

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**МАТЕРИАЛЫ**  
**IV Международной молодежной**  
**научной конференции**  
**«МАТЕМАТИЧЕСКОЕ**  
**И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ**  
**ИНФОРМАЦИОННЫХ,**  
**ТЕХНИЧЕСКИХ**  
**И ЭКОНОМИЧЕСКИХ СИСТЕМ»**

**Томск, 20–21 мая 2016 г.**

*Под общей редакцией*  
*кандидата технических наук И.С. Шмырина*

Томск  
Издательский Дом Томского государственного университета  
2016

ББК 22.17–22.19  
УДК 519.2, 519.7, 519.8  
Т78

ЧЛЕНЫ КОЛЛЕГИИ, РУКОВОДИТЕЛИ НАУЧНЫХ РЕДАКЦИЙ  
ПО НАПРАВЛЕНИЯМ:

д-р техн. наук, проф. **А.А. Глазунов** – научная редакция «Механика, математика»; д-р техн. наук, проф. **Э.Р. Шрагер** – научная редакция «Механика, математика»; д-р техн. наук, проф. **А.М. Горцев** – научная редакция «Информатика и кибернетика»; д-р техн. наук, проф. **С.П. Сущенко** – научная редакция «Информатика и кибернетика»; д-р физ.-мат. наук, проф. **В.Г. Багров** – научная редакция «Физика»; д-р физ.-мат. наук, проф. **А.И. Потекаев** – научная редакция «Физика»; д-р биол. наук, проф. **С.П. Кулижский** – научная редакция «Биология»; д-р геол.-минер. наук, проф. **В.П. Парначев** – научная редакция «Науки о Земле, химия»; канд. хим. наук, доц. **Ю.Г. Слизов** – научная редакция «Науки о Земле, химия»; д-р филол. наук, проф. **Т.А. Демешкина** – научная редакция «История, филология»; д-р ист. наук, проф. **В.П. Зиновьев** – научная редакция «История, филология»; д-р экон. наук, проф. **В.И. Канов** – научная редакция «Юридические и экономические науки»; д-р юрид. наук, проф. **В.А. Уткин** – научная редакция «Юридические и экономические науки»; д-р ист. наук, проф. **Э.И. Черняк** – научная редакция «Философия, социология, психология, педагогика, искусствознание»; д-р психол. наук, проф. **Э.В. Галажинский** – научная редакция «Философия, социология, психология, педагогика, искусствознание»

НАУЧНАЯ РЕДАКЦИЯ ТОМА:

д-р техн. наук, проф. **А.М. Горцев**, д-р физ.-мат. наук, доц. **Ю.Г. Дмитриев**, д-р физ.-мат. наук, проф. **В.В. Конев**, д-р техн. наук, проф. **А.Ю. Матросова**, д-р техн. наук, проф. **А.А. Назаров**, д-р техн. наук, проф. **К.И. Лившиц**, канд. техн. наук **И.С. Шмырин**.

**Т78** Труды Томского государственного университета. – Т. 299. Серия физико-математическая: Математическое и программное обеспечение информационных, технических и экономических систем : материалы IV Международной молодежной научной конференции. Томск, 20–21 мая 2016 г. / под общ. ред. И.С. Шмырина. – Томск : Издательский Дом Томского государственного университета, 2016. – 144 с.

ISBN 978-5-94621-570-1

Сборник содержит материалы IV Всероссийской молодежной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем», проводившейся 20–21 мая 2016 г. на базе факультета прикладной математики и кибернетики Томского государственного университета. Материалы сгруппированы в соответствии с работавшими на конференции секциями.

Для научных работников, преподавателей, аспирантов, магистрантов и студентов.

УДК 539.3.004  
ББК 22,25.22.251.22.62

# СЕКЦИЯ I. МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН И КОМПЬЮТЕРНЫХ СЕТЕЙ

## АНАЛИЗ И ВЕРИФИКАЦИЯ ИЕРАРХИЧЕСКИХ ВРЕМЕННЫХ ТИПИЗИРОВАННЫХ СЕТЕЙ ПЕТРИ

**Г.А. Белошапко, В.А. Непомнящий**

*Институт систем информатики им. А.П. Ершова СО РАН*  
ordcoder@gmail.com

### Введение

Раскрашенные сети Петри (РСП) широко используются в качестве моделей распределенных систем. Для них разработана известная система анализа CPN Tools [1]. В ИСИ СО РАН разработана система верификации раскрашенных сетей Петри [6], использующая известную систему SPIN [2]. Также разработана модификация раскрашенных сетей Петри, названная иерархическими временными типизированными сетями Петри (ИВТ-сетями) [3,4]. ИВТ-сети использовались с целью анализа и верификации распределенных систем, представленных на языке SDL [3,4]. Естественно, возникает задача применения к ИВТ-сетям мощных средств верификации и анализа, разработанных для РСП. Для этого наряду с системой анализа ИВТ-сетей можно использовать транслятор из ИВТ-сетей в РСП.

Работа частично поддержана грантом РФФИ № 14-07-00401.

### 1. Сети Петри высокого уровня

Раскрашенные сети Петри являются расширением ординарных сетей Петри. В отличие от ординарных сетей Петри, где все фишки считаются неотличимыми друг от друга, каждая фишка в раскрашенной сети обладает индивидуальностью — значением некоторого типа, которое называется цветом. Неиерархическая раскрашенная сеть Петри состоит из трех частей: структуры сети, деклараций и пометок сети. Иерархическая сеть представляет собой множество неиерархических сетей, функционирующих как единое целое.

Декларации состоят из описания множеств цветов (типов) и объявлений переменных, каждая из которых принимает значения из некоторого множества цветов. Декларации также могут содержать определение операций и функций на языке ML.

Пометка сети приписывается месту, переходу либо дуге и определяет правила распределения и перемещения фишек по сети. Каждое место имеет три разных типа пометок: имя места, множество цветов и инициализирующее выражение.

При функционировании раскрашенной сети возможность срабатывания перехода зависит не только от наличия фишек во входных местах перехода, но и от их значений. Чтобы определить срабатывание перехода, необходимо определить значения переменных перехода. При этом все вхождения одной и той же переменной замещаются одним и тем же значением. Набор значений переменных, при которых выполнена спусковая функция перехода, называется связыванием. Значение выражения на входной дуге определяет, сколько и каких фишек должно содержаться в соответствующем входном месте перехода, чтобы переход мог сработать. Выражения на выходных дугах определяют, сколько и каких фишек будет помещено в выходные места перехода, когда он работает.

Для представления времени вводится понятие глобальных часов. Значение часов представляет текущее время в модели. Начальное значение часов задается при описании сети. Множества цветов (все или часть) получают признак *timed*.

Опишем ИВТ-сети Петри. Дуги в ИВТ-сетях не имеют пометок, а вместо них переходы помечаются выражениями на языке Python, которые определяют результирующую разметку при каждом срабатывании переходов. В ИВТ-сетях присутствуют места-очереди, способные хранить неограниченное число фишек. Новая фишка в таком месте помещается в конец очереди и остается недоступной до тех пор, пока из места не будут извлечены все фишки, поступившие до нее.

В ИВТ-сетях различают два вида мест, обычные и многослойные. Многослойное место помечено специальным типом *layer*, представляющим пару из номера слоя (целого числа) и значения любого допустимого типа. Многослойное место, которое не является очередью, всегда содержит не более одной фишки, принадлежащей одному слою. Многослойное место-очередь содержит не более одной очереди в каждом слое.

ИВТ-сети являются квазибезопасными в том смысле, что каждое место может иметь не более одной фишки, принадлежащей определенному слою.

В ИВТ-сетях временной механизм связан с переходами: каждому переходу сопоставлена пара неотрицательных чисел  $d_{\min}$  и  $d_{\max}$ . Если  $t$  — момент времени, в который переход стал возможен, то его срабатывание произойдет в некоторый момент времени из интервала  $[t + d_{\min}, t + d_{\max}]$ , если условие срабатывания не будет нарушено.

## 2. Свойства сетей ИВТ-сетей Петри

Перечислим некоторые базовые свойства ИВТ-сетей. Для них рассматриваются только разметки, достижимые из начальной разметки.

Достижимость (*reachability*). Разметка  $Q$  называется достижимой из разметки  $P$ , если существует допустимая последовательность срабатываний переходов, переводящая сеть из разметки  $P$  в разметку  $Q$ .

Ограниченность (*boundedness*). Для конкретного места сети определяется нижняя  $L$  и верхняя  $U$  границы количества фишек в нем. Это означает, что во время функционирования сети данное место всегда содержит не менее  $L$  и не более  $U$  фишек. Аналогично определяется нижняя  $Lm$  и верхняя  $Um$  границы места в виде мультимножеств фишек, содержащихся в нем. То есть во время функционирования сети множество фишек в данном месте является надмножеством  $Lm$  и подмножеством  $Um$ .

Домашняя разметка (*home marking*). Разметка сети называется домашней, если она достижима из любой разметки. Аналогично, множество разметок называется домашним, если из любой разметки сети достижима хотя бы одна разметка из множества.

Живость (*liveness*). Разметка называется тупиком, если в ней нет возможных переходов. Переход называется мертвым, если он никогда не сможет сработать. Переход называется живым, если он никогда не может стать мертвым.

Справедливость (*fairness*). Свойство справедливости определено, если у сети есть последовательность бесконечных срабатываний (ПБС). Переход называется частичным, если он встречается бесконечно часто в любой ПБС. Переход называется честным, если он встречается бесконечно часто в любой ПБС, в которой он становится допустимым бесконечно часто. Переход называется справедливым, если он встречается бесконечно часто в любой ПБС, в которой он остается допустимым с некоторого момента.

## 3. Метод трансляции ИВТ-сетей в РСП

Рассмотрим разработанный метод трансляции ИВТ-сетей в РСП. Транслятор строит результирующую сеть в несколько этапов. На каждом этапе набор конструкций, специфичных для ИВТ-сетей, преобразуется в конструкции раскрашенных сетей Петри,

симулирующих их работу. Общую схему алгоритма трансляции можно представить в следующем виде:

- по файлу с ИВТ-сетью в формате PNML построить внутреннее представление сети;
- преобразовать места-очереди в места типа список;
- преобразовать многослойные места в места с типом кортеж из двух элементов;
- транслировать временные конструкции;
- сгенерировать ML-код и декларации, соответствующие выражениям на языке Python;
- перевести полученную сеть в формат, совместимый с CPN Tools.

Опишем подробнее реализацию транслятора и форматы данных. Транслятор представляет собой исполняемую программу на языке Python 2. Транслятор состоит из нескольких модулей, отвечающих за различные функции, такие как связь мест и мест-ссылок на подстраницах, генерация деклараций, поиск состояний ИВТ-сети. Одной из причин выбора Python как языка программирования является наличие встроенной библиотеки для построения синтаксического дерева Python кода, которая используется для разбора деклараций в ИВТ-сетях.

Входным файлом является файл с ИВТ-сетью в формате PNML (основанном на XML). Выходным файлом является файл с РСП в формате системы CPN Tools. Этот формат тоже основан на XML. Транслятор работает с сетью во внутреннем представлении, которое является графом с вершинами типов страница, место, переход и дуга. Ребра соединяют страницы со всеми находящимися на них местами, переходами и дугами, а также дуги с инцидентными местами и переходами.

Места-очереди транслируются в места типа список. Первый элемент списка соответствует началу очереди, последний — концу. Заметим, что место-очередь преобразуется в одно место.

Транслятор преобразует многослойное место типа  $X$  в место типа  $(\text{product int} * X)$ . Фишке в слое  $n$  со значением  $x$  многослойного места будет соответствовать фишка со значением  $(n, x)$  в РСП. Заметим, что многослойное место преобразуется в одно место, а количество дуг не увеличивается.

В ИВТ-сетях используется временная модель Мерлина, в которой время определяется локально для каждого перехода, то есть готовый переход может сработать в течении определенного интервала времени. В РСП используется глобальное время, а фишки несут временной штамп, который показывает по истечению какого момента времени они готовы к использованию. Известно, что данные модели эквивалентны.

При трансляции перехода с точечными временным интервалом вводится вспомогательное место  $T_{\text{delay}}$ , отсчитывающее задержку между готовностью перехода и возможностью его срабатывания.

Для моделирования перехода с временным интервалом  $[a, b]$ , показанном на рис. 1, вводится вспомогательное место  $T_{\text{delay}}$  и переход  $T_{\text{wait}}$ .  $T_{\text{wait}}$  служит как альтернатива переходу  $T$ . Каждый раз, когда может сработать переход  $T$ , вместо него может сработать переход  $T_{\text{wait}}$ .

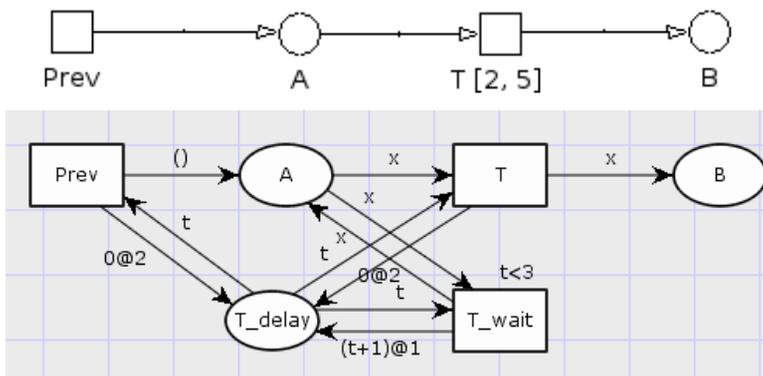


Рис. 1. Пример трансляции временной конструкции

При трансляции переходов с временной задержкой появляется не более одного дополнительного места и перехода. Таким образом суммарное количество мест и переходов увеличивается не более, чем в 3 раза.

#### 4. Анализ ИВТ-сетей Петри

Для анализа свойств ИВТ-сетей можно использовать трансляцию их в раскрашенные сети Петри. Основные сложности, возникающие при таком подходе состоят в том, что результаты анализа получаются для раскрашенной сети Петри, и необходим «обратный ход», чтобы перенести эти результаты на исходную сеть.

Поэтому возникает задача проверки набора простых свойств ИВТ-сетей без трансляции в РСП. Для этого разработаны алгоритмы проверки свойств ИВТ-сетей Петри. Набор свойств был выбран аналогично тому набору свойств, которые можно проверить с помощью системы CPN Tools у раскрашенных сетей Петри, а именно свойства достижимости, ограниченности, живости, а также свойство, является ли разметка домашней.

Реализована программа на языке Python, осуществляющая проверку данных свойств. Общая схема работы алгоритма анализа:

- по файлу с ИВТ-сетью в формате PNML построить внутреннее представление сети;
- построить граф достижимости разметок поиском в глубину из начальной разметки;
- выделить компоненты сильной связности графа достижимости разметок;
- проверить достижимость разметок;
- проверить ограниченность мест;
- проверить живость разметок и переходов;
- определить множество домашних разметок.

#### 5. Пример

С помощью транслятора была успешно преобразована ИВТ-сеть, моделирующая протокол PAR [5] в РСП.

ИВТ-сеть, моделирующая протокол PAR, содержит 78 мест и 75 переходов на 9 страницах. Сеть содержит 3 фишки в месте User\_Sender. Эти фишки моделируют сообщения, передаваемые протоколом. Полученные сообщения моделируются фишками в месте Receiver\_User.

С помощью разработанной системы трансляции были успешно верифицированы несколько свойств протокола:

1. Безопасность. В любом достижимом состоянии в каждом месте одновременно находится не более одной фишки.

2. Невозможность потери сообщений. В любом достижимом состоянии последовательность полученных сообщений является префиксом последовательности отправленных сообщений.
3. Невозможность вставки сообщений. Невозможно состояние, в котором получено больше сообщений, чем отправлено.

### Заключение

В данной работе описан метод трансляции ИВТ-сетей в РСР, на базе которого реализована система анализа и верификации ИВТ-сетей, использующая средства анализа и верификации РСР CPN Tools [1] и CPN Ver [6]. Применение этого метода проиллюстрировано примером верификации протокола PAR. В дальнейшем предполагается применение разработанных средств для анализа и верификации различных коммуникационных протоколов.

### Литература

1. *Jensen K., Kristensen L. M.* Coloured Petri Nets. Modelling and Validation of Concurrent Systems. Springer-Verlag, 2009.
2. *Holzmann G.J.* The SPIN model checker. Primer and Reference Manual. Addison-Wesley, 2004.
3. *Непомнящий В.А., Аргиров В.С., Белоглазов Д.М., Быстров А.В., Четвертаков Е.А., Чурина Т.Г.* Моделирование и верификация коммуникационных протоколов, представленных на языке SDL, с помощью сетей Петри высокого уровня // Программирование. 2008, № 6, с. 35–49.
4. *Чурина Т.Г., Аргиров В.С.* Моделирование спецификаций языка SDL с помощью модифицированных ИВТ-сетей. Препринт ИСИ СО РАН. 2005. № 124. 62 с.
5. *Таненбаум Э.* Компьютерные сети. СПб.: Питер, 2003.
6. *Стененко А.А., Непомнящий В.А.* Верификация раскрашенных сетей Петри методом проверки моделей. Препринт ИСИ СО РАН. 2015. № 178. 29 с.

## ИССЛЕДОВАНИЕ ГИБРИДНЫХ ЭВРИСТИЧЕСКИХ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ КОММИВОЯЖЕРА

**Е.А. Боронихина, В.А. Сибирякова**

*Томский государственный университет  
lin\_san@mail.ru, val349@mail.ru*

Задача коммивояжера — важная задача транспортной логистики, отрасли, занимающейся планированием транспортных перевозок. Задача состоит в определении кратчайшего гамильтонова цикла в графе. Существует несколько частных случаев задачи коммивояжера: геометрическая (планарная или евклидова), треугольная, симметричная и асимметричная задачи, а также задача с несколькими коммивояжерами — определение нескольких циклов в графе, суммарная стоимость которых будет минимальной.

Задача относится к классу NP-полных и является трансвычислительной. Чтобы гарантировать существование решения, наложим на задачу следующее ограничение: будем считать входной граф полностью связным. В случае если между некоторыми городами не существует сообщения, введем дополнительные ребра с большей длиной. Если оптимальный маршрут для данного графа существует, введенные ребра никогда не попадут в решение.

Выделяют два типа решения задачи коммивояжера: точные и эвристические [1]. К точным методам относятся алгоритм полного перебора (АПП) и метод ветвей и границ (МВГ). На поиск решения требуются большие временные затраты, но результаты, полученные данными методами, имеют 100% точность.

Эвристические методы будем разбивать на подгруппы: жадные методы — алгоритм ближайшего соседа (БС) и рекурсивного перебора (АРП), эволюционные — генетические (ГА) и метод имитации отжига (МИО), а так же поведенческие (роевые) алгоритмы — муравьиный (АМК), имитации иммунной системы (ИИС) и интеллектуальных

капель (АИК). Роевые алгоритмы, в свою очередь, можно разделить на методы, инспирированные живой и неживой природой.

Данная классификация позволяет:

- оценить риски получения неудовлетворительного результата у каждой группы алгоритмов;
- помогает выработать требования для гибридных методов;
- скомбинировать методы в оптимальную по времени и/или производительности систему.

В рамках исследования был проведена оценка временной сложности алгоритмов (табл. 1) и проведен анализ методов по различным признакам (табл. 2).

Таблица 1

Временная сложность алгоритмов

Алгоритм	Временная сложность
Полный перебор	$O(n!)$
Метод ветвей и границ	Луч. $O(n^3 \cdot \log_2 n)$
	Худ. $O(n^3 \cdot \log_2 n)$
Ближайшего соседа	$O(n)$
Рекурсивный перебор	$2O(n)$
Муравьиной колонии (МК)	$O(t \cdot m \cdot n^2)$
Искусственная иммунная система	$O(t \cdot m \cdot n_2)$
Алгоритм интеллектуальных капель	$(m \cdot n^2)$
Метод имитации отжига	$O(n^2 \cdot \log_2 n)$
Генетические алгоритмы	$O(t \cdot (m \cdot n)^2)$

Таблица 2

Сравнение алгоритмов решения задачи коммивояжера

	АПП	МВиГ	АБС	АРП	ГА	МИО	ИИС	АМК	АИК
Точный	v	v							
Эвристический			v	v	v	v	v	v	v
Предыдущий опыт	v	v			v		v	v	
Матрица стоимостей	v	v	v	v	v	\	\	v	\
Дополнительные коэффициенты		v						v	v
Случайный выбор		v	v	v	v	v		v	v
Включает плохое решение	v	v			v		v	v	v
Требует начальное решение					v	v			
Выход из локальных экстремумов	v	v			v	v			
Матрица стоимости используется только для оценки решения	v	v	\	\	v	v	v	v	v
Возможность распараллеливания					\	v	v	v	v
Количество параметров	0	0	0	1	1	4	3	4	8

где v — алгоритм полностью соответствует данному признаку, \ — частичное соответствие.

На основании этих таблиц можно сделать различные выводы. Например, алгоритмы, использующие матрицу дополнительных коэффициентов (алгоритм интеллектуальных капель или муравьиной колонии), при худшем варианте развития событий не могут выйти из локальных экстремумов.

Создание гибридной (интегрированной) системы позволит компенсировать недостатки одного алгоритма преимуществами другого.

В рамках данной работы будем опираться на одноуровневую классификацию Ванга, которая включает три разновидности гибридных алгоритмов:

1. Вложенные алгоритмы — внедрение методов друг в друга, выделяют вложение на высоком и низком уровне.
2. Препроцессор/постпроцессор — последовательная комбинация алгоритмов
3. Коалгоритмы — это параллельное выполнение методов с постоянным обменом информацией.

С использованием данной классификации, была сделана попытка создания интегрированного подхода (рис. 1) для решения задачи коммивояжера.



Рис. 1. Блок-схема интегрированного метода

Данный подход включает в себя 3 уровня и является гибридом типа препроцессор/постпроцессор с использованием высокоуровневой вложения. Первый этап генерирует начальные решения, второй обеспечивает широкий обзор решений, третий сужает область поиска.

Были предложены различные комбинации алгоритмов, реализующие данный интегрированный подход:

- Алгоритм ближайшего соседа + Метод имитации отжига + Оператор скрещивания из генетического алгоритма.
- Вырожденный алгоритм интеллектуальных капель + Кроссинговер генетического алгоритма + Алгоритм интеллектуальных капель.
- Алгоритм муравьиной колонии + Мутация генетического алгоритма + Элитная муравьиная колония.

На рис. 2 представлены результаты работы вырожденного муравьиного алгоритма и 10 итераций интегрированного подхода (последняя комбинация методов). Маршрут для 100 городов был найден с помощью 52 агентов и оператора мутации.

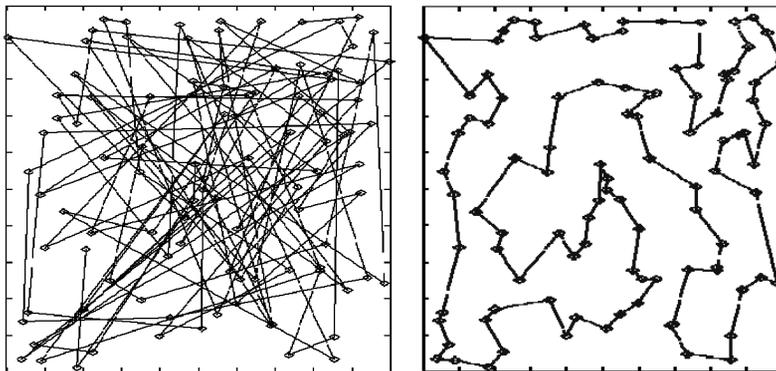


Рис. 2. 10 итераций вырожденного АМК и интегрированного подхода

Слабые места алгоритмов были компенсированы сильными сторонами других, что привело к повышению качества полученного решения.

Вычислительная сложность последней комбинации методов аналогична сложности метода муравьиной колонии.

### Заключение

Предложенный интегрированный подход имеет значительное преимущество относительно стандартных методов решения, так как способен адаптироваться к изменениям, варьировать точность в зависимости от требований и выдавать решение, приближенное к точному, за минимальное время. Это позволит создавать системы, использующие данный метод, и не требующие больших трудозатрат.

### ЛИТЕРАТУРА

1. Беспалько В.П. Образование и обучение с помощью компьютеров / В. П. Беспалько. – М.: Высш. шк., 1998. – 135 с.
2. Борознов В.О. Дополнение метода ветвей и границ для решения задачи коммивояжера // Вестн. Ростов. гос. ун-та путей сообщения. – 2007. – № 1. – С. 160–163.
3. Борознов В.О. Исследования генетических методов решения задачи коммивояжера / В.О. Борознов, О.Г. Ведерникова // Вестн. Ростов. гос. ун-та путей сообщения. – 2004. – № 1. – С. 42–45.
4. Гладков Л.А., Курейчик В.М., Курейчик В.В. Генетические алгоритмы / Л.А. Гладков. – Ростов-на-Дону: ООО «Ростиздат», 2004 г.
5. Курейчик В.В. Эволюционные методы решения оптимизационных задач. Таганрог, 1999, ТРТУ.
6. МакКоннелл Дж. Основы современных алгоритмов // Дж. МакКоннелл. – М.: Техносфера, 2004. – 368 с.
7. Романовский И.В. Алгоритмы решения экстремальных задач / И.В. Романовский. – М.: Наука, 1977. – 352 с.
8. Dorigo M., Maniezzo V., Colomi A. The Ant System: Optimization by a colony of cooperating objects // IEEE Trans. on Systems, Man, and Cybernetics. – 1996. – Part B.

## КОНФИГУРИРОВАНИЕ НЕЙРОННЫХ СЕТЕЙ КОХОНЕНА

**В.С. Веденеев, И.В. Бычков**

Челябинский государственный университет, ОАО «Челябинский цинковый завод»  
ingafen@gmail.com, bychkov@csu

### Введение

Нейронные сети Кохонена представляют собой эффективный математический аппарат, который позволяет выполнять кластеризацию и связанные с ней задачи: поиск похожего, поиск аномалий, прогнозирование. Долгое время применение сетей Кохонена ограничивалось вычислительной сложностью обучения сети. Эта проблема решена благодаря использованию многопроцессорных систем, например, GPU [1]. Это открывает широкие возможности для использования сетей Кохонена и самоорганизующихся карт Кохонена в прикладных задачах. В настоящей работе дается описание основных

параметров нейронных сетей Кохонена, самоорганизующихся карт Кохонена, описывается влияние каждого из параметров на результаты кластеризации, даются рекомендации по использованию карт и сетей Кохонена.

## 1. Постановка задачи

Нейронная сеть Кохонена — это нейронная сеть, осуществляющая обучение «без учителя» и использующая принцип «победитель получает все».

Сети Кохонена применяются в задачах кластеризации — разбиении некоторого множества данных на небольшие однородные группы (кластеры). При этом расстояние между объектами внутри одного кластера не превышает некоторую заданную величину.

Следует отметить несколько слов о том, как сеть Кохонена формирует кластер. После обучения сети, каждому из объектов заданного множества будет соответствовать один из нейронов сети, который будет являться наиболее схожим с этим объектом. Такой нейрон называют «нейроном-победителем». Множество объектов, для которых один и тот же нейрон является нейроном-победителем и евклидово расстояние до нейрона-победителя не превышает заданную величину, образуют кластер.

Таким образом, влияя на параметры обучения сети, мы можем получать несколько отличные друг от друга варианты разбиения исходного множества на кластеры. Далее мы рассмотрим эти параметры, а также их влияние на результат кластеризации.

## 2. Нейронная сеть Кохонена

Для описания параметров сети Кохонена, рассмотрим кратко алгоритм обучения.

Обозначим  $R = \{\bar{r}_i = \langle r_1, \dots, r_M \rangle, i = \overline{1, N}\}$  — множество входных векторов,  $W = \{\bar{w}_i = \langle w_1, \dots, w_M \rangle, i = \overline{1, Q}\}$  — веса нейронов.

Первоначальным этапом обучения сети является инициализация весов нейронов случайными значениями, каждое из которых принадлежит отрезку  $[-1; 1]$ . В некоторых случаях значения ограничиваются отрезком  $[0; 1]$ .

Далее рекомендуется провести нормализацию весов нейронов [2]:

$$\bar{w}_i := \bar{w}_i \frac{1}{\|\bar{w}_i\|}.$$

Обучение производится итеративно. В течение каждой итерации происходит выбор нейрона-победителя: нейрона, наиболее схожего с вектором  $\bar{r}_i$ . В качестве меры схожести может использоваться скалярное произведение, евклидово расстояние и др. [3] Затем происходит изменение весов нейрона-победителя [4]

$$\bar{w}_k(t+1) = \bar{w}_k(t) + \alpha(t)[\bar{r}_i - \bar{w}_k(t)], \quad (1)$$

где  $\alpha(t)$  — функция скорости обучения,  $t$  — дискретное время.

После изменения весов нейрона-победителя рекомендуется повторно провести нормализацию весов.

Такая итерация производится для каждого элемента  $\bar{r}_i$  из входной выборки. Когда все элементы выборки перебраны, то говорят, что прошла одна эпоха.

Для исключения «мертвых» нейронов, то есть нейронов, не участвующих в обучении, используется следующий алгоритм оптимизации: в начале обучения всем нейронам присваивается одинаковый потенциал равный  $1/Q$ , где  $Q$  — число нейронов. После прохождения этапа обучения происходит корректировка потенциалов по следующему признаку:

$$p_{i+1} = \begin{cases} p_i + \frac{1}{Q}, & i \neq j, \\ p_i - p_{\min}, & i = j, \end{cases}$$

где  $i$  — номер текущего нейрона,  $j$  — номер нейрона-победителя. При  $p_i < p_{\min}$  нейрон исключается из обучения.

### 3. Самоорганизующиеся карты Кохонена

В варианте самоорганизующихся карт Кохонена (англ. Self-Organizing Map — SOM) происходит корректировка весов не только нейрона-победителя, но и его соседей. Корректировка весов осуществляется по формуле, схожей с (1):

$$\bar{w}_k(t+1) = \bar{w}_k(t) + \alpha(t)\eta_j(d,t)[\bar{r}_i - \bar{w}_k(t)], \quad (2)$$

где  $\eta_j(d,t)$  — функция соседства,  $d$  — евклидово расстояние между нейроном победителем под номером  $j$  и вектором  $\bar{r}_i$ .

Итак, из сказанного выше следует, что пользователь нейронной сети Кохонена в своем распоряжении имеет следующие параметры:

1. Критерий, по которому заканчивается обучение сети.
2. Количество нейронов в сети.
3. Минимальное значение потенциала  $p_{\min}$ .
4. Функция скорости обучения.
5. Архитектура сети (вид соседства нейронов).
6. Функция соседства.

Обсудим далее каждый из параметров.

### 4. Влияние параметров сети на результаты кластеризации

Для оценки влияния тех или иных параметров мы будем использовать сеть Кохонена со следующими изначальными параметрами:

- фиксированное число эпох при обучении сети  $S = 100$ ;
- 48 нейронов в сети (сетка 8x6);
- функция скорости обучения  $\alpha(t) := 0.5e^{-\left(\frac{t}{SN}\right)^2}$ ; при этом время  $t$  увеличивается на единицу на каждом шаге обучения.

Меняя поочередно один из параметров, мы сможем подобрать оптимальную конфигурацию для решения поставленной пользователем сети задачи.

Для оценки результатов кластеризации в работе используется обучающая выборка, которая является приложением к конкурсу kddcup'99 (<http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>). Выборка представляет собой информацию о сетевых соединениях. Каждое соединение характеризуется фиксированным набором параметров. Обучающая выборка содержит также дополнительную информацию о том, является ли это соединение легитимным (normal) или соединение соответствует сетевой атаке (smurf, Neptune и т.д.).

Изучение полученных кластеров даст представление о воздействии отдельных параметров сети на общий результат.

#### 4.1. Критерий завершения обучения сети

Как уже отмечено выше, обучение сети происходит итеративно, эпохами. Существует 2 основных способа ограничения числа итераций при обучении. Первый состоит в простом ограничении количества эпох, т.е., по завершении  $N$ -ого количества эпох сеть считается обученной. При этом число  $N$  подбирает пользователь.

Несмотря на очевидность того, что при большом числе эпох сеть будет наиболее адаптированной, существует вероятность того, что сеть будет «переобучена» и ошибки в распознавании с некоторого момента будут только возрастать.

Второй способ состоит в том, чтобы отслеживать количество изменений в сети при каждой итерации. Если модуль максимального из изменений перестанет превышать заданную величину, то обучение будет считаться завершенным. Однако в таком случае, при неудачном подборе параметров веса нейронов могут колебаться около стационарных значений и обучение уйдет в бесконечный цикл.

Наиболее предпочтительным все же является первый способ, т.к. он требует меньше вычислительных ресурсов. Также при таком способе проще выбирать функцию соедства и функцию скорости обучения.

#### 4.2. Количество нейронов в сети

Количество нейронов в сети Кохонена соответствует числу кластеров, на которое разбивается исходное множество. При выборе числа нейронов следует учитывать, что не все из них будут образовывать кластеры.

