

Головешкин В.А.¹, Жукова Г.Н.², Ульянов М.В.³, Фомичев М.И.⁴

¹ Московский государственный университет информационных технологий, радиотехники и электроники (МИРЭА), г. Москва, Россия

² Московский политехнический университет (МПУ), г. Москва, Россия

³ Институт проблем управления РАН им. В.А. Трапезникова, г. Москва, Россия

⁴ ФКН НИУ Высшая школа экономики, г. Москва, Россия

РАСПРЕДЕЛЕНИЕ ЛОГАРИФМА СЛОЖНОСТИ ИНДИВИДУАЛЬНЫХ ЗАДАЧ КОММИВОЯЖЕРА ПРИ ФИКСИРОВАННОЙ ДЛИНЕ ВХОДА

АННОТАЦИЯ

На основе статистического анализа сложности индивидуальной задачи коммивояжера, решаемой методом ветвей и границ, показано, что распределение логарифма сложности удовлетворительно аппроксимируется нормальным распределением. Коэффициенты линейной регрессии выборки логарифма сложности на стандартное нормальное распределение использовались для оценки значений параметров аппроксимирующего нормального распределения. Даны оценки границ 90% интервала сложности.

КЛЮЧЕВЫЕ СЛОВА

Задача коммивояжера; TSP; метод ветвей и границ; MBG; ресурсная эффективность; сложность индивидуальной задачи; квантиль; квантильный коэффициент асимметрии; квантильный коэффициент эксцесса.

Goloveshkin V. A.¹, Zhukova G.N.², Ulyanov M.V.³, Fomichev M.I.⁴

¹ Moscow Technological University (MIREA), Moscow, Russia

² Moscow Polytechnic University (MPU), Moscow, Russia

³ Institute of Control Sciences V. A. Trapeznikov Academy of Sciences, Moscow, Russia

⁴ Higher School of Economics National Research University, Moscow, Russia

PROBABILITY DISTRIBUTION OF THE COMPLEXITY OF THE INDIVIDUAL TRAVELING SALESMAN PROBLEM (FIXED NUMBER OF NODES)

ABSTRACT

It is shown that the logarithm of the complexity (number of nodes in the decision tree of a branch and bound algorithm) of the individual traveling salesman problem is approximately normally distributed. We use a linear regression model (logarithm of the complexity — standard normal distribution) to estimate parameters of normal distribution, which fit the sample. Borders of the interval, which contains 90% of the sample of the logarithm of the complexity, are also given.

KEYWORDS

Traveling salesman problem; TSP; branch and bound method; B&B; complexity; efficiency; quantile; quantile skewness; quantile kurtosis.

Метод ветвей и границ [5] — один из наиболее распространенных методов точного решения задачи коммивояжера [1-3,5,7] (TSP, traveling salesman problem), корректно формулируемой как задача поиска минимального по сумме весов дуг гамильтонова цикла в полном графе. Время, которое требуется компьютерной программе, реализующей этот метод, для нахождения решения задачи коммивояжера, существенно зависит от размерности и особенностей матрицы стоимостей задачи [12,14,15]. С физическим временем, затраченным на точное решение TSP методом ветвей и границ связана такая независимая от программной реализации характеристика индивидуальной задачи как сложность. Под сложностью индивидуальной задачи мы понимаем число порожденных вершин поискового дерева решений [5] в классическом алгоритме, реализующем метод ветвей и границ [1,2].

Вычислительный эксперимент проводился для задачи коммивояжера размерности 30. Матрица стоимостей для индивидуальной задачи получается на основе исходной матрицы A ,

элементами a_{ij} которой является стоимость в долларах США наиболее дешевого перелета для 1 взрослого человека из i аэропорта в j аэропорт (данные получены на сайте kayak.com). В случае отсутствия прямого рейса в качестве стоимости перелета использовалась минимальная цена перелета с пересадками, найденная на том же сайте.

