения принадлежат интервалу: сильной связи - $r_{12}=0,86840 \in (0,7, 0,9)$, заметной (средней) связи - $r_{43}=0,5 \in (0,5, 0,7)$, умеренной связи - $r_{56}=0,3 \in (0,3, 0,5)$ по шкале Чеддока. Значения коэффициента корреляции из этих шкал являются содержательно интерпретируемыми в прикладных исследованиях [8-10]. Например, в работе [8] выделены «заметные» значения $r_{12}=0,6654$, $r_{45}=0,6221$, $r_{56}=0,4980$ коэффициентов парной корреляции из элементов, расположенных выше диагонали симметрической корреляционной матрицы R_{66} из работы [8, стр.29].

Наш выбор индексов коэффициентов корреляции $r_{12}=0,86840$, $r_{43}=0,5$, $r_{56}=0,3$ определяет три блока: 3-блоку [1:3] длины 3, 2-блоку [4:5] длины 2, половине блока [5:6] длины 1. Корреляционная матрица R_{66} является 6-блоком длины 6.

Моделирование наших трех блоков разной длины проведем тремя разными способами.

Способ 1. Так как 3-блоку соответствует блочно-диагональная матрица размерности $n=3$, то можем ее рассматривать как корреляционную матрицу R_{33} , либо как «сумму» 2-блока и половины блока [5-6]. Будем моделировать корреляционную матрицу R_{33} , применяя ОМ ГК [2,7,19]. Программа IMPC5 при $k_t=100$, т. е. при 100 обращениях к программе CORMAP вывела в файл IMPC5.OUT 9 (из 100) матриц R_{33} (Таблица 2), элементы которых удовлетворяют критерию принадлежности к интервалу (от 0.7 до 0.9) сильной связи по по шкале Чеддока. В файле IMPC5.OUT имелись все 100 матриц R_{33} , $\ell=1, \dots, 100$, выведенные для контроля работы программ IMPC5, CORMAP (CORMAT,

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

СОМА31) [19]. Одна из них - матрица R_{33} (Таблица 2, левая верхняя часть), взята нами в качестве блока длины 3 из будущей блочно-диагональной матрицы R_{66} (Таблица 2). Для моделирования следующего 2-блока используем другой способ 2.

Способ 2. В ЭТ Excel реализуем ОМ ГК при $n=2$, описанную в статье [9]. Для нашего коэффициента $r_{43}=0.5$ реализуемая модель имеет вид:

$(r_{43}=0.5) \Rightarrow R^{(t)}_{22} \Rightarrow (\Lambda_{22}, C_{22}, U^{(t)}_{m2}, Y^{(t)}_{m2}, Z^{(t)}_{m2})$, $m=20$, $t=1$. Значения 2-х коррелированных z -переменных $Z_{m2} = Y_{m2} C_{22}^T$, где i -ая строка матрицы Z_{m2} состоит из элементов

$$z_{i1} = (u_{i1} \times \sqrt{1+r}) \times (-\sin \alpha) + u_{i2} \times \sqrt{1-r} \times \cos \alpha,$$

$$z_{i2} = ((u_{i1} \times \sqrt{1-r}) \times \cos \alpha - u_{i2} \times \sqrt{1-r} \times \sin \alpha),$$

а ее столбцы удовлетворяют условиям: $z_{.1} = (z_{11}, \dots, z_{m1})^T$, $z_{.2} = (z_{12}, \dots, z_{m2})^T$, $(1/m)z_{.1}^T z_{.2} = r_{12} =$

$r_{21}=0.5$, $(1/m)z_{.1}^T z_{.1}=1$, $(1/m)z_{.2}^T z_{.2}=1$. При переносе в матрицу Z_{m6} номера (1,2) z -переменных примут новые значения (4,5). Здесь $r_{12}=r_{21}=0.5 \in (0.5, 0.7)$ -интервалу средней, заметной связи.

Способ 3. Так как значения первых 5 z -переменных были моделированы ранее с применением способов 1 и 2, то значения одной z -переменной №6 моделируем, решая Оптимизационную задачу №4. Ее решение (при $j=6$) $z_{16}, z_{26}, \dots, z_{m6}$ найдено и располагается в столбце №6 выборки (Таблица 3).

