XVI ВСЕРОССИЙСКАЯ КОНФЕРЕНЦИЯ МОЛОДЫХ УЧЕНЫХ ПО МАТЕМАТИЧЕСКОМУ МОДЕЛИРОВАНИЮ И ИНФОРМАЦИОННЫМ ТЕХНОЛОГИЯМ YM XVI

г. Красноярск, 28-30 октября 2015 г.

ПРОГНОЗИРОВАНИЕ СЛОЖНОСТИ ИНДИВИДУАЛЬНЫХ ЗАДАЧ КОММИВОЯЖЕРА

М. В. Ульянов

muljanov@mail.ru

д-р. техн. наук, проф.,

в.н.с. ИПУ им. В.А. Трапезникова РАН,

профессор кафедры алгоритмических языков ВМК МГУ

г. Москва

АКТУАЛЬНОСТЬ

Для современных наукоемких технологий актуальной и практически интересной является *проблема прогнозирования временных характеристик* для задач большой размерности и вычислительной сложности.

К этому классу задач относится, очевидно, и задача коммивояжера (TSP), принадлежащая в терминологии теории сложности к NP-трудному классу задач. Несмотря на более чем полувековые исследования этой задачи, точные методы ее решения имеют экспоненциальную сложность, а принадлежность алгоритмов, доставляющих ее точные решения, к классу NPRH, — классу алгоритмов *с сильной параметрической зависимостью* трудоемкости при фиксированной размерности, — не позволяет сегодня эффективно быстро получить априорные *прогнозы времени выполнения индивидуальной задачи*.

Разработка подходов к прогнозированию временных характеристик индивидуальных задач (конкретных проблем в терминологии Э.Л. Поста) представляет теоретический интерес в области анализа алгоритмов. А возможные решения этой задачи обладают и очевидной практической ценностью.

ЗАДАЧА КОММИОЯЖЕРА (TSP) (I)

Содержательная постановка задачи

Коммивояжер должен посетить ряд городов с целью рекламы и продажи товаров. Между каждой парой городов существует собственное транспортное сообщение. Будем называть *туром* такой порядок посещения, при котором каждый город посещается только один раз. Стоимости проезда между городами известны, задача состоит в том, чтобы найти *тур коммивояжера с минимальной стоимостью*.

Постановка в евклидовом целочисленном пространстве E_z^{n-1}

Обозначим через $\pi(k,l), l \geq k$ множество всех перестановок целых чисел $k, k+1, \cdots, l$, и рассмотрим множество перестановок $\pi(2,n)$ — оно содержит $|\pi(2,n)| = (n-1)!$ различных перестановок — ровно столько различных *туров коммивояжера* в задаче с n городами. Рассмотрим полный граф на n вершинах. *Тур — гамильтонов цикл*.

Фиксируем вершину с номером один, тогда некоторая перестановка из множества $\pi(2,n)$ задает порядок обхода вершин, начиная с первой, и после последней вершины, заданной этой перестановкой, мы снова возвращаемся в начало тура — вершину 1.

ЗАДАЧА КОММИОЯЖЕРА (TSP) (II)

Рассмотрим вектор $\mathbf{x}$ в пространстве E_z^{n-1}, компоненты $\mathbf{x}$ — это числа от 2 до n, а сам вектор ассоциирован с некоторой перестановкой из $\pi(2,n)$, и мы можем записать

$$\mathbf{x} \in \pi(2,n), \quad x_i \in \{\, 2,\ldots,n \,\},\, i = \overline{1,n-1},\, x_i \neq x_j \text{, при } i \neq j .$$

Заметим, что при этом концевые точки различных векторов $\mathbf{x}$ являются точками на положительной полусфере в E_z^{n-1} с центром в нуле и радиусом

$$r = \sqrt{\sum_{i=2}^{n} i^2} = \sqrt{\frac{n\cdot(n+1)\cdot(2n+1)}{6} - 1}, \quad \mathbf{x} \in S_z^{n-1}(\mathbf{0}, r).$$