Рассмотрим результаты обучения сети выборкой из 5000 элементов. Выборка содержит 6 типов соединений, состав выборки следующий:

Тип соединения	Количество
Normal	4756
loadmodule	1
portsweep	91
Ipsweep	51
Phf	1
Teardrop	100

Базовые параметры сети представлены выше. В таблице ниже представлено отношение числа нейронов к числу построенных кластеров.

Количество нейронов	8	9	10	12	16	20	24	28	30	32
Количество кластеров	7	9	9	11	14	17	14	21	22	27
Количество нейронов, не участвовавших в обучении	–	–	–	–	–	–	4	1	1	–

При этом стоит отметить, что рост количества кластеров происходит за счет появления небольших кластеров, соответствующих редко встречающимся элементам, которые могут являться аномалиями.

#### 4.3. Минимальное значение потенциала

Алгоритм использования потенциалов нейронов описан выше. Применение потенциалов нейронов дает уменьшение числа нейронов, не участвующих в обучении. Следствием этого является увеличение числа кластеров, т.к. в таком случае отдельные нейроны начинают участвовать в обучении и становятся нейронами-победителями для объектов из выборки.

В таблице ниже представлены результаты кластеризации выборки из 5000 элементов.

$p_{\min}$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Количество кластеров	35	34	39	37	39	38	37	37	39	43	43

#### 4.4. Функция скорости обучения

Функция обучения — монотонно убывающая функция времени [4]. В различных источниках указываются следующие виды функций:

- линейная:  $\alpha(t) = \alpha_0 \left(1 - \frac{t}{S}\right)$ ;
- экспоненциальная:  $\alpha(t) = \alpha_0 e^{-\frac{t}{S}}$ ;
- обратно-пропорциональная времени:  $\alpha(t) = \frac{\alpha_0}{t}$ ;
- «оптимизированная»:  $\alpha(t+1) := \frac{\alpha(t)}{1 + \alpha(t)h_{c_j}}$ .

Функция скорости обучения введена для того, чтобы обеспечивать значительное изменение весов нейронов в начале обучения для уменьшения затрачиваемого времени, а также для обеспечения точной подгонки весов в конце обучения. Изменения весов в конце обучения должно быть минимальным. Таким образом обеспечивается баланс между двумя фазами обучения: быстрого изменения весов и точной подгонки.

Выбор вида функции обеспечит требуемый баланс: линейная функция равномерно убывает и может использоваться лишь в хорошо разделимых кластерах, т.к. при большом значении множителя  $\alpha_0$  начало процесса обучения будет происходить достаточно интенсивно, но точной подгонки весов уже не будет обеспечено. Вторая «крайность» будет наблюдаться при малых значениях  $\alpha_0$ . В таком случае картина будет противоположной — в начале обучения не будет обеспечена быстрая сходимость и точной подгонки при малых изменениях весов может не хватить, и необходимая точность не будет достигнута.

Более подходящей является экспоненциальная функция, которая обеспечивает быстрое обучение в начале и длительную стадию точной подгонки.

Также немаловажное значение играет выбор начальной скорости обучения  $\alpha_0 \in [0; 1]$ . Высокая начальная скорость обучения дает достаточно быструю сходимость, при ее использовании может потребоваться гораздо меньшее число эпох в обучении. Однако, при этом существует опасность, что веса нейронов не смогут стабилизироваться в центрах ожидаемых кластеров, что повлечет за собой большой рост ошибок при кластеризации. По мнению авторов, оптимальное значение  $\alpha_0 \approx 0.5$ .

#### 4.5. Архитектура сети

Обсудим параметры, которые относятся к самоорганизующимся картам Кохонена. Как было отмечено выше, основное отличие сети Кохонена от карт Кохонена в том, что в картах Кохонена нейроны соединены логическими связями. При обучении сети происходит корректировка весов не только «нейрона-победителя», но и его «соседей». При этом нейроны могут иметь по 4 или 6 соседей, образуя тем самым четырехугольные или шестиугольные ячейки.

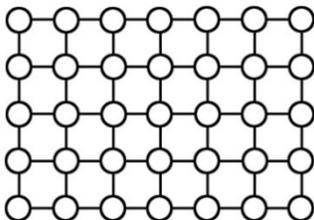


Рис. 1

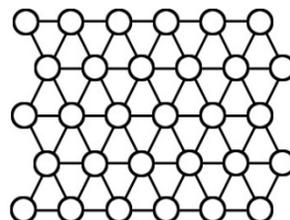


Рис. 2

Шестиугольные ячейки более точно отображают расстояние между объектами, что является более предпочтительным при графическом отображении карт Кохонена (Self-Organizing Map — SOM), т.к. геометрическое расстояние между центрами шестиуголь-

ников является одинаковым. В квадратных же ячейках расстояние между центрами по диагонали и через общую сторону будет отличаться.

Карты Кохонена дают менее точное разбиение исходного множества на кластеры. При этом, благодаря удобному графическому представлению, они дают своему пользователю представление о множестве объектов: оценка расстояния между группами объектов, распределение объектов по кластерам и т.д. Таким образом выполняется сведение многомерного пространства в двумерное.

#### 4.6. Функция соседства

Для функции соседства должно выполняться:  $h_{e_j} \rightarrow 0$  при  $t \rightarrow \infty$ . Существует несколько часто используемых вариантов:

– Сужающиеся области вокруг нейрона-победителя, внутри которых происходит корректировка весов нейронов. Размеры области изменяются дискретно с ростом времени.

– Функция Гаусса:  $h_{e_j} := \alpha(t)e^{-\frac{d^2}{2\sigma^2(t)}}$ , где  $\sigma(t)$  — монотонно-убывающая функция.

#### Заключение

Комбинирование параметров сети Кохонена, самоорганизующихся карт Кохонена позволяет получать различные желаемые результаты: точная кластеризация, получение представления о многомерной выборке данных в двумерном виде, поиск аномалий и др. При правильном конфигурировании, сети и самоорганизующиеся карты Кохонена являются удобным инструментом для решения практических задач.

#### ЛИТЕРАТУРА

1. Боголюбов Д.П., Чанкин А.А., Стемиковская К.В. Реализация алгоритма обучения самоорганизующихся карт Кохонена на графических процессорах. // Промышленные АСУ и контроллеры, № 10, 2012. С. 30–35.
2. Яхьяева Г.Э. Нечеткие множества и нейронные сети: учебное пособие. 2-е-е изд. М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2011. 203 с.
3. Кохонен Т. Ассоциативные запоминающие устройства. М.: Мир, 1982. pp. 30–35
4. Кохонен Т. Самоорганизующиеся карты ; пер. 3-го англ. изд. М.: Бинум. Лаборатория знаний, 2013. 655 с.

## КОМПЛЕКСНАЯ СИСТЕМА ОБРАБОТКИ МЕТЕОРОЛОГИЧЕСКОЙ И АЭРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ ПРИ ОБЕСПЕЧЕНИИ ПОДГОТОВКИ СТАРТОВ И ПУСКОВ РАКЕТ КОСМИЧЕСКОГО НАЗНАЧЕНИЯ

**Н.Т. Дарибаева**

Филиал «Восход» Московского авиационного института  
nurz\_dar7@mail.ru

#### Введение

Существенное влияние на ход выполнения работ с ракетой космического назначения (РКН) в период ее подготовки, транспортировки на стартовый комплекс и пуска оказывают явления погоды, связанные с развитием конвекции. Специалисты метеорологической службы космодрома постоянно контролируют опасные явления погоды с целью предупреждения должностных лиц руководства подготовкой и пуском РКН с максимальной заблаговременностью, тем самым сводя к минимуму воздействия явлений погоды, таких как ливневый дождь, шквал, гроза и град на технические средства [3,4].

Для оперативной деятельности специалистов метеорологической службы возникает необходимость в разработке комплексной системы обработки метеорологических

и аэрологических параметров. Система должна решать ряд задач, таких как предупреждение о критических значениях параметров и прогнозирование конвективных явлений, хранение базы данных.

### **1. Информационная модель**

Входными данными системы являются:

- архив метеорологических и аэрологических данных в формате txt (температура и влажность воздуха, атмосферное давление, скорость и направление ветра, точка росы);
- архив метеонаблюдений в формате xls (дата наблюдения, название явление, продолжительность явления);
- данные, снимаемые с карт барической топографии для поверхностей 500 (5,5 км), 700 (3 км), 850 (1,5 км над уровнем Земли): температура воздуха на изобарической поверхности, температура точки росы, дефицит точки росы.

Выходные данные:

- осреднённые значения метеорологических параметров;
- информация о допустимости значений метеорологических и аэрологических параметров, сообщение об их критичности;
- график зависимости скорости ветра от высоты;
- отчет о данных, полученных при высотно-ветровом радиозондировании атмосферы;
- общее количество и общая продолжительность того или иного явления погоды за рассматриваемый период;
- средняя продолжительность гроз и ливневых дождей;
- повторяемость гроз в различные часы суток в процентах;
- значения индексов неустойчивости;
- прогноз о состоянии атмосферы.

### **2. Разработка программного продукта**

В качестве среды проектирования выбрана среда Rad Studio (Rad Studio Delphi 2007). Среда разработки обладает множеством достоинств, среди которых обширная библиотека классов, возможность быстрой разработки приложений, встроенный отладчик, мощная подсистема работы с базами данных. Для создания базы данных выбрана СУБД Microsoft SQL Server 2008 — одна из наиболее мощных систем управления базами данных архитектуры клиент-сервер, которая в полной мере решает поставленные задачи, позволяет импортировать Excel-файлы в базу данных, представлять отчеты в формате Microsoft Word и Excel [5,6].

Программа содержит четыре подсистемы: подсистему онлайн отображения метеорологических данных (рис. 1), высотно-ветрового зондирования, базу данных и подсистему прогнозирования.

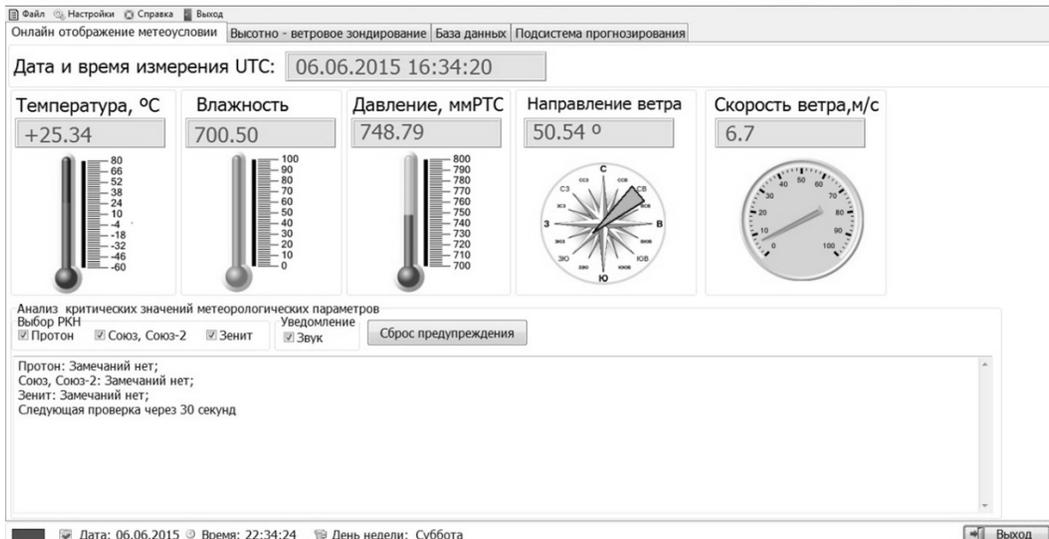


Рис. 1. Интерфейс подсистемы онлайн-отображения метеорологических данных

Подсистема высотно-ветрового зондирования служит для мониторинга значений скорости ветра на высотах (рис. 2). Анализ производится на основе данных файлов ТАЕ-03. В файлах данного типа содержится вся информация, полученная при радиозондировании.

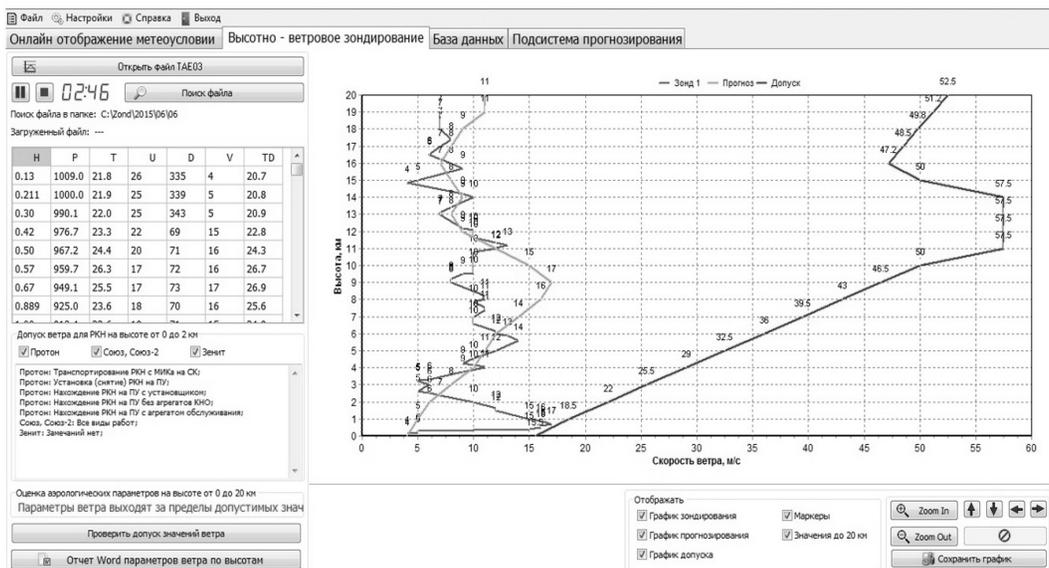


Рис. 2. Интерфейс подсистемы высотно-ветрового зондирования

В подсистеме «База данных» хранится информация о погодных явлениях на территории комплекса Байконур с 1956 по 2013 годы (рис. 3). База данных метеорологической информации позволяет оперативно представлять в удобном для использования виде необходимую метеорологическую информацию, отображать ее в виде графиков, диаграмм, использовать весь спектр статистической информации при разработке прогнозов погоды и совершенствовании методов прогнозирования для более качественно-го обеспечения работ по подготовке и пуску РКН.

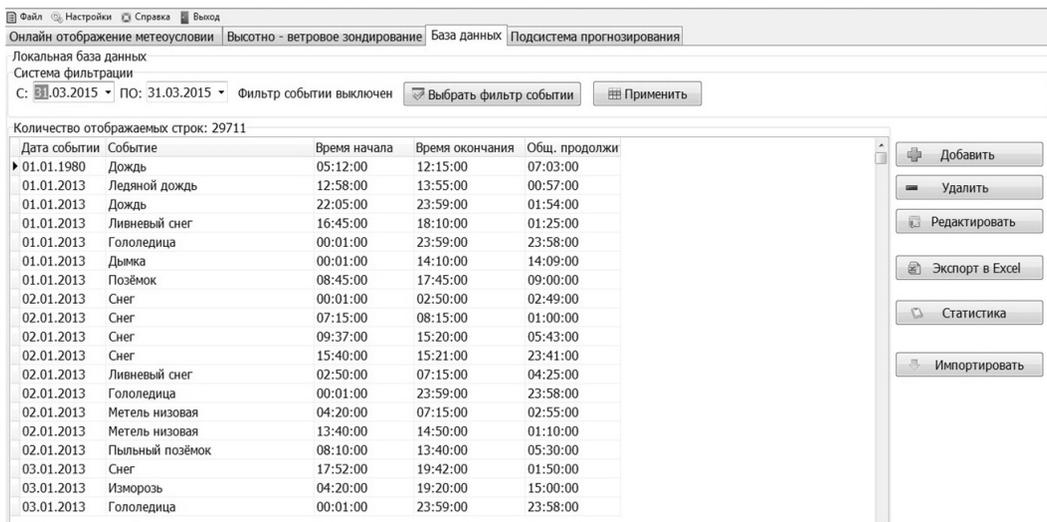


Рис. 3. Интерфейс подсистемы «База данных»

Подсистема прогнозирования выдает прогнозы на основе индексов неустойчивости, таких как LIFT, KIND, TOTL, CAPE, SWEAT (рис. 4).

LIFT объединяет влажность и вертикальный градиент в одном индексе, оценивает степень стабильности, которая измеряется разностью между температурами поднимающейся частицы и окружающей среды на уровне 500 гПа., °C; KIND характеризует температурно-влажностное состояние атмосфере в слое 850–700 гПа., °C; TOTL — комплексная характеристика статической стабильности и влажности на уровне 850 гПа., °C; CAPE — конвективная потенциальная энергия атмосферы, Дж/кг [1,2].

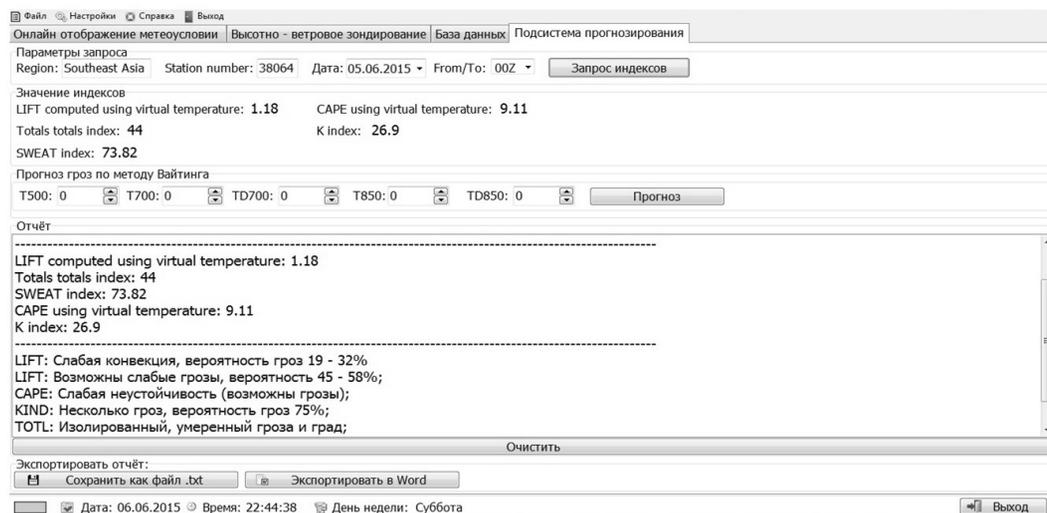


Рис.4. Интерфейс подсистемы прогнозирования конвективных явлений

Подсистема прогнозирования позволяет получать значения индексов неустойчивости, описывающих состояние атмосферы в срок 00.00 и 12. ВCB для аэрологической станции Кызыл-Орда. Особенно информативными для диагноза опасных явлений погоды являются индекс TOTL [3].

## Заключение

Основное назначение комплексной системы хранения и обработки заключается в оперативном представлении метеорологической и аэрологической информации должностным лицам руководства подготовкой и пуском ракет космического назначения (РКН). Предупреждение о приближающейся грозе с максимальной заблаговременностью крайне важно на всех этапах подготовки и пуска РКН. Информация о параметрах ветра на высотах при обеспечении пусков позволяет понимать степень воздействия воздушных течений на ракету космического назначения в полете.

## ЛИТЕРАТУРА

1. *Andersson, T., Andersson, M., Jacobsson, C., Nilsson, S.*: Thermodynamic indices for forecasting thunderstorms in southern Sweden, *Meteorol. Mag.*, 118, P. 141–146, 1989.
2. *Johns R.H., Doswell C.A.*: Severe local storms forecasting. *Wea. Forecasting*, 7 – 1992. – P. 558–612.
3. *Золотухина О.И., Громницкая А.А., Дарибаева Н.Т.* Климатология и гляциология Сибири. Материалы международной научной конференции / под общ. ред. В.П. Горбатенко, В.В. Севастьянова, Томск, 2015. С. 56–59.
4. *Золотухина О.И., Горбатенко В.П., Вареник П.А.* Характеристики ветра в дни пусков ракет космического назначения на космодроме «Байконур» // Труды ГГО им. А.И. Воейкова. 2015. № 576. С. 114–129.
5. *Архангельский А.Я.* Приемы программирования в Delphi на основе VCL. – М. ООО "Бином-Пресс", 2009 г. – 944 с.
6. *Корняков В.* Программирование документов и приложений MS Office в Delphi, «БХВ – Петербург», 2003.

## ИСПОЛЬЗОВАНИЕ ГРАФА ДЛЯ ВИЗУАЛЬНОГО ЗАДАНИЯ БАЗОВЫХ ПРАВИЛ ДОКУМЕНТООБОРОТА

**М.В. Дудкина**

*Томский государственный университет  
masha\_gold@mail.ru*

### Введение

Зачастую предприятию, независимо от его размеров и сферы деятельности, по мере своей работы приходится обрабатывать множество различной документации. Вся деятельность по организации движения документации на предприятии с момента ее создания или получения до завершения исполнения и составляет документооборот предприятия.

Документация предприятия состоит из множества различных типов документов. Для каждого типа требуется иметь четкие представления о стандартном процессе формирования документа и последовательности его обработки, что составляет базовые правила обработки документа. В значительной степени они зависят от назначения документа.

Из-за большого количества этапов, которые должен пройти документ в процессе обработки, и сложных связей между этими этапами, перед компанией встает задача упрощения представления информации о последовательности обработки документации для уменьшения возможных ошибок. Наиболее простым и наглядным представлением является представление в виде графа.

### 1. Постановка задачи

Для разъяснения специфики предметной области рассмотрим подробнее процесс обработки одного типа документа. С момента поступления документ должен пройти последовательность стадий, каждая из которых соответствует некоторой подзадаче обработки. Как только эта подзадача достигнута, документ отправляется на следующую стадию. Таким образом, для каждой стадии должен быть определен переход в одну или несколько следующих стадий, куда может быть направлен документ.

Совокупность стадий и переходов между ними может быть представлена в виде ориентированного графа, вершины которого соответствуют стадиям, а дуги — возможным переходам. Такое представление наглядным образом показывает всю необходимую информацию. При достаточно большой сложности обработки документа в графе могут появляться противоположно направленные дуги между двумя стадиями, что соответствует возможности возврата в исходную стадию. Также в каждой стадии дополнительно может быть определен список действий над документом. Это могут быть такие действия, как: просмотреть, удалить, добавить запись и другие.

Кроме того, имеется список сотрудников, которые могут выполнять те или иные действия из списка возможных в каждой стадии или осуществлять перевод документа между стадиями. Каждый сотрудник может выполнять только назначенные ему действия или переводы документа. Для назначения какого-либо действия или перевода, сотруднику предоставляется право на его выполнение.

Таким образом, имеет смысл добавление в систему программного модуля, который на основе имеющейся информации о всевозможных стадиях, переходах между стадиями, действиях в каждой стадии и сотрудниках позволяет задавать последовательность обработки документа в виде графа. А так же наглядно показывать и регулировать доступные права для каждого сотрудника, что в значительной степени увеличивает контроль над всей системой.

Целью данной работы является упрощение задания базовых правил документооборота с помощью визуального представления в виде графа.

## **2. Реализация отображения**

Пусть имеется информация о последовательности обработки документа. Она содержит список стадий, которые должен пройти документ в процессе своей обработки, список действий в каждой стадии и множество переходов между этими стадиями. Две стадии могут быть связаны переходом как в одну сторону, так и в обе. Переходов из каждой стадии может быть несколько. В каких-то стадиях список действий может совпадать, а в каких-то — сильно отличаться. Вся эта информация отвечает стандарту обработки конкретного типа документа и должна быть учтена при построении графа.

Для визуального представления необходимо, чтобы вся важная информация имела свое простое и понятное обозначение. Это значительно упрощает контроль над системой, так как сразу видны ошибки, если они есть. В данном случае ошибками могут являться, например, лишние или недостающие переходы, отсутствие одной из стадий. Зная весь процесс обработки, пользователь легко увидит некорректную информацию.

Рассмотрим представление данной информации в виде графа (рис. 1): стадии соответствуют вершинам графа, переходы между стадиями — направленным дугам. Для отображения действий припишем каждой стадии соответствующий ей список действий, который можно скрывать и разворачивать по требованию пользователя. Таким образом, вся необходимая информация о последовательности обработки документа учтена и визуализирована.

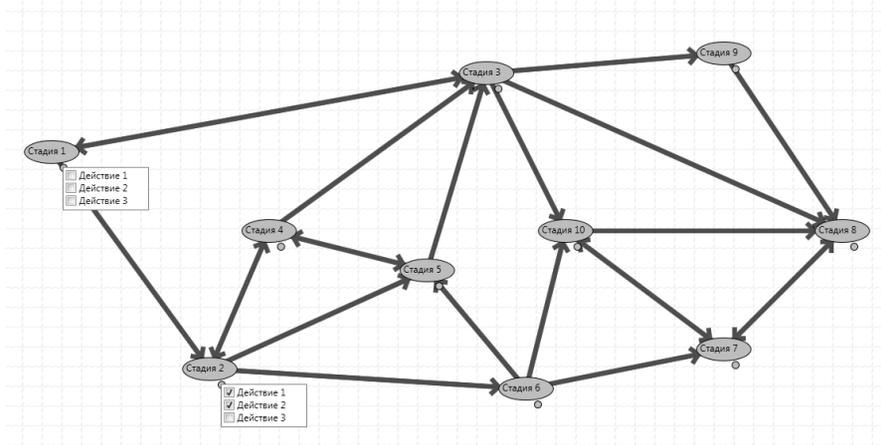


Рис. 1. Пример графа обработки документа

Каждый сотрудник, в зависимости от его квалификации и должности, имеет набор прав, иными словами — действий, которые он может осуществлять над документом. Они не зависят от последовательности обработки документа и, зачастую, пользователь имеет права, которые не могут быть определены в текущей последовательности.

Права могут быть двух видов:

- на выполнение операции с документом, которая находится определённой в стадии;
- на перевод документа в одну из последующих стадий.

Определим их отображение таким образом: право на действие представляется как отмеченное имя действия в списке действий у соответствующей стадии, а право на перевод документа — как выделенная дуга от одной стадии к другой, при этом переходы, на которых у пользователя нет права, отображаются такой же, но невыделенной дугой. Для каждого сотрудника может быть определен частичный граф, содержащий только дуги, соответствующие его правам. Такой частичный граф показывает все возможности сотрудника и позволяет пользователю ограничить его права. Невыделенные дуги не входят в частичный граф сотрудника, но отображаются для сохранения вида всей последовательности.

У сотрудника могут быть права, которых нет в текущей последовательности, но их необходимо отображать для контроля всех прав. Дуги, соответствующие таким правам, отмечаются цветом, в данном случае, темно-серым (рис. 2).

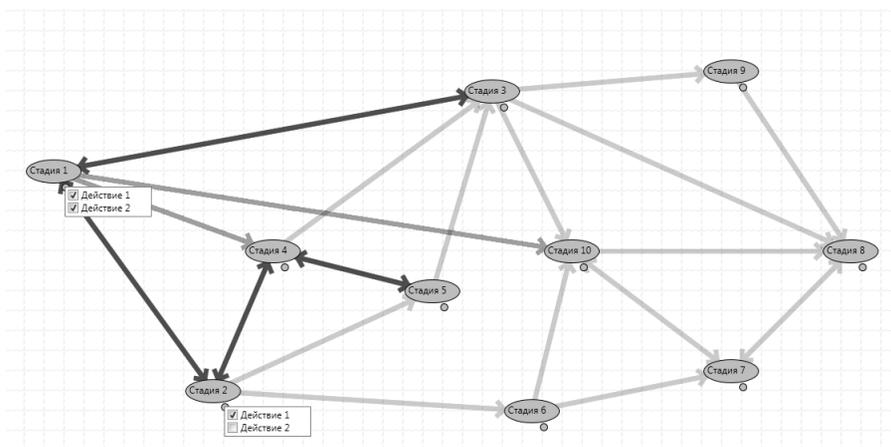


Рис. 2. Пример частичного графа сотрудника

Выбирая определенного сотрудника, пользователь видит все его возможные действия и переходы, при необходимости может добавить новые или убрать лишние.

В каждой последовательности обработки документа имеет смысл проверить ее выполнимость, то есть, существует ли хотя бы одно розданное право на каждое действие или переход. Это гарантирует, что ситуации, в которой никто из сотрудников не сможет выполнить какое-либо действие или переход, не случится. Для этого в модуле реализована возможность выделения группы сотрудников, при которой частичные графы этих сотрудников накладываются друг на друга, отображая все выделенные права. Если существует такая группа сотрудников, выделяя которые получим исходный граф, то можем говорить о выполнимости последовательности обработки документа.

### Заключение

Таким образом, с помощью разработанного модуля пользователь имеет возможность визуализировать информацию о процессе обработки документа. Это делает работу с системой намного проще и понятнее. Пользователь может контролировать текущее состояние системы и быстро находить ошибки в последовательности обработки документа или неправильно розданные права, что значительно уменьшает затраченное время на решение такой задачи по сравнению с традиционными методами.

### ЛИТЕРАТУРА

1. Документооборот: [Электронный ресурс] // Википедия. URL: [https://ru.wikipedia.org/wiki/Делопроизводство/](https://ru.wikipedia.org/wiki/Делопроизводство) (Дата обращения: 07.04.16)
2. Граф\_(математика): [Электронный ресурс] // Википедия. URL: [https://ru.wikipedia.org/wiki/Граф\\_\(математика\)/](https://ru.wikipedia.org/wiki/Граф_(математика)) (Дата обращения: 13.04.16)

## СОЗДАНИЕ АВТОМАТИЗИРОВАННОГО ЭЛЕКТРОННОГО СПРАВОЧНИКА ДЛЯ РАСЧЕТА ТЕРМОДИНАМИЧЕСКИХ И ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ СЛОЖНЫХ ХИМИЧЕСКИХ СИСТЕМ

**А.И. Коробейников, С.Н. Калякин**

*Институт химии и химической технологии СО РАН  
ФГУП «Горно-химический комбинат»  
artkor777@rambler.ru, snk@icct.ru*

### Введение

В настоящее время разработаны практически все необходимые теоретические методы для приближенных расчетов термодинамических (далее — ТД) и физико-химических (далее — ФХ) свойств индивидуальных веществ, их смесей, растворов, химических реакций, фазовых переходов и т.д. Проблема в том, что информация о данных методах и величинах констант находится в разрозненном состоянии, а разработанные методы сложны, трудоёмки и почти всегда требуют от специалиста, проводящего данные расчёты, хорошего знания физической химии и высшей математики. Кроме того, отсутствует доступное программное обеспечение, позволяющее автоматически проводить подобные расчёты, в том числе оценивать значения ТД параметров или констант при заданном давлении и температуре.

В данной статье описана разрабатываемая компьютерная программа (далее — программа), предназначенная для решения вышеуказанных проблем.

### 1. Описание программы

Данная программа создаётся для вычисления ТД и ФХ свойств различных химических систем и процессов и предназначена для максимальной автоматизации расчётов

при решении различных химико-технологических задач, анализе возможности и условий протекания химических реакций, определении ТД параметров веществ и др.

В качестве интерфейса (изображён на рис. 1) и базы данных (далее — БД) для констант и другой необходимой информации по веществам выбран MS Excel, математические вычисления производятся при помощи алгоритма, реализованного в MS Visual Basic for Applications.

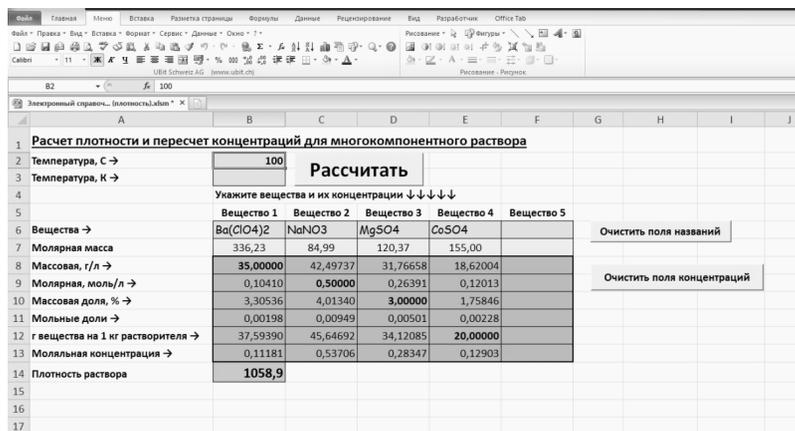


Рис. 1. Интерфейс модуля программы для пересчета концентраций и определения плотности многокомпонентного раствора

При проведении всех расчётов программа производит автоматический поиск веществ, их констант и прочих параметров в БД на основе строчной записи формул веществ. Вся внесенная в БД информация может легко редактироваться и дополняться путём внесения изменений в соответствующий документ MS Excel. После редактирования в MS Excel все дальнейшие вычисления будут проводиться с учётом внесенных в БД изменений, при этом изменений в программном коде не потребуется.

На данный момент программа разрабатывается в виде отдельных независимых модулей, каждый из которых предназначен для решения одной определенной задачи: определение молярной массы вещества, расчёт плотности и т. д. В дальнейшем планируется объединение всех модулей в единый электронный автоматизированный справочник с взаимосвязанными расчетами, так как это значительно расширит функциональность программы.