Оптимизационная задача №4 при $m=20$, $n=6$ решена с применением процедуры Solver. В окне (Рисунок 1) надстройки «Поиск решения» $n-1$ линейным уравнениям и 1 нелинейному уравнению (*).

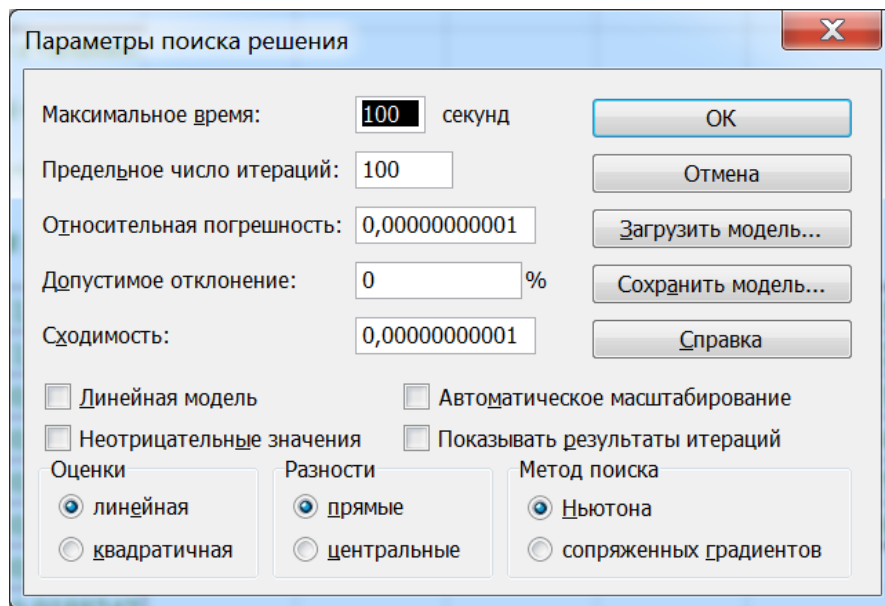


Рисунок 1 - Параметры процедуры «Поиск решений» в ЭТ Excel для решения оптимизационной задачи №4

Таблица 1

Начальные значения

1	2	3	4	5	6	7	8	9	10
1,8709901	1,0777906	0,4807108	0,6358906	0,3608648	0,4196449	0,0895122	0,9188242	-1,068565	1,5590402
11	12	13	14	15	16	17	18	19	20
1,0751419	-0,143386	0,0885203	0,831682	0,479433	-0,246619	-0,815438	1,4226047	1,1394391	0,1165249
	The solution of the problem №4 (n=6)	Setpoints			initial values	setpoints			
1	6	6,00			1	-0,306434	6,00		
2	0,5195166	0,5195166			2	-7,669074	0,5195166		

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

3	0,6030152	0,6030152		3	-4,846662	0,6030152		
4	0,8308359	0,8308359		4	-5,985789	0,8308359		
5	0,0146184	0,0146184		5	4,4017983	0,0146184		
6	1	1		6	0,8098015	1		

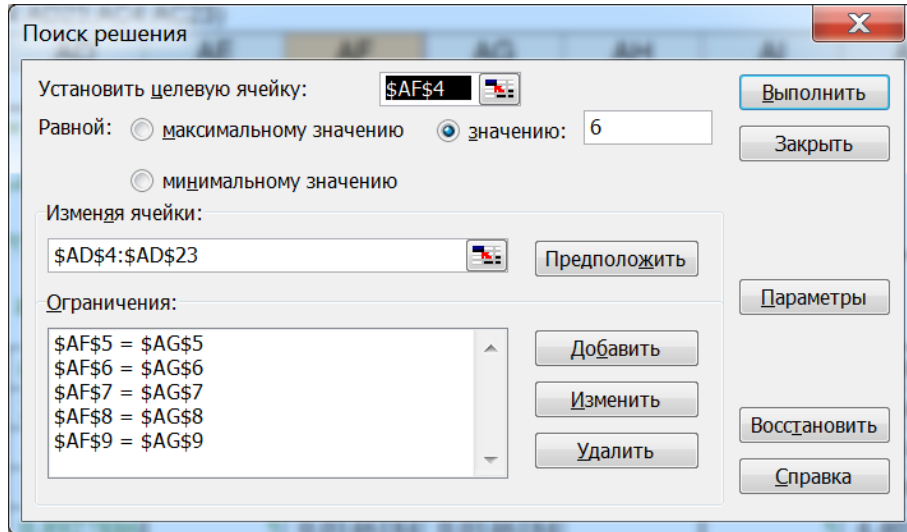


Рисунок 2 - Окно процедуры «Поиск решений» в ЭТ Excel с введенными формулами целевой функции оптимизационной задачи №4 (n=6) и 5 функций ограничений