Определим отображение $N \times N \underset{c}{\to} R^+, (i,j) \to c(i,j), (i,i) \to \infty$, которое каждой упорядоченной паре вершин графа ставит в соответствие стоимость ребра, инцидентного этой паре. В построенном формализме задача коммивояжера в пространстве E_z^{n-1} имеет следующую постановку

$$\mathbf{x} = (x_1, \cdots, x_{n-1}) \in \pi(2,n), \quad f(\mathbf{x}) = c(1, x_1) + \sum_{i=1}^{n-2} c(x_i, x_{i+1}) + c(x_{n-1}, 1)$$

$$\mathbf{x}^* = \underset{\mathbf{x} \in \pi(2,n)}{\arg\min}\, f(\mathbf{x})$$

МЕТОД ВЕТВЕЙ И ГРАНИЦ (I)

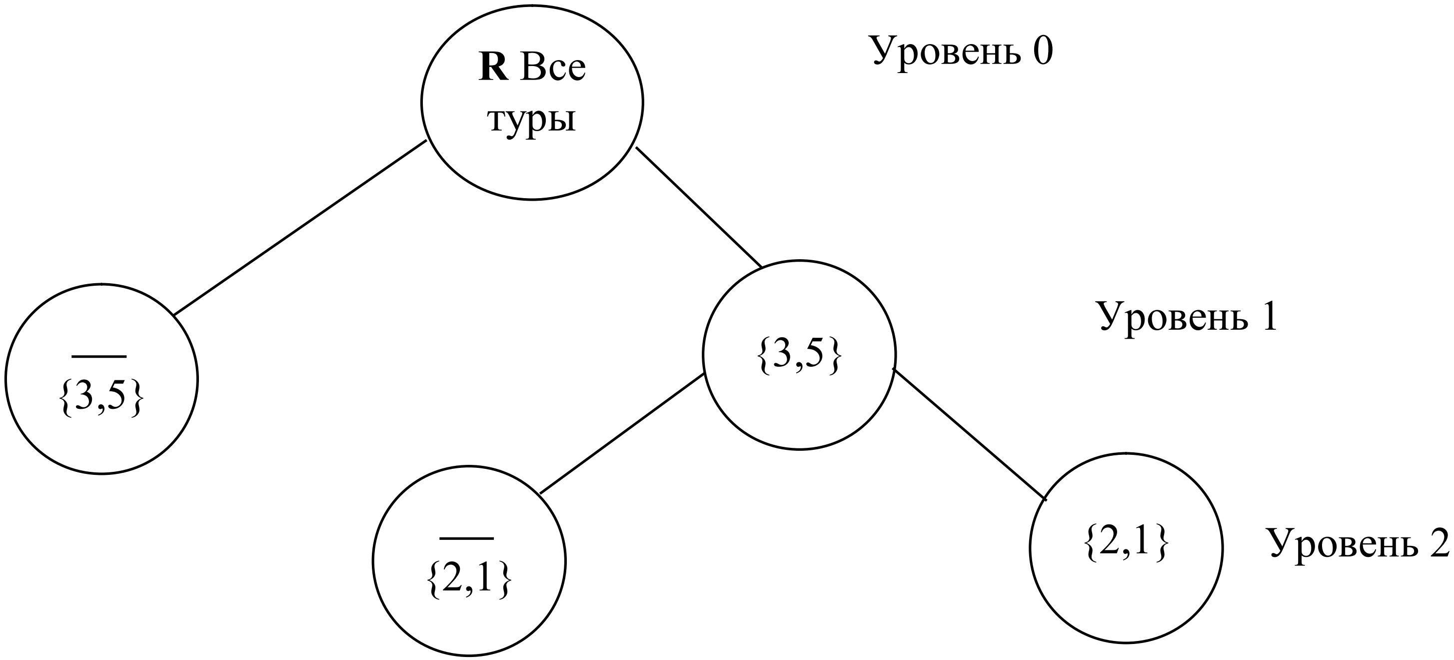
Общая идея метода ветвей и границ предполагает разделение всего множества допустимых решений на подмножества с целью дальнейшего сокращения перебора — это *процедура ветвления*.

С каждым таким подмножеством должна быть связана оценка (нижняя граница при поиске минимума), обеспечивающая отсечение тех подмножеств, которые заведомо не содержат оптимального решения — это *процедура построения границ*.

Классическая реализация для TSP (Little J. D. C., Murty K. G., Sweeney D. W., and Karel C. An algorithm for the traveling salesman problem // Operations Research. v11 (1963), pp. 972–989.)

