УДК 519.632.4

COMPARATIVE ANALYSIS OF THE COMPUTING OPPORTUNITIES OF THE OPTIMIZATION PROCEDURES FOR DIFFERENT MATHEMATICAL PROCESSORS ON THE PROBLEMS OF THE SPLITTING MIXTURES OF PROBABILISTIC DISTRIBUTIONS

Tuluchenko H., Starun N., Malomuzh T.

In this paper, we compare the computing *opportunities* ***of optimization procedures for the package Optimization of the Maple mathematical processor, the package Direct Search, which is compatible with the MP Maple, and package Search of Solutions from the table processor Excel for the problems solving of the splitting of mixtures from normal and log-normal probability distributions. The test cases show the advantages of the procedures for the Direct Search package.***

Keywords: mixture of probability distributions, optimization, Maple, Direct Search package.

Formulation of the problem. When describing heterogeneous data it is expedient to use mixtures of probabilistic distributions as mathematical models. Often, for estimating the parameters of the distribution of the mixture, use the method of maximum likelihood, which is to find the global maximum of the objective function of likelihood. Formal solution to this problem can be used as standard optimization methods, so problem-oriented EM-algorithm. The complexity of the problem lies in the fact that in the general case the likelihood function has many local maxima. Because of this, known standard methods of optimization and EM-algorithm in real-world tasks are not sufficiently effective.

Analysis of basic research and publications. An overview of the methods of splitting mixtures of probabilistic distributions is given, for example, in a monograph [2]. The capabilities of the Direct Search package for resolving the nonlinear conditional optimization problems described in [1]. The DirectSearch package is compatible with Maple Computer Mathematics.

Formulating the goals of the article. Show the benefits of using the Search Search package of Direct Search when solving the problems of splitting the mixtures of logarithmically normal probabilistic distributions in comparison with other procedures of modern optimization packages.

Main part. Let the investigated random variable X have an empirical distribution, which is shown in Fig. 1. Sample volume $N = 175$.

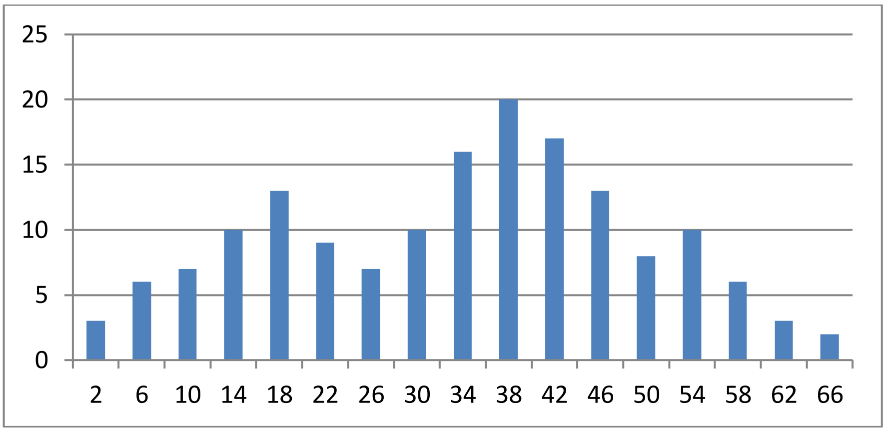

Fig. 1. Histogram of empirical absolute frequencies

From Fig. 1 it is obvious that the investigated distribution law is a mixture of three logarithmically normal distribution laws:

$$f(x)=\begin{cases} 0, & x\le 0 \\ \sum_{j=1}^{3} w_j \cdot \frac{1}{\sqrt{2\pi}\cdot x\cdot\sigma_j}\cdot e^{-\frac{(\ln x - a_j)}{2\cdot\sigma_j^2}}, & x>0 \end{cases}, \tag{1}$$

where w_j – specific weight j-th components in the mixture; $w_j \ge 0$, $\sum_{j=1}^{3} w_j = 1;\ j=\overline{1;3}$;

a_j – mathematical hope;

σ_j – mean square deviation.

