

Использование квантильных коэффициентов асимметрии и эксцесса для оценки сложности решения задачи коммивояжера

В.А. Головешкин, Г.Н. Жукова, М.В. Ульянов, М.И. Фомичев

Аннотация — Исследуется сложность индивидуальных задач коммивояжера, т.е. число порожденных вершин поискового дерева в классическом методе ветвей и границ. Вероятностное распределение логарифма сложности аппроксимируется нормальным распределением. На основе экспериментальных данных рассчитаны значения параметров линейного преобразования, обеспечивающих минимальное среднеквадратическое отклонение выборочных квантилей логарифма сложности от соответствующих квантилей стандартного нормального распределения, получена формула зависимости этих параметров от числа вершин графа.

Ключевые слова— задача коммивояжера, квантильный коэффициент асимметрии, квантильный коэффициент эксцесса, коэффициент асимметрии, коэффициент эксцесса, сложность.

I. ВВЕДЕНИЕ

Задача коммивояжера (TSP, traveling salesman problem) заключается в поиске оптимального (например, самого дешевого) гамильтонова цикла (тура) в полном ориентированном взвешенном графе. Формально задача может быть решена полным перебором всех возможных вариантов, однако при большом числе вершин графа это невозможно сделать за разумное время. Одним из наиболее распространенных способов оптимизации перебора является метод ветвей и границ МВГ, требующий обычно рассмотрения гораздо менее чем $(n-1)!$ туров в графе с n вершинами [1]-[6]. Однако сейчас еще нет формулы, по которой можно вычислить заранее точное количество туров, которые придется рассмотреть в каждой конкретной задаче. Существуют некоторые подходы к оценке и прогнозированию числа вершин поискового дерева [7]-[10], но для их применения требуется в течение какого-то времени проводить вычисления, чтобы на основе информации о начальном этапе расчетов строить прогноз.

Статья получена 31.10.2016.

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

Г.Н. Жукова работает в Московском политехническом университете (МПУ) (e-mail: gnzh@mail.ru)

В.А. Головешкин работает в Институте прикладной механики РАН, Московском государственном университете информационных технологий, радиотехники и электроники (МИРЭА) (e-mail: nikshevolog@yandex.ru)

М.В. Ульянов работает в Институте проблем управления РАН им. В.А. Трапезникова (e-mail: muljanov@mail.ru)

М.И. Фомичев является студентом магистратуры Высшей школы экономики (ФКН НИУ ВШЭ) (e-mail: michan94@yandex.ru)

Время, которое требуется компьютерной программе, реализующей метод ветвей и границ, для нахождения решения задачи коммивояжера, существенно зависит от размерности и особенностей матрицы стоимостей задачи [17]-[19], а также от параметров процессора, на котором производится вычисления, от особенностей операционной системы и т.п., поэтому вместо физического времени мы будем использовать связанные с ним независимые от ЭВМ характеристики, а именно — сложность индивидуальной задачи. Понятие сложности задачи коммивояжера введено Д.Э. Кнудом в [9] и определяется как общее число порожденных вершин поискового дерева решений в классическом методе ветвей и границ [1]. Мы будем использовать помимо числа порожденных вершин поискового дерева, и такие характеристики как максимальное (в динамике вычислений) число вершин дерева решений, и максимальное количество листьев дерева. Все эти величины не зависят от параметров ЭВМ, а определяются только матрицей стоимости индивидуальной задачи коммивояжера и выбором алгоритма решения TSP.

Предлагается оценивать сложностные характеристики TSP, рассматривая их как случайные величины, на основе использования характеристик асимметрии и эксцесса, как традиционных, выражаемых через центральные моменты, так и определяемых с помощью квантилей. Обычные коэффициенты асимметрии и эксцесса (см. [11]) используются для выбора наиболее подходящего типа распределения, удовлетворительно описывающего исследуемые характеристики. Далее с помощью квантильных коэффициентов асимметрии и эксцесса уточняется тип распределения.