Корень поискового дерева решений дерева представляет множество R всех $(n-1)!$ возможных туров в задаче с n городами. Ветви, выходящие из вершины, определяются выбором одного ребра, например, ребра (k,l). Идея авторов — специальный алгоритм выбора ребра (k,l), которое, вполне вероятно, входит в оптимальный тур. Множество R разделяется на два множества $\{k,l\}$ и $\{\overline{k,l}\}$. Во множество $\{k,l\}$ входят все туры из R, проходящие через ребро (k,l), а во множество $\{\overline{k,l}\}$ — туры, не содержащие это ребро.

Пример фрагмента дерева:



С вершиной дерева связывается нижняя граница стоимости. Предположим, уже получен конкретный полный тур T со стоимостью $s(T)$. Если нижняя граница, для некоторой вершины поискового дерева, больше, чем $s(T)$, то до конца процесса поиска не нужно рассматривать эту и все следующие за ней вершины — редукция дерева решений.

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ (I)

Для каждой фиксированной длины входа (25-45) выполнялось 10 000 псевдослучайных генераций матрицы стоимости, для каждой из которых запускалась классическая реализация алгоритма. Для каждого запуска замерялось:

— время выполнения программной реализации алгоритма;

— порожденное число вершин поискового дерева решений.

Эксперименты проводились на стационарном компьютере со следующими характеристиками:

— процессор: Intel i7 3770K 3800 ГГц

— оперативная память: Kingston KHX1600C9D3P1 16 ГБ

— материнская плата: GIGABYTE GA-Z77X-D3H

— операционная система: Fedora 21 workstation

Алгоритмы реализованы на языке C++.

Версия компилятора: gcc 4.9.2 20150212 (Red Hat 4.9.2-6) (GCC).

(Результаты получены М.И. Фомичевым)