В созданных модулях были выявлены возможные системные ошибки и разработаны алгоритмы их обработки, например, оповещение пользователя о вводе некорректных значений или об отсутствии запрашиваемых веществ в БД.

## 2. Возможности разрабатываемой программы

Программа разрабатывается под широкий круг задач, с которыми сталкиваются пользователи различной квалификации и сферы деятельности.

Программа может использоваться для проведения типичных расчётов, возникающих в инженерно-технологической, методической и аналитической практике, а именно:

- пересчёт концентраций веществ, при различных способах их выражения, в т. ч. для многокомпонентных растворов с возможностью теоретического расчёта плотности;
- уравнивания химических реакций;
- расчёт молярной массы произвольных веществ, в т. ч. кристаллогидратов;
- расчёт навесок при приготовлении растворов.

Для многокомпонентных водных растворов программа позволит производить расчёты следующих ФХ параметров:

- активности воды;
- коэффициентов активности электролитов;
- плотности;
- вязкости;
- теплоёмкости;
- теплопроводности;
- коэффициентов диффузии;
- поверхностного натяжения;
- температур кипения и замерзания.

Для расчёта ТД параметров индивидуальных веществ, бинарных растворов и химических реакций:

- определение энтальпии, энтропии и энергии Гиббса индивидуальных веществ для произвольной температуры, в различных агрегатных состояниях и в водных растворах различной концентрации;
- уравнивание химических реакций и определение их энтальпии, энтропии и энергии Гиббса и констант равновесия.

Дополнительно в БД планируется добавить и систематизировать (без использования в расчётах) все другие доступные ТД и ФХ данные по химическим системам, реакциям, полуреакциям и т.д.

Уже разработаны и доведены до рабочего состояния следующие части (модули) программы:

- пересчёт концентраций веществ;
- уравнивание химических реакций;
- расчёт молярной массы произвольных веществ, в т. ч. кристаллогидратов;
- расчёт навесок при приготовлении растворов.
- расчёт плотностей;
- в БД внесены константы для расчёта различных ФХ параметров многокомпонентных растворов.

В дальнейшем планируется объединение всех перечисленных модулей в единую программу, так как это позволит добиться синергетического расширения функциональности и более точных вычислений. Например, совместное использование формул модифицированного метода Эзрохи для плотности и формул для выражения концентраций для бинарных растворов (растворимое вещество – растворитель), позволяет решать любые задачи, связанные с пересчётом концентраций и с определением плотности для сколь угодно сложных систем (несколько растворимых веществ – растворитель). Это, в свою очередь, исключает для пользователя необходимость предварительно пересчитывать концентрации в те единицы измерения, которые используются в формулах для расчёта ТД и ФХ величин.

### **Заключение**

В среде MS Excel и VBA разрабатывается программа для автоматизации проведения теоретических расчетов по общей химии, химической термодинамике и физической химии. Программа интегрирована с информационной базой данных по химической термодинамике и физико-химическим параметрам химических систем, которая может быть легко дополнена или исправлена. Были разработаны и оптимизированы различные алгоритмы, в т.ч. разработан оригинальный метод уравнивания химических реакций и метод решения систем нелинейных уравнений при использовании модифицированного метода Эзрохи. Некоторые отдельные части программы уже доведены до рабочего состояния в виде самостоятельных модулей.

### **ЛИТЕРАТУРА**

1. База данных «Термические Константы Веществ» Химического факультета МГУ им. Ломоносова

2. *Верятин У.Д., Маширев В.П. и др.* Термодинамические свойства неорганических веществ. Справочник. // М.: Атомиздат, 1965.
3. *Глушко В.П.* Термические константы веществ. // Справ. изд. – М.: Химия, 1978.
4. *Зайцев И.Д., Асеев Г.Г.* Физико-химические свойства бинарных и многокомпонентных растворов неорганических веществ // Справ. изд. – М.: Химия, 1988.
5. *Никольский Б.П.* Справочник химика. Том первый. // Справ. изд. – Л.: Химия, 1966.
6. *Никольский Б.П.* Справочник химика. Том второй. // Справ. изд. – Л.: Химия, 1964.
7. *Никольский Б.П.* Справочник химика. Том третий. // Справ. изд. – Л.: Химия, 1965.

## **ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННОГО ХРАНИЛИЩА ДЛЯ АВТОМАТИЗИРОВАННОЙ ДИАЛОГОВОЙ СИСТЕМЫ МЕДИЦИНСКОЙ ДИАГНОСТИКИ**

**О.И. Мизюркина**

*Томский государственный университет*  
olga.mizyurkina@mail.ru

### **Введение**

Медицинская диагностика — процесс установления диагноза, то есть заключение о сущности болезни и состоянии пациента, выраженное в принятой медицинской терминологии. Диагностика основывается на всестороннем и систематическом изучении больного, включающем в себя различные виды обследования пациента. На основе полученных результатов исследований специалист переходит к задаче постановки диагноза, решение которой подчиняется пошаговой инструкции, или так называемому медицинскому алгоритму.

В последнее время всё большую популярность набирают так называемые «Check-Up-диагностики». При такой диагностике пациенту назначается определенный перечень лабораторных исследований, сформированный с учетом возраста, пола и некоторых других факторов. Имея на руках результаты анализов, пациент получает консультацию ряда узких специалистов, необходимых именно для него, на основании чего делается вывод о состоянии здоровья пациента. Такой подход не требует от пациентов большого количества времени, что является основной причиной распространения «Check-Up-диагностики».

В данной статье описан процесс проектирования клиентской части автоматизированной системы, предназначенной для расшифровки лабораторных исследований и которая может быть использована на первых этапах «Check-Up-диагностики». Проектирование выполняется с использованием языка объектно-ориентированного моделирования UML (Unified Modeling Language).

### **1. Постановка задачи**

Проектируемая система, если рассматривать её как черный ящик, должна принимать от пользователя несколько видов данных, сохранять их в базу данных, после вычислений сервера получить результат диагностики и направить его клиенту в удобном формате.

Рассматриваемая система должна решать следующие задачи.

1. Поддерживать пользователей типа «пациенты».
2. Хранить информацию о каждом пользователе независимо от других.
3. Не должна требовать от пользователя неточных данных, т.е. данных, которые могут быть оценены более точно, если бы эту оценку проводил специалист в области медицины.
4. Иметь возможность получать от пользователя результаты лабораторных исследований и персональных данных.
5. Поддерживать хранение результатов лабораторных исследований трёх видов (ОАК, б/х крови, ОАМ).

6. Хранить в базе справочник заболеваний с описанием их симптомов.

Все взаимодействия пользователя с системой осуществляются посредством веб-приложения. Основным мотивом к выбору данного подхода при реализации взаимодействия является то, что при использовании веб-приложений пользователю не нужно устанавливать специализированные программные обеспечения, а нужен лишь браузер и наличие доступа в интернет.

## 2. Проектирование

Проектирование с использованием языка UML представляет собой многоуровневый спуск от самой общей и абстрактной концептуальной модели к логической, а затем и к физической модели соответствующей автоматизированной системы. Для проектирования UML предоставляет ряд графических средств — диаграмм, каждая из которых используется для уточнения и конкретизации структуры.

### 2.1. Концептуальная модель

Концептуальная модель представляет собой абстрактную модель, в которой отображается структура проектируемой системы. Первоначально необходимо выделить личность или организацию, взаимодействующую с системой. Так как проектируемая система ориентирована исключительно на пациентов, то и актером в системе будет являться пациент.

Поле выделений актеров осуществляется моделирование работы системы. В терминах UML построение обобщенной модели функционирования системы носит название *диаграмма вариантов использования* (*диаграмма прецедентов*). Основным направлением действия пациента является получение результатов диагностики, которое в свою очередь может быть детализировано, т.е. разбито на несколько шагов (рис. 1).

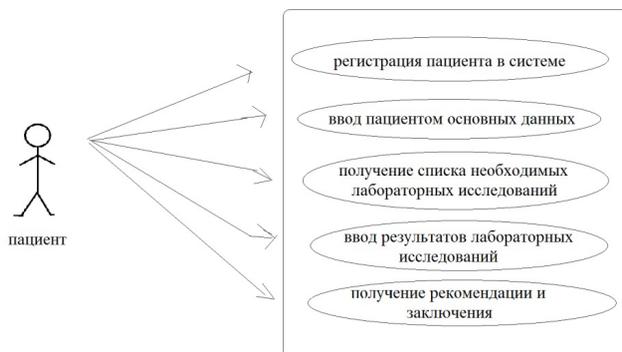


Рис. 1. Диаграмма вариантов использования

Следующий этап детализации — разложение определенной деятельности на её составные части, т.е. описание поведение системы в рамках одного прецедента. Диаграмма, отображающая такую детализацию, называется *диаграммой видов деятельности*. При построении диаграмма используется понятие дорожек. Дорожка — это разновидность пакетов, описывающая связанную совокупность работ. Каждая дорожка представляет собой некоторую сущность и в будущем может быть выделена в отдельный класс. Варианты использования некоторых прецедентов проектируемой системы похожи между собой.

Заключительный этап концептуального моделирования — разработка *модели объектов*, которая отображает выполнение процессов в системе. Одними из основных компонентов модели объектов являются сущности, которые отражают в себе всё то, что используется при реализации процессов, проходящих в системе. Модель объектов

строится для каждого варианта использования и из этого формируется набор классов, которые будут необходимы для функционирования системы.

## 2.2. Логическая модель

На основании всех диаграмм, которые были описаны на этапе концептуального моделирования, осуществляется переход к проектированию *логической модели*, которая становится на уровень ближе к программному коду. Логическая модель — это графическое представление сущностей, выделенных на этапе концептуального проектирования, с указанием их отношений между собой. Помимо связей сущностей, на этапе логического моделирования выделяются и атрибуты сущностей, отображающие свойства объектов или субъектов системы.

Основной диаграммой этапа моделирования логической структуры является *диаграмма классов* (рис. 2). Для её построения первоначально выделяются сущности, с которыми дальше будет осуществляться вся работа. Следующее, что необходимо выполнить, это выделить отношения объектов. Отношение — это семантическая связь между двумя и более объектами. Чаще всего отношения являются графическим изображением действий одного класса относительно другого. Основные виды отношений — это ассоциация, агрегация, композиция, обобщение, отношение зависимости. После выделения отношений сущностей следует выделять атрибуты каждой сущности. Атрибут — свойство класса (объекта). Их чаще можно представить как существительные.

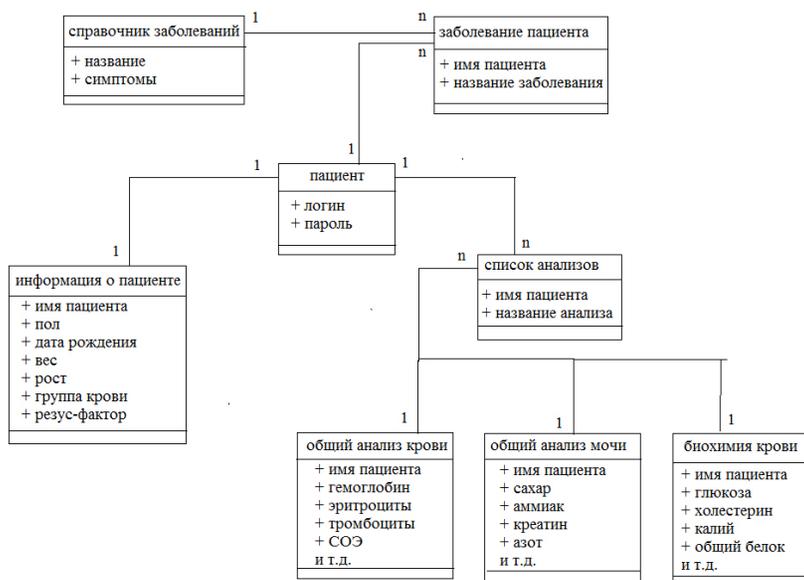


Рис. 2. Диаграмма классов

На основании полученной диаграммы классов, осуществляется переход к проектированию базы данных, т.е. составлению *ER-диаграммы* (рис. 3). В каждый класс добавляется уникальный ключ, для осуществления взаимодействия классов друг с другом. В каждом классе, где встречается атрибут «имя пациента», он заменяется уникальным ключом пациента. Аналогично для атрибута «название заболевания» в классе «заболевание пациента». Это необходимо потому, что именно посредством первичных и внешних ключей таблицы в базе данных будут связываться друг с другом. Устанавливаются некоторые ограничения (триггеры) на таблицы, позволяющие в дальнейшем упростить процессы добавления или удаления записей связанных по внешним ключам.

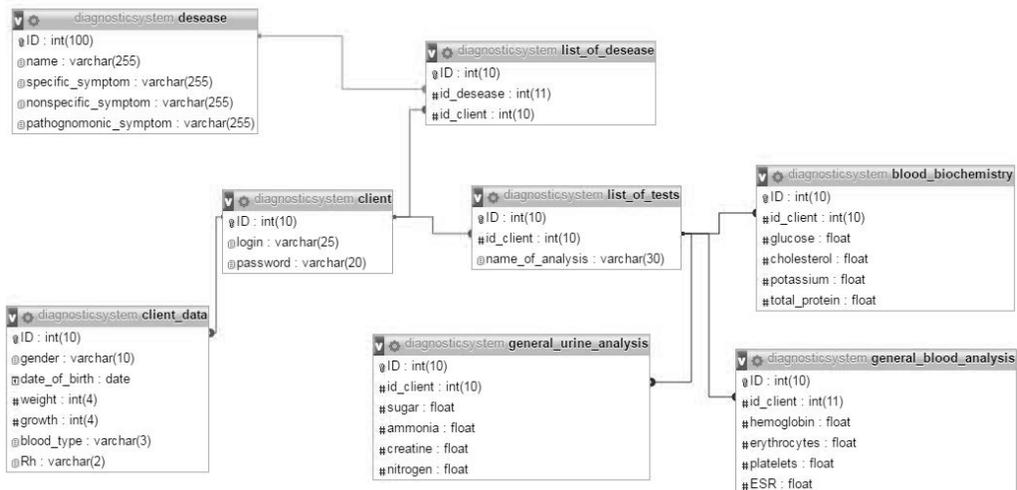


Рис. 3. ER-модель

### 2.3. Физическая модель

Физическая модель является отображением реального развертывания системы, т.е. на данном этапе описываются все компоненты (библиотеки, классы, пакеты и пр.), а также все узлы, используемые при работе системы. Разбиение программной системы на структурные компоненты описывает *диаграмма компонентов* — иерархия подсистем. Из-за простоты системы, исключается необходимость наследования классов друг от друга, а достаточно лишь реализации каждого отдельного класса в отдельном файле. Физическая архитектура системы представляется с помощью *диаграммы развертывания*, на которой отображается вся аппаратная конфигурация. Согласно постановке задачи, будущая система должна взаимодействовать с пользователем посредством веб-приложения. В связи с этим в диаграмме появляются такие компоненты, как веб-сервер и браузер (рис. 4).

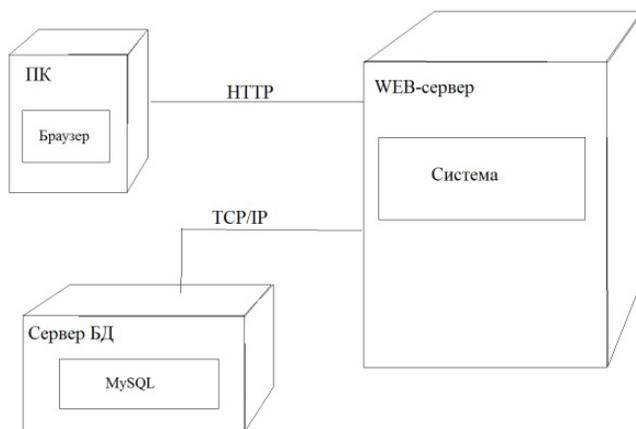


Рис. 4. Диаграмма развертывания

### Заключение

В статье рассмотрена клиентская часть автоматизированной системы, предназначенной для расшифровки лабораторных исследований. Результаты работы системы предоставляются только в виде ознакомительной рекомендации пациенту по здоровью.

В дальнейшем могут использоваться пользователем для обращения в медицинское учреждение и прохождения осмотров специалистами разных областей.

## ЛИТЕРАТУРА

1. Буч Г., Рамбо Д., Якобсон И. Язык UML. Руководство пользователя // 2-е изд.: Пер. с англ. Мухин Н. – М.: ДМК Пресс, 2006. – 496 с.

2. UML [Электронный ресурс]: Материал из Википедии – свободной энциклопедии: Версия 68502747 / Авторы Википедии // Википедия, свободная энциклопедия. – URL: <http://ru.wikipedia.org/wiki/UML> (дата обращения: 10.05.2016)

## ОБУЧАЮЩАЯ ПРОГРАММА НА ТЕМУ «ДЕРЕВО ВАН ЭМДЕ БОАСА»

**Т.С. Овчинникова**

*Томский государственный университет*

*tanyalastochkina@mail.ru*

Деревья поиска — это структура хранения данных. Целью создания является быстрый поиск информации в больших базах данных. С распространением использования интернета в середине 1990-х годов сильно повысилось их значение и важность. Существуют различные деревья поиска. Некоторые из них приведены далее.

- AVL-дерево. Изобретено в 1968 г. Адельсоном-Вельским Георгием Максимовичем и Ландасом Евгением Михайловичем.
- B-дерево. Изобретено в 1970 г. Рудольфом Байером и МакКрейтом.
- Красно-черное дерево. Изобретено в 1972 г. Рудольфом Байером. Название «красно-черное дерево» структура данных получила в 1978 г. в статье Л. Гимпаса и Р. Седжвика.
- Дерево ван Эмде Боаса. Изобретено в 1975 г.
- Рандомизированное дерево. Основателем считается Майкл О. Рабин, исследовавший рандомизированный тест простоты чисел в 1977 г.
- Декартово дерево. Изобретено в 1996 г. Раймундом Сейделем и Сецилией Арагон.

Остановимся на дереве ван Эмде Боаса. Дерево ван Эмде Боаса — ассоциативный массив, который позволяет хранить целые числа в диапазоне  $[0, U)$ , где  $U = 2^k$ , т.е. числа, состоящие не более чем из  $k$  бит. Главная особенность этой структуры — выполнение всех операций за время  $O(\log \log U)$  независимо от количества хранящихся в ней элементов.

Введём основные понятия и обозначения дерева ван Эмде Боаса (vEB):

- 1) множество  $\{0, 1, 2, \dots, u - 1\}$  будем называть *универсумом* или генеральной совокупностью значений, которые могут храниться в дереве, а  $u$  — размером универсума;
- 2) обозначим  $vEB(u)$  — дерево ван Эмде Боаса с размером универсума  $u$ ;
- 3) для хранения динамического множества универсума  $\{0, 1, 2, \dots, u - 1\}$  используем массив  $A[0..u]$  из  $u$  бит. Элемент  $A[x]$  хранит 1, если значение  $x$  находится в динамическом множестве, и 0 — в противном случае;
- 4) *кластером* назовем  $\sqrt{u}$ -битовый подмассив  $A$ . Для данного значения  $x$  бит  $A[x]$  находится в кластере с номером  $\lceil x/\sqrt{u} \rceil$ ;
- 5) введем массив  $summary[0..\sqrt{u} - 1]$ , где  $summary[i]$  содержит 1 тогда и только тогда, когда подмассив  $A[i\sqrt{u}..(i+1)\sqrt{u} - 1]$  содержит 1;

б) размер универсума может быть любой точной степенью 2. И, когда  $\sqrt{u}$  не является целым числом (т.е.  $u$  представляет собой нечетную степень 2:  $u = 2^{2k+1}$  для некоторого целого  $k \geq 0$ ), мы будем делить  $\lg u$  бит числа на старшие  $\lceil \lg u/2 \rceil$  бит и младшие  $\lfloor \lg u/2 \rfloor$  бит. Обозначим  $2^{\lceil \lg u/2 \rceil} = \uparrow\sqrt{u}$  — «верхний квадратный корень», и  $2^{\lfloor \lg u/2 \rfloor} = \downarrow\sqrt{u}$  — «нижний квадратный корень», так что  $u = \uparrow\sqrt{u} \downarrow\sqrt{u}$ . Если  $u$  представляет собой чётную степень 2 ( $u = 2^{2k}$  для некоторого  $k$ ), то  $\uparrow\sqrt{u} = \downarrow\sqrt{u} = u$ .

Для дальнейшего изучения дерева, необходимо разобраться в протоструктуре дерева. Протоструктура — это структура, созданная ван Эмде Боасом, которая существовала до появления дерева. Она поможет понять суть структуры деревьев.

Для начала на битовый вектор  $A$  происходит наложение структуры бинарного дерева. Каждый внутренний узел содержит 1 только тогда, когда некоторый лист его поддерева содержит 1. Другими словами, бит, хранящийся во внутреннем узле, представляет собой логическое ИЛИ его дочерних узлов.

После выполнения наложения дерева с большей степенью станут образовываться кластеры. На рис. 1(а) показано дерево битового вектора  $A\{2,3,4,5,7,14,15\}$ , но уже со степенью  $\sqrt{u}$ . Как и ранее, каждый внутренний узел хранит результат логического ИЛИ битов в его поддереве. На рис. 1(б) показано, что узлы можно рассматривать как массив  $summary[0..\sqrt{u}-1]$ , где  $summary[i]$  содержит 1 тогда и только тогда, когда подмассив  $A[i\sqrt{u}..(i+1)\sqrt{u}-1]$  содержит 1. Этот битовой подмассив мы и назвали кластером.

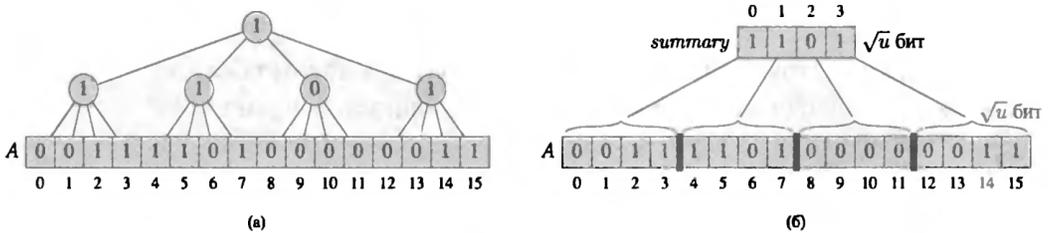


Рис. 1. Дерево со степенью  $\sqrt{u}$ , наложенное на тот же битовый вектор, что и на рис. 4

Теперь перейдем к рассмотрению протоструктуры ван Эмде Боаса (proto-vEB( $u$ )). Каждый элемент хранит следующие атрибуты:

- 1) универсум  $u$ ;
- 2) указатель  $summary$  на структуру proto-vEB( $\sqrt{u}$ );
- 3) массив  $cluster[0..\sqrt{u}-1]$  из  $\sqrt{u}$  указателей, каждый из которых указывает на proto-vEB( $\sqrt{u}$ )-структуру.

На рис. 2 показана полностью развернутая протоструктура, представляющая множество  $\{2,3,4,5,7,14,15\}$ . Структура proto-vEB(16) указывает на четыре структуры proto-vEB(4). Каждая структура proto-vEB(4) указывает на две структуры proto-vEB(2). Каждая структура proto-vEB(2) содержит только массив  $A[0..1]$  из двух битов.

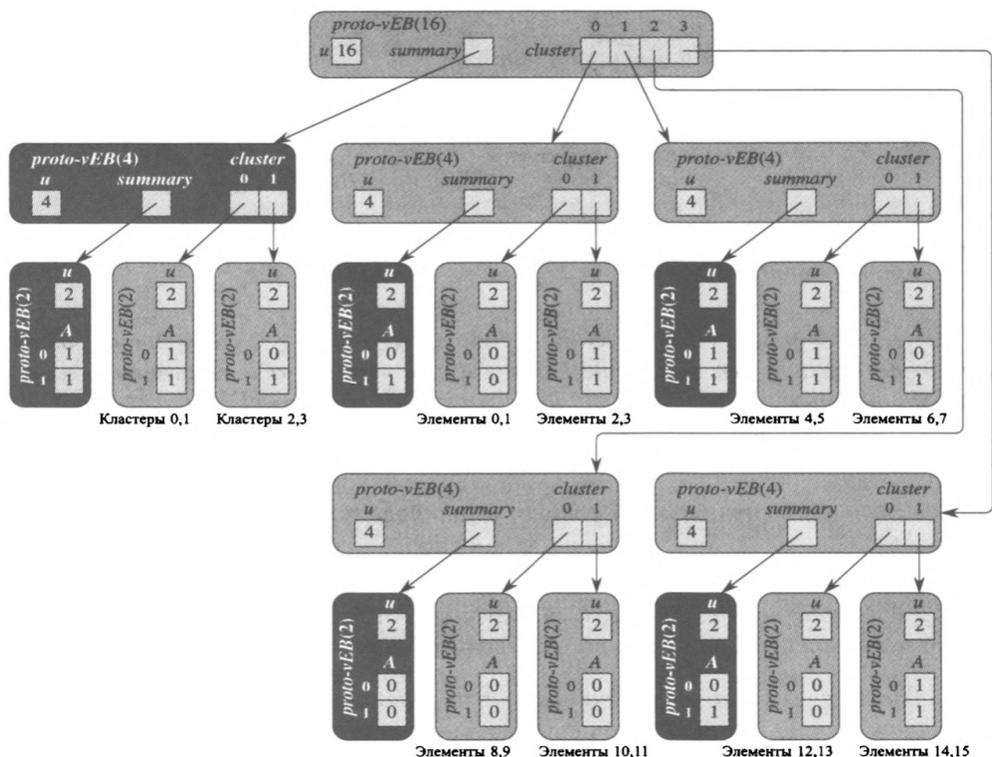


Рис. 2. Структура proto-vEB(16), представляющее множество {2,3,4,5,7,14,15}

Дерево ван Эмде Боаса модифицирует протоструктуру ван Эмде Боаса. vEB-дерево содержит два атрибута, отсутствующие в протоструктуре –  $min$  и  $max$ . Кроме того, элемент, хранящийся в  $min$ , не появляется ни в одном из рекурсивных деревьев  $vEB(\sqrt[3]{u})$ -деревьях, на которые указывает массив  $cluster$ . Таким образом, элементы, хранящиеся в vEB-дереве  $V$ , представляют собой  $V.min$  плюс все элементы, рекурсивно хранящиеся в  $vEB(\sqrt[3]{u})$ -деревьях, на которые указывает  $V.cluster[0..\sqrt[3]{u}-1]$ . Заметим, что когда vEB-дерево содержит два или более элементов, мы рассматриваем  $min$  и  $max$  по-разному: элемент, хранящийся в  $min$ , не появляется ни в одном из кластеров, в отличие от элемента  $max$ , который располагается в кластере (кроме случая, когда vEB-дерево содержит единственный элемент, так что минимальный и максимальный элементы совпадают.)

Поскольку базовый размер 2, vEB(2)-дерево не нуждается в массиве  $A$ , который имеет соответствующая структура proto-vEB(2). Вместо этого его элементы можно определить из атрибутов  $min$  и  $max$ . В vEB-дереве без элементов, независимо от его размера универсума  $u$ , оба атрибута,  $min$  и  $max$ , равны NULL.

В результате, в каждом узле дерева содержатся:

- 1) универсум  $u$ , причем  $u = 2^{2^k}$  для некоторого числа  $k \geq 1$ , так что  $u \geq 4$ . Если же  $u = 2$ , то дерево является протоструктурой ван Эмде Боаса базового размера, и содержит массив  $A[0..1]$  из двух битов;
- 2) атрибут  $summary$ ;
- 3) массив  $cluster[0..\sqrt[3]{u}-1]$ ;

- 4) значение  $min$  минимального элемента дерева  $vEB$ ;
- 5) значение  $max$  максимального элемента дерева  $vEB$ .

Если множество  $\{2,3,4,5,7,14,15\}$  представить в виде дерева ван Эмде Боаса, то оно будет содержать следующие исходные данные:

- массив  $A = \{0011110100000011\}$ ;
- универсум  $u = 16$ ;
- $min = 2$ ;
- $max = 15$ ;
- $summary = \{1101\}$ .

Структура дерева представлена на рис. 3. Поскольку наименьшим элементом является 2, атрибут  $V.min$  равен 2, и несмотря на то что  $high(2) = 0$ , этот элемент 2 отсутствует в  $vEB(4)$ -дереве, на которое указывает  $V.cluster[0]$ : обратите внимание, что  $V.cluster[0].min = 3$ , так что 2 в этом дереве отсутствует. Аналогично, поскольку  $V.cluster[0].min = 3$ , и 2, и 3 являются единственными элементами в  $V.cluster[0]$ , кластеры  $vEB(2)$  в  $V.cluster[0]$  пусты.

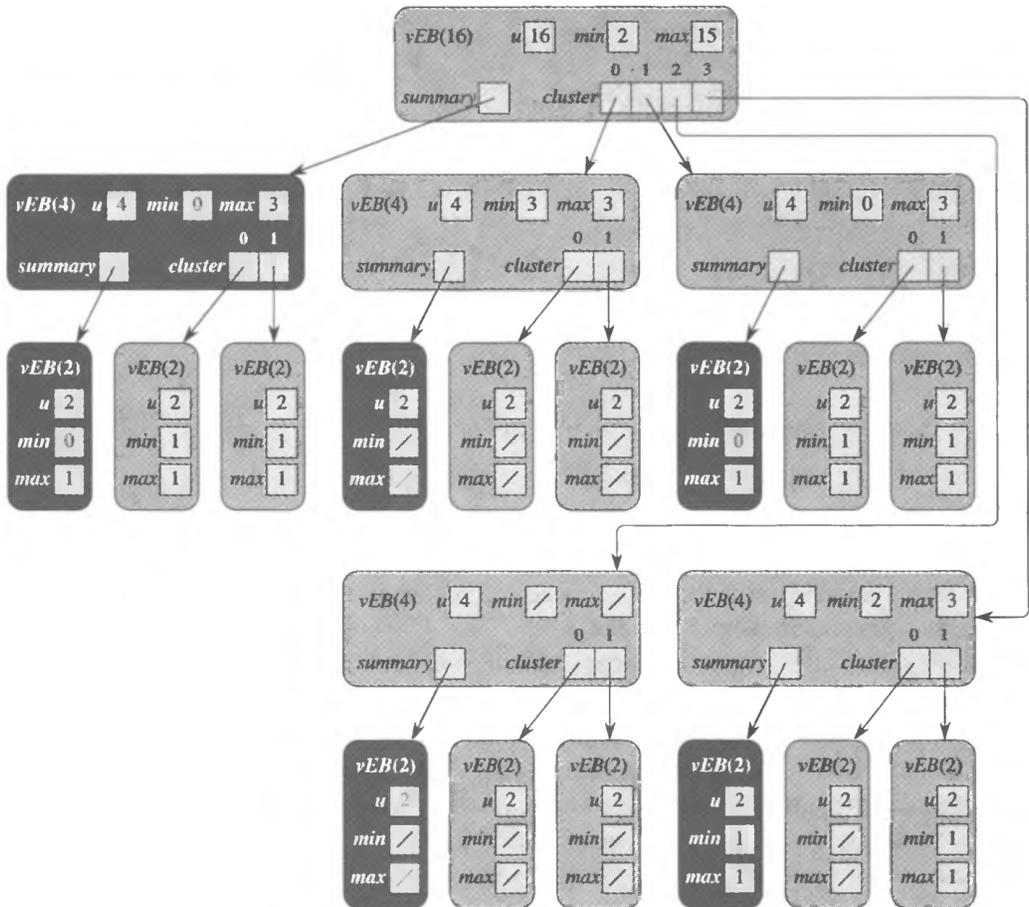


Рис. 3.  $vEB(16)$ -дерево. Оно хранит множество  $\{2,3,4,5,7,14,15\}$ . Косые черты указывают значение NULL.

## Применение дерева ван Эмде Боаса

Кроме очевидного применения дерева ван Эмде Боаса возможны следующие применения:

1. Сортировка последовательности из  $N$  чисел за  $O(n \log \log U)$ . Для этого последовательно выполняем операцию поиска следующего значения. Заметим, что такая реализация удаляет повторяющиеся элементы. Основное преимущество этой сортировки в том, что по асимптотике этот алгоритм обгоняет даже цифровую сортировку.

2. Нахождение наидлиннейшей возрастающей подпоследовательности за  $O(n \log \log U)$ .

3. Алгоритм Дейкстры поиска минимума в графе за  $O(E \log \log U)$ . Реализация алгоритма Дейкстры с помощью двоичной кучи, как известно, работает за  $O(E \log V)$ , где  $V$  — количество вершин в графе, а  $E$  — количество ребер между ними. Но если теперь вместо кучи применить дерево ван Эмде Боаса, то получим асимптотику  $O(E \log \log U)$ .

Минус всех этих алгоритмов заключается в том, что для слишком больших значений  $U$  дерево ван Эмде Боаса будет занимать большее количество памяти (грубая оценка —  $O(U)$ ).

В обучающей программе можно ознакомиться с алгоритмами вставки и удаления, полным описанием дерева и визуализацией дерева. В режиме визуализации обучающийся может по шагам посмотреть как меняется дерево при добавлении новой вершины. На рис.4 показан пример добавления вершины «5» в построенное дерево (слева — до, справа — после добавления этого узла).