Для исследования сложности решения TSP была проведена серия экспериментов, в которых матрица стоимости задачи коммивояжера заполнялась псевдослучайными целыми числами, затем полученная индивидуальная задача решалась методом ветвей и границ без предварительного вычисления рекорда, при этом подсчитывалось число порожденных вершин дерева решений, а также максимальное число вершин дерева и максимальное количество листьев. Таким образом используемые характеристики являются инвариантными относительно ЭВМ и языка программирования характеристиками индивидуальной задачи коммивояжера. Эти экспериментальные данные и послужили основой для последующих исследований,

связанных с изучением распределения сложности индивидуальных задач коммивояжера.

Рассматривались как симметричная, так и асимметричная задачи коммивояжера, задаваемые матрицей стоимости, элементами которой были равномерно распределенные псевдослучайные целые числа в диапазоне от 1 до 2^{15} . Поисковое дерево решений хранилось в памяти в виде бинарной кучи. Для каждого фиксированного n (числа вершин графа) выполнялось 10^4 (а в некоторых случаях 10^5) псевдослучайных генераций матрицы стоимостей порядка n . Эксперименты проводились при значениях n от 10 до 49.

Полученные на основе данных вычислительных экспериментов распределения сложностных характеристик, являющиеся приближениями неизвестных распределений этих характеристик, используются для прогнозирования параметров распределения тех же характеристик при большем числе вершин, что позволяет прогнозировать, в частности, среднее значение характеристик, квантили и т.д. Подчеркнем, что полученные нами формулы имеет смысл использовать для прогноза при n порядка 10^2 , а не при $n \rightarrow \infty$.

II. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ

Сложность, максимальное число порожденных вершин, количество листьев и т.д. рассматриваются как случайные величины, их функции распределения аппроксимируются некоторыми функциями распределения непрерывного типа с конечным четвертым моментом. Для подбора вида распределения использован метод К. Пирсона, в соответствии с которым идентификация типа распределения проводится на основе коэффициентов асимметрии γ_1 и эксцесса γ_2 (см. [11]-[13]).

Пусть наблюдается случайная величина X , имеющая математическое ожидание EX , дисперсию DX , третий и четвертый центральные моменты $\mu_m(X) = E(X - EX)^m$, $m = 3, 4$, тогда коэффициент асимметрии γ_1 и коэффициент эксцесса γ_2 определяются по формулам [11]

$$\gamma_1 = \frac{\mu_3(X)}{(DX)^{3/2}}, \quad \gamma_2 = \frac{\mu_4(X)}{(DX)^2} - 3. \quad (1)$$

В соответствии с методом К. Пирсона в системе координат «коэффициент асимметрии – коэффициент эксцесса» изобразим некоторые классические распределения, а также точки, соответствующие значениям коэффициентов асимметрии и эксцесса экспериментальных данных (см. рис. 1). Данные представляют собой положительные целые числа, равные числу порожденных вершин при решении асимметричной задачи коммивояжера размерности от 20 до 49 программной реализацией классического метода ветвей и границ (от 10^4 до 10^5 значений для каждой размерности задачи).

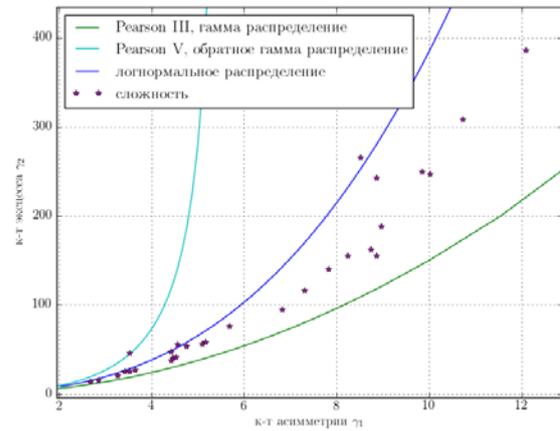


Рис. 1 Карта коэффициентов асимметрии и эксцесса

В системе координат «коэффициент асимметрии – коэффициент эксцесса» звездочки, соответствующие экспериментальным данным (на легенде обозначены как сложность), попадают в область распределений шестого типа системы Пирсона, расположенную между зеленой линией гамма распределения и голубой линией обратного гамма распределения. Гамма распределение является типом III системы Пирсона, обратное гамма распределение — V тип, бета распределение второго типа (Beta prime distribution) — VI тип. В область VI типа попадает и логнормальное распределение, поскольку при некоторых значениях параметров коэффициенты асимметрии и эксцесса бета II распределения и логнормального совпадают.