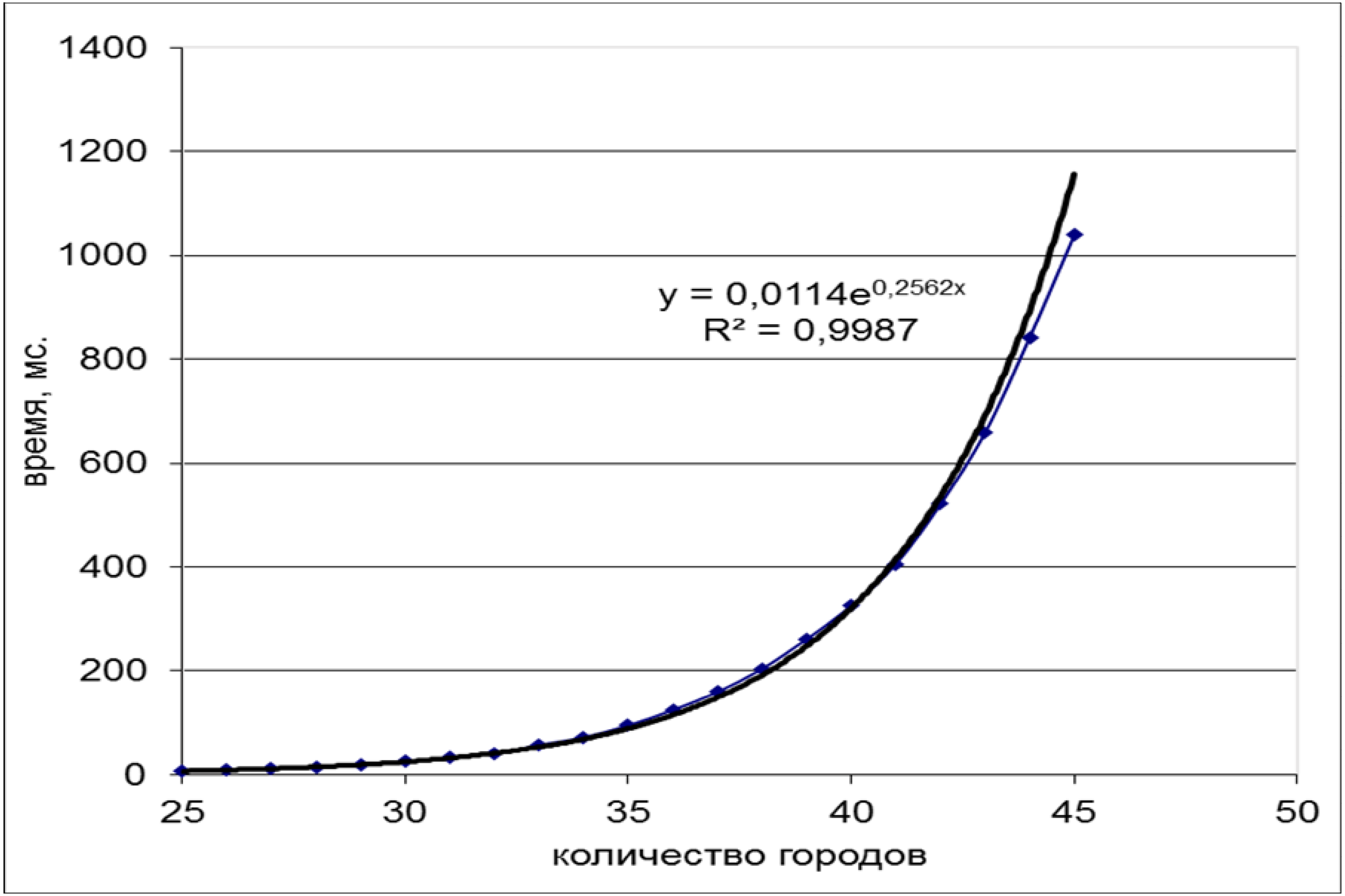


Рис. 1. Зависимость среднего времени расчёта тура от длины входа

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ (III)