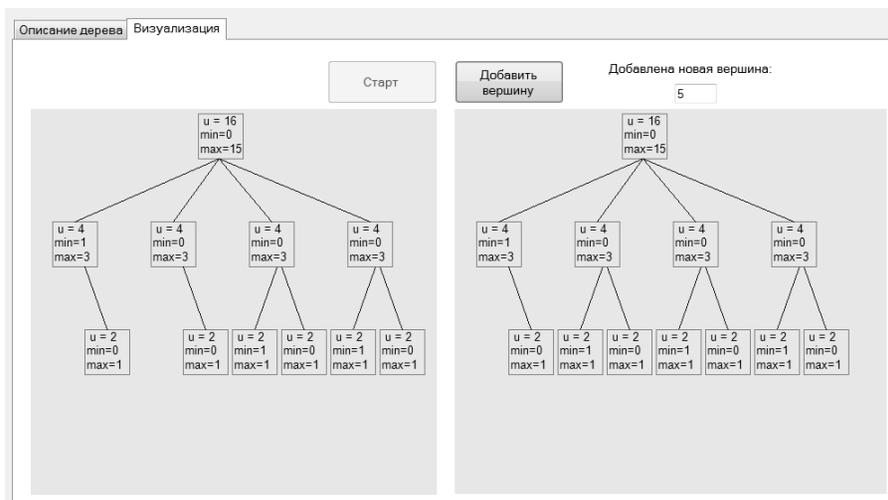


Рис. 4. Обучающая программа

В ходе данной работы изучено дерево ван Эмде Боаса и создана обучающая программа, позволяющая изучить дерево. Алгоритм построения дерева был запрограммирован на языке C++ в визуальной среде Visual Studio.

## Литература

1. Алгоритмы: построение и анализ : [пер. с англ.] / Т. Кормен [и др.]. –3-е изд. – М. [и др.] : Вильямс, 2013. – 1328 с. : ил.
2. Дерево ван Эмде Боаса [Электронный ресурс] // Хабрхабр – свободная энцикл – Электрон. Дан. – [Б. м.], 2011 – URL: <https://habrahabr.ru/post/125499/>.
3. Дерево ван Эмде Боаса [Электронный ресурс] // Википедия : свободная энцикл. – Электрон. дан. – [Б. м.], 2016 – URL: [https://en.wikipedia.org/wiki/Van\\_Emde\\_Boas\\_tree](https://en.wikipedia.org/wiki/Van_Emde_Boas_tree)

## СЕКЦИЯ II. ПРИКЛАДНОЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

### ОЦЕНИВАНИЕ ВЕРОЯТНОСТНЫХ ХАРАКТЕРИСТИК С УЧЕТОМ ИНФОРМАЦИИ О КАЧЕСТВЕННОМ ПРИЗНАКЕ

**Е.В. Белая**

*Томский государственный университет*

*elena-belaya-fpmk@ya.ru*

#### **Введение**

Оценка статистических данных выборочным методом основывается на соответствии свойств выборочных совокупностей генеральным. Добиться повышения точности статистических оценок можно за счет привлечения дополнительной информации.

Влияние априорной информации на оценку может варьироваться в зависимости от типа используемой информации. Задачи с использованием априорной информации широко востребованы во многих направлениях: технике, экономике, транспорте, медицине и т.д. Поэтому данная тема стала активно развиваться на западе. Практическое значение использования дополнительной информации заключается в том, что она помогает получить более точное описание интересующей характеристики генеральной совокупности на основе выборочных данных или сократить объем выборки, необходимой для достижения заданной точности оценок.

Оценки, построенные с учетом априорной информации, могут быть использованы в задачах выборочного контроля качества, социологии, маркетинге, исследованиях потребительских предпочтений, клинических и демографических исследований. Также тема будет актуальна, например, при необходимости сбора первичной информации в «пилотных» исследованиях или в случае, когда необходимо установить качество всей партии или определить среднюю продолжительность работы лампочек, что можно осуществить только разрушающим контролем.

Существуют различные подходы использования априорной информации, в том числе она может быть использована в качестве догадки, доли объектов или множества пропорций оценок. В [1–4] изучаются некоторые свойства оценок при оценивании среднего значения переменной, которые предлагают использовать информацию о множестве пропорций. В [5] рассмотрены различные типы использования априорной информации.

Адаптивные комбинированные оценки, учитывающие априорную догадку и непараметрическую оценку функционала, были представлены в [7–12]. Также интересен метод, использующий априорную информацию о некоторых функционалах плотности [13], который имеет при получении комбинированных оценок регрессии, как в случае конечного объема выборок  $N$ , так и в асимптотическом случае.

Априорная информация о доле объектов, обладающих конкретным значением признака, может позволить добиться более высокой точности при оценке генеральной совокупности. Рассмотрим этот метод, используя подход, предложенный в [7–11]. При этом будем использовать дополнительную информацию о качественном признаке.

Такие методы позволяют получить достаточно надежные результаты при меньших затратах на его реализацию, а также наиболее выгодно, в том числе экономически, исследовать генеральную совокупность.

Целью работы является построение оценки и изучение влияния априорной информации о доле объектов, обладающих конкретным значением признака, на дисперсию, проанализировать асимптотические свойства, а также свойства при конечном объеме

наблюдений путем имитационного моделирования и исследовать ее влияние на генеральную совокупность.

### 1. Постановка задачи

Пусть имеется совокупность объектов  $O_1 \dots O_N$  объема  $N$ . Каждый объект обладает парой признаков  $(X, Y)$ , признак  $X$  является качественным, а  $Y$  – количественным. Предположим, что доля объектов, у которых одно из значений  $X$  задано, известна и равна  $P$ , также имеется выборка без возвращения  $(x_i, y_i)$ ,  $i = \overline{1, n}$ . Необходимо оценить математическое ожидание  $M\varphi(Y)$  некоторой функции  $\varphi$  от случайной величины  $Y$  с учетом знания доли  $P$ . Общепринятой оценкой  $M\varphi(Y)$  является выборочное среднее

$$\hat{J} = \frac{1}{n} \sum_{i=1}^n \varphi(y_i). \quad (1)$$

Эта оценка является несмещенной и обладает наименьшей дисперсией в классе несмещенных оценок

$$D\hat{J} = \frac{N-n}{nN} D\varphi(y) = \frac{1}{n} \left(1 - \frac{n}{N}\right) D\varphi(y). \quad (2)$$

Будем наблюдать значение признака  $X$  через бинарную переменную. Если  $X$  обладает свойством  $A$ , то  $x_i = 1$ , в противном случае  $x_i = 0$ . Введем индикаторную функцию  $I$  и посчитаем количество объектов  $v(A) = \sum_{i=1}^n I_A(x_i)$ , у которых признак  $X$  обладает свойством  $A$ , их доля равна

$$\hat{P} = \frac{v(A)}{n}. \quad (3)$$

Далее предлагается оценка, учитывающая совместно всю имеющуюся информацию, содержащуюся в  $\hat{J}$ ,  $\hat{P}$  и  $P$ .

### 2. Комбинированная оценка

Оценку, учитывающую совместно всю имеющуюся информацию, возьмём в виде, предложенном в [7–9]:

$$\hat{J}(\lambda) = \hat{J} - \lambda(\hat{P} - P), \quad (4)$$

где весовой коэффициент  $\lambda$  выбран из минимума дисперсии [6]

$$D\hat{J}(\lambda) = M[\hat{J}(\lambda) - J]^2 = M(\hat{J}(\lambda) - J)^2 - 2\lambda M(\hat{J}(\lambda) - J)(\hat{P} - P) + \lambda^2 M(\hat{P} - P)^2$$

и равен

$$\lambda = \frac{M(\hat{J}(\lambda) - J)(\hat{P} - P)}{M(\hat{P} - P)^2}.$$

Таким образом, минимум дисперсии при оптимальном коэффициенте  $\lambda$  равен

$$D\hat{J} = M(\hat{J}(\lambda) - J)^2 - \frac{M((\hat{J}(\lambda) - J)(\hat{P} - P))^2}{M(\hat{P} - P)^2}. \quad (5)$$

Второе слагаемое в (5) указывает на величину уменьшения дисперсии с учетом априорной информации при известном коэффициенте  $\lambda$ .

### 3. Адаптивная оценка

Оптимальный коэффициент  $\lambda$  неизвестен, поэтому на практике использование оценки (4) затруднено. Оценку  $\hat{\lambda}$  можно получить методом подстановки. В качестве адаптивной оценки возьмём  $\hat{J}(\hat{\lambda})$ , где

$$\hat{\lambda} = \frac{\frac{1}{n} \sum_{i=1}^n \varphi(y_i) I_A(x_i) - \frac{1}{n} \sum_{i=1}^n \varphi(y_i) P}{P(1-P)}. \quad (6)$$

Исследуем асимптотические свойства оценки  $\hat{J}(\hat{\lambda})$ . При такой замене нарушаются оптимальные свойства оценок, поэтому требуется ответить на вопрос, при каких условиях адаптивные комбинированные оценки являются предпочтительнее обычной оценки  $\hat{P}$  по величине дисперсии.

Предположим, что  $N \gg n$  и  $D\varphi(Y) < \infty$ . Покажем, что при достаточно большом объёме  $n$  свойства адаптивной оценки практически совпадают со свойством оптимальной оценки. В силу центральной предельной теоремы можем записать

$$P\left(\sqrt{n}(\hat{J}(\lambda) - J) < t\right) \approx \Phi(t) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^t e^{-\frac{x^2}{2\sigma^2}} dx, \quad (7)$$

где  $\sigma^2 = D\varphi(Y) - \frac{\iint \varphi(y) I_A(x) dF(x, y) - JP}{P(1-P)} < \infty$ . Далее имеем

$\sqrt{n}(\hat{J}(\hat{\lambda}) - J) = \sqrt{n}(\hat{J}(\lambda) - J) + \sqrt{n}(\hat{J}(\hat{\lambda}) - \hat{J}(\lambda))$ . Покажем, что  $\sqrt{n}(\hat{J}(\hat{\lambda}) - \hat{J}(\lambda)) \Rightarrow 0$  ( $\Rightarrow$  — знак слабой сходимости) при  $n \rightarrow \infty$ . Так как  $\sqrt{n}(\hat{J}(\hat{\lambda}) - \hat{J}(\lambda)) = \sqrt{n}(\hat{P} - P)(\lambda - \hat{\lambda})$  и  $\sqrt{n}(\hat{P} - P)$  слабо сходится к гауссовской случайной величине ( $M = 0$ ,  $D = \hat{P}$ ), а  $\hat{\lambda}$  сходится по вероятности к  $\lambda$  (в силу второй теоремы непрерывности [11, глава 6]), то  $\sqrt{n}(\hat{J}(\hat{\lambda}) - J) \approx \sqrt{n}(\hat{J}(\lambda) - J)$ .

### ЛИТЕРАТУРА

1. Singh R., Pankaj Chauhan, Nirmala Sawan, Florentin Smarandache // Ratio Estimators in Simple Random Sampling Using Information on Auxiliary Attribute // Pakistan Journal of Statistics and Operation Research; Vol 4. No. 1, Jan 2008, pp. 47–53.
2. Jhaji, H.S., Sharma, M.K. and Grover, L.K. (2006): A family of estimators of population means using information on auxiliary attributes. Pak. J. Statist., 22 (1), 43–50.
3. Naik, V.D. and Gupta, P.C. (1996): A note on estimation of mean with known population proportion of an auxiliary character. Jour. Ind. Soc. Agr. Stat., 48 (2), 151–158.
4. Ray, S. K. and Singh, R.K. (1981): Difference-cum-ratio type estimators. J. Ind. Statist. Assoc., 19, 147–151.
5. Дмитриев Ю.Г., Устинов Ю.К. Статистическое оценивание распределений вероятности с использованием дополнительной информации. — Томск: Изд-во Том. ун-та, 1988. — 194 с.
6. Дмитриев Ю.Г., Тарасенко П.Ф. Использование априорной информации в статистической обработке экспериментальных данных. // Известия вузов. Физика. — 1992. № 9. С. 136–142.
7. Дмитриев Ю.Г. Комбинированная оценка линейного функционала // Материалы X Российской конференции с международным участием «Новые информационные технологии в исследовании сложных структур» Томск. Изд-во НТЛ, 9–13 июня 2014. С. 91.
8. Dmitriev Yu., Tarassenko P., Ustinov Yu. On Estimation of Linear Functional by Utilizing a Prior Guess. // Dudin A. et al. (Eds.): ITMM 2014, CCIS 487, pp. 82–90, 2014.
9. Дмитриев Ю.Г., Тарасенко П.Ф. О комбинированных оценках линейного функционала // (ITMM 2014) Материалы XIII Международной научно-практической конференции им. А.Ф. Терпугова (20–22 ноября 2014) — Томск: Изд-во Том. ун-та, 2014. Ч. 1 — С. 25–29.
10. Белая Е.В., Дмитриев Ю.Г., Логвинова А.И. О точности оценок с априорной догадкой при конечном числе наблюдений // Известия вузов. Физика. 2015. Т. 58, № 11/2. С. 230–234.

11. Белая Е.В. О точности комбинированной оценки среднего при известной дисперсии // Математическое и программное обеспечение информационных, технических и экономических систем: материалы III Всероссийской молодежной научной конференции. Томск, 22–23 мая 2015 г. – Томск: Издательский Дом Томского государственного университета, 2015. С. 58–60.
12. Дмитриев Ю.Г., Скрипин С.В. О комбинированной оценке вероятности безотказной работы по полной выборке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2012. – № 4. – С. 32–38.
13. Дмитриев Ю.Г. Непараметрическое условное оценивание функционалов плотности распределения // Математическое моделирование и теория вероятностей / Под ред. И.А. Александрова, А.М. Бубенчикова, В.Н. Берцуна, Ю.К. Устинова. – Томск: Изд-во «Пеленг», 1998. – С. 169–177.
14. Боровков А.А. Математическая статистика. М.: Наука, 2007. 704 с.
15. Дмитриев Ю.Г., Кошевая Т.О. О комбинированных оценках вероятности // Известия вузов. Физика. – 2015. – Т. 58, № 11/2. – С. 242–246.
16. Dmitriev Yu., Tarassenko P. On Adaptive Estimation Using a Prior Guess // Applied Methods of Statistical Analysis. Nonparametric Approach – AMSA-2015, Novosibirsk, Russia, 14–15 September, 2015: Proceedings of the International Workshop. – 2015. – P. 49–55.
17. Дмитриев Ю.Г., Тарасенко П.Ф. О к-адаптивных оценках вероятности. // Информационные технологии и математическое моделирование (ИТММ-2015). Ч. 1: Материалы XIV Международной конференции имени А.Ф. Терпугова, 18–22 ноября 2015 г. . Изд-во Томского университета, 2015. С. 40–45.
18. Дупин-Барковский И.В., Смирнов Н.В. Теория вероятностей и математическая статистика в технике (общая часть). М., 1955.
19. Беляев Ю.К. Вероятностные методы выборочного контроля. М., 1975.
20. Кендалл М., Стьюарт А. Теория распределений. Пер. с англ., М., 1966.
21. Arcos A., Rueda M., Martínez M.D., Roman Y. Incorporating the auxiliary information available in variance estimation // Applied Mathematics and Computation 160 (2005) 387–399.
22. Subhash K., Cem K. A two parameter variance estimator using auxiliary information // Applied Mathematics and Computation 226 (2014) 117–122.

## ОЦЕНИВАНИЕ СИММЕТРИЧНЫХ РАСПРЕДЕЛЕНИЙ С НЕИЗВЕСТНЫМ ЦЕНТРОМ

**Е.А. Бударина**

*Томский государственный университет*  
budarinajane@gmail.com

В современном мире объем имеющихся статистических данных огромен, более того каждую секунду они продолжают непрерывно поступать. Обработка имеющихся данных включает в себя статистическое оценивание параметров, проверку статистических гипотез, прогнозирование и т.д. Улучшить обработку статистических данных можно за счет использования разнообразной дополнительной априорной информации, которой может располагать исследователь. Так, в работах [1,2] предложены способы учета априорной информации в статистическом оценивании распределений. В частности, рассматривалась задача оценки функции распределения случайной величины с учетом симметрии распределения с неизвестным центром. В данной работе рассматривается ситуация, когда центр симметрии не оценивается по выборке, а задается исследователем в качестве априорной догадки. Предлагается структура оценки с весовым коэффициентом, использующим эту информацию, и анализируется среднеквадратическая ошибка в зависимости от отклонения априорной догадки от истинного значения. Строится адаптивная оценка функции распределения путем оценивания весового коэффициента по исходным данным.

### 1. Постановка задачи и структура оценки

Пусть  $x_1, x_2, \dots, x_n$  — случайная выборка размера  $n$  из распределения  $F(x)$ , которая принадлежит классу всех непрерывных симметричных функций распределения  $F \in \Omega(\theta) = \{G : G(x) = 1 - G(2\theta - x), x \in R^1\}$ . Предположим, что центр симметрии  $\theta$  неизвестен и положим его равным  $a$ . Возьмем в качестве оценки  $F(x)$  взвешенную функцию вида:

$$F_n^G(x) = (1 - \lambda)F_n(x) + \lambda(1 - F_n(2a - x)) = F_n(x) - \lambda(F_n(x) - 1 + F_n(2a - x)), \quad (1)$$

где  $F_n(x)$  — эмпирическая функция распределения;  $\lambda$  — параметр, который находится из условия минимума среднеквадратической ошибки (СКО)  $S^2 = M(F_n^G(x) - F(x))^2$  и равен

$$\lambda_0 = \frac{F(x)(1 - F(x) - F(2a - x)) + F(\min(x, 2a - x))}{2F(\min(x, 2a - x)) + 1 - F(x) - F(2a - x) + (n - 1)(1 - F(x) - F(2a - x))^2}. \quad (2)$$

Оценка (1) является, вообще говоря, смещенной, и СКО оценки есть:

$$S^2 = DF_n(x) - \frac{[M(F_n(x) - F(x))(F_n(x) - (1 - F_n(2a - x)))]^2}{M(F_n(x) - (1 - F_n(2a - x)))^2}, \quad (3)$$

где  $DF_n(x) = \frac{F(x)(1 - F(x))}{n}$  — дисперсия эмпирической функции распределения.

Второе слагаемое в (3) показывает, насколько уменьшается СКО оценки (1) при использовании априорной догадки центра симметрии.

В частности, если  $a = \theta$ , то  $\lambda_0 = \frac{1}{2}$ ,  $S^2 = DF_n(x) - \frac{1}{2}F(\min(x, 2\theta - x))$ . Если  $a \neq \theta$ ,

то

$$S^2 = DF_n(x) - \frac{\frac{1}{n}[F(x)(1 - F(x) - F(2a - x)) + F(\min(x, 2a - x))]^2}{2F(\min(x, 2a - x)) + 1 - F(x) - F(2a - x) + (n - 1)(1 - F(x) - F(2a - x))^2}. \quad (4)$$

Оптимальный коэффициент  $\lambda$  в (2) неизвестен. Однако представляет интерес изменения СКО (3) при оптимальном  $\lambda$  и различных отклонениях априорной догадки от истинного центра симметрии. С этой целью рассмотрим иллюстративный пример.

## 2. Иллюстративный пример

Для анализа сравнения СКО (3) с дисперсией эмпирической функции распределения рассмотрим отношение:

$$E = \frac{S^2}{DF_n(x)}. \quad (5)$$

Возьмем в качестве  $F(x)$  стандартный нормальный закон с нулевым математическим ожиданием и единичной дисперсией  $N(0,1)$ .

Пусть у исследователя имеется одна априорная догадка. Пронаблюдаем поведение СКО оценки (1) на примере нормального стандартного распределения при различных значениях центра симметрии.

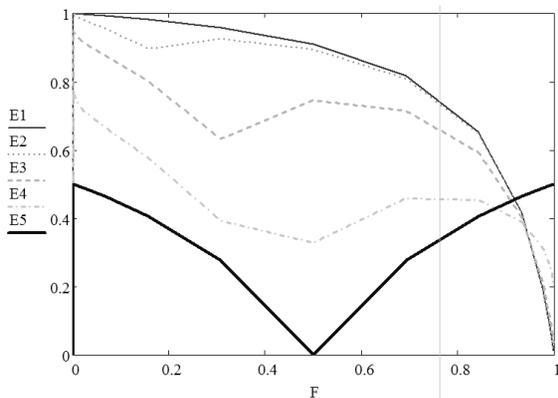


Рис. 1. Зависимость отношения  $E$  от отклонения априорной догадки от истинного значения

Кривая E1 показывает поведение  $E$  при отклонении априорной догадки от истинного на 100 единиц; E2 — при отклонении на 1 единицу; E3 — на 0.5 единицы; E4 — на 0.1 единицы; E5 — нет отклонения.

На графике видно, что при любом из рассмотренных отклонениях центра симметрии от истинного существует выигрыш в точности в смысле СКО.

### 3. Построение адаптивной оценки

Предположим, что  $F(x)$  неизвестна. Тогда оценим  $\lambda$  по имеющейся выборке и введем адаптивную оценку:

$$\lambda_n = \frac{F_n(x)(1 - F_n(x) - F_n(2a - x)) + F_n(\min(x, 2a - x))}{2F_n(\min(x, 2a - x)) + 1 - F_n(x) - F_n(2a - x) + (n - 1)(1 - F_n(x) - F_n(2a - x))^2}, \quad (6)$$

$$\hat{F}_n^G(x) = F_n(x) - \lambda_n(F_n(x) - 1 + F_n(2a - x)). \quad (7)$$

Анализ оценки (7) будет проведен автором в последующих работах.

### Заключение

В настоящей работе рассмотрена оценка функции с учетом симметрии при неизвестном центре, проиллюстрирован выигрыш в точности при использовании данной оценки по сравнению с эмпирической функцией распределения. Также построена адаптивная оценка функции распределения путем оценивания весового коэффициента по исходным данным.

### ЛИТЕРАТУРА

1. *Shuster E.U.* Estimating the distribution function of a symmetric distribution. *Biometrika* (1975), 62, 3, p. 631–635.
2. *Дмитриев Ю.Г., Устинов Ю.К.* Статистическое оценивание распределений вероятностей с использованием дополнительной информации, Издательство Томского университета, Томск, 1988.
3. *Shuster E.U.* On the Goodness-of-Fit Problem for Continuous Symmetric Distributions. *Journal of the American Statistical Association*, September 1973, Volume 58, Number 343, p. 713–715.

## РЕГРЕССИОННЫЕ МОДЕЛИ ЗАРАБОТНОЙ ПЛАТЫ В РОССИЙСКОЙ ФЕДЕРАЦИИ

**Ю.Г. Дмитриев, П.А. Савченко**

*Томский государственный университет*  
dmit70@mail.ru, sapi95@ramble.ru

### Введение

Одна из важнейших микроэкономических проблем, стоящих перед концепцией человеческого капитала — оценка влияния, оказываемого на величину текущих доходов (заработной платы) различными формами человеческого капитала: продолжительностью обучения, общим профессиональным опытом и так называемым специфическим человеческим капиталом — продолжительностью работы на определенной фирме. Решение данной задачи осуществляется на основе различных модификаций модели Дж. Минцера [1], впервые обоснованной в 1958 году:

$$\ln W = \beta_0 + \beta_1 SCH + \beta_2 EXP + \beta_3 EXP^2 + \beta_4 TEN + \beta_5 TEN^2,$$

где

- $SCH$  — число лет обучения, скорректированное по достигнутому уровню образования (начальное и неполное среднее — 8 лет, полное среднее — 10 лет, профессиональное-техническое — 11,5 лет, среднее специальное — 13 лет, высшее — 15 лет; послевузовское (аспирантура) — 18 лет);
- $EXP$  — потенциальный опыт на рынке труда (рассчитывается по условной формуле  $EXP = \text{возраст} - SCH - 6$  лет);

- $TEN$  – «специфический человеческий капитал» или профессиональный опыт, накопленный на данном рабочем месте (или с данным работодателем);  $W$  — заработная плата по основному месту работы;  $\beta_i$  – коэффициенты при соответствующих переменных, характеризующие норму отдачи от инвестиций в образование, профессиональный опыт и «специфический человеческий капитал» ( $\beta_0, \beta_1, \beta_2, \beta_4$  имеют знак «плюс»;  $\beta_3$  — знак «минус»).

Основным недостатком модели Минцера, отмечаемым многими исследователями, является постоянство коэффициента  $\beta_1$ , т.е. при увеличении длительности обучения  $SCH$  на один год заработная плата индивида должна меняться одинаково вне зависимости от того, какое образование он уже имеет. Однако можно предположить, что обучение в аспирантуре увеличивает заработную плату сильнее, чем обучение в техникуме [2].

В настоящей статье исследуется применимость модели Минцера (и её модификаций) для данных о жителях РФ за 2014 г., а также строится непараметрическая модель зависимости заработной платы от разных факторов по тем же данным и проводится сравнение полученных результатов.

### 1. Постановка задачи

Пусть  $n$  раз измерены значения факторов  $x_1, x_2, x_3, \dots, x_m$  и соответствующие значения переменной  $Y$ , тогда предполагается, что

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,m} + \varepsilon_i, \quad i = \overline{1, n}$$

(второй индекс при  $x$  относится к номеру фактора, а первый — к номеру наблюдения). Также предполагается, что выполняются условия Гаусса-Маркова:

- 1)  $M\varepsilon_i = 0, \quad \forall i = \overline{1, n},$
- 2)  $M\varepsilon_i \varepsilon_j = \begin{cases} \sigma^2, & i = j, \\ 0, & i \neq j, \end{cases} \quad \forall i, j = \overline{1, n},$
- 3)  $D\varepsilon_i = \sigma^2, \quad \forall i = \overline{1, n}.$

На основании метода наименьших квадратов несмещенная оценка для вектора параметров  $\beta$  в матричной форме примет вид [3]:  $\hat{\beta} = (X^T X)^{-1} X^T Y$ , где  $X$  — матрица планирования эксперимента размерности  $n \times (m + 1)$ :

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix}.$$

Для проверки качества уравнения регрессии будет использоваться коэффициент детерминации  $R^2 = 1 - \frac{\|Y - \hat{Y}\|^2}{\|Y - \bar{Y}\|^2}$  (где  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$ ), который показывает качество подгонки

регрессионной модели к наблюдаемым значениям. Если  $R^2 = 0$ , то регрессия  $Y$  на  $x_1, x_2, x_3, \dots, x_k$  не улучшает качество предсказания по сравнению с тривиальным предсказанием  $\hat{y}_i = \bar{y}$ . Другой крайний случай  $R^2 = 1$  означает точную подгонку: все  $\varepsilon_i = 0$ , то есть все точки наблюдений лежат на регрессионной плоскости.

Еще одной важной характеристикой является средняя ошибка аппроксимации, которая показывает среднее отклонение расчетных значений от фактических [3]:

$$A = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|.$$

Непараметрическая регрессия, в отличие от параметрических подходов, использует модель, которая не описывается конечным числом параметров [4].

Идея ядерного сглаживания состоит в представлении последовательности весов  $\{W_i(x)\}$ ,  $i = \overline{1, n}$ , где форма весовой функции  $W_i(x)$  описывается посредством функции плотности со скалярным параметром  $h$  (ширина окна), который регулирует размер и форму весов около  $x$ . Эту функцию формы принято называть ядром  $K$ . Ядро — ограниченная, вещественная функция  $K$  с единичным интегралом.

Для сглаживания в многомерном случае в основном применяют произведение одномерных ядер. В этом случае ядерная оценка регрессии имеет вид:

$$\hat{r}(x) = \sum_{i=1}^n Y_i W_i(x) = \frac{\sum_{i=1}^n Y_i \prod_{j=1}^m K\left(\frac{x_j - X_{ij}}{h}\right)}{\sum_{i=1}^n \prod_{j=1}^m K\left(\frac{x_j - X_{ij}}{h}\right)},$$

где  $j$  — номер фактора.

## 2. Выбор оптимальных параметров для ядерной регрессии

Обычно в качестве ядерных функций рассматривают равномерное, треугольное, Епанечниково, Гауссово (нормальное) ядро. Важно отметить, что областью определения первых трех ядер является отрезок  $[-1, 1]$ , в то время как последнее имеет бесконечный носитель. Следовательно, при использовании равномерного, треугольного или Епанечникова ядра оценка будет использовать информацию в ограниченном окне в окрестности  $X$ , а оценка, использующая гауссово ядро, будет использовать информацию из всех наблюдений. Исходя из этого, для расчета ядерной оценки регрессии заработной платы будет использоваться нормальное ядро [4]:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

Ключом к проведению качественного непараметрического оценивания является выбор оптимальной ширины окна для поставленной задачи. Хотя ядерная функция остается важной, всё же главная её роль состоит в обеспечении дифференцируемости и гладкости получающихся оценок. Ширина окна, с другой стороны, определяет поведение оценки в конечных выборках, что ядерная функция сделать просто не в состоянии. Так как данные в имеющейся выборке масштабированы, то в данной работе параметр размытости ищется как минимум средней ошибки аппроксимации:

$$A = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i(h)}{y_i} \right| \rightarrow \min_h,$$

если  $A \cdot 100\% \leq 8\%$ , то модель считается адекватной [5].

## 3. Результаты регрессионного анализа

В настоящей работе регрессионный анализ строится по данным 23 волны Российского мониторинга экономического положения и здоровья населения (РМЭЗ), где объем выборки равен 50 респондентам, проживающим в одном городе (Ленинградская область, Волосовский район). В табл. 1 представлены оценки параметров модели Минцера в общем виде ( $\ln W = Y$ ,  $EXP^2 = EXP2$ ,  $TEN^2 = TEN2$ ), в табл. 2 — её характеристики, которые были вычислены в программе Mathcad Prime 3.0:

Таблица 1

## МНК оценки параметров

	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
$\hat{\beta}$	9,341	0,044	0,00542	-0,0003195	0,022	-0,000352

Таблица 2

## Характеристики модели №1 (Минцера)

	$R^2$	$A$
Характеристики модели	0,122	20,7%

В результате анализа полученных в табл. 1, 2 результатов делаем вывод, что общая модель Минцера не применима к имеющимся данным.

В табл. 3, 4 приведены результаты анализа модернизированной модели Минцера, в которой из набора факторов исключаются квадраты  $EXP^2$ ,  $TEN^2$ :

$$\ln W_i = \beta_0 + \beta_1 SCH_i + \beta_2 EXP_i + \beta_3 TEN_i + \varepsilon_i, \quad i = \overline{1, n}.$$

Таблица 3

## МНК оценки параметров

	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
$\hat{\beta}$	9,497	0,052	0,011	0,008305

Таблица 4

## Характеристики модели №2

	$R^2$	$A$
Характеристики модели	0,114	5,687%

Согласно полученным результатам можно сделать вывод, что модель Минцера без квадратов факторов ( $EXP^2$  — потенциальный опыт работы,  $TEN^2$  — профессиональный опыт) можно применять к имеющимся данным.

В данной статье также рассматривается непараметрический анализ данных по выборке  $(X, Y)$ , где  $Y$  — вектор логарифмов заработной платы  $\ln W$ , а  $X$  — матрица факторов размерности  $n \times m$ , где  $n = 50$  — количество наблюдений,  $m = 3$  — количество факторов ( $SCH, EXP, TEN$ ). На рис. 1 представлен график зависимости ошибки аппроксимации  $A$  от ширины окна  $h$ . По этому графику видно, что существует минимум функции, который достигается в искомом оптимальном  $h$ .

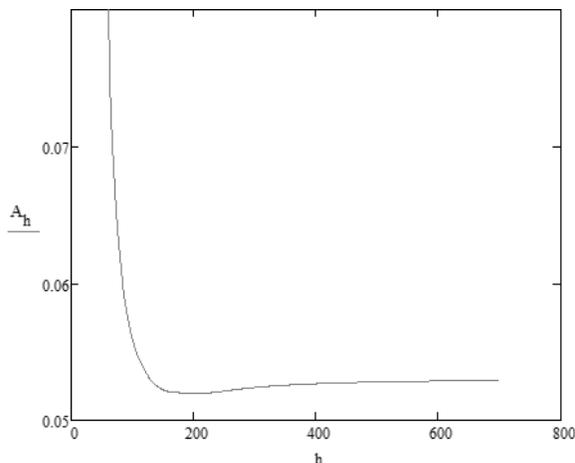


Рис. 1. График зависимости ошибки аппроксимации от ширины окна

В табл. 5 представлены вычисленные значения оптимальной ширины окна и минимальной ошибки аппроксимации для имеющихся данных:

Таблица 5

**Характеристики ядерной оценки**

	$h$	$A$
Характеристики $\hat{f}(x)$	206	5,193%

**Заключение**

В работе рассмотрены параметрическая и непараметрическая регрессионные модели для анализа заработной платы в РФ от различных факторов. Сравнение результатов проведенного анализа показало, что ядерная оценка зависимости заработной платы является более приемлемой для имеющихся данных. Полученная ядерная оценка может применяться для восстановления данных по заработной плате при опросе в данной области путем прогнозирования.