Далее тип и параметры распределения уточнялись в соответствии с методом, основанным на квантилях (см. [14]-[16]). Метод аналогичен подходу К. Пирсона, но асимметрия и эксцесс распределений выражаются через квантили, что позволяет получать более точные и робастные оценки.

Обозначим $E_i = p_{i/8} = F^{-1}(i/8)$, $i = 1, 2, \dots, 7$ октили непрерывного распределения $F(x)$ и рассмотрим четверку (Q, R, S, T) [16], где

$Q = E_4$ – медиана,

$R = (E_6 - E_2)/2$ – половина интерквантильного размаха,

$S = (E_6 - 2E_4 + E_2)/(E_6 - E_2)$ – квантильный коэффициент асимметрии,

$T = (E_7 - E_5 + E_3 - E_1)/(E_6 - E_2)$ – квантильный коэффициент эксцесса.

Эта четверка рассматривается как квантильный аналог набора $(\mu, \sigma, \gamma_1, \gamma_2)$, использующегося в методе моментов (математическое ожидание, среднеквадратическое отклонение, коэффициенты асимметрии и эксцесса). Квантильные коэффициенты асимметрии и эксцесса инвариантны относительно линейных преобразований, как и обычные коэффициенты асимметрии и эксцесса.

В системе координат (γ_1, γ_2) и (S, T) изобразим линии, соответствующие логнормальному распределению, а также гамма распределению и обратному гамма распределению (см. рис. 1 и 2)

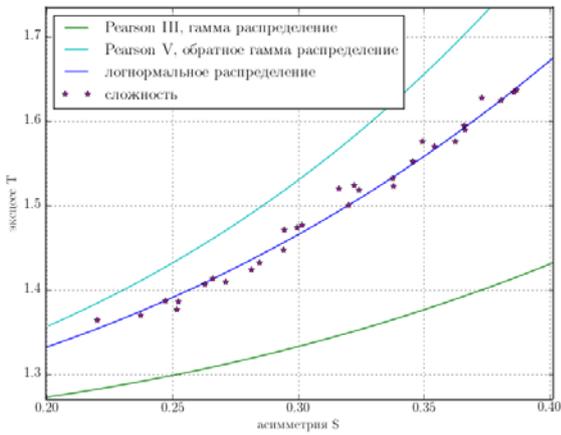


Рис. 2 Карта квантильных коэффициентов асимметрии и эксцесса

На карте системы функций К.Пирсона точки с координатами, равными выборочным значениям коэффициентов асимметрии и эксцесса сложности TSP, попадают в ту часть области распределения бета второго рода, которая расположена ниже линии логнормального распределения. В системе координат квантильных мер асимметрии и эксцесса экспериментальные данные в основном попадают в область распределения бета второго рода, но ближе к линии логнормального распределения (см. рис. 1).

На основании выборочных квантильных мер асимметрии и эксцесса можно считать удовлетворительным приближением для вероятностного распределения сложности TSP (при равномерном распределении элементов матрицы стоимости) логнормальное распределение. В таком случае распределение логарифма сложности близко к нормальному, что и иллюстрируется рис. 3 и 4.

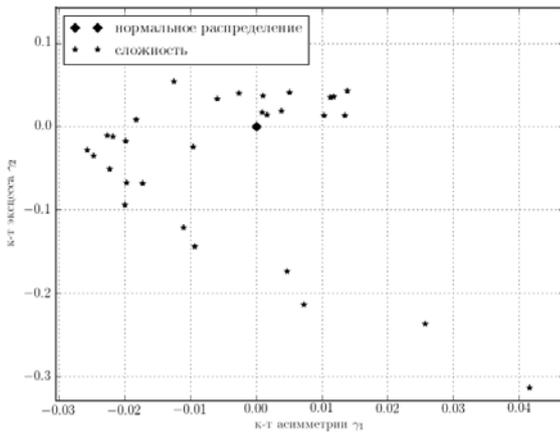


Рис. 3 Логарифм сложности в системе (γ_1, γ_2)