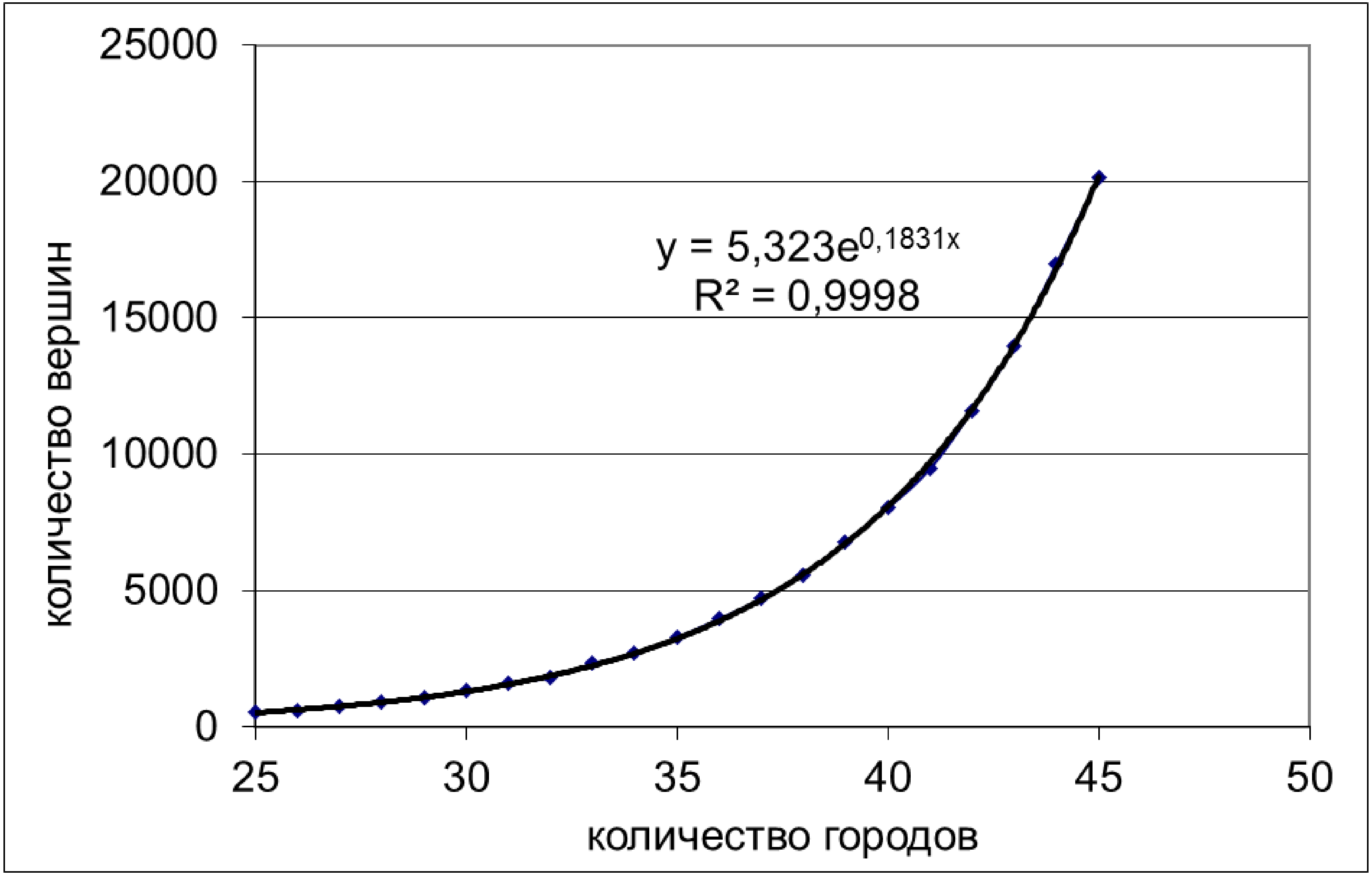


Рис. 2. Зависимость среднего числа вершин поискового дерева решений от длины входа

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ (IV)