В качестве элементов матрицы стоимостей индивидуальной TSP использовались нормально распределенные случайные величины $X_{ij} \approx N(a_{ij}, \sigma_{ij})$, параметр $\sigma_{ij} = Ka_{ij}$, где $K = 0.1$, $K = 0.5$, а также равномерно распределенные на отрезке $[0.5a_{ij}, 1.5a_{ij}]$ случайные величины (с округлением до целых чисел), что порождает индивидуальные задачи трех типов.

Для каждого типа задач коммивояжера объем выборки (число экспериментов) для получения статистически значимых результатов был выбран равным 10^5 . Полученные при проведении вычислительных экспериментов значения сложности индивидуальных задач были использованы для подбора вероятностного распределения, удовлетворительно описывающего распределение величины m — десятичного логарифма сложности. Расчеты показали, что в качестве приближения для вероятностного распределения m можно использовать нормальное распределение.

Для предварительного анализа данных для выборки каждого из трех типов были вычислены выборочные коэффициенты асимметрии и эксцесса [4,11,13], как для всей выборки объема 10^5 , так и для ее первой и второй половины и для каждой четверти. На рис. 1 в системе координат коэффициент асимметрии-коэффициент эксцесса черным ромбом изображается точка, соответствующая выборочным коэффициентам асимметрии и эксцесса для выборки объема 10^5 , синий и голубой кружок соответствуют первой и второй половине выборки, треугольники – четвертям выборки. Нормальные случайные величины имеют равные нулю коэффициенты асимметрии и эксцесса, что изображает фиолетовый ромб в начале координат. Для большей наглядности на рис. 1 также изображены точки, соответствующие выборкам из нормального распределения, полученным генератором псевдослучайных чисел (желтые точки для выборок объема 10^5 , голубые – $2.5 \cdot 10^4$).

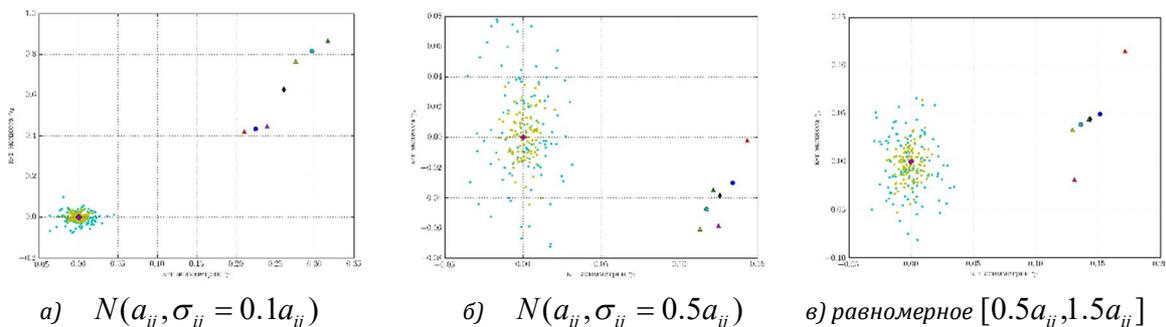


Рис. 1. Коэффициенты асимметрии и эксцесса

Выборочный коэффициент асимметрии больше 0.1, а коэффициент эксцесса достигает 0.9, что вызывает сомнение в том, что нормальное распределение удовлетворительно описывает выборку. Однако квантильные выборочные коэффициенты асимметрии и эксцесса [13] показывают на достаточную близость исследуемого распределения к нормальному (см. рис. 2). Ввиду того, что квантили практически не меняются при добавлении в выборку очень малого числа больших значений, а традиционные коэффициенты асимметрии и эксцесса очень чувствительны к наличию в выборке редких, но больших значений, принимаем гипотезу о нормальном распределении логарифма сложности.

Приведем формулы квантильных коэффициентов асимметрии S и эксцесса T :

$$S = (E_6 - 2E_4 + E_2)/(E_6 - E_2), \quad T = (E_7 - E_5 + E_3 - E_1)/(E_6 - E_2),$$

где $E_i = p_{i/8} = F^{-1}(i/8)$, $i = 1, 2, \dots, 7$ октили непрерывной функции распределения $F(x)$ [8-10]. Выборочные квантильные коэффициенты асимметрии и эксцесса вычисляются по аналогичным формулам, вместо E_i используются выборочные квантили.