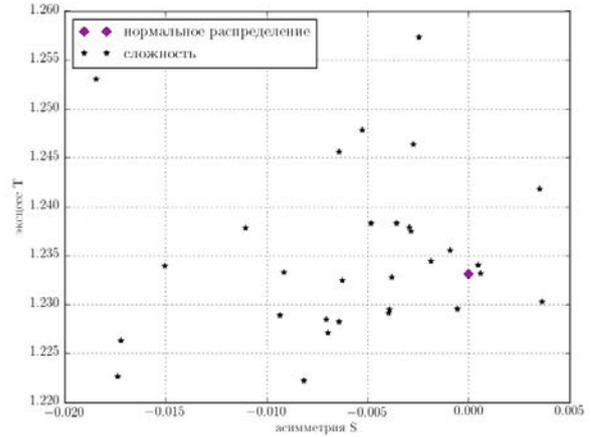


Рис. 4 Логарифм сложности в системе (S, T)

III. ЛИНЕЙНОЕ ПРЕОБРАЗОВАНИЕ

Обозначим линейное преобразование $Y = kX + b$, где X — наблюдаемая случайная величина, $q_k = F_s^{-1}(k/20)$, $i = 1, 2, \dots, 19$ выборочные квантили Y , F_s^{-1} — обратная функция к выборочной функции распределения, $Q_k = F^{-1}(k/20)$, $i = 1, 2, \dots, 19$ квантили стандартного нормального распределения, F^{-1} — обратная функция к стандартной нормальной функции распределения.

Поскольку квантильные коэффициенты асимметрии и эксцесса подобно традиционным инвариантны относительно линейных преобразований, то к натуральному логарифму экспериментальных данных применяется линейное преобразование с параметрами k и b , при которых среднее квадратическое отклонение выборочных квантилей q_k от Q_k минимально:

$$\sum_{k=1}^{19} (Q_k - q_k)^2 \rightarrow \min \quad (2)$$

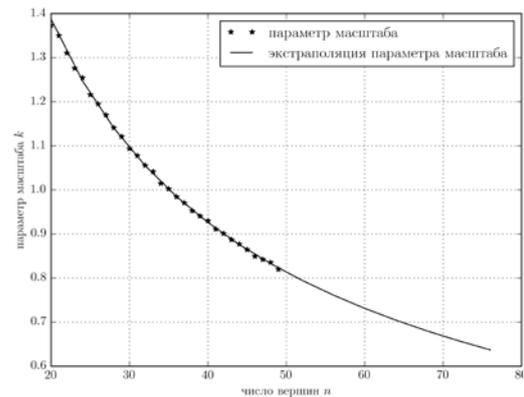


Рис. 5 Параметр масштаба для логарифма сложности

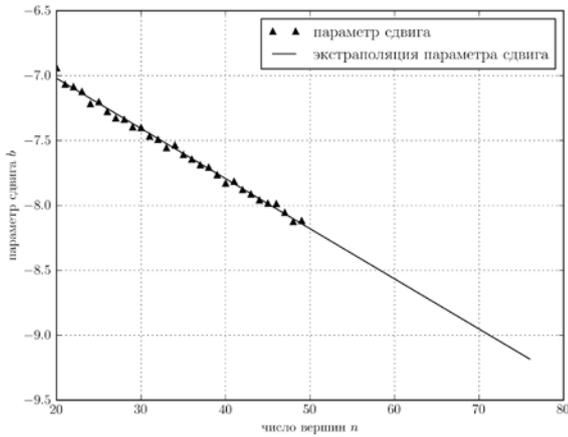


Рис. 6 Параметр сдвига для натурального логарифма сложности

Коэффициент регрессии выборочных квантилей логарифма сложности на квантили стандартного нормального распределения принимал значения от 0.999 до 1.000, что свидетельствует об очень тесной линейной связи.

IV. ЭКСТРАПОЛЯЦИЯ ПАРАМЕТРОВ ЛИНЕЙНОГО ПРЕОБРАЗОВАНИЯ

Как видно на рис. 6, параметр сдвига b линейно убывает с ростом числа вершин n , линейная регрессия b на n приводит к формуле

$$b = -0.0386n - 6.2483, \quad (3)$$

коэффициент регрессии -0.997 и среднеквадратическое отклонение 0.0005 показывают наличие тесной линейной зависимости.