**ЛИТЕРАТУРА**

1. Саградов А.А. Экономическая демография: учебное пособие // М.: Инфра, Москва, 2010, с. 135–137.
2. Демидова А.А. Непараметрический анализ зависимости заработной платы россиян от образования и опыта работы // Москва, ВШЭ
3. Магнус Я.Р. Эконометрика: начальный курс // Я.Р. Магнус, П.К. Катыхев, А.А. Пересецкий // Москва: Изд-во «Дело».
4. Анатольев С. Непараметрическая регрессия // журнал «Квантиль», 2009.
5. Расин, Джеффри Непараметрическая эконометрика: вводный курс // журнал «Квантиль», 2008.

**ИССЛЕДОВАНИЕ ТАРИФНЫХ СТАВОК ПРИ СТРАХОВАНИИ АВТОТРАНСПОРТА: СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ**

**Е.В. Жукова, Г.М. Кошкин**

*Томский государственный университет*  
water\_melon93@mail.ru, kgm@mail.tsu.ru

**Введение**

Страхование — одно из наиболее гибких и стабильных видов деятельности по защите интересов субъекта, нарушаемых в результате непредвиденных негативных явлений. Эффективность финансовой деятельности страховой компании зависит от правильного расчета нетто-премий для различных видов страхования различных категорий населения [1]. В данной работе проводится анализ влияния количества заключенных договоров и страховых выплат на изменение страховых тарифов и прибыли страховой компании в автостраховании с помощью статистического моделирования.

**1. Постановка задачи**

Объектами страхования являются автомобили, застрахованные по КАСКО. Рассмотрим два отдела страховой компании (А и В). Количество заключенных договоров и страховых случаев соответствует данным реальных отделов компании “Росгосстрах”. Другие данные, связанные со страховыми суммами и выплатами, являются конфиденциальной информацией, поэтому анализ динамики КАСКО проводится с помощью статистического моделирования. По реальным средним величинам страховых сумм и выплат, которые равны 910739 руб. и 53634 руб. соответственно [2], в работе моделируются выборки по страховым суммам и выплатам, подчиняющиеся нормальному закону. Задача состоит в анализе динамики прибыли страховой компании при изменении различных факторов.

## 2. Основные понятия

Пусть  $n$  и  $m$  — количество договоров и выплат соответственно. Тогда вероятность наступления страхового случая по одному договору страхования есть  $q = \frac{m}{n}$ . Далее, пусть  $S_i$  — страховая сумма при заключении  $i$ -го договора страхования,  $Sb_i$  — возмещение, выплачиваемое по  $i$ -му договору ( $i = \overline{1, n}$ ),  $V_i = \frac{Sb_i}{S_i}$  — относительная величина страхового возмещения по  $i$ -му договору, если состоялся страховой случай.

Введем индикатор  $J_i$ :  $J_i = 1$ , если произошел страховой случай по  $i$ -му договору,  $J_i = 0$  — в противном случае. Пусть  $T_H$  — нетто-ставка. Рассмотрим случайные величины  $z_i$  ( $i = \overline{1, n}$ ):  $z_i = (V_i J_i - T_H) S_i$ . Рассчитаем ставку  $T_H$  из условия финансовой устойчивости компании с заданным уровнем надежности  $\gamma$  [3]:

$$P \left\{ \sum_{i=1}^n (V_i J_i - T_H) S_i < 0 \right\} \geq \gamma. \quad (1)$$

Согласно методике Росстрахнадзора, тарифная ставка определяется формулой [4]

$$T_H = \beta_1 + \alpha_\gamma \sqrt{\frac{\alpha_2(S)(\beta_2 - \beta_{1,2})}{n\alpha_{1,2}(S) - \alpha_{\gamma,2}\mu_2(S)}}, \quad (2)$$

где  $S_i$  — страховая сумма при заключении  $i$ -го договора страхования (денежная сумма, на которую компания страхует автомобиль),  $Sb_i$  — возмещение, выплачиваемое по  $i$ -му договору при условии, что состоялся страховой случай,

$$\beta_i = \frac{q \cdot \alpha_i(Sb)}{\alpha_i(S)}, \quad (3)$$

$$\alpha_k(S) = \frac{1}{n} \sum_{i=1}^n (S_i)^k \quad (4)$$

— оценка  $k$ -го начального момента заключаемой страховой суммы,

$$\alpha_k(Sb) = \frac{1}{S} \sum_{i=1}^s (Sb_i)^k \quad (5)$$

— оценка начального момента  $k$ -го порядка страхового возмещения,

$$\mu_2(S) = \alpha_2(S) - \alpha_1^2(S) = D(S) \quad (6)$$

— соответствующая оценка дисперсии, которая позволяет найти  $\sigma(S) = \sqrt{\mu_2(S)}$  и характеризует разброс страховых сумм от среднего значения,  $\Phi^{-1}(\gamma) = \alpha_\gamma$  — квантиль уровня  $\gamma$ ,  $\Phi^{-1}$  — обратная функция стандартного нормального распределения. Возьмем заданный уровень безопасности  $\gamma = 0,95$  [5].

## 3. Результаты расчетов

По смоделированным выборкам найдем вероятности наступления страхового случая  $q$ ,  $\alpha_1$  и  $\alpha_2$  — моменты страховой суммы и страхового возмещения, которые приведены в табл. 1.

Таблица 1

Вероятности наступления страхового случая и моменты в отделах А и В

	$q$	$\alpha_1(S)$ , руб	$\alpha_2(S)$ , руб	$\alpha_1(Sb)$ , руб	$\alpha_2(Sb)$ , руб	$\mu_2(S)$ , руб
<b>А</b>	0,022	867432	$7,83 \cdot 10^{11}$	524411	$3,41 \cdot 10^{11}$	30850646599
<b>В</b>	0,038	681147	$4,95 \cdot 10^{11}$	339310	$1,48 \cdot 10^{11}$	31347355673

Из табл. 1 следует, что у отдела В вероятность наступления страхового случая выше. В отделе А в среднем застрахованы более дорогие автомобили, чем в отделе В, и выплаты в нем соответственно выше.

Таблица 2

Тарифные ставки отделов А и В

Отдел	$T_H$
А	0,023
В	0,036

Таким образом, в отделе В ставка принимает большее значение, что объясняется большей вероятностью наступления страхового случая. Отметим, что в отделе А договоры страхования доступнее для клиентов, и можно предположить, что по этой причине в данном отделе большее число страхователей.

Для определения прибыли необходимо знать общие сборы каждого из отделов, рассчитываемые следующим образом:

$$M = \sum_{i=0}^{n-1} T(Sb) \cdot S_i, \quad (7)$$

где  $S_i$  — страховая сумма при заключении  $i$ -го договора страхования, а  $T(Sb)$  — полученная ранее нетто-ставка (табл. 2).

Далее, для отделов были рассчитаны сборы исходя из полученных ранее нетто-ставок (табл. 3).

Таблица 3

Сборы и прибыль отделов А и В (в рубл.)

А	В
<b>Сборы, М</b>	
12333973	6915605
<b>Выплаты, V</b>	
7503989	3773924
<b>Прибыль, P</b>	
4829984	3141681

Следовательно, прибыль у отдела А составляет 41% от сборов, у отдела В — 45%.

#### 4. Динамика прибыли компании

Проверим, как будет меняться прибыль компании при изменении количества заключенных договоров. Для этого добавляем в выборки новые данные по договорам, при этом, для того чтобы вероятность наступления страхового случая оставалась фиксированной, также добавляем новые страховые выплаты.

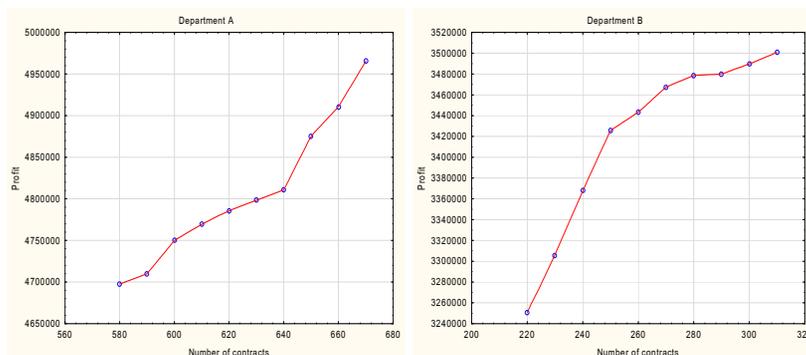


Рис. 1. Зависимость прибыли компании от количества заключенных договоров

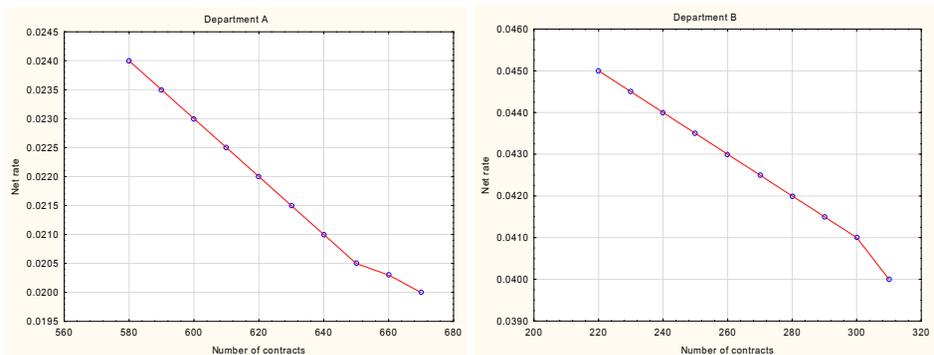


Рис. 2. Динамика тарифов страхования при изменении количества договоров

Как видим, прибыль компании увеличивается за счет роста числа страхователей, не смотря на снижение нетто-ставки тарифов.

Проанализируем, как будет меняться прибыль компании при изменении количества страховых случаев при фиксированном числе договоров (т.е. вероятности наступления страхового случая также изменяются). Для этого добавляем в выборки новые данные по выплатам.

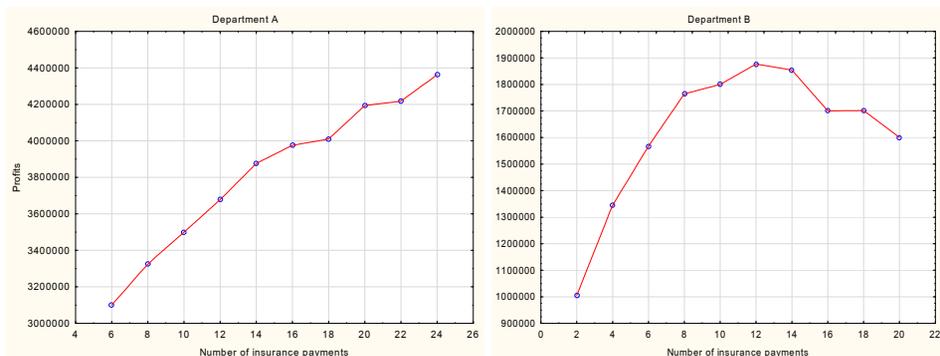


Рис. 3. Зависимость прибыли компании от количества страховых выплат

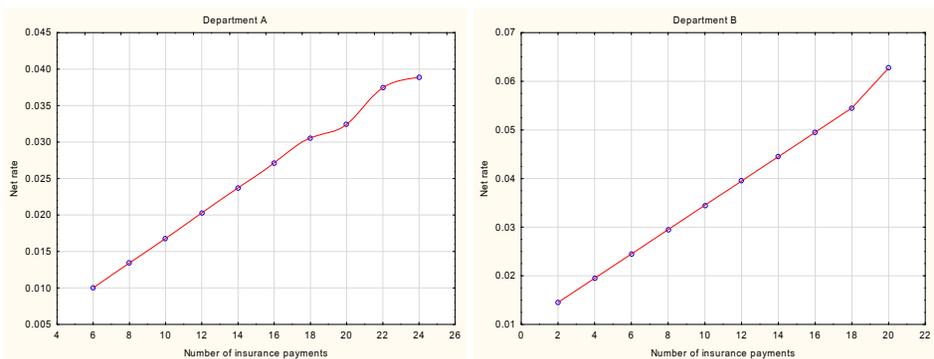


Рис. 4. Динамика тарифов страхования при изменении количества страховых случаев

Из рис. 4 следует, что прибыль отдела А растет. Это происходит вследствие того, что вероятность наступления страхового случая увеличивается, нетто-ставка возрастает, следовательно, возрастают и страховые взносы. Увеличение прибыли происходит также ввиду того, что значительных страховых выплат в отделе А не было. Однако в отделе В прибыль, наоборот, упала. Это объясняется тем, что в отделе В было несколько значительных выплат.

## Заключение

В настоящей работе на основе статистического моделирования рассчитаны нетто-ставки для двух отделов. Также проведен анализ динамики прибыли страховой компании при изменении различных факторов. Результаты анализа представлены графически. Можно сделать вывод о том, что в отделе А более выгодные условия как для страхователя, так и для страховщика, так как услуги страхования в нем дешевле, чем в отделе В, и прибыль для страховщика выше.

## ЛИТЕРАТУРА

1. Кошкин Г.М. Введение в математику страхования жизни. – Томск: Томский государственный университет, 2004. – 112 с.
2. Морозов А.В., Нефедов П.В. Динамика рынка страхования КАСКО [Электронный ресурс]//URL:<http://narinfo.ru/page/id/30868> (дата обращения: 14.05.2016).
3. Бирючев О.И. Франшиза как один из методов оптимизации расходов на рынке страхования.// Финансы, 2001, №10, с. 51–55.
4. Фалин Г.И., Фалин А.И. Введение в актуарную математику. – Москва: Финансовый актуарный центр МГУ им. М.В. Ломоносова, 1994. – 153 с.
5. Евсеева О. Комментарий к методикам расчета тарифных ставок по рисковым видам страхования.// Страхование дело. – 1993, № 8, с. 17–19.

## МЕТОД ЦЕНООБРАЗОВАНИЯ PSM ДЛЯ СЛУЧАЯ ЦЕНЗУРИРОВАННЫХ ВЫБОРОК МАЛОГО ОБЪЕМА

**Е.С. Журко, Ж.Н. Зенкова**

*Томский государственный университет*  
helena1993@mail.ru, thankoff@fpmk.tsul.ru

Цена и ценовая политика для предприятия — существенный элемент маркетинговой деятельности. Именно поэтому разработке ценовой стратегии и цен должно уделяться самое пристальное внимание со стороны руководства любого предприятия, желающего наиболее эффективно и долговременно развивать свою деятельность на рынке, т.к. любой ложный или недостаточно продуманный шаг немедленно отражается на динамике продаж и рентабельности. Также цена, как правило, оказывает самое значимое влияние на принятие потребителем решения о покупке, а, следовательно, на достижение фирмой стратегических целей и на объемы получаемой прибыли. При этом только одного анализа затрат чаще всего бывает недостаточно, т.к. подобный подход не учитывает существующие у потребителей предпочтения относительно выпускаемого фирмой товара.

Существует несколько методов ценообразования, ориентированных на потребительские предпочтения, одним из наиболее распространенных, простых в использовании и доступных является метод Price Sensitivity Meter (PSM). Данный метод также используется при определении цены на товар-новинку, для которого ещё не существует статистики продаж.

PSM-метод базируется на ценовых предпочтениях представителей целевой аудитории, зафиксированных после того, как товар был ими предварительно опробован, использован или протестирован на протяжении некоторого периода времени. Естественно полагать, что подобное маркетинговое исследование приводит к большим затратам. Поэтому для снижения стоимости проведения опроса необходимо свести число опрошиваемых респондентов к минимуму. Также на практике потребители нередко затрудняются дать точные ответы на вопросы, задаваемые им в рамках исследования, что приводит к появлению неполных, цензурированных данных, а значит, к необходимости модификации метода PSM.

Суть метода PSM [2] заключается в следующем. Респондентам предлагается ответить на четыре основных вопроса:

- 1) При какой цене  $X_1$  товар кажется Вам настолько дешевым, что возникают сомнения относительно его качества?
- 2) При какой цене  $X_2$  товар кажется Вам подходящей для совершения покупки?
- 3) При какой цене  $X_3$  Вы сочтете товар дорогим, но всё же купите?
- 4) Начиная с какой цены  $X_4$  товар покажется Вам настолько дорогим, что Вы не станете его покупать?

В результате полученных ответов формируется случайная выборка  $(X_{i1}, X_{i2}, X_{i3}, X_{i4})$ ,  $i = \overline{1, N}$ ,  $i$ -й элемент которой представляет собой четыре уровня цен, указанных  $i$ -м респондентом. По каждому вектору  $\{X_{ij}\}$ ,  $j = \overline{1, 4}$ , строится эмпирическая функция распределения (э.ф.р.) по формуле:

$$F_N^{(j)} = \frac{1}{N} \sum_{i=1}^N I_{[0, x]}(X_{ij}), \quad (1)$$

где

$$I_{[0, x]}(y) = \begin{cases} 0, & \text{если } y \notin [0, x]; \\ 1, & \text{если } y \in [0, x] \end{cases}$$

— индикаторная функция. Для  $j = 1, 2$  строится оценка функции выживания

$$S_N^{(j)}(x) = 1 - F_N^{(j)}(x), \quad (2)$$

далее эти функции, которые называются кривыми ценовой чувствительности, отображаются на одном графике (рис. 1). В качестве рекомендуемого диапазона цен рассматривается отрезок от оптимального до ожидаемого уровня.

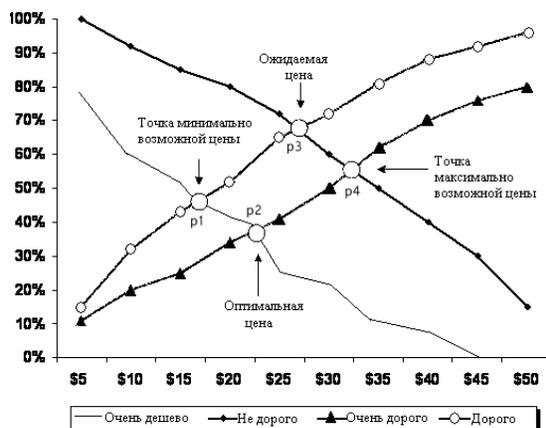


Рис. 1. Метод PSM

Изложенный метод базируется на полных данных, когда респонденты готовы дать точные значения цен в ответах на заданные вопросы. На практике же часто возникают ситуации, когда потребитель затрудняется дать точные ответы на поставленные вопросы, а может указать лишь некоторые интервалы. В данном случае в массиве данных появляются неполные, цензурированные значения [5]. Для выборки, включающей такие данные, применение формулы (1) недопустимо, в данном случае вместо э.ф.р. может быть использована оценка Каплана-Мейера [5], которая для случая однократного цензурирования справа имеет вид:

$$F_N^{(j)-CR}(t) = \left\{ \begin{array}{ll} \frac{1}{N} \sum_{i=1}^N I_{[0,t)}(X_i) \bar{I}_i, & 0 \leq t \leq T_1, \\ \frac{r}{N} + \frac{1}{(1-g)N} \sum_{i=1}^N I_{[T_1,t)}(X_i) \bar{I}_i, & (N-r)(1-g) > 0; \\ \frac{r}{N}, & (N-r)(1-g) = 0 \end{array} \right\} T_1 < t \leq T, \quad (3)$$

где  $(X, I) = \{(X_1, I_1), \dots, (X_N, I_N)\}$  — цензурированная выборка (ц.в.) объёма  $N$ , построенная по следующей схеме: количество неполных значений в интервале  $(T_1, T]$  — с.в., численно равная доле  $g$ ,  $0 < g < 1$ , от числа респондентов, выбравших при анкетировании уровень цены, больший, чем значение цензуры  $T_1$ , для  $i = \overline{1, N}$ ,

$$I_i = \begin{cases} 0, & X_i - \text{полное наблюдение;} \\ 1, & X_i - \text{наблюдение до цензурирования, } X_i = T_1, \end{cases}$$

$\bar{I}_i = 1 - I_i$ , где  $r$  — число полных значений в интервале  $[0, T_1]$ .

Оценка (3) является асимптотически несмещенной, непараметрической, что было показано в работе [3], при этом

$$\lim_{N \rightarrow \infty} NDF_N^{CR}(t) = F(t)(1 - F(t)) + \frac{g(F(t) - p)(1 - F(t))}{(1 - p)(1 - g)} I_{(T_1, T)}(t),$$

где  $p = F(T_1) \in (0, 1)$ ,  $DF_N^{CR}(t)$  — дисперсия оценки  $F_N^{CR}(t)$ . Таким образом, при  $g = 0$  дисперсия оценки совпадает с дисперсией э.ф.р., которая определяется формулой [2]

$$DF_N(x) = \frac{F(x)(1 - F(x))}{N}.$$

Очевидно, что при появлении цензурирования дисперсия оценки растёт, причем с ростом доли цензурирования и расширением области неопределенности дисперсия увеличивается значительно, а, следовательно, происходит существенная потеря точности оценивания.

На практике часто используются опросы малого числа потребителей, так как сбор информации при большом числе респондентов зачастую труднореализуем и затратен, поэтому возникает необходимость апробирования метода PSM при малом объёме выборки. В данной работе этот вопрос решался с помощью имитационного моделирования.

Рассматривалось однократно I типа справа прогрессивное цензурирование. В качестве оценок кривых ценовой чувствительности использовалась оценка (3), третий вектор значений подвергался либо правому цензурированию, либо оставался неизменным. Для оценки  $F_N^{(3)}(x)$  использовалась формула (3), а для не цензурированных данных  $(F_N^{(4)}(x), S_N^{(1)}(x), S_N^{(2)}(x))$  — формула (1).

В процессе имитационного моделирования фиксировалось значение  $T_1 = 0.3$ , параметр моделирования  $M = 10000$ , при этом изменялся объём выборки  $N$  ( $N = 6, 7, \dots, 12$ ) и доля  $g$  ( $g = 0.1, 0.2, \dots, 0.9$ ). Далее также увеличивалось и  $T_1$ . Выборки генерировались из равномерного распределения с разными параметрами: для первой выборки —  $R_{(0,1)}(x)$ , для второй —  $R_{(0.05, 1.05)}(x)$ , третья и четвёртая — соответственно из  $R_{(0.25, 1.25)}(x)$  и  $R_{(0.4, 1.4)}(x)$ . Значение точек пересечения рассматриваемых ф.р. (истинные значения искомым ценовым диапазонов):  $p_1 = 0.625$ ;  $p_2 = 0.7$ ;  $p_3 = 0.65$ ;

$p_4 = 0.725$ . На каждом шаге  $i = 1, \dots, M$  был получен набор цен  $\{\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4\}_i$ , который впоследствии усреднялся. Так как цензурирование никак не изменяет значения цен  $\hat{p}_2$  и  $\hat{p}_4$ , то исследовались только смещения оценок  $\hat{p}_1$  и  $\hat{p}_3$  от истинных значений цен. Зависимость усреднённых наборов оценок  $\hat{p}_1$  и  $\hat{p}_3$  при  $T_1 = 0.3$  и  $g = 0.3$  от  $N$  изображена на рис. 2. Отметим, что наблюдается существенное смещение кривых, полученных из цензурированной выборки от кривых, где цензурированные данные отсутствуют. При этом при увеличении  $N$  качество результатов улучшается.

СКО значений усредненных цен для малых  $N$  представлено на рис. 3. Для выявления наименьшего допустимого значения объёма выборки исследовались эмпирическое среднеквадратическое отклонение СКО полученных наборов значений точек пересечений. Очевидно, что СКО точки  $p_1$  больше при левом цензурировании, чем при правом, а СКО точки  $p_3$  — при правом, чем при левом (рис. 4). Данный факт наблюдается и при изменении значения цензуры  $T_1$ .

Результат моделирования показал, что при  $N = 10$  значение СКО является сравнительно небольшим и несущественно отличается от СКО, полученного при дальнейшем увеличении  $N$ . Поведение СКО точки  $p_1$  при различных объёмах выборки в зависимости от доли цензурирования  $g$  при  $T_1 = 0.7$  представлено на рис. 5.

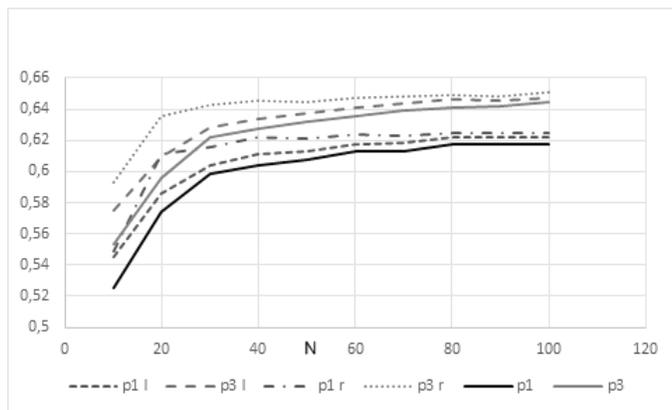


Рис. 2. Зависимость рекомендуемых цен от объёма выборки для  $T_1 = 0.3$ ,  $g = 0.3$

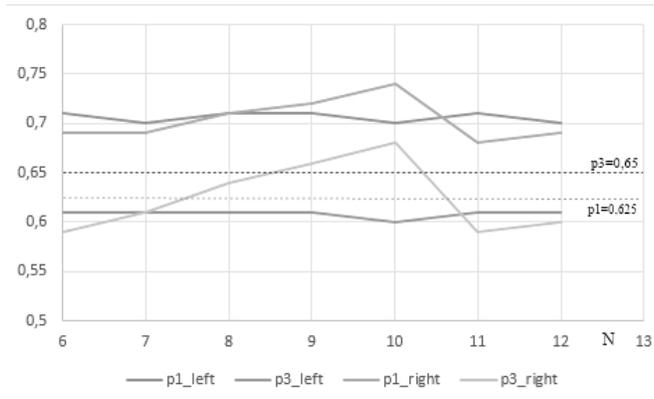


Рис. 3. Зависимость рекомендуемых цен от объёма выборки для  $T_1 = 0.7$ ,  $g = 0.8$

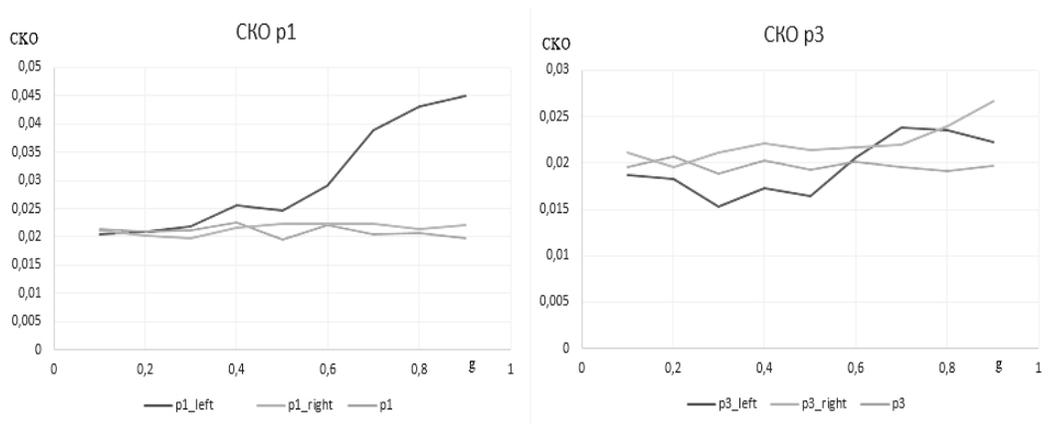


Рис. 4. Графики SKO цен  $p_1$  и  $p_3$  при левом и правом цензурировании

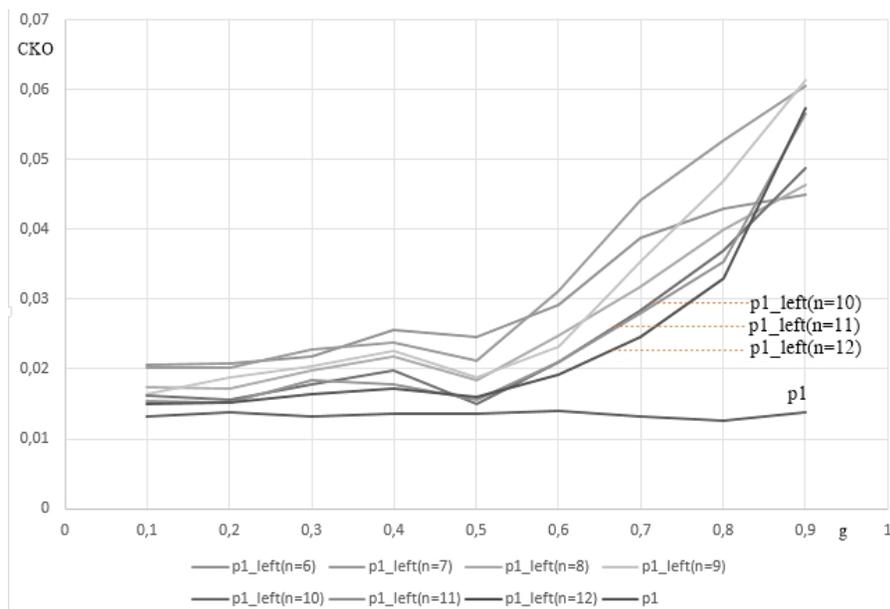


Рис. 5. Графики SKO цены  $p_1$  при левом цензурировании в зависимости от  $g$

Таким образом, путём имитационного моделирования было показано, что при малом количестве опрошенных респондентов возникает искажение результатов метода PSM, что в особенности относится к точке минимально возможной цены  $p_1$ , которая оказывается сильно завышенной относительно своего истинного значения (примерно на 12%), при этом точка  $p_3$  оказывается заниженной приблизительно на 8%.

## ЛИТЕРАТУРА

1. Боровков А.А. Математическая статистика. Новосибирск: Наука; Изд-во Института математики, 1997. – 772 с.
2. Зенкова Ж.Н., Краковецкая И.В. Моделирование по неполным данным в логистике и маркетинге/ Логистические системы в глобальной экономике: материалы Междунар. науч.-практ. конф.: в 2 ч. Ч. 1. Научно-исследовательский сектор / Сиб. гос. аэрокосмич. ун-т. – Красноярск, 2013. – 332 с. С. 98–105.
3. Анализ надежности технических систем по цензурированным выборкам / В.М. Скрипник, А.Е. Назин, Ю.Г. Благовещенский. М.: Радио и связь, 1988. – 184 с.
4. Статистические задачи для случайных подстановок с цензурированными данными / Г.И. Ивченко, М. В. Солдаткина. – СПб. : Питер, 2012. – 104 с.

# ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА

А.А. Замщикова, Ж.Н. Зенкова  
Томский государственный университет  
zamashikova.n@mail.ru, thankoff@fpmk.tsu.ru

## Введение

В современных логистике и маркетинге широко используются различные методы прогнозирования временных рядов. Связанно это с тем, что на основе полученных прогнозов принимаются различные управленческие решения при планировании и оптимизации производства, поставок, запасов, распределения продукции и т.д.

В данной работе рассматривается метод Сингулярного Спектрального Анализа SSA (Singular Spectrum Analysis), также известный как метод «Гусеница». Метод SSA появился в 80-х годах 20 века, он основан на преобразовании одномерного временного ряда в многомерный ряд с последующим применением к полученному многомерному временному ряду метода главных компонент.

Метод SSA позволяет выделять тренд, обнаруживать периодичность, корректировать сезонность, сглаживать, подавлять шум и многое другое, он не требует стационарности ряда, знания модели тренда, а также сведений о наличии в ряде периодических составляющих и их периодах.

## Алгоритм прогнозирования методом «Гусеница»

Пусть имеется исходный вещественный временной ряд  $x = (x_i)_{i=\overline{1,N}}$  длины  $N$ . Предположим, что ряд  $x$  — ненулевой (т.е. всегда существует хотя бы один номер  $i$ , для которого  $x_i \neq 0$ ) [1].

### Первый этап алгоритма: Разложение

**Шаг 1: Вложение.** Исходный временной ряд преобразуется в последовательность многомерных векторов. Для этого задается целое число  $L$  — длина окна или длина «гусеницы», при этом  $1 < L < N$  и  $L \leq \frac{N}{2}$ . Далее строятся  $K = N - L + 1$  векторов вложения

$X_i = (x_i, \dots, x_{i+L-1})$ ,  $i = \overline{1, K}$ , затем —  $L$ -траекторная матрица  $X$ , состоящая из векторов  $X_i$  в качестве столбцов:

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_L & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{L+1} & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{L+2} & \dots & x_{K+2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_{2L-1} & \dots & x_N \end{pmatrix}, \quad (1)$$

где  $K = N - L + 1$ . Полученная матрица является ганкелевой.

**Шаг 2: Сингулярное разложение.** Вычисляется матрица

$$S = \frac{1}{K} X \cdot X^T, \quad (2)$$

затем рассматриваются  $\lambda_1, \lambda_2, \dots, \lambda_L$  — собственные числа матрицы  $S$ ,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$ , и  $U_1, U_2, \dots, U_L$  — ортонормированная система собственных векторов матрицы  $S$ , соответствующих этим собственным числам, при этом

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_L \end{pmatrix}, U = (U^1, U^2, \dots, U^L) = \begin{pmatrix} u_1^1 & u_1^2 & \dots & u_1^L \\ u_2^1 & u_2^2 & \dots & u_2^L \\ \dots & \dots & \dots & \dots \\ u_L^1 & u_L^2 & \dots & u_L^L \end{pmatrix}.$$

Тогда матрица  $S$  может быть представлена в виде:

$$S = U \Lambda U^T. \quad (3)$$

**Шаг 3: Переход к главным компонентам.** Здесь матрица  $U$  представляется как матрица перехода к главным компонентам исходного ряда:

$$G = U^T X = (G_1, G_2, \dots, G_L), \quad (4)$$

где величины  $G_i = (U^i)^T X$ ,  $i = 1, 2, \dots, L$ , называются главными компонентами исходного вектора.