Квантильные коэффициенты асимметрии и эксцесса выборок логарифма сложности

примерно такие же, какие наблюдаются у выборок того же объема из стандартного нормального распределения (рис. 2), обозначения такие же, как на рис. 1.

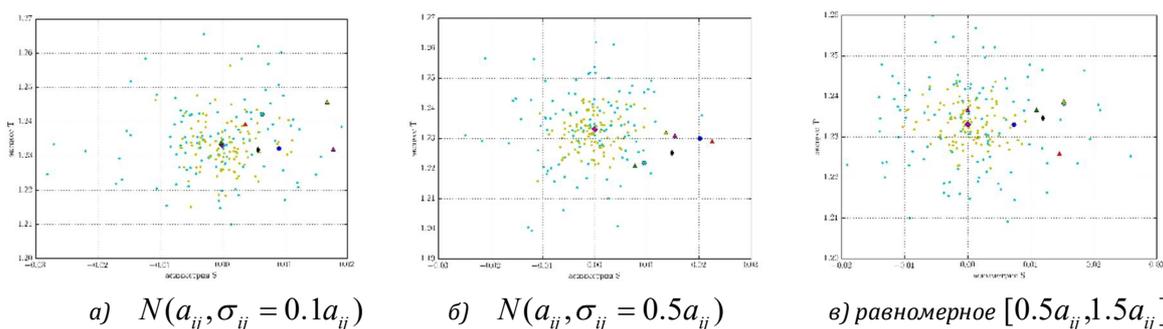


Рис. 2. Квантильные коэффициенты асимметрии и эксцесса

Как видно на рис. 2, выборочные квантильные коэффициенты асимметрии и эксцесса исследуемых выборок принимают значения из области, в которую попадают те же коэффициенты для выборок того же объема, взятых из стандартного нормального распределения.

Графическое представление данных - гистограммы и графики

Вычислим параметры k и b линейной регрессии выборочных квантилей на квантили стандартного нормального распределения

$$q_p^{st\ norm} = k\bar{q}_p + b$$

Значения параметров k и b , а также коэффициента регрессии r приведены в табл. 1.

Табл. 1 Параметры линейной регрессии стандартного нормального распределения на выборки

| Выборка | Регрессия ст. норм. на выборку | | | Регрессия выборки на ст. норм. | | |
|-------------------------------------|--------------------------------|-------|--------|--------------------------------|--------|--------|
| | k_0 | b_0 | r_0 | k_1 | b_1 | r_1 |
| $N(a_{ij}, 0.1a_{ij})$ | 0.288 | 2.846 | 0.9997 | 3.468 | -9.868 | 0.9997 |
| $N(a_{ij}, 0.5a_{ij})$ | 0.360 | 2.636 | 0.9998 | 2.778 | -7.324 | 0.9998 |
| равномерн. $[0.5a_{ij}, 1.5a_{ij}]$ | 0.352 | 2.904 | 0.9997 | 2.839 | -8.244 | 0.9997 |

Отметим, что линии регрессии стандартного нормального распределения на выборки и выборок на стандартное нормальное распределение практически совпадают.

Применим к выборке линейное преобразование $L: Lx = kx + b$ с вычисленными значениями параметров k_1 и b_1 (см. табл. 1.), затем для преобразованной выборки построим гистограммы с n интервалами, в каждый из которых значение случайной величины со стандартным нормальным распределением попадает с вероятностью $1/n$. Для сравнения изобразим на гистограмме плотность стандартного нормального распределения (см. рис. 3, красная линия).

Как видно, гистограмма удовлетворительно аппроксимирует плотность стандартного нормального распределения в случае каждой из трех выборок.