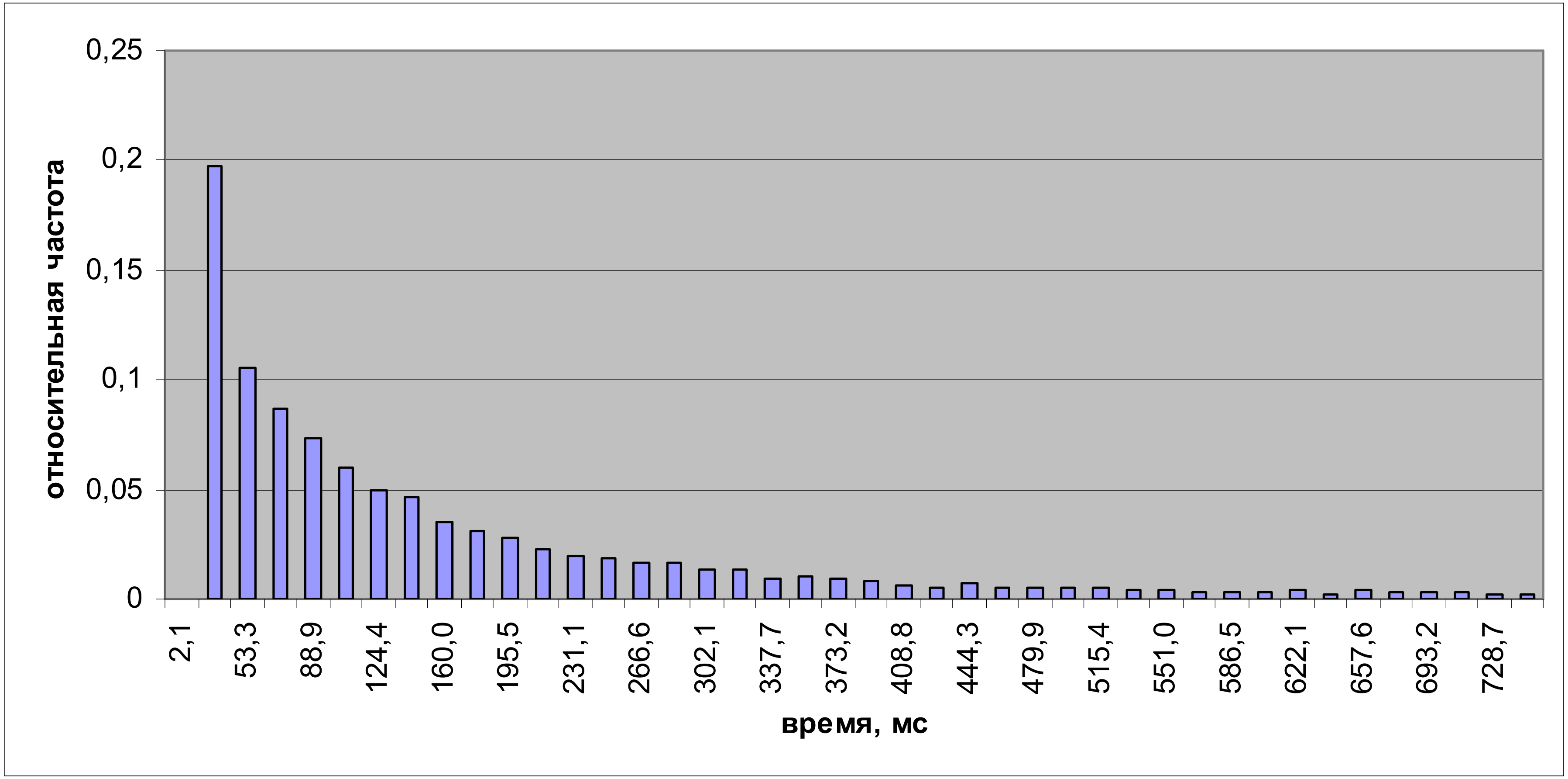


Рис. 3. Относительные частоты времен выполнения для задачи с n=45

(первые 42 полусегмента из 500 по результатам 10000 экспериментов).

Выборочный квантиль 0,95 находится в точке 728,70 мс, за которой лежит только 500 (5%) наблюденных времен из 10 000, при максимальном значении в 8888,74 мс.

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ (V)

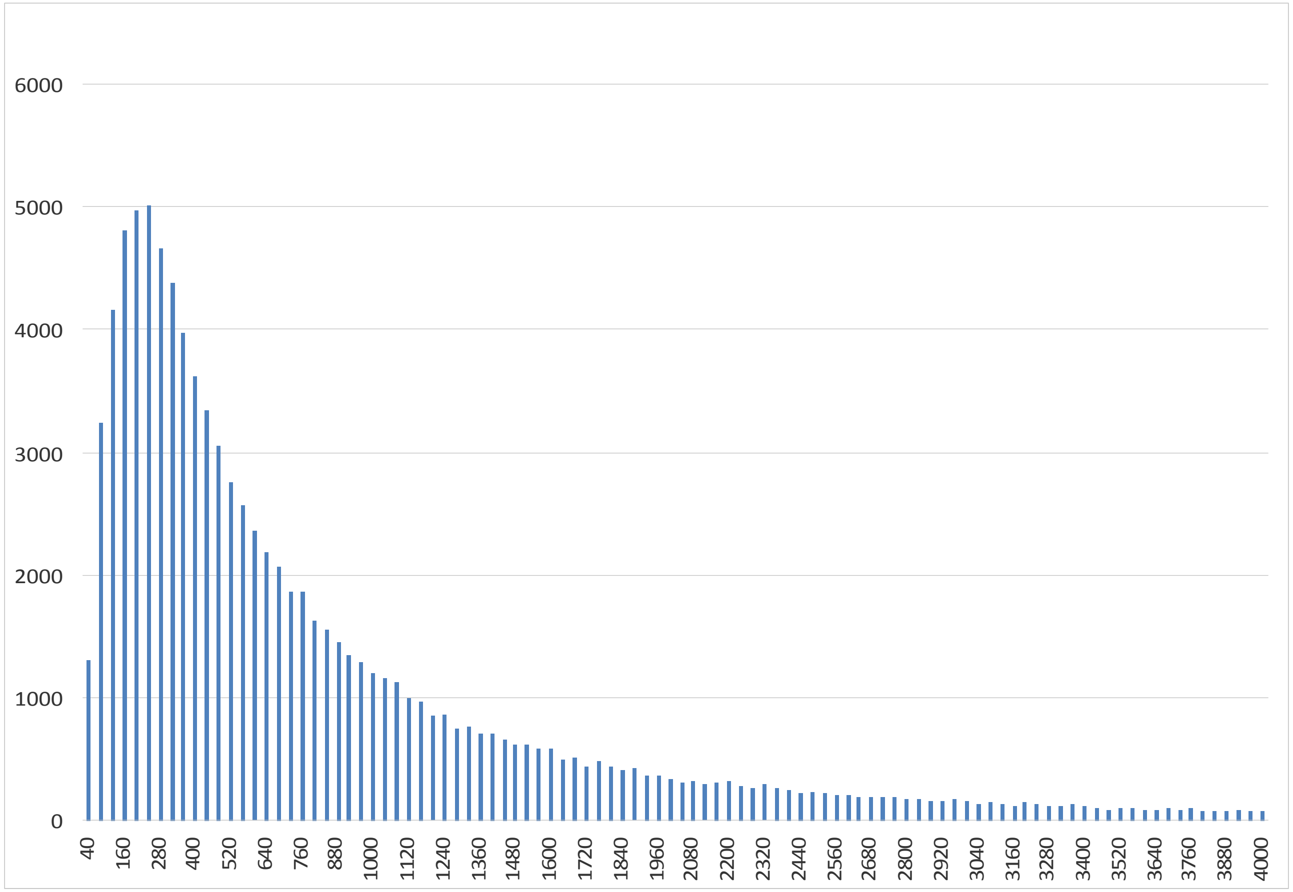


Рис. 3а. Частоты времен выполнения для задачи с n=45

(первые 100 полусегментов с шагом 40 мс. по результатам 100 000 экспериментов).