The effectiveness of procedures from different mathematical packages will be compared when solving two optimization problems with target functions:

logarithm of maximum likelihood function:

$$\ln L = \sum_{i=1}^{s} \ln\left(\int_{x_{i-1}}^{x_i} f(x)dx \right) \Rightarrow \max, \tag{2}$$

observed value of the Pearson criterion:

$$\chi^2 = \sum_{i=1}^{s} \frac{\left(n_i^{експ} - n_i^{теор}\right)^2}{n_i^{теор}} \Rightarrow \min, \tag{3}$$

where s – number of partial intervals; $n_i^{експ}$, $n_i^{теор}$ – absolute experimental and theoretical frequencies respectively; $x_0 = -\infty;\ x_s = +\infty;$.

When solving problems (2) and (3) we use the initial approximations from the Table 1 and restrictions:

$$\sigma_j \geq 0;\ w_j \geq 0,\ \sum_{j=1}^{3} w_j = 1;\ j = \overline{1;3}. \qquad (4)$$

Table 1

Initial approximations for distribution mix parameters (1)

j	a_j	σ_j	w_j
1	2,9	0,7	0,40
2	3,6	0,5	0,45
3	4,0	1	0,15

To solve the described optimization problems we use:

1) conjugate gradient method (MSG), which is part of the Excel Processor Tablet Decision Solution Package;
2) the NLPSolve function, which is part of the Optimization package of the Maple Mathematical Processor, and implements a number of methods that are focused on solving nonlinear conditional optimization problems;
3) the Search function, which is part of the Direct Search package, and is focused on locating an extremum for a nonlinear conditional optimization problem with given search point origin coordinates;
4) GlobalSearch, which is part of the Direct Search package, and is focused on finding a global extremum for a nonlinear conditional optimization problem without assigning an initial search point.

The method of conjugate gradients will be implemented with the parameters: estimates - quadratic; differences - central; All other settings are the default.

Conclusions In the test cases, the complexity of the target functions (2) and (3) exceeded the capabilities of the GlobalSearch procedure and the conditional global extremum for them was never found.

In tasks with the target function (2), the procedures NLPSolve and Search resulted in the same solution, which provided a more accurate approximation of the empirical distribution than the solution obtained by the conjugate gradient method in Excel.

In problems with the objective function (3), all investigated procedures led to different solutions. The solution obtained with the Search procedure provided the best accuracy of the empirical distribution approximation (Fig. 2).

Table 2

Optimal values of the parameters of the mixture of distributions (1)

Function	a_1	a_2	a_3	σ_1	σ_2	σ_3	w_1	w_2	w_3
Target function - logarithm of maximum likelihood function (2)									
MSG (Excel)	2,631	3,801	3,962	0,950	0,317	1,010	0,46	0,54	0,00
Maple	2,260	3,468	4,000	0,976	0,423	0,155	0,296	0,438	0,266
Target function - observed value of Pearson criterion (3)									
MSG (Excel)	2,903	3,797	3,958	0,749	0,250	0,978	0,41	0,59	0,00
NLPSolve (Maple)	2,329	2,900	3,781	0,763	0,266	0,223	0,14	0,71	0,15
Search (Maple)	2,984	3,693	4,069	0,778	0,169	0,062	0,46	0,40	0,14