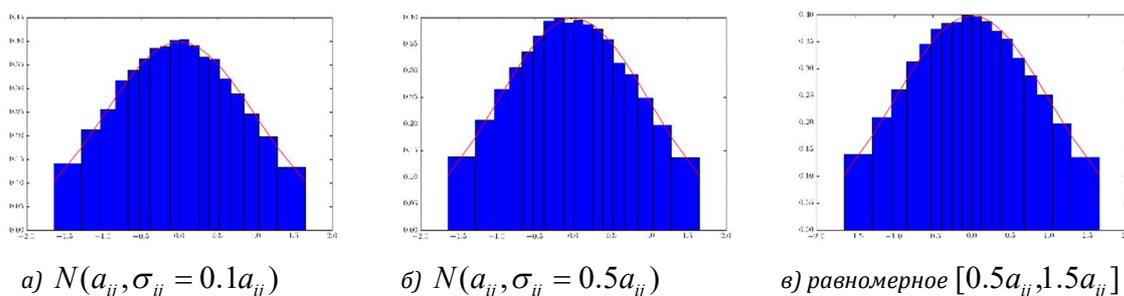


Рис. 3. Гистограммы линейно преобразованных выборок

Изобразим на графике (рис. 4) точки $(q_{st\ norm}, \bar{q})$, где $q_{st\ norm}$ квантиль стандартного нормального распределения, \bar{q} - квантиль выборки, элементы которой равны сложности решения

индивидуальной задачи коммивояжера с элементами платежной матрицы вида $N(a_{ij}, \sigma_{ij} = 0.1a_{ij})$, уровень квантилей принимает значения из (0,1).

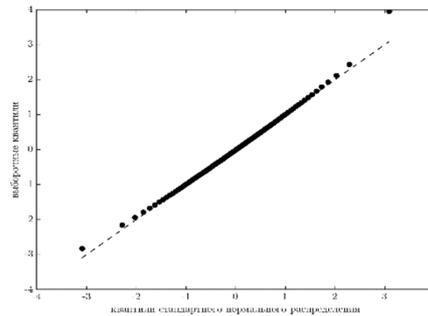


Рис. 4. Q-Q plot (график квантилей)

Как видно из рис. 4, самые большие и самые маленькие значения в выборке наблюдаются несколько чаще, чем это характерно для нормального распределения, в целом же нормальное распределение можно считать удовлетворительным приближением.

Уточнение параметров распределения выборки

Определив, что нормальное распределение достаточно хорошо описывает распределение логарифма сложности индивидуальных задач коммивояжера при их решении классическим алгоритмом методом ветвей и границ (при фиксированной длине входа), подберем параметры распределения, согласующиеся с выборкой.

Грубую оценку параметров можно провести по методу моментов, точечные оценки и доверительные интервалы приведены в табл. 2.

Более робастные оценки можно получить с использованием квантилей, в качестве оценки параметра a используем выборочную медиану. Параметр σ оценим, используя линейное преобразование выборки $L : Lx = kx + b$, при котором интерквантильный размах $E_6 - E_2$ преобразованной выборки равен интерквантильному размаху стандартного нормального распределения. Учитывая, что при линейном преобразовании новые выборочные квантили E_i^* равны

$$E_i^* = kE_i + b,$$

получим

$$k = \frac{E_6^* - E_2^*}{E_6 - E_2}. \tag{1}$$

В качестве оценки параметра σ используем k , значения, вычисленные для каждой из трех выборок также приведены в табл. 2.

Табл. 2 Оценки параметров нормального распределения

| выборка | a | | | σ | | |
|-------------------------------------|--------------------|-------------------|--------------------|--|-------------------|-------------------------------------|
| | Выборочное среднее | Доверит. интервал | Выборочная медиана | Выборочное среднеквадратическое отклонение | Доверит. интервал | отношение интерквантильных размахов |
| $N(a_{ij}, 0.1a_{ij})$ | 2.849 | (2.847, 2.850) | 2.842 | 0.291 | (0.290, 0.292) | 0.289 |
| $N(a_{ij}, 0.5a_{ij})$ | 2.638 | (2.636, 2.640) | 2.631 | 0.359 | (0.358, 0.361) | 0.359 |
| равномерн. $[0.5a_{ij}, 1.5a_{ij}]$ | 2.906 | (2.904, 2.908) | 2.899 | 0.352 | (0.351, 0.353) | 0.352 |

Как видно из табл. 2, значения параметров, вычисленные по методу моментов и с помощью квантилей, очень близки.