Минимальное время — 1,922 мс., максимальное — 126481,00 мс.

ПОСТАНОВКА ЗАДАЧИ

Пусть:

A — матрица индивидуальной задачи коммивояжера (несимметричная постановка),

$C_X(A)$ — сложность индивидуальной задачи — число порожденных вершин поискового дерева решений в классическом алгоритме (Knuth, D. E. 1975. Estimating the efficiency of backtracking programs. Mathematics of Computing, v. 29, pp. 121–136).

Постановка задачи:

I. Найти:

— полиномиально вычисляемую характеристику A — $\lambda(A)$,

— и функцию $C(\cdot)$:

$$C(\lambda(A)) \in [C_X(A) - \varepsilon, C_X(A) + \varepsilon], \quad \varepsilon << C_X(A).$$

II. Идентифицировать функцию плотности, аппроксимирующую распределение частот наблюдаемых времен, и определить зависимость ее параметров от длины входа в целях вероятностного прогнозирования.

ИДЕИ ПО ПОСТАНОВКЕ I

1. Разработка на основе применения метода классов входов к задаче коммивояжера обобщенных представлений индивидуальных задач — введение такого обобщающего описания A^* (в виде целочисленной матрицы стоимостей), которое представляет класс $\widetilde{A}$ индивидуальных задач с близкой сложностью

$$C_X\left(A \in \widetilde{A}\right) \in \left[C_X\left(A^*\right) - \varepsilon,\, C_X\left(A^*\right) + \varepsilon\right], \quad \varepsilon << C_X\left(A^*\right)$$

2. Попытка методами корреляционного анализа определить характеристику $\lambda(A)$, путем вычисления для матриц индивидуальных задач и их обобщенных представлений:

a) стандартных характеристик:

— определитель; энтропия; спектр; сингулярные числа и т.д.

b) дополнительных и специально разработанных характеристик для выявления матриц, порождающих большие по объему поисковые деревья решений.

— рассматриваются возможности применения аппаратов двумерного Фурье- и вейвлет- анализа, модулярные преобразования и т.д.

ИДЕИ ПО ПОСТАНОВКЕ II

Идентификации функций плотности распределения вероятностей, аппроксимирующей экспериментальные данные — времена выполнения и мощности порожденных индивидуальными задачами деревьев решений.

Дополнительный интерес представляет выявление зависимости параметров таких распределений от длины входа, что позволит осуществлять вероятностный прогноз на большие размерности.

Основная проблема: *«угадать» семейство распределений*!

Будет использован подход, развиваемый Г.Н. Жуковой, базирующийся на первоначальной идентификации вида распределения на основе коэффициентов асимметрии и эксцесса известных распределений и точечных оценок этих коэффициентов по данным выборки (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.).

$$\gamma_1 = \frac{\mu_3(X)}{(DX)^{3/2}} = \frac{E(X-EX)^3}{\left(E(X-EX)^2\right)^{3/2}}, \qquad \gamma_2 = \frac{\mu_4(X)}{(DX)^2} = \frac{E(X-EX)^4}{\left(E(X-EX)^2\right)^2}.$$