Натуральный логарифм параметра масштаба k линейно убывает с ростом натурального логарифма числа вершин n , линейная регрессия $\ln k$ на $\ln n$ приводит к формуле

$$\ln k = -0.58 \ln n + 2.07, \quad k = 7.94n^{-0.58}. \quad (4)$$

коэффициент регрессии -0.9997 и среднеквадратическое отклонение 0.002 также свидетельствуют о тесной линейной зависимости.

Таким образом, распределение линейного преобразования натурального логарифма сложности близко к стандартному нормальному:

$$7.94n^{-0.58}X - 0.0386n - 6.2483 \approx N(0,1) \quad (5)$$

$$\hat{q}_k \approx 0.126n^{0.58}(Q_j + 0.0386n + 6.2483), \quad (6)$$

где \hat{q}_k и Q_k соответственно выборочная квантиль натурального логарифма сложности и квантиль стандартного нормального распределения.

Полученную формулу (6) используем для расчета квантилей распределения натурального логарифма сложности при больших n . На рис. 7 показаны квантили 5%, 50% и 95%, на основе которых можно оценивать сложность решения задачи методом ветвей и границ при числе вершин больше 40.

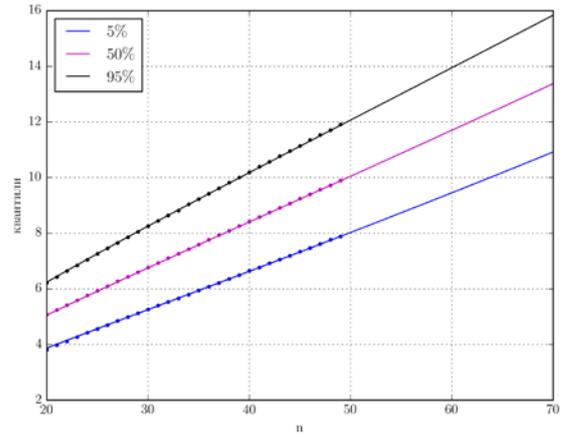


Рис. 7 Прогноз для квантилей натурального логарифма сложности

Параметры уравнения (6) были рассчитаны по выборке объема 10^5 натуральных логарифмов сложности индивидуальных задач коммивояжера.

Аналогичные формулы получены для сложности, максимального числа вершин дерева решений и для максимального числа листьев в очереди при объеме выборки 10^4 для симметричной и асимметричной TSP:

$$\hat{q}_k \approx An^\alpha(Q_j + Bn + C), \quad (7)$$

параметры уравнения, полученные по выборке объема 10^4 , приведены в таблице 1.

Таблица 1. Параметры уравнения (7).

		α	A	B	C
асимметричная	сложность	0.5651	0.1336	0.0426	6.1230
	максимальное количество вершин дерева решений	0.5680	0.1321	0.0432	6.0632
	максимальное количество листьев в очереди	0.5564	0.1382	0.0600	4.7168
симметричная	сложность	0.5644	0.1307	0.0864	7.7553
	максимальное количество вершин дерева решений	0.5676	0.1281	0.0873	7.7501
	максимальное количество листьев в очереди	0.5611	0.1310	0.1063	6.3532

V. ЗАКЛЮЧЕНИЕ

Оценка сложности решения задачи коммивояжера получена путем последовательного применения метода Пирсона, квантильных коэффициентов асимметрии и эксцесса и линейного преобразования, минимизирующего среднеквадратическое отклонение выборочных квантилей от квантилей стандартного нормального распределения.

Установлено наличие тесной линейной связи квантилей логарифма сложности и стандартного нормального распределения. Получена формула (7) для

оценки квантилей распределения натурального логарифма сложности для большего числа вершин.