Оценка границ интервала, в который попадает более 90% выборки

Примерно 99.7% выборки (достаточно большого объема) из нормального распределения

попадает в интервал $(a-3\sigma, a+3\sigma)$, 99.99% попадает в интервал $(a-4\sigma, a+4\sigma)$, далее будем называть эти интервалы соответственно 3σ и 4σ . Границы интервалов 3σ и 4σ , минимальное и максимальное значения в выборке, а также доля выборки, попавшей в эти интервалы, приведены в табл. 3. Для удобства приведены сами значения сложности, а не их десятичные логарифмы. Интервал 90% обозначает интервал между 5% и 95% квантилями, так что примерно 5% выборки попадает левее этого интервала и 5% правее.

Табл. 3. Границы выборки и интервалов 90%, 3σ и 4σ

| выборка | Минимум в выборке | Максимум в выборке | 90% интервал | 3σ интервал | 4σ интервал |
|-------------------------------------|------------------------|--------------------|--------------|--------------------|--------------------|
| | $N(a_{ij}, 0.1a_{ij})$ | 69 | 248428 | (200;2100) | (100;5300), 99,58% |
| $N(a_{ij}, 0.5a_{ij})$ | 57 | 34531 | (100;1700) | (30;5200), 99,79% | (20;12000), 99,99% |
| равномерн. $[0.5a_{ij}, 1.5a_{ij}]$ | 57 | 92869 | (200;3100) | (70;9200), 99,74% | (30;20600), 99,98% |

Сравнивая максимальное значение сложности в выборке с правой границей 90% интервала, видим отличие до 100 раз, правая граница интервала 3σ почти в 3 раза больше правой границы 90% интервала и примерно в 2 раза меньше правой границы интервала 4σ . Таким образом, основная часть выборки сложности принимает значения до 2000-3000, и лишь около 5% выборки превосходит 2000, причем больше 5000 в выборке наблюдается менее чем в 0.5% случаев, а более 10000 бывает реже 0.1% (в случае нормально распределенных элементов).

Отметим, что правые границы 90% интервала и интервалов 3σ и 4σ в случае равномерно распределенных элементов матрицы стоимости примерно в 2 раза больше, чем при нормально распределенных стоимостях перелетов.

Уточнение параметров распределения

Будем рассматривать «слишком большие» и «слишком маленькие» значения, не попавшие в интервал 3σ и 4σ , как исключения из правила, погрешность эксперимента и т.п. Исключим такие значения из выборки и по полученной новой выборке вычислим медиану, среднее и среднеквадратическое отклонение, а также отношение интерквантильных размахов выборки и стандартного нормального распределения, коэффициенты асимметрии и эксцесса, в том числе квантильные. Результаты расчетов приведены в табл. 4 и 5.

Табл. 4 Характеристики усеченной выборки (медиана, среднее, среднеквадратическое отклонение)

| выборка | μ | | a | | σ | |
|-------------------------------------|----------|-----------|----------|-----------|----------|-----------|
| | исходная | усеченная | исходная | усеченная | исходная | усеченная |
| $N(a_{ij}, 0.1a_{ij})$ | 2.842 | 2.841 | 2.849 | 2.845 | 0.291 | 0.283 |
| $N(a_{ij}, 0.5a_{ij})$ | 2.631 | 2.630 | 2.638 | 2.635 | 0.359 | 0.355 |
| равномерн. $[0.5a_{ij}, 1.5a_{ij}]$ | 2.899 | 2.898 | 2.906 | 2.904 | 0.352 | 0.347 |

Табл. 5 Характеристики усеченной выборки (коэффициенты асимметрии и эксцесса)

| выборка | К-т асимметрии | | К-т эксцесса | | Квантильный к-т асимметрии | | Квантильный к-т эксцесса | |
|-------------------------------------|----------------|-------|--------------|-------|----------------------------|-------|--------------------------|-------|
| | исходн. | усеч. | исходн. | усеч. | исходн. | усеч. | исходн. | усеч. |
| $N(a_{ij}, 0.1a_{ij})$ | 0.26 | 0.09 | 0.63 | -0.18 | 0.006 | 0.006 | 1.232 | 1.229 |
| $N(a_{ij}, 0.5a_{ij})$ | 0.13 | 0.07 | -0.04 | -0.20 | 0.015 | 0.015 | 1.225 | 1.226 |
| равномерн. $[0.5a_{ij}, 1.5a_{ij}]$ | 0.14 | 0.08 | 0.04 | -0.17 | 0.012 | 0.012 | 1.234 | 1.236 |