**Второй этап: Восстановление ряда по выбранным главным компонентам**

Зная главные компоненты, можно полностью восстановить траекторную матрицу следующим образом:

$$X = \begin{pmatrix} U^1 & U^2 & \dots & U^L \end{pmatrix} \cdot \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_L \end{pmatrix} = \sum_{i=1}^L U^i G_i.$$

При этом, зная траекторную матрицу, можно всегда перейти к искомому временному ряду  $x = (x_i)_{i=1, \dots, N}$ .

Рассмотрим первые  $r$  из  $L$  компонент собственных векторов. Введем обозначения:

$$\tilde{U} = \begin{pmatrix} U^1 & U^2 & \dots & U^L \end{pmatrix}, \tilde{G} = \tilde{U}^T X.$$

Тогда восстановленная траекторная матрица определяется формулой:

$$\tilde{X} = \tilde{U}^T \tilde{G} = \sum_{i=1}^r U^i G_i. \quad (5)$$

Она имеет размерность  $L \times K$ . Матрица  $\tilde{X}$  позволяет получить восстановленный исходный временной ряд

$$\tilde{x}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{X}_{s, s-i+1}, & 1 \leq s \leq L, \\ \frac{1}{L} \sum_{i=1}^L \tilde{X}_{i, s-i+1}, & L \leq s \leq K, \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{X}_{i+s-K, K-i+1}, & K \leq s \leq N, \end{cases} \quad (6)$$

который называется *SSA-сглаживанием исходного временного ряда*  $(x_i)_{i=1, \dots, N}$  по первым  $r$  компонентам из  $L$ .

**Третий этап: Прогнозирование**

На этом этапе нам необходимо построить числовой ряд  $x = (x_i)_{i=N+1, M}$ , который называется прогнозом исходного временного ряда  $x = (x_i)_{i=1, N}$  на  $M$  шагов вперёд. Для этого рассмотрим некоторый набор главных компонент  $G_i$ ,  $i = 1, 2, \dots, L$ , обозначим:

$$U_* = \begin{pmatrix} u_1^1 & u_1^2 & \dots & u_1^r \\ u_2^1 & u_2^2 & \dots & u_2^r \\ \dots & \dots & \dots & \dots \\ u_{L-1}^1 & u_{L-1}^2 & \dots & u_{L-1}^r \end{pmatrix}, Q = \begin{pmatrix} X_{N-L+2} \\ X_{N-L+3} \\ \dots \\ X_N \end{pmatrix}, \tilde{u} = (u_L^1, u_L^2, \dots, u_L^r).$$

Тогда прогноз определяется как

$$x_{N+1} = \tilde{u} (U_*^T U_*)^{-1} U_*^T Q. \quad (7)$$

Прогнозирование следующих значений может быть реализовано двумя способами. В основе первого способа лежит изменение соответствующим образом матрицы  $Q$ . Второй способ заключается в полном или частичном повторе алгоритма метода SSA. В этом случае, помимо матрицы  $Q$ , будут также меняться матрицы  $\tilde{u}$  и  $U_*$ .

### Прогнозирование стоимости цен на нефть методом SSA

Метод SSA был опробован для прогнозирования цен на нефть за период 03.04.2014 – 12.05.2016. Прогноз строился на период 03.05.2016 – 12.05.2016, итоговые прогнозные значения сравнивались с реальными.

При выборе входных параметров использовались два подхода.

1) Несмотря на то, что строгих правил при выборе длины гусеницы  $L$  нет, существуют некоторые рекомендации [3], например, если взять  $L$  слишком маленькой, то произойдёт смешивание содержательных частей временного ряда. При этом мы знаем, что  $L$  не может быть больше, чем  $N/2$  (так как SVD-разложения ряда размерности  $N$  при выборе длины гусеницы  $L$  и  $N-L+1$  эквивалентны), и чем больше  $L$ , тем более подробным будет разбиение ряда. Из этих соображений и в зависимости от поведения графика значений реального и сглаженного ряда была выбрана длина гусеницы  $L = 34$ .

При выборе  $r$  рассматривался тот факт, что собственные числа  $\lambda_1, \lambda_2, \dots, \lambda_L$  играют роль доли информационного содержания главных компонент от общего информационного содержания исходного временного ряда. По рис. 1 и 2 можно увидеть, что содержательный смысл несут, в основном, первые две главные компоненты, а оставшиеся, вероятно, шум. Следовательно, зафиксируем  $r = 2$ .

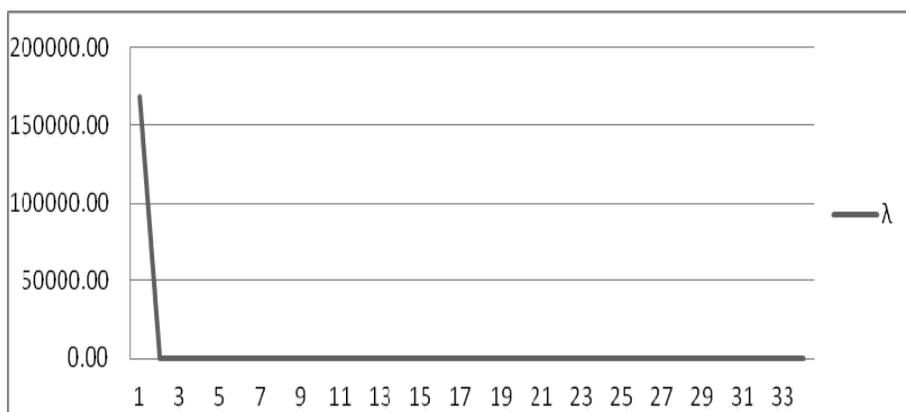


Рис. 1. График первых 34 собственных чисел ковариационной матрицы

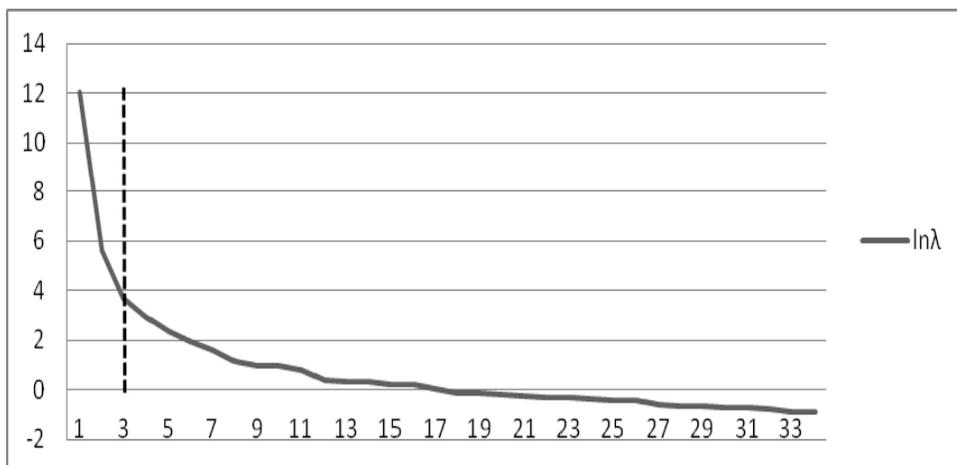


Рис. 2. График логарифма первых 34 собственных чисел ковариационной матрицы

2) Второй подход был выбран из соображений минимума отклонения между модельными и реальными значениями. Была разработана программа, которая подбирает такие пары  $r$  и  $L$  на каждом шаге прогнозирования, которые удовлетворяют вышеуказанному правилу. При этом ошибка прогнозирования, считается по формуле:

$$\varepsilon = \frac{1}{M} \sum_i \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\%, \quad (7)$$

где  $M$  — число шагов прогнозирования,  $x_i$  — реальные цены, а  $\hat{x}_i$  — прогнозные значения. Результаты моделирования при оптимальных параметрах представлены на рис. 3.



Рис. 3. Значение реальных и модельных цен на нефть за период 03.04.2014 – 12.05.2016, в долларах за баррель



Рис. 4. Значение реальных и модельных цен на нефть за период 01.04.2016 – 01.05.2016, в долларах за баррель



Рис. 5. Значение реальных и спрогнозированных цен на нефть за период 03.05.2016 – 12.05.2016, в долларах за баррель

Заметим, что ошибка прогнозирования при первом подходе составила  $\varepsilon_1 = 4.629\%$ , что меньше 5%, следовательно, можно говорить о высоком качестве модели, однако при использовании оптимальных параметров  $L$  и  $r$  ошибка составила  $\varepsilon_2 = 0.162\%$ , что в 28.7 раз меньше  $\varepsilon_1$ . Следовательно, при использовании метода гусеницы имеет смысл предварительно оптимизировать параметры  $L$  и  $r$  в смысле минимизации отклонений реальных и модельных данных.

### Заключение

В данной работе был описан метод SSA «Гусеница», алгоритм применялся к реальным данным о значениях цен на нефть за период 03.04.2014 – 12.05.2016. Было реализовано сглаживание и прогнозирование временного ряда, произведено исследование зависимости качества прогнозирования от выбора входных параметров. В результате выбран набор параметров алгоритма, дающий очень маленькую ошибку прогнозирования. Процесс подбора параметров был полностью автоматизирован.

### ЛИТЕРАТУРА

1. Голяндина Е.Э. Метод «Гусеница»-SSA: прогноз временных рядов: Учеб. пособие. – СПб., 2004. – 52 с.
2. Солнцев В.Н., Данилов Д.Л., Жиглявский А.А. Главные компоненты временных рядов: метод «Гусеница», С.-Петербургский государственный университет, 1997. 307 с. <http://www.gistatgroup.com/gus/>

3. Леонтьева Л.Н. Многомерная гусеница, выбор длины и числа компонент // Машинное обучение и анализ данных. – 2011. – № 1. – С. 2–10.

4. Нефть Brent — котировки за прошедший период [Электронный ресурс] – Электрон. дан. — URL: <http://ru.investing.com/commodities/brent-oil-historical-data>

## **ВЛИЯНИЕ ВИДА РАСПРЕДЕЛЕНИЯ СТОИМОСТИ ОБОРОТНЫХ СРЕДСТВ НА ПОКАЗАТЕЛИ ОБОРАЧИВАЕМОСТИ ПО ИНТЕРВАЛЬНЫМ ДАННЫМ**

**Ж.Н. Зенкова, О.Б. Макеева**

*Томский государственный университет*  
oksi-mak-tsu@mail.ru, thankoff@fpmk.tsu.ru

В работе исследуется проблема влияния вида распределения стоимости вложений в оборотные средства предприятия на коэффициент оборачиваемости и оборот. Задача обобщается на случай интервально-цензурированных данных о размерах вложений. Рассматривается случай с использованием дополнительной информации о квантиле функции распределения. С помощью имитационного моделирования показано, что учёт информации о квантиле позволяет улучшить качество рассматриваемых статистических процедур в смысле снижения среднеквадратического отклонения и восполнить потери информации, возникающие при цензурировании, независимо от вида распределения.

### **Введение**

На сегодняшний день одной из актуальных проблем при оценке эффективности управления предприятием является анализ и прогноз стоимости его оборотных активов [1-4]. Их эффективное использование зависит от того, насколько точно определена потребность в оборотных средствах, при этом и в случае занижения, и в случае завышения величины вложений в оборотный капитал возникают риски, чреватые отрицательными последствиями для организации. Однако специфика исходных данных изначально не позволяет получить адекватную оценку среднего размера вложений, так как фактически аналитики работают со значениями моментного ряда, отображающими текущую стоимость активов, которая может существенно меняться даже в течение одного дня, при этом подобные колебания могут не фиксироваться, а в традиционных методиках расчетов никак не учитываются. В работах [3,4] были предложены новые подходы к оцениванию показателей оборачиваемости, когда исходная выборка рассматривалась как цензурированная, в [4-6] исследовался вопрос привлечения дополнительной информации при оценивании стоимости вложений. Заметим, что привлечение дополнительной информации приводит к существенному улучшению качества статистических процедур.

В данной работе с помощью имитационного моделирования исследовано влияние вида распределения стоимости вложений в оборотные средства на показатели оборачиваемости по интервальным данным с учётом априорной информации о квантиле функции распределения (ф.р.).

### **1. Непараметрический алгоритм Тёрнбулла и его модификация с учётом известного квантиля в оценивании показателей оборачиваемости**

Пусть стоимость оборотных средств предприятия  $\tau > 0$  есть случайная величина (с.в.) с ф.р.  $F(t)$ . Исходная выборка объёма  $N$  состоит из интервалов  $(L_i, R_i]$ ,  $i = \overline{1, N}$ , т.е. известно, что  $i$ -ое наблюдение попало в промежуток  $(L_i, R_i]$ , но его точное значение не известно. Для оценивания  $F(t)$  рассмотрим разбиение  $0 < \tau_0 < \tau_1 < \dots < \tau_m$ , состоя-

шее из всех неповторяющихся упорядоченных границ интервалов  $L_i$  и  $R_i$ ,  $i = \overline{1, N}$ . Для  $j = \overline{1, m}$  и  $i = \overline{1, N}$  определяем вес [7,8]

$$\alpha_{ij} = \begin{cases} 1, & \text{если } (\tau_{j-1}, \tau_j) \subseteq (L_i, R_i], \\ 0, & \text{если } (\tau_{j-1}, \tau_j) \not\subseteq (L_i, R_i]. \end{cases}$$

Для произвольного шага  $k \geq 0$ :

1. Вычисляем вероятность попадания с.в.  $\tau$  в интервал  $(\tau_{j-1}, \tau_j]$  по формуле  $p_j^k = S^{(k)}(\tau_{j-1}) - S^{(k)}(\tau_j)$ ,  $j = \overline{1, m}$ .

2. Находим значение  $d_j^{(k)} = \sum_{i=1}^N \left( \alpha_{ij} p_j^{(k)} / \sum_{s=1}^m \alpha_{is} p_s^{(k)} \right)$ ,  $j = \overline{1, m}$ .

3. Для  $j = \overline{1, m}$  определяем  $Y_j^{(k)} = \sum_{s=j}^m d_s^{(k)}$  и  $S^{(k+1)}(\tau_j) = \prod_{i=1}^j \frac{Y_i^{(k)} - d_i^{(k)}}{Y_i^{(k)}}$ , при этом

$S^{(k+1)}(\tau_0) = 1$ . Алгоритм повторяем до тех пор, пока для всех  $j = \overline{1, m}$  не выполнится условие  $|S^{(k+1)}(\tau_j) - S^{(k)}(\tau_j)| \leq 10^{-7}$ . Начальное значение определим как  $p_j^0 = \frac{1}{m}$ ,  $j = \overline{1, m}$ .

В итоге значение ф.р.  $F(t)$  в точке  $\tau_j$ ,  $j = \overline{0, m}$ , можно оценить как

$$F_N(\tau_j) = 1 - S^{(k+1)}(\tau_j), \quad (1)$$

при этом для  $t \in [0, \tau_0]$   $F_N(t) = 0$ . В [9] показывается состоятельность оценки (1), при этом доказано, что при определенных условиях оценка Тёрнбулла является равномерно строго состоятельной.

Найдём оценку средней стоимости запаса

$$E \tau = \int_0^{+\infty} x dF(t), \quad (2)$$

подставив в интеграл (2) оценку (1) [9]. Получим:

$$\hat{E} \tau = \int_0^{+\infty} x dF_N(t) = \sum_{i=1}^m \tau_i (F_N(\tau_i) - F_N(\tau_{i-1})). \quad (3)$$

Пусть известно, что ф.р.  $F(t)$  удовлетворяет свойству:

$$F(x_q) = q, \quad (4)$$

где  $x_q$  — квантиль уровня  $q$ , притом  $x_q$  и  $q$  известны. Тогда модифицированная с учётом (4) оценка ф.р. строится с помощью проектирования оценки (1) в априорный класс по формуле [6]:

$$F_N^q(x) = q \cdot \left( \left( \frac{F_N(x)}{F_N(x_q)} \right) \vee 0 \right) \wedge 1 + (1-q) \left( \left( \frac{F_N(x) - F_N(x_q)}{1 - F_N(x_q)} \right) \vee 0 \right) \wedge 1. \quad (5)$$

Методом подстановки найдём модифицированную оценку математического ожидания:

$$\bar{X}^q = \int_{-\infty}^{+\infty} x dF_N^q(x) = \frac{q}{N \cdot F_N^q(x_q)} \cdot \sum_{i=1}^N X_i \cdot I_{(x_i < x_q)} + \frac{(1-q)}{N(1 - F_N^q(x_q))} \cdot \sum_{i=1}^N X_i \cdot I_{(x_i \geq x_q)}. \quad (6)$$

## 2. Исследование влияния вида распределения запасов на показатели оборачиваемости с помощью моделирования

Т.к. аналитическое исследование свойств оценки Тёрнбулла является весьма проблематичным, в данной работе использовалось имитационное моделирование. Генерировалось  $M = 10000$  выборок объёма  $N$  из трёх распределений: экспоненциального, нормального, логнормального. Затем по каждому из распределений строились интервально-цензурированные выборки: каждое значение покрывалось интервалом, при этом расстояния от исходного значения до левой и правой границ моделировались из разных распределений. Далее находились следующие оценки: э.ф.р. по исходной полной выборке; модификация э.ф.р. с учётом свойства (4) с применением проектора (5); оценка Тёрнбулла по интервально-цензурированной выборке; модификация оценки Тёрнбулла с учётом свойства (4). На рис. 1–9 приведены итоговые графики поведения среднеквадратической ошибки (с.к.о.) каждой оценки при различных значениях известного квантиля для всех трёх распределений.

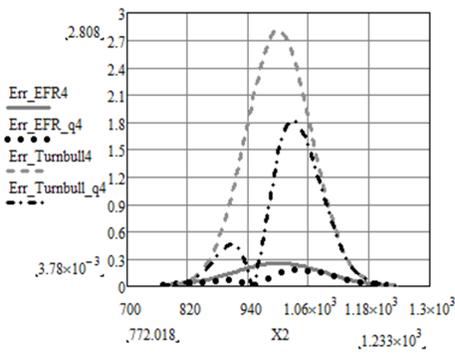


Рис. 1. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для  $F(t) = N(1000, 100)(x)$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.3$

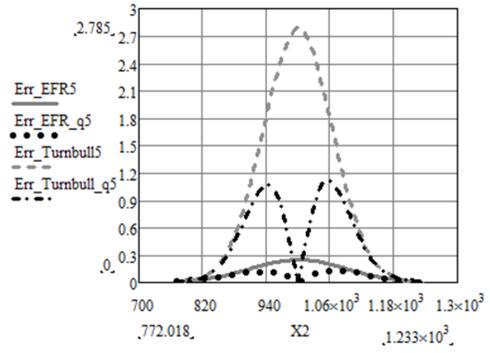


Рис. 2. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для  $F(t) = N(1000, 100)(x)$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.5$

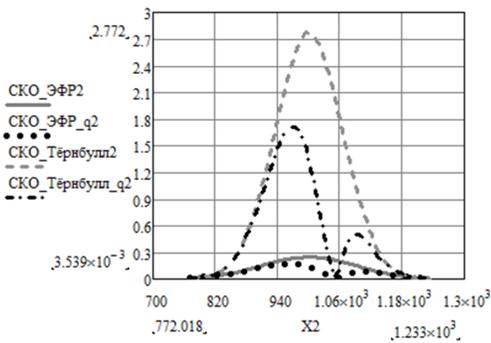


Рис. 3. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для  $F(t) = N(1000, 100)(x)$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.7$

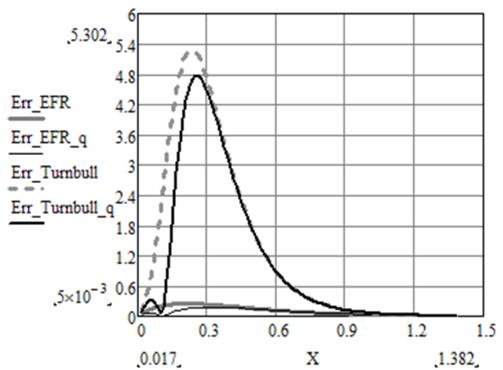


Рис. 4. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для экспоненциального распределения,  $\lambda = 0.3$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.3$

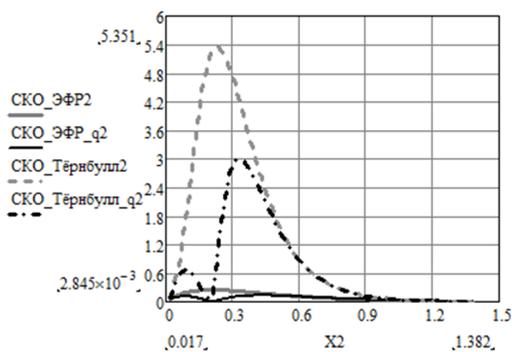


Рис. 5. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для экспоненциального распределения,  $\lambda = 0.3$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.5$

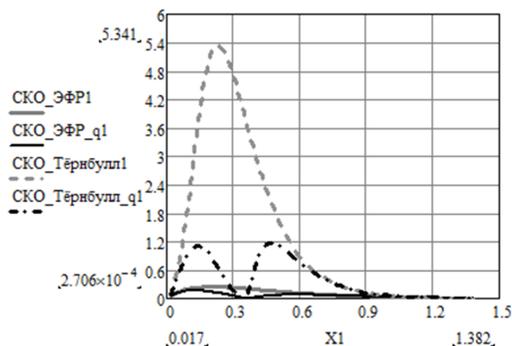


Рис. 6. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для экспоненциального распределения,  $\lambda = 0.3$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.7$

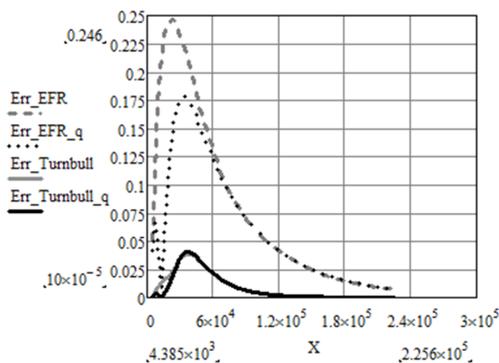


Рис. 7. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для логнормального распределения  $F(t) = \text{LogN}(10,1)(x)$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.3$

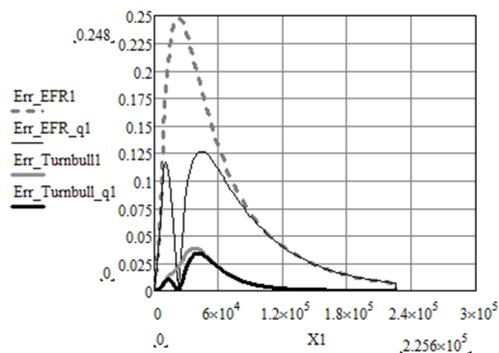


Рис. 8. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для логнормального распределения  $F(t) = \text{LogN}(10,1)(x)$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.5$

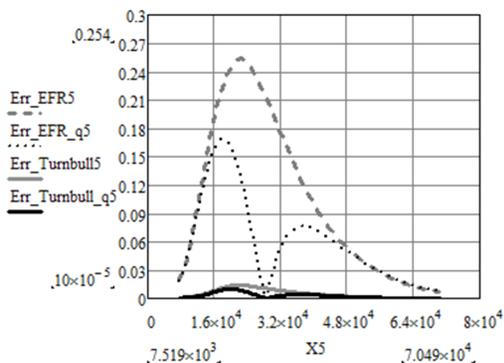


Рис. 9. С.к.о.  $F_N(t)$ ,  $F_N^q(x)$ ,  $F_N^T(x)$  и  $F_N^{Tq}(x)$  для логнормального распределения  $F(t) = \text{LogN}(10,1)(x)$ ,  $N = 50$ ,  $M = 10000$ ,  $q = 0.7$

Приведём результаты расчетов показателей оборачиваемости для нормального и логнормального распределения, которые рассчитываются по формулам (7), (8).

$$O = \frac{\text{количество дней за определенный период}}{KO \text{ за данный период}}, \quad (7)$$

где

$$KO = \frac{\text{Суммарный объем реализации за период (год, месяц)}}{\text{Размер среднего месячного (недельного, дневного) запаса}}. \quad (8)$$

При этом размер суммарного объема реализации за период положим равным 100000 тыс. руб.

Результаты вычислений для нормального и логнормального распределений приведены в табл. 1 и 2 соответственно.

Таблица 1

**Результаты вычислений показателей оборачиваемости различными методами для выборки объема  $N = 50$  из нормального закона распределения  $N(1000, 100)$ ,  $q = 0.7$ ,  $\bar{X} = 1000$ .**

Средние запасы	Значение, тыс.руб./мес.	KO, раз/год	Оборот O, дни
Арифметическое среднее	$\bar{X}_1 = 999.858$	100.0142	3.65
Среднее с учётом информации о квантиле	$\bar{X}_1^q = 999.96$	100.00	3.65
Среднее по интервально-цензурированной выборке	$\bar{X}_1^T = 1037.782$	96.359	3.79
Среднее по интервально-цензурированной выборке с учётом информации о квантиле	$\bar{X}_1^{Tq} = 1002.661$	99.723	3.66

Таблица 2

**Результаты вычислений показателей оборачиваемости различными методами для выборки объема  $N = 50$  из логнормального распределения  $\text{LogN}(10, 1)$  и  $q = 0.7$ ,  $\bar{X} = 24959, 26$**

Средние запасы	Значение, тыс.руб./мес.	KO, раз/год	Оборот O, дни
Арифметическое среднее	$\bar{X}_2 = 24661.331$	4.05	90.12
Среднее с учётом информации о квантиле	$\bar{X}_2^q = 24789.38$	4.03	90.57
Среднее по интервально-цензурированной выборке	$\bar{X}_2^T = 24086.048$	4.15	87.95
Среднее по интервально-цензурированной выборке с учётом информации о квантиле	$\bar{X}_2^{Tq} = 24825.297$	4.03	90.57

Заметим, что информация о квантиле позволяет существенно снизить значения с.к.о. как в случае полной, так и интервально-цензурированной выборки для всех исследованных распределений. При этом использование более точной оценки ф.р. при подсчёте среднего уровня запаса позволяет добиться более качественных расчетов показателей оборачиваемости.

### Заключение

Таким образом, в данной работе исследовано влияние вида распределения на среднеквадратическое отклонение оценки Тёрнбулла, с помощью имитационного моделирования показано, что привлечение дополнительной информации существенно снижает с.к.о., что, в свою очередь, позволяет получить более качественную оценку среднего уровня вложений в оборотные средства, а значит, и более точные значения показателей оборачиваемости оборотных средств предприятия.

## ЛИТЕРАТУРА

1. Бауэрсокс Д., Клосс Д. Логистика: Интегрированная цепь поставок: пер. с англ. – М.: ЗАО «Олимп-Бизнес», 2010. – 640 с.
2. Зенкова Ж.Н. Логистический подход в управлении предприятием. Учебно-методический комплекс, Томский государственный университет, 2012.
3. Макеева О.Б., Зенкова Ж.Н. Применение методов обработки цензурированных данных при анализе оборачиваемости. Вестник науки Казахского агротехнического университета им. С.Сейфуллина, №3 (82). Астана. – 2014. С. 21–30.
4. Зенкова Ж.Н., Макеева О.Б. Использование информации о квантиле при анализе оборачиваемости оборотных средств / Материалы III Всероссийской молодежной конференции «Математическое и программное обеспечение информационных, технических и экономических систем», Томск, 22-23 мая 2015. Томск: Издательский Дом Томского государственного университета, 2015. С. 82–87
5. Зенкова Ж.Н. Статистическая обработка данных с учётом симметрии распределения. – Германия: LAP LAMBERT; Academic Publishing GmbH&Co., 2011. –181 с.
6. Дмитриев Ю.Г., Устинов Ю.К. Статистическое оценивание распределений вероятностей с использованием дополнительной информации – Томск: Изд-во ТГУ, 1988. 194 с.
7. Giolo S.R. Turnbull's estimator for interval-censored data // Technical Report, August, 2004
8. Zhao, Guolin M.A. Nonparametric and Parametric Survival Analysis of Censored Data with Possible Violation of Method Assumptions, 2008, 55 p.
9. Chang M.N., Yang G.L. Strong consistency of a nonparametric estimator of the survival function with doubly censored data // The Annals of Statistics, 1987, Vol. 15, No. 4, – С. 1536–1547
10. Боровков А.А. Математическая статистика / А.А. Боровков. – Новосибирск: Наука; Изд-во Института математики, 1997. – 772 с.

## ИССЛЕДОВАНИЕ ЦЕНЫ СПРОСА НА ПРОДУКЦИЮ ЗАО «НАТУРАЛЬНОЕ МЫЛО» МЕТОДОМ ШТОЦЕЛЯ ПО ЦЕНЗУРИРОВАННЫМ ДАННЫМ

**Ж.Н. Зенкова, Н.Р. Цыбулькиова**

*Томский государственный университет*  
tsybulnikova94@gmail.com, thankoff@fpmk.tsu.ru

### Введение

При появлении нового товара или услуги и его внедрении на потребительский рынок очень важно определить цену, по которой он будет продаваться и с которой производитель будет согласен выйти на целевого потребителя [1–5]. Эта задача влечет за собой некоторый ряд проблем, ведь на новинку ещё нет накопленной статистики продаж, а есть лишь предположения и интуитивные догадки, основанные на анализе себестоимости, опросах потребителей и т.д. Результаты этих опросов имеют неоднозначные значения, т.к. потенциальные покупатели готовы платить не конкретно ту сумму, которую они указали при анкетировании, а какой-то диапазон цен из некоторого интервала значений. Эти неточности при неграмотной статистической обработке могут привести к финансовым потерям и убыткам предприятия.

Существует достаточно много различных методов ценообразования, базирующихся на ценовых предпочтениях потребителей. Одним из них является метод Жана Штоцеля, который носит название «Психологическая цена». Этот метод достаточно прост, а результатом его работы является некий ценовой диапазон, полученный путем комплексного анализа всего массива данных о ценовых предпочтениях.

В данной работе детально был изучен метод Штоцеля. Основной задачей работы было наиболее точно определить цену на товар-новинку и минимизировать затраты на сбор статистических данных при опросе потребителей.

### 1. Метод ценообразования Ж. Штоцеля

Французский психолог и социолог Жан Штоцель предложил этот метод в середине XX в. Данный метод очень прост и заключается в следующем. Потребителям предлагают ответить на два вопроса:

1. Ниже какого уровня цены  $P_1$  товар кажется вам настолько дешевым, что возникает подозрение о его низком качестве и о том, что товар — подделка?
2. Начиная с какого уровня цены  $P_2$  товар кажется настолько дорогим, что вопрос о покупке даже не встает?

В результате получаем двумерную выборку  $(P_{i1}, P_{i2})$ ,  $i = \overline{1, N}$ , где  $N$  — количество опрошенных потребителей. Отметим, что цены рассматриваются как случайные величины. Далее каждый элемент выборки обрабатывается отдельно, по каждому вектору  $\{P_{ij}\}$ ,  $j = \overline{1, 2}$ , строится эмпирическая функция распределения (э.ф.р.) по формуле:

$$F_N^{(j)}(p) = \frac{1}{N} \sum_{i=1}^N I_{[0, p)}(P_{ij}), \quad (1)$$

где

$$I_{[0, p)}(y) = \begin{cases} 0, & \text{если } y \notin [0, p), \\ 1, & \text{если } y \in [0, p), \end{cases}$$

— индикаторная функция.

Известно, что э.ф.р. есть скачкообразная непрерывная справа функция с высотой скачка  $\frac{1}{N}$ . Однако в методе Ж. Штоцеля э.ф.р. «сглаживается», соединяя между собой точки, при этом получаем ломаную линию, возрастающую по оси  $Y$  от нуля до единицы. Фактически, в данном случае используется проектор в класс непрерывных ф.р., предложенный в [6].

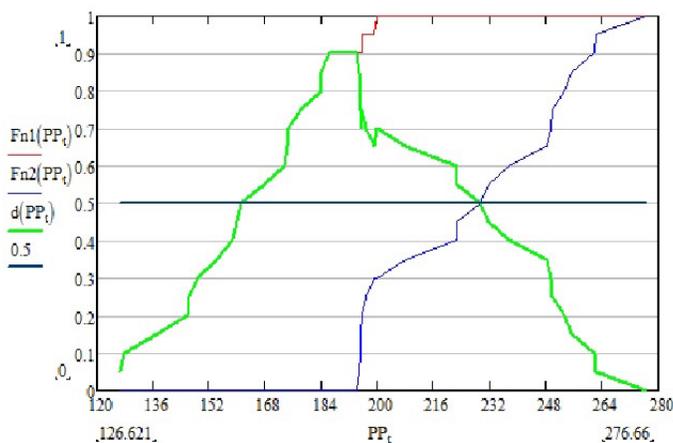


Рис. 1. Пример реализации метода Штоцеля

В качестве рекомендуемого диапазона цен принято рассматривать те значения цены, при которых функция  $d(p) = 0.5$ , где

$$d(p) = F_N^{(1)}(p) - F_N^{(2)}(p). \quad (2)$$

Путём имитационного моделирования была показана нормальность [7] ценовых значений, полученных в результате работы метода Штоцеля для  $N > 30$ .