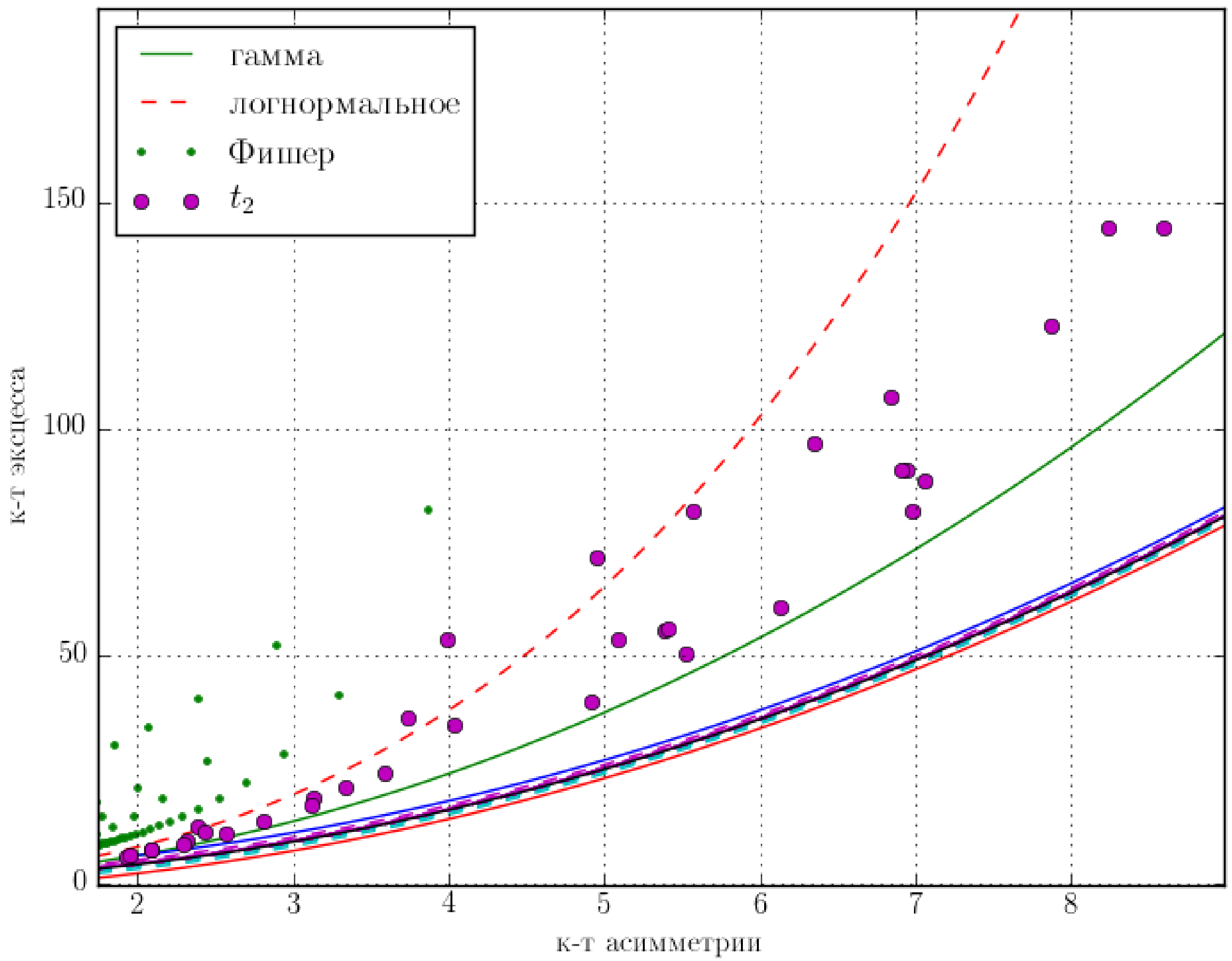


Рис.4. Коэффициенты асимметрии и эксцесса для некоторых распределений и экспериментальных данных (t).

Библиографический список

Прогнозирование для индивидуальных задач коммивояжера:

1. **Cornuéjols, G.; Karamanov, M.; and Li, Y.** 2006. Early estimates of the size of branch-and-bound trees. INFORMS Journal on Computing. 2006, v. 18, No. 1, pp. 86—96.

2. **Lobjois, L., M. Lemaitre.** Branch-and-bound algorithm selection by performance prediction. American Association for Artificial Intelligence, Menlo Park, CA. 1998.

Публикации авторов:

3. **Жукова Г.Н.** Идентификация распределения по коэффициентам асимметрии и эксцесса // Автоматизация и современные технологии, 2015, (статья принята к публикации в №10, 2015 г.).

4. **Ульянов М.В., Фомичев М.И.** Ресурсные характеристики способов организации дерева решений в методе ветвей и границ для задачи коммивояжера // Бизнес – информатика, 2015, (статья принята к публикации в №4 журнала в 2015 г.)

БЛАГОДАРЮ ЗА ВНИМАНИЕ!