Как видно из табл. 4. и 5. «усеченная выборка» имеет более близкие к теоретическим значения традиционных коэффициентов асимметрии и эксцесса, квантильные коэффициенты асимметрии и эксцесса практически одинаковы у исходной и усеченной выборки.

Таким образом, удовлетворительным приближением для вероятностного распределения

логарифма сложности индивидуальных задач коммивояжера классическим алгоритмом метода ветвей и границ можно считать нормальное распределение с параметрами a и σ , равными соответственно выборочной медиане и отношению интерквартильных размахов (1) или коэффициентам b_0 и k_0 линейной регрессии стандартного нормального распределения на выборку (см. табл. 1).

Заключение

Показано, что логарифм сложности решения задачи коммивояжера методом ветвей и границ (в случае платежной матрицы с нормально или равномерно распределенными элементами) можно считать нормально распределенным, оценки параметров можно получать на основе параметров линейной регрессии стандартного нормального распределения на выборку.

Поскольку тип аппроксимирующего распределения логарифма сложности оказался одинаковым в случае нормально и равномерно распределенных элементов платежной матрицы, возможно, класс распределений элементов платежной матрицы, приводящий к нормальному распределению логарифма сложности, содержит еще какие-то типы распределений.

Работа выполнена при поддержке гранта РФФИ №16-07-160.

Литература

1. Dantzig G. B., Fulkerson R., Johnson S. Solution of a large scale traveling salesman problem. Technical Report P-510. RAND Corporation, Santa Monica, California, USA. – 1954.
2. Dantzig, G. B., Fulkerson D. R., Johnson S. M. «On a linear programming, combinatorial approach to the traveling-salesman problem» // Operations Research 1959. №7, pp. 58–66.
3. Eastman, W. L. Linear Programming with Pattern Constraints. Ph.D. Thesis. Department of Economics, Harvard University, Cambridge, Massachusetts, USA. – 1958.
4. Johnson N.L., Kotz S., and Balakrishnan N. Continuous Univariate Distributions. Vol. 2, Wiley, 1995.
5. Knuth, D. E. «Estimating the efficiency of backtracking programs» // Mathematics of Computing, 1975. Vol. 29, pp. 121–136.
6. Land, A. H., Doig A. G. «An automatic method of solving discrete programming problems» // Econometrica 1960. №28, pp. 497–520.
7. Little, J. D. C., Murty K. G., Sweeney D.W., Karel C. «An algorithm for the traveling salesman problem» // Operations Research, 1963. №11, pp. 972–989.
8. Moors J. J. A. A quantile alternative for kurtosis // The Statistician, 1988. Vol. 37, pp. 25–32.
9. Moors J. J. A., Coenen V. M. J., Heuts R. M. J. «Limiting distributions of moment- and quantile-based measures for skewness and kurtosis». School of Economics and Management, Tilburg University, Res. Mem. FEW 620, 1993.
10. Moors J. J. A., Wagemakers R. Th. A., Coenen V. M. J., Heuts R. M. J., Janssens M. J. B. T. «Characterizing systems of distributions by quantile measures» // Statistica Neerlandica, 1996. Vol. 50, № 3, pp. 417–430.
11. Pearson, K. «Contributions to the Mathematical Theory of Evolution. III. Regression, Heredity and Panmixia» // Philosophical Transactions of the Royal Society of London, 1896. 187, pp. 253–318.
12. Головешкин В.А., Жукова Г.Н., Ульянов М.В., Фомичев М.И. «Сравнение ресурсных характеристик традиционного и модифицированного метода ветвей и границ для TSP» // Современные информационные технологии и ИТ-образование, 2015. Т. 2, № 11. С. 151–159.
13. Крамер Г. Математические методы статистики. — М.: Мир, 1975. — 648 с.
14. Ульянов М.В., Фомичев М.И. «Ресурсные характеристики способов организации дерева решений в методе ветвей и границ для задачи коммивояжера» // Бизнес – информатика, 2015. № 4 (34). С. 38–46.
15. Ульянов М.В. Ресурсно-эффективные компьютерные алгоритмы. Разработка и анализ. — М.: ФИЗМАТЛИТ, 2008. — 304 с.