Таблица 2

Модельная блочно-диагональная корреляционная матрица R_{66} вида 6-блок длины 6

	1	2	3	4	5	6
1	1,0000	0,8593	0,9793	0,0402	0,1308	0,0260
2	0,8593	1,0000	0,8668	0,0230	-0,0508	0,0302
3	0,9793	0,8668	1,0000	0,0929	0,1308	0,0415
4	0,0402	0,0230	0,0929	1,0000	0,5000	0,0007
5	0,1308	-0,0508	0,1308	0,5000	1,0000	0,3000
6	0,0260	0,0302	0,0415	0,0007	0,3000	1,0000

Таблица 3

Модельная 6-мерная выборка $Z^{(t,\ell)}_{m6} = [Z^{(t,\ell)}_{m3}, Z^{(t+1,\ell)}_{m2}, Z^{(L)}_{m1}]$, $m=20$, $t=1, \ell=1, L=1$ имеющая блочно-диагональную корреляционную матрицу R_{66} вида 6-блок длины 6

Values of the z-variables: R_{33}			Values of the z-variables: $r=0.5$		Values of the z-variables: $r=0.3$
1	2	3	4	5	6
-1,7486	-0,84905	-1,3589	-0,246022	-0,679569	0,91677645

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

-1,0243	-1,5789	-1,1286	0,7968352	1,6937153	2,20985766
-1,3308	-1,6115	-1,276	0,2050141	-0,675413	-0,8257356
-0,24354	0,27442	-0,46036	-0,462647	0,8734136	1,65419513
-0,70003	-1,17	-0,93425	-1,094766	0,2263316	0,89278863
-1,7524	-1,3215	-1,8374	-0,537787	-0,648685	-0,2132969
-0,076051	0,42324	-0,24024	-1,559075	-2,603269	0,28231785
0,52023	0,9166	0,77113	-0,174967	-0,377692	0,9585942
0,71842	0,95725	0,4863	0,2105531	0,4973616	-0,6155958
1,0105	1,4394	1,2665	1,4363995	0,2918035	1,36358425
0,089902	0,17971	-0,052251	-0,19993	-1,362517	0,76500826
0,7184	1,3014	0,81813	-0,761445	-0,858966	0,04143056
0,78228	0,61429	0,72225	-0,999071	1,3300653	1,5221544
-0,04121	-0,50514	0,15556	1,6866936	1,3768869	0,57586001
0,19502	-0,25865	-0,061358	-1,282744	-0,63549	1,08462824
2,2074	1,6004	2,1099	0,7766785	0,1409087	0,71161068
1,1805	0,59905	1,0222	0,3144525	0,4285449	-0,2826876
-0,43276	-0,45891	-0,10734	0,2432353	-0,257921	0,94200162
-0,76328	-0,1077	-0,83147	2,3675579	0,7293749	0,45710988
0,69032	0,52486	0,93619	-0,718969	0,5111114	0,82691956
1,000000	1,000000	1,000000	1,000000	1,000000	1,000000