## 2. Применение метода Штоцеля при анализе цены продукции ЗАО «Натуральное мыло»

Применим данный метод к результатам работы фокус-группы по продукции ЗАО «Натуральное мыло». Потребителям предлагалось протестировать натуральное мыло в течении 10 дней, после чего в процессе работы фокус-группы респонденты ответили

на вопросы, описанные в методе Штоцеля. Отметим, что участниками маркетингового исследования были женщины в возрасте от 35 до 45 лет, одна участница возраста до 30 лет. Данные, собранные в результате опроса, представлены в табл. 1.

Таблица 1

Диапазоны цен, полученных в ходе опроса, руб./ед.

$P_1$	$P_2$	$P_1$	$P_2$	$P_1$	$P_2$	$P_1$	$P_2$
145	239	251	310	277	362	307	403
172	253	253	313	280	364	308	408
173	254	255	315	284	365	311	408
175	256	256	322	286	367	313	424
186	257	257	333	288	367	313	427
186	260	262	339	291	368	314	430
186	274	262	341	293	373	320	438
199	277	264	350	294	374	322	439
201	279	266	350	295	377	323	474
202	283	266	352	300	381	327	474
220	284	271	354	301	391	336	508
222	293	274	355	303	394	343	581
236	308	276	359	306	399	361	655

В результате применения метода Штоцеля к имеющимся данным диапазон рекомендуемых цен имеет вид [303;359] руб. за единицу товара. При этом, если цена равна 303 рубля, то 29% потребителей считают ее слишком низкой, и не будут покупать мыло, так как будут считать его не качественным, а 21% потребителей не будут покупать, так как считают цену слишком высокой.

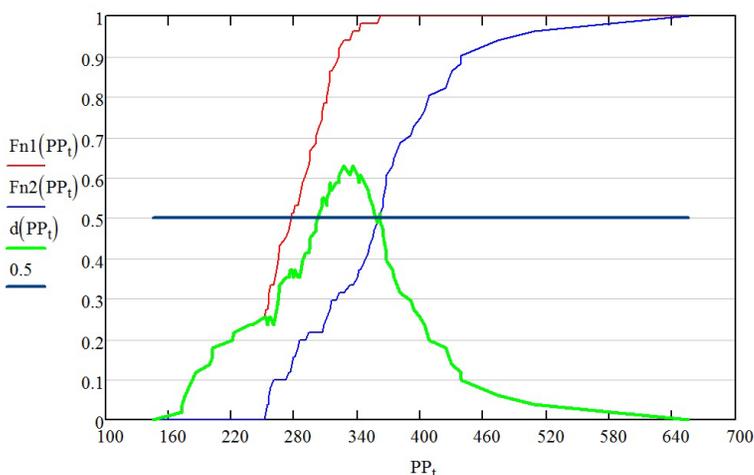


Рис. 2. Метод Штоцеля для данных о ценовых предпочтениях на продукцию ЗАО «Натуральное мыло»

При цене 359 рублей лишь 2% потребителей не считают товар качественным, при этом 48% откажутся от покупки, так как будут считать её слишком дорогой. Заметим, что функция  $d(P)$  имеет два одинаковых максимальных значения в точках  $P_1 = 327$  и  $P_2 = 336$ . В первой максимальной точке функции  $d(P_1 = 327) = 0.63$  получаем, что 94% потребителей считают товар качественным, при этом 31% потребителей откажутся от покупки из-за высокой цены. Во второй точке максимума функции  $d(P_2 = 336) = 0.63$  получаем, что 96% потребителей считают товар качественным, при этом 33% потребителей откажутся от покупки из-за высокой цены.

Таким образом, в целях захвата большего объема рынка фирме рекомендуется назначить цену в 336 рублей за единицу продукции, так как при этом большее количество потребителей считают товар качественным.

### Заключение

Таким образом, в данной работе с помощью метода Штоцеля получен ценовой диапазон для продукции ЗАО «Натуральное мыло», найдена цена, при которой достигается баланс доли потребителей, считающих данный товар слишком дорогим, а также доли потребителей, для которых выбранная цена будет являться показателем низкого качества продукции. Фирме даны практически значимые рекомендации по ценовой политике относительно данного товара.

### Литература

1. *Зенкова Ж.Н., Краковецкая И.В.* Моделирование по неполным данным в логистике и маркетинге/ Логистические системы в глобальной экономике: материалы Междунар. науч.-практ. конф. (14–15 марта 2013 г., Красноярск): в 2 ч. Ч.1. Научно-исследовательский сектор / Сиб. гос. аэрокосмич. ун-т. – Красноярск, 2013. – С. 98–105.
2. *Гелюта И.Ф.* Ценообразование в системе маркетинга зарубежных компаний- Мировая экономика [Электронный ресурс]: дис. канд. экон. наук / И.Ф. Гелюта. – М., 2002. – 211 с. – Электрон. версия печат. публ. – Доступ из „Электронная библиотека: Диссертации“.
3. Ценообразование [Электронный ресурс] // Википедия: свободная энцикл. – Электрон. дан. – [Б. м.], 2014. – URL: <http://ru.wikipedia.org/wiki/Ценообразование> (дата обращения: 25.04.2014).
4. *Панкрухин А.П.* Маркетинг: учебник. – М.: ИКФ Омега-Л, 2002. – С. 272–273.
5. *Котлер Ф.* Основы маркетинга. Краткий курс. : Пер. с англ. – М.: Прогресс, 1991. – 475 с.
6. *Дмитриев Ю.Г., Устинов Ю.К.* Статистическое оценивание распределений вероятностей с использованием дополнительной информации – Томск: Изд-во ТГУ, 1988. 194 с.
7. *Кобзарь А.И.* Прикладная математическая статистика. – М.: Физматлит, 2006. – 238 с.

## ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ НЕПАРАМЕТРИЧЕСКОЙ ПРОЦЕДУРЫ CUSUM ДЛЯ ОБНАРУЖЕНИЯ РАЗЛАДКИ В АВТОРЕГРЕССИОННЫХ МОДЕЛЯХ С НЕИЗВЕСТНЫМ КОНЕЧНЫМ ПАРАМЕТРОМ

**В.В. Конев, Б.Н. Назаренко**

*Томский государственный университет*  
vvkonev@mail.tsu.ru, nazarenkoB@sibmail.com

### Введение

В практической деятельности человека часто встречается задача об обнаружении изменений свойств некоторого случайного процесса - так называемая задача о разладке. В качестве примеров можно привести технологические процессы (разладка — появление брака), процессы в земной коре (разладка — землетрясения), радиолокационные процессы (разладка — появление движущегося вражеского объекта на фоне шума в виде атмосферы или гидросферы), медицинские процессы — кардиограммы и энцефалограммы (разладка — появление аномалии) [3].

Во всех этих случаях требуется обнаружить момент изменения свойств (момент разладки) как можно эффективнее. В качестве критериев эффективности чаще всего берут среднее время между ложными тревогами и среднее время запаздывания [1–3]. Эти характеристики напрямую связаны с понятием статистической ошибки первого и второго рода. Для идеальной процедуры обнаружения разладки желательно бесконечная первая характеристика и нулевая вторая.

На практике особую важность представляют непараметрические процедуры обнаружения разладки, так как почти всегда точные распределения помех неизвестны, а вместе с ними и распределение самого случайного процесса. Поэтому применение классических подходов, основанных на отношениях правдоподобия и им подобным, невозможно.

В данной работе представлены результаты имитационного моделирования непараметрического обнаружения разладки и выяснение их характеристик.

## 1. Постановка задачи

Используется модель с не полностью известным конечным параметром. Данную ситуацию можно формализовывать разными способами. Мы выбрали вариант, когда конечный вектор лежит на известной прямой, проходящей через начальный вектор параметров, причем величина скачка параметров после модели не меньше, чем заданный порог  $\Delta$ . Точнее, пусть до разладки  $n < v$  (где  $v$  — момент разладки) процесс имеет вид

$$x_n = \theta_1 x_{n-1} + \theta_2 x_{n-2} + \theta_3 x_{n-3} + \varepsilon_n, \quad (1)$$

а после разладки:

$$x_n = (\theta_1 + \lambda e_1) x_{n-1} + (\theta_2 + \lambda e_2) x_{n-2} + (\theta_3 + \lambda e_3) x_{n-3} + \varepsilon_n. \quad (2)$$

Здесь векторы  $\theta = (\theta_1, \theta_2, \theta_3)^T$  и  $e = (e_1, e_2, e_3)^T$  известны, а величина скачка  $\lambda$  неизвестна, но предполагается, что она больше заданного порога  $\Delta$ . Шумы  $\{\varepsilon_n\}_{n \geq 1}$  — независимые стандартные гауссовские случайные величины.

Помимо базовой модели (1) – (2) будем использовать модель с мультипликативными шумами: до разладки  $n < v$  процесс имеет вид

$$x_n = (\theta_1 + \delta_n^1) x_{n-1} + (\theta_2 + \delta_n^2) x_{n-2} + (\theta_3 + \delta_n^3) x_{n-3} + \varepsilon_n, \quad (3)$$

а после разладки:

$$x_n = (\theta_1 + \lambda e_1 + \delta_n^4) x_{n-1} + (\theta_2 + \lambda e_2 + \delta_n^5) x_{n-2} + (\theta_3 + \lambda e_3 + \delta_n^6) x_{n-3} + \varepsilon_n. \quad (4)$$

Здесь  $\{\delta_n^i\}_{n \geq 1}^{i=1..6}$  — последовательности независимых гауссовских случайных величин с нулевым математическим ожиданием и дисперсией  $\sigma^2$ .

Для моделирования были использованы следующие значения параметров:

$$\theta = \begin{bmatrix} 0.4 \\ -0.2 \\ 0.3 \end{bmatrix}, \quad e = \begin{bmatrix} -0.826 \\ 0.34 \\ 0.45 \end{bmatrix}, \quad \Delta = 0.3, \quad \lambda = 0.4, \quad \sigma = 0.2.$$

## 2. Построение решающих процедур

Для достижения свойства непараметричности требуется отказаться от принятия решения о разладке в каждый момент времени, вместо этого следует разбить все наблюдения на циклы случайной длины, в некотором смысле выравнивающие исходные наблюдения и делающие их однородными в смысле информации о разладке [4]. Введём последовательность моментов остановки:

$$\tau_0 = 0, \quad \tau_n = \inf \left\{ N > \tau_{n-1} : \sum_{k=\tau_{n-1}+1}^N [e_1 x_{k-1} + e_2 x_{k-2} + e_3 x_{k-3}]^2 \geq H \right\}, \quad n \geq 1,$$

где  $H$  — заданный параметр. На основе каждого цикла  $[0, \tau_1]$ ,  $(\tau_{n-1}, \tau_n]_{n > 1}$  строим статистику вида:

$$z_n = -1 + \frac{2}{\Delta H} \sum_{k=\tau_{n-1}+1}^{\tau_n} \gamma_n(k) [x_k - \theta_1 x_{k-1} - \theta_2 x_{k-2} - \theta_3 x_{k-3}] [e_1 x_{k-1} + e_2 x_{k-2} + e_3 x_{k-3}],$$

$$\gamma_n(k) = \begin{cases} 1, & k < \tau_n; \\ \frac{H - \sum_{k=\tau_{n-1}+1}^{\tau_n-1} [e_1 x_{k-1} + e_2 x_{k-2} + e_3 x_{k-3}]^2}{[e_1 x_{\tau_{n-1}} + e_2 x_{\tau_{n-2}} + e_3 x_{\tau_{n-3}}]^2}, & k = \tau_n. \end{cases}$$

Базовая решающая процедура имеет вид:

$$N(H, C) = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n+1} \sum_{i=k}^n z_i \geq C \right\}. \quad (5)$$

Помимо этой базовой процедуры будем еще применять её квантованный аналог. Для этого проквантуем статистики  $z_n$ :

$$z_n^{\text{quant}} = \begin{cases} -1, & z_n < \delta, \\ w, & z_n \geq \delta, \end{cases}$$

и к полученным квантованным статистикам снова применим процедуру CUSUM:

$$N^{\text{quant}}(H, C) = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n+1} \sum_{i=k}^n z_i^{\text{quant}} \geq C \right\}. \quad (6)$$

Для моделирования будут использоваться следующие значения параметров:  $w = 2$ ,  $\delta = 0.6$ .

### 3. Результаты моделирования

В качестве основных характеристик процедур обнаружения разладки будем использовать две:  $T_0 = E_0 N(H, C)$  — среднее число временных периодов между ложными тревогами,  $T_1 = \sup_{v \geq 1} \text{ess sup } E^{(v)} \left[ (N - n_v + 1)^+ | \mathfrak{F}_{\tau_{n-1}} \right]$  — среднее число временных периодов запаздывания после периода с разладкой. Здесь  $n_v$  — номер цикла, в котором произошла разладка,  $\mathfrak{F}_{\tau_n} = \sigma(x_0, \dots, x_{\tau_n})$ ,  $E_0$  — математическое ожидание, когда разладка не происходила;  $E^{(v)}$  — когда разладка произошла в момент времени  $v$ .

Для практического применения процедур обнаружения разладки желательно выполнение двух свойств:

1. Логарифмическая асимптотика, то есть среднее время запаздывания растёт в асимптотике не хуже логарифма среднего времени между ложными тревогами.
2. Контролируемость — выполнение условий вида:  $T_0 \geq e^C$ ,  $T_1 \leq C$  в некотором диапазоне изменения контролируемого нами параметра  $C$ . Очевидно, из контролируемости следует логарифмическая асимптотика. Это свойство важно на практике, так как оно позволяет по требуемому качеству процедуры определить необходимое значение параметра  $C$ .

При моделировании рассматриваются 3 варианта:

1. К базовой модели (1) – (2) применяется процедура (5). Результаты моделирования показаны на рис. 1 – 3.
2. К зашумленной модели (3) – (4) применяется процедура (5). Результаты моделирования показаны на рис. 4 – 6.
3. К базовой модели (1) – (2) применяется квантованная процедура (6). Результаты моделирования показаны на рис. 7 – 9.

На их основе можно сделать следующие выводы:

1. Среднее время между ложными тревогами для всех трёх вариантов имеет экспоненциальную границу снизу (рис. 1, 4, 7).

2. Начиная с  $C > 2$  для первого и второго варианта и с  $C > 2.5$  для третьего варианта, среднее время запаздывания имеет линейную границу сверху (рис. 2, 5, 8).
3. Первый и второй пункты позволяют сказать, что процедуры (5) и (6) контролируемы в смысле, указанном ранее.
4. Рис. 3, 6 и 9 говорят о логарифмической асимптотике среднего времени запаздывания относительно среднего времени между ложными тревогами.
5. На основе результатов для второго варианта можно утверждать, что процедура (5) устойчива к небольшим неточностям в модели в виде мультипликативных шумов у параметров модели.

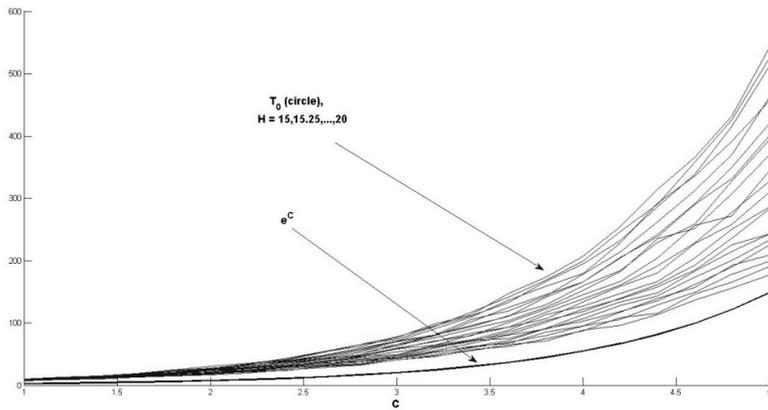


Рис. 1. Среднее число циклов между ложными тревогами для модели (1) – (2)

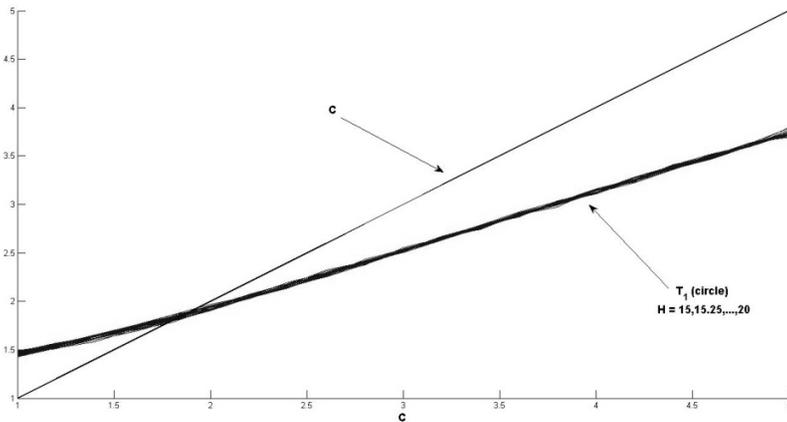


Рис. 2. Среднее число циклов запаздывания для модели (1) – (2)

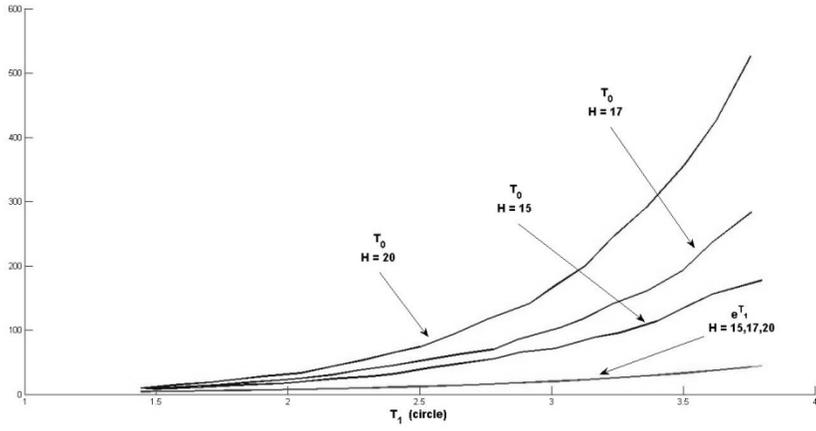


Рис. 3. Сравнение  $T_0$  и  $T_1$  для модели (1) – (2)

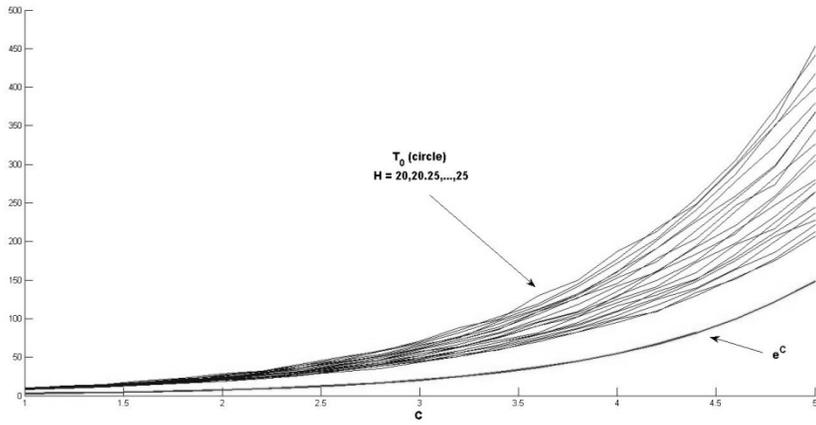


Рис. 4. Среднее число циклов между ложными тревогами для модели (3) – (4)

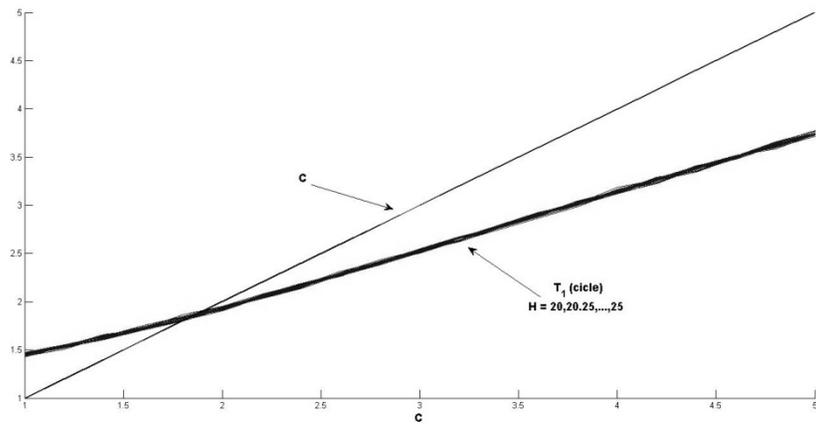


Рис. 5. Среднее число циклов запаздывания для модели (3) – (4)

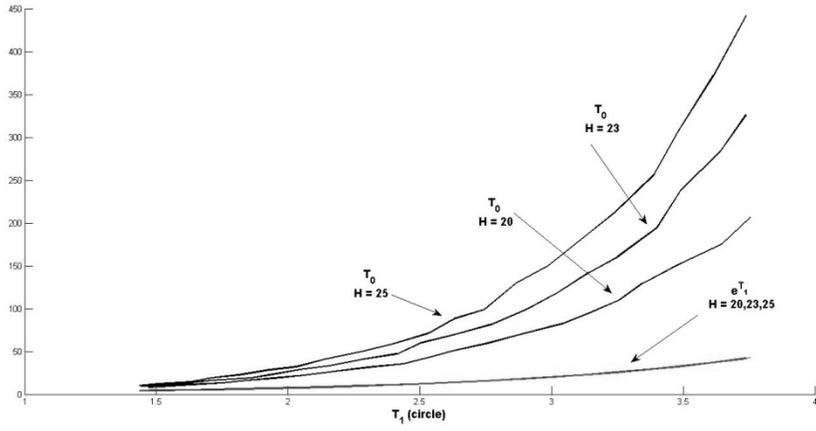


Рис. 6. Сравнение  $T_0$  и  $T_1$  для модели (3) – (4)

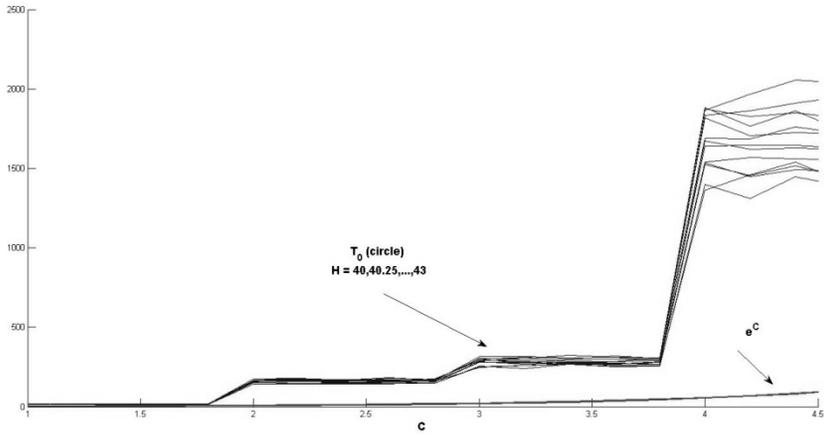


Рис. 7. Среднее число циклов между ложными тревогами для модели (1) – (2) при применении квантованной процедуры (6)

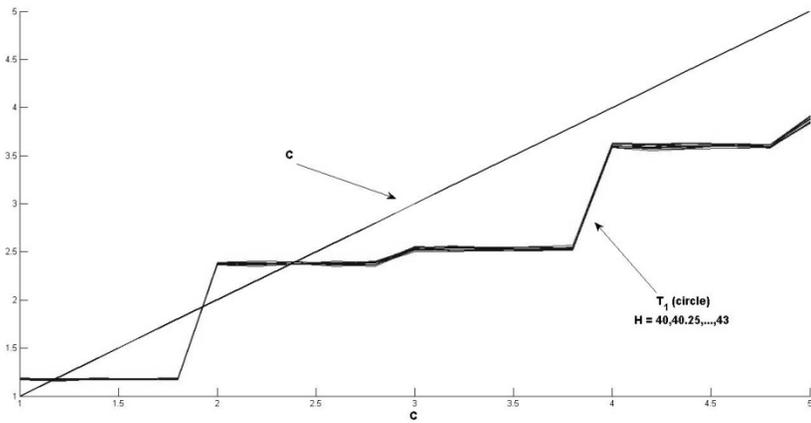


Рис. 8. Среднее число циклов запаздывания для модели (1) – (2) при применении квантованной процедуры (6)

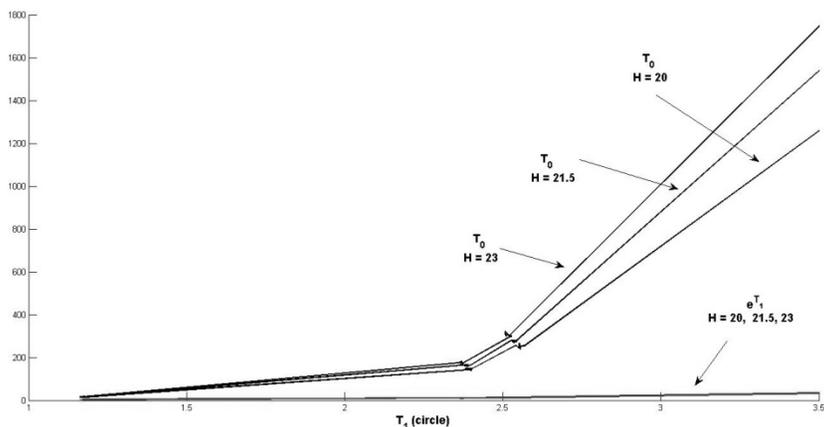


Рис. 9. Сравнение  $T_0$  и  $T_1$  для модели (1) – (2) при применении квантованной процедуры (6)

### Заключение

В данной работе приведено экспериментальное исследование непараметрической процедуры CUSUM и её квантованного аналога к модели авторегрессионного процесса. Результаты моделирования говорят в пользу логарифмической асимптотики указанных процедур и их контролируемости. К тому же наблюдается устойчивость к небольшим погрешностям в модели в виде мультипликативных шумов у параметров.

### ЛИТЕРАТУРА

1. *T.L. Lai*. Information bounds and quick detection of parameter changes in stochastic systems// IEEE Transaction of Information Theory, vol. 44, №7, 1998.
2. *G. Lorden*. Procedures for reacting to a change in distribution// The Annals of Mathematical Statistics, vol. 42, №. 6, 1971. – pp. 1897–1908.
3. *A. Tartakovsky, I. Nikiforov, M. Basseville*. Sequential analysis: Hypothesis testing and changepoint detection, Englewood Clifs, NJ: Prentice-Hall, 2014
4. *В.В. Конева*. Последовательные оценки параметров стохастических динамических систем. Томск: Изд-во Томск. ун-та, 1985. с. 268

## УЛУЧШЕННЫЕ ОЦЕНКИ ДОЛИ В ВЫБОРОЧНОМ КОНТРОЛЕ КАЧЕСТВА

**Т.О. Кошевая**

*Томский государственный университет*  
tomakoshevaya@gmail.com

### Введение

Существует несколько методов контроля качества продукции. Можно выделить тотальный (или сплошной) контроль, который подразумевает проверку всей партии, а также выборочный контроль, при котором исследуется лишь часть поступившей продукции. Тотальный контроль обходится слишком дорого и не целесообразен, за исключением, когда поступившая партия маленькая или каждая единица продукции дорогая и важная. Выборочный контроль ставит перед нами вопрос: сколько проверять? Ведь чем больше продукции мы подвергнем контролю, тем эффективнее будет проверка, но и обойдется она дороже. Если, например, нам нужно применить разрушающий контроль, и таким образом уничтожить часть продукции, то экономически целесообразно отобрать небольшой объём наблюдений, но при этом гарантировать точность оценивания. Поэтому часто исследователи привлекают дополнительную информацию, основанную на своём опыте или знании, с целью сокращения объёма экспериментальных

данных или повышения точности оценивания при фиксированном объеме наблюдений [1]. В данной работе, в отличие от [2–5], предлагаются комбинированные оценки доли брака (или доли продукции с интересующим признаком) в партии, учитывающие совместно обычную статистическую оценку и несколько априорных догадок в форме некоторых значений искомой доли.

### 1. Постановка задачи

Пусть имеется некая партия объема  $N$ , из которой случайным образом производится выборка без возвращения объема  $n$ . Под  $P$  будем понимать долю объектов в партии с дефектом (или с интересующим нас признаком). Пусть  $p_i, i = \overline{1, m}$  — заданные априорные догадки, которые выступают в качестве возможного значения неизвестной доли  $P$ . Требуется оценить долю  $P$  в партии, учитывая совместно статистическую оценку доли  $\hat{P} = v_n / n$  и априорные догадки  $p_i$ .

Рассмотрим комбинированную оценку вида

$$\hat{P}_\lambda = \hat{P} - \sum_{i=1}^m \lambda_i (\hat{P} - p_i). \quad (1)$$

Из условий минимума среднеквадратических ошибок (СКО)  $S^2(\lambda_i) = M[\hat{P}_i - P]^2$  выбраны весовые коэффициенты  $\lambda_i$ , где

$$\hat{P}_i = \hat{P} - \lambda_i (\hat{P} - p_i). \quad (2)$$

и имеют вид

$$\lambda_i = (1 + n\Delta_i^2 / \sigma^2)^{-1}. \quad (3)$$

Здесь  $\sigma^2 = P(1-P)(N-n)/(N-1)$ ,  $\Delta_i = P - p_i$  — величина отклонения  $i$ -ой априорной догадки от истинного значения искомой доли. Весовой коэффициент  $\lambda_i$  показывает, какое влияние оказывает  $i$ -я априорная догадка в комбинированной оценке (2) и изменяется в пределах  $0 < \lambda_i \leq 1$ . СКО оценки (1) определяется выражением

$$S^2 = \frac{\sigma^2}{n} \left( 1 - \sum_{i=1}^m \lambda_i \right)^2 + \left( \sum_{i=1}^m \lambda_i \Delta_i \right)^2. \quad (4)$$

Формула (4) характеризует потенциальную точность оценки (1) при заданных значениях  $\Delta_i, \sigma^2$ .

### 2. Адаптивная оценка

Поскольку весовые коэффициенты  $\lambda_i$  в (1) неизвестны, то получим комбинированную оценку доли брака в партии путем замены  $\lambda_i$  на статистические оценки  $\hat{\lambda}_i$ , построенные по исходной выборке. Однако здесь возникает вопрос, является ли адаптивная комбинированная оценка предпочтительнее обычной оценки  $\hat{P}$  по величине СКО и при каких условиях? Подставляя в (3) вместо неизвестного  $P$  оценку  $\hat{P}$ , в качестве оценки возьмем  $\hat{\lambda}_i = (1 + n\hat{\Delta}_i^2 / \hat{\sigma}^2)^{-1}$ , где  $\hat{\Delta}_i = \hat{P} - p_i$  — оценка отклонения,  $\hat{\sigma}^2 = \hat{P}(1-\hat{P})(N-n)/(N-1)$  — оценка дисперсии. Подставляя  $\hat{\lambda}_i$  в (1), получаем адаптивную комбинированную оценку

$$\hat{P}_i = \hat{P} - \sum_{i=1}^m \hat{\lambda}_i (\hat{P} - p_i).$$

СКО  $S_1^2$  оценки  $\hat{P}_1$  можно вычислить, применяя гипергеометрический закон с параметрами  $(n, P, N)$ .

### 3. Сравнение СКО оценок

Введем отношения для сравнения СКО оценок

$$E = D\hat{P} / D\hat{P} = 1, E_0^m = S^2(\lambda), E_1^m = S_1^2 / D\hat{P}, \quad (5)$$

где верхний индекс  $m$  показывает количество априорных догадок. Априорные догадки зададим в виде  $p_i = (2i - 1) / 2m, i = \overline{1, m}$  и рассмотрим эти отношения. Графики отношений СКО в зависимости от  $P, n, m$  и  $N$  приводятся на рисунках.