References

1. Dantzig G. B., Fulkerson R., Johnson S. Solution of a large scale traveling salesman problem. Technical Report P-510. RAND Corporation, Santa Monica, California, USA. – 1954.
2. Dantzig, G. B., Fulkerson D. R., Johnson S. M. «On a linear programming, combinatorial approach to the traveling-salesman problem» // Operations Research 1959. №7, pp. 58–66.
3. Eastman, W. L. Linear Programming with Pattern Constraints. Ph.D. Thesis. Department of Economics, Harvard University, Cambridge, Massachusetts, USA. – 1958.
4. Johnson N.L., Kotz S., and Balakrishnan N. Continuous Univariate Distributions. Vol. 2, Wiley, 1995.
5. Knuth, D. E. «Estimating the efficiency of backtracking programs» // Mathematics of Computing, 1975. Vol. 29, pp. 121–136.
6. Land, A. H., Doig A. G. «An automatic method of solving discrete programming problems» // Econometrica 1960. №28, pp. 497–520.
7. Little, J. D. C., K. G. Murty, D.W. Sweeney, C. Karel «An algorithm for the traveling salesman problem» // Operations Research, 1963. №11, pp. 972–989.
8. Moors J. J. A. A quantile alternative for kurtosis // The Statistician, 1988. Vol. 37, pp. 25–32.
9. Moors J. J. A., Coenen V. M. J., Heuts R. M. J. «Limiting distributions of moment- and quantile-based measures for skewness and kurtosis». School of Economics and Management, Tilburg University, Res. Mem. FEW 620, 1993.
10. Moors J. J. A., Wagemakers R. Th. A., Coenen V. M. J., Heuts R. M. J., Janssens M. J. B. T. «Characterizing systems of distributions by quantile measures» // Statistica Neerlandica, 1996. Vol. 50, № 3, pp. 417–430.
11. Pearson, K. «Contributions to the Mathematical Theory of Evolution. III. Regression, Heredity and Panmixia» // Philosophical Transactions of the Royal Society of London, 1896. 187, pp. 253–318.

12. Goloveshkin V.A., Zhukova G.N., Ul'yanov M.V., Fomichev M.I. «Sravnenie resursnykh kharakteristik traditsionnogo i modifitsirovannogo metoda vetvey i granits dlya TSP» // *Sovremennye informatsionnye tekhnologii i IT-obrazovanie*, T. 2, № 11, 2015, S. 151-159.
13. Kramer G. *Matematicheskie metody statistiki*. M.: Mir, 1975. — 648 s.
14. Ul'janov M.V., Fomichev M.I. «Resursnye harakteristiki sposobov organizacii dereva reshenij v metode vetvej i granic dlja zadachi kommivojazhera» // *Biznes – informatika*, 2015. № 4 (34). S. 38–46
15. Ul'janov M.V. *Resursno-jeffektivnye komp'juternye algoritmy. Razrabotka i analiz*. — M.: FIZMATLIT, 2008. — 304 s.

Поступила: 10.10.2016

Об авторах:

Головешкин Василий Адамович, д.т.н., профессор Московского государственного университета информационных технологий, радиотехники и электроники (МИРЭА), nikshevolog@yandex.ru;

Жукова Галина Николаевна, к.ф.-м.н., доцент Московского политехнического университета (МПУ), galinanzhukova@gmail.com;

Ульянов Михаил Васильевич, д.т.н., профессор, ведущий научный сотрудник Института проблем управления РАН им. В.А. Трапезникова, muljanov@mail.ru;

Фомичев Михаил Игоревич, студент магистратуры ФКН НИУ Высшая школа экономики, michan94@yandex.ru.