Полученная матрица имеет вид (Таблица 3). Для нее вычисляем ее собственные числа (Рисунок 3) и собственные векторы (Таблица 4). Вычисленный при реализации ПСЗ собственные числа имеют вид $\Lambda_{nn}=\text{diag}(\lambda_1, \dots, \lambda_n)=\text{diag}(2.8265, 1.5731, 0.99974, 0.44297, 0.13931, 0.018362)$. Динамика изменения долей λ -переменных: $\Delta_j=\lambda_j/6$, накопленных долей λ -переменных: $D(j)=(\lambda_1+\dots+\lambda_j)/6$, значений a -переменных $a_j=\lambda_j/\lambda_{j-1}$ не показывают отличия нашего спектра от спектров реальных данных из работ [10,11,16-18]. Функциональные параметры спектра имеют следующие значения. $f_1=6$, $f_2=11,67925571$, $f_3=153,9320335$, $f_4=0,471083176$, $f_5=0,00503703$, $f_6=16,39378645$, $\varphi=0,435095994$. их значения также не показывают уникальности нашего спектра. Полученная в результате применения 3-х разных способов моделирования наша модельная 6-мерная выборка $Z^{(t,\ell)}_{m6}=[Z^{(t,\ell)}_{m3}, Z^{(t+1,\ell)}_{m2}, Z^{(L)}_{m1}]$, $m=20$, $t=1$, $\ell=1$, $L=1$, представляет собой результат комбинации разных случайных преобразований и дискретных итерационных вычислений как в прямых, так и в обратных задачах. Обсуждение этих вопросов приведено в параграфе «Арифметика» прямой задачи и Алгоритмы обратной задачи» статьи [11,стр.253].

Выводы

Наши модельные выборки применимы для тестирования методов классификации (кластер-

анализа) многомерных объектов с разными подмножествами признаков, различающихся степенью тесноты парных связей. Объекты представлены в виде таблицы объект-свойство. При заданных значениях средних x_j^{cp} признаков (в разных единицах измерения) и при заданных значениях дисперсий s_j^2 наши значения z -переменных (Таблица 3) определяют исходные значения $x_{ij}^\circ=z_{ij}s_j+x_j^{cp}$, где $x_j^{cp}=(x_{1,j}^\circ+\dots+x_{20,j}^\circ)/20$, $s_j^2=(x_{1,j}^2+\dots+x_{20,j}^2)/20$, $x_{ij}=x_{ij}^\circ-x_j^{cp}$, $i=1, \dots, 20$, $j=1, 2$, вычислялись ранее по другим реальным данным. Анализируемый массив данных $\{x_{ij}\}$, $i=1, \dots, M$, $j=1, \dots, n$, $M=K \times 20$, $n=n_1+2k+1$, $n_1=3, 4, \dots, 10$, $k=1, \dots, 10, K \geq 1$, может содержать тысячи и сотни тысяч наблюдений с десятками и сотнями характеристик наблюдений. Некоторые наблюдения с их n характеристиками могут быть выделены в виде отдельной выборки, имеющей блочно-диагональную корреляционную матрицу R_{66} одного из рассмотренных выше видов. При этом для наших модельных выборок для измерения тесноты связей между свойствами многомерных объектов не надо использовать корреляционную матрицу R_{nn} , а использовать меры сходства и различия между количественными данными [3-5]. Случаи возможного применения наших модельных выборок как близких к «идеальным» образцам, можно найти в [1,3-5].

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	ПИИЦ (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

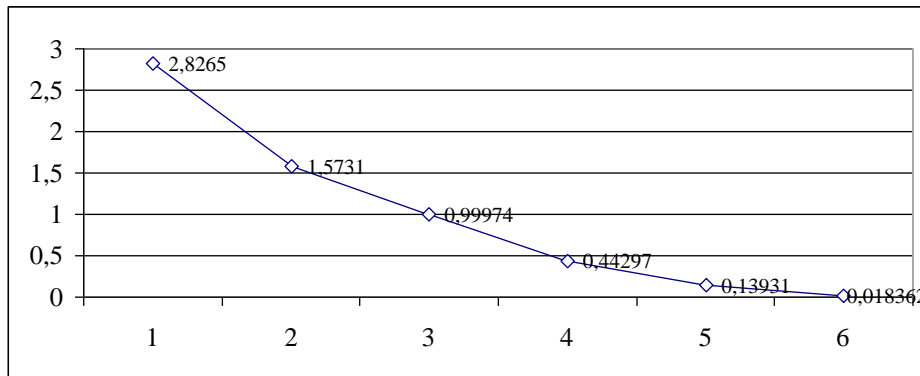


Рисунок 3 $\Lambda_{66}=\text{diag}(2.8265,1.5731,0.99974,0.44297,0.13931, 0.018362)$.