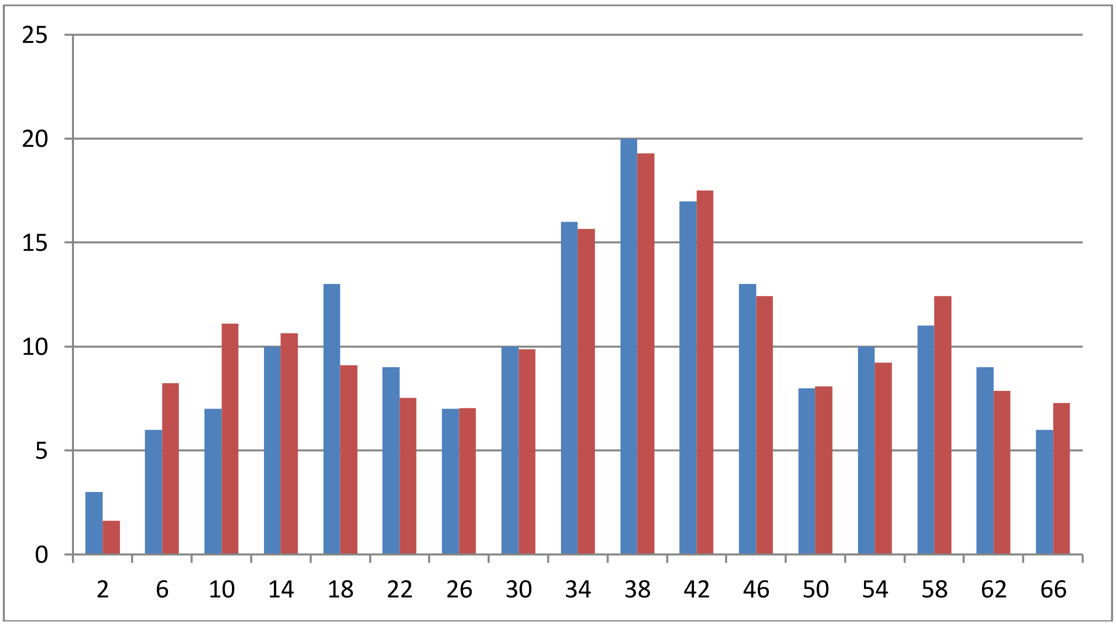

Fig. 2. A combined histogram of absolute empirical and theoretical frequencies

Література

1. Моисеев С.Н. Универсальный метод оптимизации без использования производных с квадратичной сходимостью [Электронный ресурс] / С.Н. Моисеев. – Воронеж, 2011. – 25 с. – Режим доступа: http://www.twirpx.com/file/1534928/.
2. Сиротин В.П. Декомпозиция распределений в моделировании социально-экономических процессов. Монография. / В.П. Сиротин, М.Ю. Архипова. – М.: изд-во МГУЭСИ, 2011. – 146 с.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЫЧИСЛИТЕЛЬНЫХ ВОЗМОЖНОСТЕЙ ПРОЦЕДУР ОПТИМИЗАЦИИ РАЗНЫХ МАТЕМАТИЧЕСКИХ ПРОЦЕССОРОВ НА ЗАДАЧАХ РАСЩЕПЛЕНИЯ СМЕСЕЙ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ

Тулученко Г.Я., Старун Н.В., Маломуж Т.В.

В работе сравниваются вычислительные возможности процедур оптимизации пакета Optimization математического процессора Maple, пакета Direct Search, совместимого с МП Maple, и пакета Поиск решений табличного процессора Excel при решении задач расщепления смесей нормальных и логарифмически нормальных вероятностных распределений. На тестовых примерах показаны преимущества процедур пакета Direct Search.

Ключевые слова: смесь вероятностных распределений, оптимизация, Maple, пакет Direct Search.

ПОРІВНЯЛЬНИЙ АНАЛІЗ ОБЧИСЛЮВАЛЬНИХ МОЖЛИВОСТЕЙ ПРОЦЕДУР ОПТИМІЗАЦІЇ РІЗНИХ МАТЕМАТИЧНИХ ПРОЦЕСОРІВ НА ЗАДАЧАХ РОЗЩЕПЛЕННЯ СУМІШЕЙ ЙМОВІРНІСНИХ РОЗПОДІЛІВ

Тулученко Г.Я., Старун Н.В., Маломуж Т.В.

У роботі порівнюється обчислювальні можливості процедур оптимізації пакету Optimization математичного процесору Maple, пакету Direct Search, який сумісний із МП Maple та пакету Пошук розв'язків табличного процесору Excel при розв'язанні задач розщеплення сумішей нормальних та логарифмічно нормальних ймовірнісних розподілів. На тестових прикладах показані переваги процедур пакету Direct Search.

Ключові слова: суміш ймовірнісних розподілів, оптимізація, Maple, пакет Direct Search.