БИБЛИОГРАФИЯ

- [1] J. D. C. Little, K. G. Murty, D.W. Sweeney, and C. Karel, "An algorithm for the traveling salesman problem," *Operations Research*, vol. 11, pp. 972–989, 1963.
- [2] G. B. Dantzig, R. Fulkerson, and S. Johnson, "Solution of a large scale traveling salesman problem," RAND Corp., Santa Monica, CA, Tech. Rep. P-510, 1954.
- [3] G. B. Dantzig, D. R. Fulkerson, and S. M. Johnson, "On a linear programming, combinatorial approach to the traveling-salesman problem," *Operations Research*, vol. 7, pp. 58–66, 1959.
- [4] W. L. Eastman, "Linear Programming with Pattern Constraints," Ph.D. thesis, Dept. Economics, Harvard Univ., Cambridge, MA, 1958.
- [5] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, pp. 497–520, 1960.
- [6] A. S. Manne and H. M. Markowitz, "On the solution of discrete programming problems," RAND Corp., Santa Monica, CA, Tech. Rep. P-711, 1956.
- [7] C. Cotta, J. Aldana, A. Nebro, and J. Troya, "Hybridizing genetic algorithms with branch and bound techniques for the resolution of the TSP", in *Artificial Neural Nets and Genetic Algorithms*, D. Pearson, N. Steele, R. Albrecht, Eds. Wien New York: Springer–Verlag, 1995, pp. 277-280.
- [8] G. Carpaneto and P. Toth, "Some new branching and bounding criteria for the asymmetric traveling salesman problem," *Management Science*, vol. 26, 1980, pp. 736–743.
- [9] D. E. Knuth, "Estimating the efficiency of backtracking programs," *Mathematics of Computing*, vol. 29, 1975, pp. 121–136.
- [10] G. Cornuéjols, M. Karamanov, and Y. Li, "Early estimates of the size of branch-and-bound trees," *INFORMS J. Comp.*, vol. 18, No. 1, 2006, pp. 86–96.
- [11] Г. Крамер, "Математические методы статистики", М.: Мир, 1975, 648 с.
- [12] N.L. Johnson, S. Kotz, and N. Balakrishnan, "Continuous Univariate Distributions," vol. 2, Wiley, 1995.
- [13] K. Pearson, "Contributions to the Mathematical Theory of Evolution. III. Regression, Heredity and Panmixia," *Phil. Trans. Royal Soc. London*, vol. 187, 1896, pp. 253–318.
- [14] J. J. A. Moors, "A quantile alternative for kurtosis," *The Statistician*, vol. 37, 1988, pp. 25–32.
- [15] J. J. A. Moors, V. M. J. Coenen, and R. M. J. Heuts, "Limiting distributions of moment- and quantile-based measures for skewness and kurtosis", School of Economics and Management, Tilburg University, Res. Mem. FEW 620, 1993
- [16] J. J. A. Moors, R. Th. A. Wagemakers, V. M. J. Coenen, R. M. J. Heuts, and M. J. B. T. Janssens, "Characterizing systems of distributions by quantile measures", *Statistica Neerlandica*, vol. 50, No 3, pp. 417–430, Nov. 1996.
- [17] М.В. Ульянов, М.И. Фомичев, "Ресурсные характеристики способов организации дерева решений в методе ветвей и границ для задачи коммивояжера", *Бизнес – информатика*, №4, 2015.
- [18] М.В. Ульянов, "Ресурсно-эффективные компьютерные алгоритмы. Разработка и анализ." М.: ФИЗМАТЛИТ, 2008. 304 с.
- [19] В.А. Головешкин, Г.Н. Жукова, М.В. Ульянов, М.И. Фомичев "Сравнение ресурсных характеристик традиционного и модифицированного метода ветвей и границ для TSP," *Современные информационные технологии и ИТ-образование*, Т. 2, № 11, 2015, 614 с.

The estimation of the complexity of solving a particular travelling salesman problem by quantile-based measures for skewness and kurtosis

Goloveshkin V. A., Zhukova G.N., Ulyanov M.V., Fomichev M.I.

Abstract—The complexity of solving a particular travelling salesman problem is studied. Complexity is a number of nodes of the decision tree, when a particular problem is being solved by the branch and bound algorithm. A probability distribution of the logarithm of the complexity of a particular TSP is approximately normal. Parameters of the linear transformation of a sample of the logarithm of the complexity into a standard normally distributed sample are obtained by the method of least squares for the sample quantiles. The formulas for the parameters are also given.

Keywords—traveling salesman problem, TSP, branch and bound method, B&B, complexity, efficiency, quantile, quantile skewness, quantile kurtosis.

Goloveshkin V. A. Moscow Technological University (MIREA), Moscow, Russia

Zhukova G.N. Moscow Polytechnic University (MPU), Moscow, Russia

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

Fomichev M.I. Higher School of Economics National Research University, Moscow, Russia