Таблица 4

Матрица собственных векторов $C^{(l)}_{66}$, $l=1$

0,58156	-0,053757	0,0013104	-0,15522	0,37669	0,70208
0,55012	-0,14634	-0,014766	0,23357	-0,78811	0,0076311
0,58471	-0,030640	0,14994	-0,062920	0,38804	-0,70882
0,075943	0,59749	0,51014	0,60132	0,11126	0,055137
0,096659	0,70104	-0,0012154	-0,65701	-0,25772	-0,033362
0,046556	0,35539	-0,85983	0,35235	0,0.87247	0,21339

Итак мы реализовали модель вида:

$$R^{(l)}_{66} \Rightarrow (\Lambda^{(l)}_{66}, C^{(l)}_{66}, U^{(l)}_{m3}, U^{(t+1)}_{m2}, Y^{(l)}_{m3}, Y^{(t+1)}_{m2}, Z^{(t,l)}_{m3}, Z^{(t+1,l)}_{m2}, Z^{(L)}_{m1}), \quad m=20, \quad l=1, \dots, k_l < \infty, \quad t=1, \dots, k_t < \infty, \quad L=1, \dots, k_L < \infty.$$

Здесь матрица $R^{(l)}_{66}$ -смоделированная матрица (Таблица 2), для нее решается ПСЗ $R^{(l)}_{66} \Rightarrow (\Lambda^{(l)}_{66}, C^{(l)}_{66})$ методом Холецкого, вычисляются ее спектр $\Lambda^{(l)}_{66}$ (Рисунок 3) и матрица собственных векторов $C^{(l)}_{66}$ (Таблица 4). Спектр $\Lambda^{(l)}_{66}=\text{diag}(2.8265, 1.5731, 0.99974, 0.44297, 0.13931, 0.018362)$ имеет l -параметры $f_1=6, f_2=11, 67925571, f_3=153, 9320335, f_4=0,471083176, f_5=0,00503703, f_6=16, 39378645, \varphi=0,435095994$. Этот факт получен по схеме:

$\Lambda_{66} \Rightarrow (f_1, f_2, f_3, f_4, f_5, f_6, \varphi) = (f_1=6, f_2=11, 67925571, f_3=153, 9320335, f_4=0,471083176, f_5=0,00503703, f_6=16, 39378645, \varphi=0,435095994)$. Динамика убывания значений элементов нашего спектра на Рисунок 3 показывает его непохожесть на

спектры реальных данных из работ [1,2,6-11,14, 16-18]. Это отличие связано с блочно-диагональной структурой корреляционной матрицы (Таблица 2). Выборки $U^{(l)}_{m3} \in N_s(0, I_{33}), U^{(t+1)}_{m2} \in N_s(0, I_{22})$, используемые при моделировании матриц $Y^{(l)}_{m3}, Y^{(t+1)}_{m2}, Z^{(t,l)}_{m3}, Z^{(t+1,l)}_{m2}$, должны генерироваться независимо: одна из выборки $U^{(l)}_{mn} \in N_s(0, I_{mn})$, другая – из выборки $U^{(t+1)}_{mn} \in N_s(0, I_{mn})$, где номера t и $t+1$ указывают на наличие 2 разных матриц выборок, не обязательно следующих одна за другой – номер $t+1$ может быть равным $t+7$ или иным. Моделирование $R^{(l)}_{33}$ -выборки производится с применением программы IMPC4 из ППП «Спектр» [19], получаем выборки $U^{(l)}_{m3} \in N_s(0, I_{33}), Y^{(l)}_{m3}$. Выборку $U^{(t+1)}_{m2} \in N_s(0, I_{22})$ моделируем отдельно от $U^{(l)}_{m3} \in N_s(0, I_{33})$. Преобразование выборки $U^{(t+1)}_{m2} \in N_s(0, I_{22})$ в выборки $Y^{(t+1)}_{m2}, Z^{(t+1,l)}_{m2}$ проводится по формулам из [9, стр.183].

References:

- (2017) Matrices -Matrix Depot: A Test Matrix Collection Available: www.matrixdepotjl.readthedocs.io/en/latest/matrices.html (Accessed: 10.12.2017).
- Zhanatauov S.U. (1987) Obratnaya model' glavnykh komponent i ee primenenie. Diss. na soiskanie uch. step.. kand. fiz.-mat. nauk:05.13.11:zashchishchena 8.12.1987: utv.

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHII (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 3.860	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

- 1.06.1988/Zhanatauov Sapargali Utepovich-Vychislitel'nyy tseñtr Sibirs kogo otdeleniya AN SSSR, Novosibirsk, 1987, 302 p.
- Duran B.S., Odell P.L. (1974) Cluster Analysis: a survey. Springer-Verlag, 1974.-137 p.
 - Byul A., Cefel P. (2002) SPSS: the art of information processing. analysis of statistical data and the restoration of hidden patterns. - SPb, 2002.
 - (2017) CRM Segmentation and Clustering Using SAS(R) Enterprise Miner(TM) Izdatel' SAS Institute. ISBN 1599948311, 9781599948317
 - Zhanatauov S.U. (1989) Modelirovanie odnoy zamecha tel'noy ekstremal'noy sovokupnosti //Sistem noe modelirovanie-14, -Novosibirsk. 1989, p.27- 33.
 - Zhanatauov S.U. (2013) Obratnaya model' glavnykh komponent:-monografiya. -Almaty: Kazstatin form, 2013.- 201 p.
 - Zhanatauov S.U. (2017) A model of calculation risk changing of the interest rate "yield to maturity date" for foreign currency bonds of the republic of Kazakhstan. International scientific journal Theoretical&Applied Science. 2017, №8, vol.52, p.19-36. <http://www.T-Science.org>.
 - Zhanatauov SU. (2017) Theorem on the Λ -samples. International scientific journal Theoretical&Applied Science. 2017, № 9, vol.53, p.177-192. <http://www.T-Science.org>.
 - Zhanatauov SU (2017) Optimization problem of modeling missing elements of the spectrum of the correlation matrix. International scientific journal Theoretical &Applied Science. 2017,№10, vol.54, p.189-198. <http://www.T-Science.org>.
 - Zhanatauov SU (2017) The optimization problem with linearized equations f-parameters (f1,f2,f3,f4,f5,f6)-spectrum. International scientific journal Theoretical &Applied Science. 2017, №11, vol.55, p.251-267. <http://www.T-Science.org>.
 - Zhanatauov S.U. (2016) Model and histogram to adequacy of variables (C, Λ)-samples and real multidimensional sample. International Scientific Journal Theoretical &Applied Science.2016, № 11, vol. 43, p. 53-61.
 - Zhanatauov S.U. (2014) The (C, Λ ,Y)-sample is adequate to real multidimensional sample. Proced. Int. conf. "Leadership in Education, Business and Culture". 25 april 2014, Almaty-Seatle, ICET USA. Leadership Iternational Conference "Leadership on Education, Business and Culture». p.151.
 - Zhanatauov S.U. (2017) Modelirovanie mnogomernykh vyborok znacheniy priznakov zernovoy kul'tury. "II mezhdun. nauchno-prakt.konf. «Evropa i tyurkskiy mir: nauka, tekhnika i tekhnologii".Izmir (Turtsiya), 29-31 maya 2017. www.regionaca.org.
 - Hotelling H. (1933) Analysis of a complex of statistical variables into principal compo nents. -J. Educ. Psychol.,1933, vol.24, p. 417-441, p. 498-520.
 - Zhanatauov S.U. (2013) Kognitivnaya karta i model' sotsial'no-ekonomicheskikh fak torov kar'ernoy uspeshnosti shkol'nikov munitsipal'nykh shkol SShA. Sibirskiy pedagogicheskiy zhurnal. 2013, №6, p.28-33.
 - Zhanatauov S.U. (2014) Analiz budushchikh debitorskoy i kreditorskoy zadolzhen nostey munitsipalitetov gorodov. Ekonomicheskii analiz: teoriya i praktika. M.:№2 (353), 2014, p. 54-62.
 - Zhanatauov S.U. (2015) Kognitivnaya karta i kognitivnaya model' analiza glavnykh komponent (telekommunikatsionnaya otrasl').Natsional'naya assotsiatsiya uchenykh (NAU).IX Mezhd. nauch.-prakt.konf:«Otechestvennaya nauka v epokhu izmeneniy: postulaty proshlogo i teorii novogo vremeni».Rossiya, Ekaterinburg,16-17 maya 2015. p. 55-58.
 - Zhanatauov S.U. (1988) O funktsional'nom napolnenii PPP "Spektr". Sistemnoe modelirovanie -13 .-Novosibirsk, 1988, p.3-11.