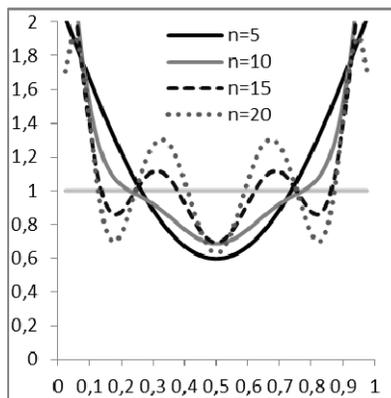
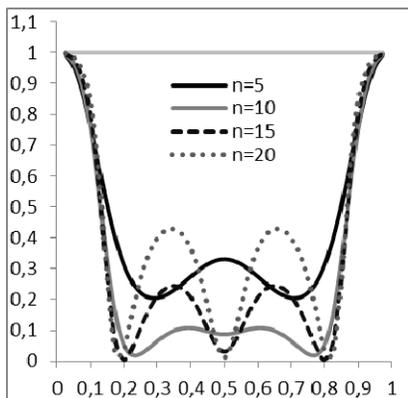


Рис. 1. Зависимость  $E_0^m$  от  $P$  и  $n$  при  $m = 3, N = 40$

Рис. 2. Зависимость  $E_1^m$  от  $P$  и  $n$  при  $m = 3, N = 40$

Рис. 1 показывает потенциальный выигрыш по СКО (при известных коэффициентах  $\lambda_i$ ). С увеличением числа наблюдений, как видно, выигрыш увеличивается по величине. Рис. 2 показывает, как переход к адаптивной оценке  $\hat{P}_1$  ухудшает выигрыш по СКО, а так же указывает области значений  $P$  проигрыша в точности.

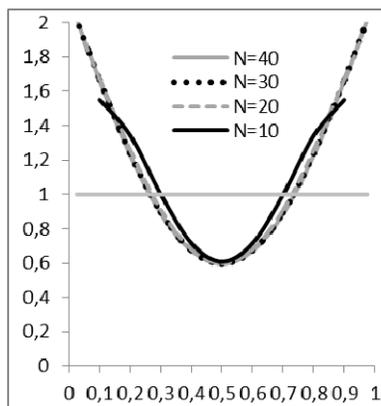
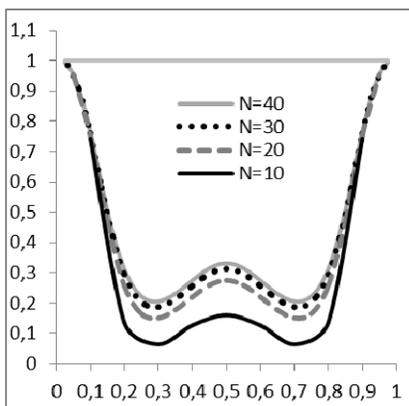


Рис. 3. Зависимость  $E_0^m$  от  $P$  и  $N$  при  $m = 3, n = 5$

Рис. 4. Зависимость  $E_1^m$  от  $P$  и  $N$  при  $m = 3, n = 5$

Рис. 3 показывает, как изменяется потенциальная точность оценивания при фиксированном объеме наблюдений  $n$  с увеличением объема партии  $N$ . Чем больше доля выборки относительно партии, тем больше выигрыш в оценивании. Рис. 4 демонстрирует, что увеличение объема партии относительно выборки практически не влияет на выиг-

рыш в оценивании, т.е. привлечение априорной догадки даёт нам возможность оценить партию из 40 единиц продукции на основании выборки объёма  $n = 5$  на том же уровне, что и партию из 10 единиц продукции.

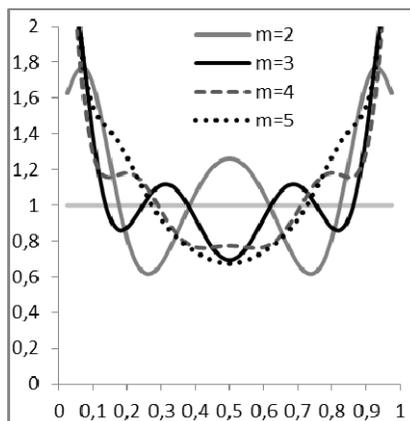
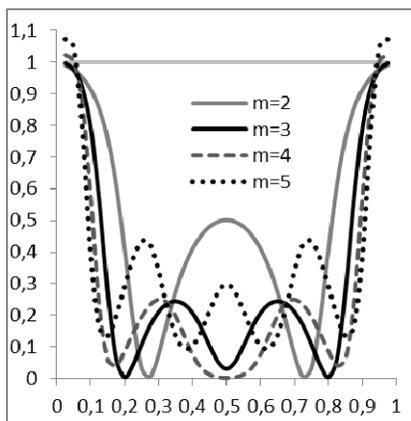


Рис. 5. Зависимость  $E_0^m$  от  $P$  и  $m$ , при  $N = 40$ ,  $n = 15$

Рис. 6. Зависимость  $E_1^m$  от  $P$  и  $m$ , при  $N = 40$ ,  $n = 15$

Из рис. 5 и 6 видно, что с увеличением числа априорных догадок выигрыш увеличивается в более широком диапазоне изменения  $P$ .

Автор выражает благодарность и признательность своему научному руководителю Дмитрию Юрию Глебовичу за помощь в выполнении данной работы.

#### ЛИТЕРАТУРА

1. Дмитриев Ю.Г., Тарасенко П.Ф. Использование априорной информации в статистической обработке экспериментальных данных. // Известия вузов. Физика. – 1992. – № 9. С. 136–142.
2. Dmitriev Yu., Tarassenko P., Ustinov Yu. On Estimation of Linear Functional by Utilizing a Prior Guess // A. Dudin et al. (Eds.): ITMM 2014, CCIS 487. 2014. – P. 82–90.
3. Дмитриев Ю.Г., Тарасенко П.Ф. О к-адаптивных оценках вероятности // ITMM 2015. Ч. 1. С. 40–45.
4. Кошечая Т.О., Дмитриев Ю.Г. Комбинированные оценки доли // ITMM 2015. Ч. 1. С. 50–55.
5. Дмитриев Ю.Г., Кошечая Т.О. О комбинированных оценках вероятности // Известия вузов. Физика. – 2015. – Т. 58, № 11/2. – С. 242–246.
6. Dmitriev Yu., Tarassenko P. On Adaptive Estimation Using a Prior Guess // Applied Methods of Statistical Analysis. Nonparametric Approach – AMSA-2015, Novosibirsk, Russia, 14-15 September, 2015: Proceedings of the International Workshop. – 2015. – P. 49–55.

## НЕПАРАМЕТРИЧЕСКИЕ АЛГОРИТМЫ ИДЕНТИФИКАЦИИ, ПРОГНОЗИРОВАНИЯ И УПРАВЛЕНИЯ В ARX-ПРОЦЕССАХ

**В.Ю. Луков**

*Томский государственный университет*  
lukov\_vadim@rambler.ru

### Введение

Во многих прикладных задачах требуется исследовать влияние на выход стохастического объекта некоторого множества факторов (входов). Однако может возникнуть ряд трудностей, а именно: тип зависимости предсказываемой переменной от факторов может быть неизвестен, условие независимости векторов наблюдений над выходом и входами может нарушаться. Такие задачи естественно рассматривать с точки зрения непараметрической статистики. В данной работе рассматриваются непараметрические алгоритмы идентификации ARX-процессов. Предложенные алгоритмы применяются для идентификации, прогнозирования и управления процессов реальных данных.

## 1. Постановка задачи

Пусть случайная последовательность  $Y_1, \dots, Y_n$  генерируется моделью

$$Y_t = \psi(Y_{t-1}, \dots, Y_{t-k}, X_{t-1}, \dots, X_{t-q}) + \xi_t, \quad t = 1, \dots, n, \quad (1)$$

где  $Y_t$  — выход модели в момент времени  $t$ ,  $X_t$  — экзогенный случайный фактор, который не зависит от  $\xi_t$ ,  $\xi_t$  — последовательность независимых одинаково распределенных случайных величин с  $E\xi_1 = 0$  и  $E\xi_1^2 < \infty$ ,  $\psi$  — неизвестная функция. Пусть функция  $\psi$  ограничена, плотность распределения  $\xi_t$  на  $R^1$  неотрицательна с  $E\xi_1^3 = 0$  и  $E\xi_1^4 < \infty$ . В этом случае, процесс (1) является цепью Маркова с коэффициентами сильного перемешивания [1]

$$\alpha(\tau) \leq c_0^\tau, \quad 0 < \rho_0 < 1, \quad c_0 > 0. \quad (2)$$

Ставится задача восстановления зависимости  $\psi$  для модели (1).

## 2. Непараметрическая идентификация

Обозначим  $Y_{t,k} = (Y_{t-1}, \dots, Y_{t-k})$ ,  $X_{t,q} = (X_{t-1}, \dots, X_{t-q})$ . Известно, что функция регрессии может быть использована как модель  $\psi$  в (1):

$$r(y, x) = E(Y_t | Y_{t,k} = y, X_{t,q} = x) = E(Y_t | y, x) = \frac{a(y, x)}{p(y, x)} = \int Y_t f(Y_t | y, x) dY_t, \quad (3)$$

где  $(y, x) \in R^{k+q}$ ,  $a(y, x) = \int z f(z, y, x) dz$  — базовый функционал,  $p(y, x)$  — плотность распределения случайного вектора  $(Y_{t,k}, X_{t,q})$ ,  $f(z, y, x)$  — плотность распределения вектора  $(Y_t, Y_{t,k}, X_{t,q})$  [1].

Тогда мы можем оценить неизвестную функцию  $\psi$  следующим аналогом оценки Надарая-Ватсона:

$$\psi_n(z) = \frac{\sum_{i=1}^n Y_i \cdot K\left(\frac{z - Z_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{z - Z_i}{h_n}\right)}. \quad (4)$$

где  $K\left(\frac{y - Z_t}{h_n}\right)$  —  $(k+q)$ -мерная плотность, (произведение одномерных плотностей стандартного нормального закона),  $h_n \in R^1$  — параметр размытости,  $h_n \downarrow 0$ ,  $Z_t = (Y_{t-1}, \dots, Y_{t-k}, X_{t-1}, \dots, X_{t-q})$ .

Введём следующие обозначения:

$$\begin{aligned} h_n^{(y)} &= (h_{n,1}^{(y)}, \dots, h_{n,k}^{(y)}), \quad h_n^{(x)} = (h_{n,1}^{(x)}, \dots, h_{n,q}^{(x)}), \\ K_k\left(\frac{y - Y_{i,k}}{h_n^{(y)}}\right) &= K\left(\frac{y_1 - Y_{i,1}}{h_{n,1}^{(y)}}\right) \times \dots \times K\left(\frac{y_k - Y_{i,k}}{h_{n,k}^{(y)}}\right), \\ K_q\left(\frac{x - X_{i,q}}{h_n^{(x)}}\right) &= K\left(\frac{x_1 - X_{i,1}}{h_{n,1}^{(x)}}\right) \times \dots \times K\left(\frac{x_q - X_{i,q}}{h_{n,q}^{(x)}}\right). \end{aligned}$$

В соответствии с (4) непараметрическая оценка для  $r(y, x)$  в точке  $(y, x)$  запишется в виде:

$$r_n(y, x) = \psi_n = \frac{\sum_{i=s+1}^n Y_i \frac{1}{\prod_{j=1}^k h_{n,j}^{(y)}} K_k \left( \frac{y - Y_{i,k}}{h_n^{(y)}} \right) \frac{1}{\prod_{j=1}^q h_{n,j}^{(x)}} K_q \left( \frac{x - X_{i,q}}{h_n^{(x)}} \right)}{\sum_{i=s+1}^n \frac{1}{\prod_{j=1}^k h_{n,j}^{(y)}} K_k \left( \frac{y - Y_{i,k}}{h_n^{(y)}} \right) \frac{1}{\prod_{j=1}^q h_{n,j}^{(x)}} K_q \left( \frac{x - X_{i,q}}{h_n^{(x)}} \right)}, \quad (5)$$

где  $s = \max(k, q)$ .

По методологии [2–5] находится главная часть оптимальной среднеквадратической ошибки оценки  $r_n(y, x)$ .

### 3. Статистическое моделирование

Применим алгоритм (5) для прогноза цен акций ОАО «Газпром» за период 01.01.2015 – 31.12.2015, используя модель (6). В качестве экзогенных факторов возьмём курс доллара, цену на нефть и цену на газ.

$$Y_t = f(Y_{t-1}, X_{t-1}) + \xi_t, \quad (6)$$

Качество моделей будем характеризовать относительной ошибкой идентификации  $\delta$  и абсолютной ошибкой прогнозирования  $\eta$ :

$$\delta = \frac{1}{n} \cdot \sum_{i=1}^n \frac{|Y_i - \psi_n|}{Y_i} \cdot 100\%, \quad \eta = \frac{1}{30} \cdot \sum_{i=n+1}^{n+30} |Y_i - \psi_n|.$$

Рассмотрим модель с зависимостью от курса доллара. Статистика (5) переписывается в виде:

$$\psi_n = \frac{\sum_{j \geq 2} Y_j \cdot K \left( \frac{y - Y_{j-1}}{h_n^{(y)}} \right) \cdot K \left( \frac{x - USD_{j-1}}{h_n^{(x)}} \right)}{\sum_{j \geq 2} K \left( \frac{y - Y_{j-1}}{h_n^{(y)}} \right) \cdot K \left( \frac{x - USD_{j-1}}{h_n^{(x)}} \right)}. \quad (7)$$

Оптимальные параметры размытости  $h_n^{(y)}$ ,  $h_n^{(x)}$  находятся путём минимизации относительной ошибки идентификации (см. рис. 1),

$$\varepsilon(h) = \frac{1}{30} \cdot \sum_{i=n-30}^n \frac{|Y_i - \psi_n|}{Y_i} \cdot 100\%. \quad (8)$$

В этом случае минимум ошибки достигается на  $h_n^{(y)} = 0.3$ ,  $h_n^{(x)} = 2$ .

Реальные значения цен и результаты непараметрической идентификации модели (6) представлены на рис. 2.

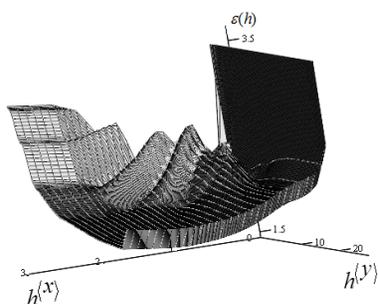


Рис. 1. Зависимость ошибки  $\varepsilon$  от параметров размытости

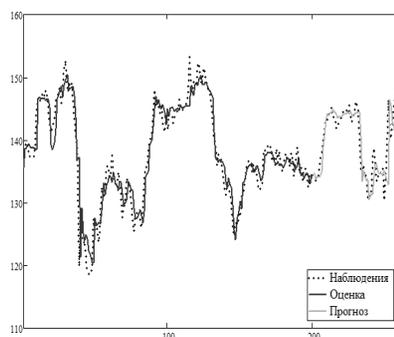


Рис. 2. Наблюдения  $Y$  и их оценка  $\psi_n(0.3, 2)$

Для моделей с зависимостью от цены на нефть и цены на газ были проведены аналогичные исследования. Сравнительный анализ моделей приведен в табл. 1.

Таблица 1

Сравнение ARX-моделей

	USD	OIL	GAS
$h_n^{(y)}, h_n^{(x)}$	2 18.9	0.9 13.5	1.1 7.3
$\delta$ (%)	1.745	1.948	1.919
$\eta$ (руб.)	1.687	2.102	1.998

#### 4. Управление

Пусть выходы  $Y$  модели (6) должны быть на следующих заданных уровнях  $Y_1^* = \max(Y_2, \dots, Y_N)$ ,  $Y_2^* = \frac{1}{N} \sum_{i=2}^N Y_i$  путем управления контролируемого фактора  $X$ .

Согласно [8] аналогично (5) построим оценку фактора  $X$  следующим образом:

$$X^*(g, n) = \frac{\sum_{i=2}^N X_j \cdot K\left(\frac{Y^*(g) - Y_i}{h^*}\right) \cdot K\left(\frac{Y_{n-1} - Y_{i-1}}{h_y}\right)}{\sum_{i=2}^N K\left(\frac{Y^*(g) - Y_i}{h^*}\right) \cdot K\left(\frac{Y_{n-1} - Y_{i-1}}{h_y}\right)}. \quad (9)$$

Качество оценивания будем определять относительной ошибкой управления, определяемой формулой:

$$A = \frac{1}{30} \cdot \sum_{i=n-30}^n \frac{|\hat{\Psi}_n - Y^*(g)|}{Y^*(g)} \cdot 100\%, \quad (10)$$

где  $g = 1, 2$ , а  $\hat{\Psi}_n$  получается путем подстановки оценки  $X^*(g, n)$  в формулу (7).

Рассмотрим модель с зависимостью от курса доллара при требуемом уровне  $Y_1^*$ . В этом случае минимум (10) достигается при  $h^* = 2.5$ ,  $h_y = 1.9$ . Исходные значения цен и результаты управления приведены на рис. 3.

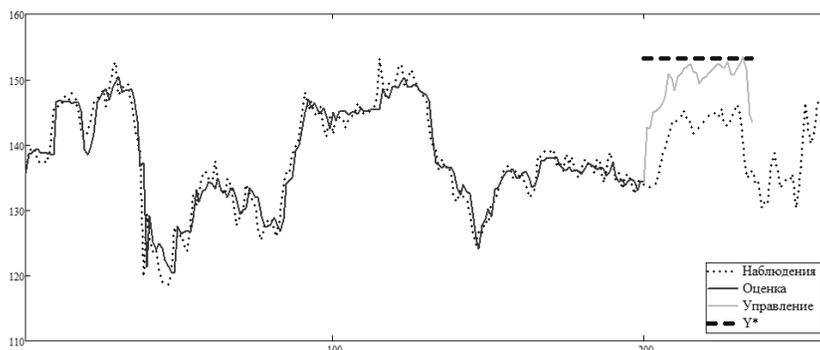


Рис. 3. Наблюдения  $Y$ , их оценка  $\psi_n$  и управление  $\hat{\psi}_n$

Сравнительный анализ для  $Y_2^*$  и для остальных моделей приведен в табл. 2.

Таблица 2.

Сравнение моделей с управлением

Требуемый уровень	Характеристики	USD	OIL	GAS
$Y_1^*$	Оптимальные параметры размытости для $X^*(1, n)$ .	2.5 1.9	1.4 1.8	0.1 1.2
	Ошибка управления $A$ , %.	3.122	1.982	3.845
$Y_2^*$	Оптимальные параметры размытости для $X^*(2, n)$ .	0.3 0.2	0.1 0.2	0.4 0.1
	Ошибка управления $A$ , %.	3.448	3.019	3.536

### Заключение

В настоящей работе исследовалась проблема оценивания неизвестной функции, определяемой нелинейными ARX-процессами с использованием ядерных регрессионных оценок. Также рассмотрены методы управления выходом путём использования контролируемых входов.

Среди ARX-процессов модель с зависимостью от курса доллара показала наилучший результат как по ошибке идентификации, так и по ошибке прогнозирования. Модели с управлением также показали свою работоспособность. Наилучшее качество показала модель с зависимостью от цены на нефть.

Автор выражает особую благодарность своему научному руководителю, Кошкину Геннадию Михайловичу за оказание помощи и поддержки на всех этапах работы.

### ЛИТЕРАТУРА

1. Китаева А.В., Кошкин Г.М. Непараметрическая полурекуррентная идентификация в широком смысле процессов сильного перемешивания. Проблемы передачи информации. 2010. Т. 46. Вып.1. С. 25–41.
2. Koshkin G.M. Asymptotic properties of functions of statistics and their application to nonparametric estimation. Automation and Remote Control. 1990. Т. 51(3). С. 345–357.
3. Васильев В.А., Кошкин Г.М. Непараметрическая идентификация авторегрессий. Теория вероятностей и ее применения. 1999. Т. 43(3). С. 507–517.
4. Kitaeva A.V., Koshkin G.M. Recurrent nonparametric estimation of functions from functionals of multidimensional density and their derivatives. Automation and Remote Control. 2009. Vol. 70(3). P. 389–407.
5. Kitaeva A.V., Koshkin G.M. Semi-recursive nonparametric identification in the general sense of a nonlinear heteroscedastic autoregression. Automation and Remote Control. 2010. Vol. 71(2). P. 257–274.
5. Кошкин Г.М., Глухова И.Ю. Непараметрическая идентификация нелинейных ARX-процессов. Вестн. Том. гос. ун-та. Управление, вычислительная техника и информатика. 2012. № 3. С. 55–61.
6. Китаева А.В., Кошкин Г.М. Ядерные оценки базовых функционалов по зависимым наблюдениям. Известия Томского политехнического университета. 2009. Т. 314. № 2. С. 26–31.
7. Koshkin G.M., Lukov V.Yu., Piven I.G. Nonparametric Algorithms of Identification and Prediction in the ARX-Models // Proceedings. The Second International Symposium on Stochastic Models, in Reliability Engineering, Life Science, and Operations Management / Ilia Frenkel and Anatoly Lisnianski (Eds.). SMRLO 2016, February 15-18, 2016, Beer Sheva, Israel. Conference Publishing Services The Institute of Electrical and Electronics Engineers, Inc. 2016. P. 620–623.
8. Koshkin G.M., Tarasenko F.P. Nonparametric algorithms for identifying and control of continuous-discrete stochastic objects. 8-th IFAC-IFORS Symposium on Identification and System Parameter Estimation. Beijing: Pargamon Press, 1988. №2. P. 882–887.

# МАТЕМАТИЧЕСКИЕ МОДЕЛИ ЗАВИСИМОСТИ ПРОИЗВОДСТВА И ПОТРЕБЛЕНИЯ ОТ ВОЗРАСТНОЙ СТРУКТУРЫ НАСЕЛЕНИЯ

Ю.Е. Пальянова

Томский государственный университет  
pje0360@gmail.com

## Введение

Население является не только производственной силой общества, но и основным условием экономического роста. Сложившаяся демографическая ситуация в России характеризуется сужением воспроизводства населения, сокращением абсолютной численности лиц трудоспособного возраста и ростом доли старших возрастных групп. В связи с этим актуально исследование социально-экономических последствий изменения динамики населения, в том числе при обосновании приоритетов демографической и миграционной политике.

Целью данной исследовательской работы является определение экономических факторов, положительно влияющих на рост валового регионального продукта и валовых внутренних сбережений таких регионов России, как Томская и Новосибирская области. Данные взяты с сайта Федеральной службы Государственной статистики [1].

В ходе работы были поставлены задачи: провести анализ возрастных особенностей производства и потребления и их влияние на величину сбережений и темпы экономического роста.

## 1. Исследование с помощью регрессионной модели Н. Леффа

В 1958 г. Э. Коулом и Э. Гувером было выдвинуто предположение, согласно которому большая численность населения в младших возрастах увеличивает потребительские запросы и снижает внутренние сбережения. Одной из первых попыток эмпирического обоснования данного положения стала модель Н. Леффа:

$$\frac{GDS}{Y} = \alpha + \beta_1 \frac{Y}{N} + \beta_2 \left(\frac{Y}{N}\right)^2 + \gamma \left(\frac{Y}{N}\right)^* + \delta_1 D_1 + \delta_2 D_2 + \varepsilon, \quad (1)$$

где GDS — валовые внутренние сбережения;  $Y$  — валовой внутренний продукт;  $N$  — численность населения;  $\left(\frac{Y}{N}\right)^*$  — годовой темп роста среднедушевого ВВП;  $D_1$  — показатель демографической нагрузки детьми (отношение численности населения в возрасте 0–14 лет к численности населения в возрасте 15–64 года);  $D_2$  — показатель демографической нагрузки пожилыми (отношение численности населения в возрасте 65 лет и старше к численности населения в возрасте 15–64 года);  $\alpha$  — свободный член, выражающий влияние неучтенных факторов;  $\varepsilon \sim N(0, \sigma^2)$  — ошибка модели.

Данная модель является нелинейной по  $\left(\frac{Y}{N}\right)^2$ , но она линейна по параметрам, т.е.

при дальнейшем анализе нужно использовать значения факторной переменной в квадрате [2].

### 1.1. Исследование для Томской области

Рассмотрим период с 1996 по 2013 гг.

Значение коэффициентов и  $t$ -критерия для модели (1)

Параметр	Значение коэффициента	Значение $t$ -критерия
Свободный член	1,059298	4,87533
ВРП на душу населения	- 0,002088	- 5,59217
ВРП на душу населения в квадрате	0,000005	5,20379
Годовой темп роста среднедушевого ВРП	0,002196	2,3383
Коэффициент демографической нагрузки детьми	- 0,001014	- 0,40991
Коэффициент демографической нагрузки пожилыми	- 0,014847	- 2,14876

*Значимость коэффициентов.*

Для определения значимости воспользуемся  $t$ -критерием Стьюдента. Рассмотрим гипотезу  $H_0 : \alpha, \beta_1, \beta_2, \gamma, \delta_1, \delta_2 = 0$ , т.е. эти коэффициенты не являются статистически значимыми. Гипотеза будет принята в том случае, если  $|t| > |t_T|$ , где  $t_T = 1,78$  — табличное значение  $t$ -критерия с уровнем значимости  $\alpha = 0,05$  и числом степеней свободы равным 16;  $t$  — значение, вычисленное для каждого из коэффициентов.

Таким образом, значимыми являются все коэффициенты, кроме  $\delta_1, \delta_2$  (коэффициенты демографической нагрузки).

*Общее качество уравнения регрессии.*

Коэффициент детерминации модели, который отражает долю изменения зависимой переменной, объясняемую рассмотренной моделью  $R^2 = 0,9545$ ; скорректированный коэффициент детерминации  $R^2 = 0,8522$ .

Далее воспользуемся  $F$ -критерием Фишера. Выдвигаем гипотезу  $H_0$  о неадекватности модели, которая выполняется в том случае, если  $F < F_{кр}$ , где  $F_{кр}(1;16) = 4,49$  — табличное значение  $F$ -критерия с уровнем значимости  $\alpha = 0,05$ ;  $F = 50,34$  — значение  $F$ -критерия для рассматриваемой модели. Таким образом, гипотеза  $H_0$  отклоняется. Делаем вывод об адекватности модели.

Также подсчитав ошибку аппроксимации  $A = 0,05\%$ , можно сделать вывод об адекватности модели.

При дальнейшем пересчёте модели без незначимых коэффициентов можно наблюдать сильную корреляционную зависимость между такими переменными, как ВРП на душу населения и ВРП на душу населения в квадрате:

$$r_{\frac{Y}{N}, \left(\frac{Y}{N}\right)^2} = 0,97; r_{\left(\frac{Y}{N}\right)^2, \frac{Y}{N}} = 0,97.$$

Полученные коэффициенты парной корреляции могут свидетельствовать о наличии эффекта мультиколлинеарности, то есть объясняющие переменные являются зависимыми. Для того, чтобы убедиться в наличие мультиколлинеарности найдем определитель обратной матрицы произведения транспонированной матрицы независимых переменных  $\left( \frac{Y}{N}, \left(\frac{Y}{N}\right)^2, \left(\frac{Y}{N}\right)^* \right), D_1, D_2$  на исходную матрицу независимых переменных:

$$\det \left[ \left( X^T X \right)^{-1} \right] = 5,65353 \cdot 10^{-24}.$$

Полученный определитель близок к 0, что свидетельствует о наличии мультиколлинеарности.

## 1.2. Исследование для Новосибирской области

Рассмотрим период с 1996 по 2013 гг.

Таблица 2

Значение коэффициентов и  $t$ -критерия для модели (1)

Параметр	Значение коэффициента	Значение $t$ -критерия
Свободный член	0,898063	1,03422
ВРП на душу населения	-0,00176	-1,523
ВРП на душу населения в квадрате	-0,000005	1,388
Годовой темп роста среднедушевого ВРП	0,001088	0,49
Коэффициент демографической нагрузки детьми	-0,00299	-1,088
Коэффициент демографической нагрузки пожилыми	-0,01979	-0,816

*Значимость коэффициентов.*

Для определения значимости воспользуемся  $t$ -критерием Стьюдента. Рассмотрим гипотезу  $H_0: \alpha, \beta_1, \beta_2, \gamma, \delta_1, \delta_2 = 0$ , т.е. эти коэффициенты не являются статистически значимыми. Гипотеза будет принята в том случае, если  $|t| > |t_T|$ , где  $t_T = 1,78$  — табличное значение  $t$ -критерия с уровнем значимости  $\alpha = 0,05$  и числом степеней свободы равным 16;  $t$  — значение, вычисленное для каждого из коэффициентов.

Таким образом, ни один из коэффициентов не является значимым.

*Общее качество уравнения регрессии.*

Коэффициент детерминации модели, который отражает долю изменения зависимой переменной, объясняемую рассмотренной моделью  $R^2 = 0,28$ ; скорректированный коэффициент детерминации  $R^2 = 0,09$ .

Далее воспользуемся  $F$ -критерием Фишера. Выдвигаем гипотезу  $H_0$  о неадекватности модели, которая выполняется в том случае, если  $F < F_{кр}$ , где  $F_{кр}(1;16) = 4,49$  — табличное значение  $F$ -критерия с уровнем значимости  $\alpha = 0,05$ ;  $F = 0,78$  — значение  $F$ -критерия для рассматриваемой модели. Таким образом, гипотеза  $H_0$  принимается. Делаем вывод о неадекватности модели.

Но, подсчитав ошибку аппроксимации  $A = 3,53\%$ , можно сделать вывод об адекватности модели. В этом случае модель неудовлетворительно описывает имеющиеся экспериментальные данные.

Для того, чтобы убедиться в наличии мультиколлинеарности, найдём определитель обратной матрицы произведения транспонированной матрицы независимых переменных  $\left( \frac{Y}{N}, \left( \frac{Y}{N} \right)^2, \left( \frac{Y}{N} \right)^*, D_1, D_2 \right)$  на исходную матрицу независимых переменных:

$$\det \left[ \left( X^T X \right)^{-1} \right] = 5,16094 \cdot 10^{-25}.$$

Полученный определитель близок к 0, что свидетельствует о наличии мультиколлинеарности.

### 1.3. Тест Уайта

Суть этого теста состоит в следующем. Наличие гетероскедастичности говорит о том, что дисперсии ошибок зависят от регрессоров, в то время как она должна проявляться в остатках регрессии исходной модели [3].

*Результаты теста Уайта для Томской области.*

$H_0$ : гомоскедастичность присутствует.  $H_1$ : гомоскедастичность отсутствует. Тестовая статистика = 15,0255.  $\chi^2(0,05;10) = 27,587 > 15,0255$ , отсюда следует, что выдвигаемая гипотеза принимается, в модели отсутствует гетероскедастичность.

*Результаты теста Уайта для Новосибирской области.*

$H_0$ : гомоскедастичность присутствует.  $H_1$ : гомоскедастичность отсутствует. Тестовая статистика = 5,1729.  $\chi^2(0,05;10) = 27,587 > 5,1729$ , отсюда следует, что выдвигнутая гипотеза принимается, в модели отсутствует гетероскедастичность.

#### 1.4. Тест Харке-Бера

Оценки качества уравнения регрессии производились из условия того, что шум был распределен по нормальному закону. Убедимся в этом, используя тест Харке-Бера — это статистический тест, который позволяет проверить ошибки наблюдений на нормальность путем сравнения их коэффициентов асимметрии и эксцесса с моментами гауссовского распределения, у которого  $S = 0$ ,  $K = 3$  [4].

$$H_0: S = 0, K = 3$$

$$H_1: S \neq 0, K \neq 3$$

*Результаты теста Харке-Бера для Томской области.*

Тестовая статистика = 3,276.  $\chi^2(0,05;16) = 26,296 > 3,276$ , отсюда следует, что выдвинутая гипотеза принимается, остатки в модели нормальные.

*Результаты теста Харке-Бера для Новосибирской области.*

Тестовая статистика = 5,946.  $\chi^2(0,05;16) = 26,296 > 5,946$ , отсюда следует, что выдвинутая гипотеза принимается, остатки в модели нормальные.

#### 1.5. Выводы для регрессионной модели Н. Леффа

При проверке значимости коэффициентов, общего качества уравнения регрессии и проверки на наличие гомоскедастичности, а также нормальности остатков в регрессионной модели Н. Леффа можно сделать вывод о том, что модель не подходит для конкретных наборов данных (Томская область и Новосибирская область).

Это свидетельствует о том, что данную регрессионную модель нужно использовать с осторожностью, учитывая наличие сильной зависимости между такими переменными как валовой региональный продукт на душу населения и валовой региональный продукт на душу населения в квадрате, а также ВРП содержится в зависимой переменной  $\left(\frac{GDS}{Y}\right)$ .

Для дальнейшей возможности оценить степень зависимости валовых внутренних сбережений от различных факторов, мной было рассмотрено логарифмическое обобщение модели Н. Леффа.

### 2. Логарифмическая модель для эмпирических результатов Н. Леффа

Артур Голдбергер оценил пару уравнений вида

$$y_1 = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \varepsilon_1,$$

$$y_2 = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \varepsilon_1,$$

где  $y_1 = \ln\left(\frac{S}{Y}\right)$ ,  $S$  — внутренние сбережения,  $Y$  — ВРП;  $y_2 = \ln\left(\frac{S}{N}\right)$  — логарифм сбережений на душу населения;  $a_0, b_0$  — свободные члены, выражающие влияние неучтенных факторов;  $x_1 = \ln\left(\frac{Y}{N}\right)$  — логарифм дохода на душу населения;  $x_2 = g$  — темп роста дохода на душу населения;  $x_3 = \ln D_1$  — логарифм процентного соотношения населения младше 14 лет;  $x_4 = \ln D_2$  логарифм процентного соотношения населения старше 65 лет.

Сделав замену  $\frac{S}{N} \equiv \frac{S}{Y} \frac{Y}{N}$ , получим модель

$$y_2 = a_0 + (1 + a_1)x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \varepsilon_1. \quad (2)$$

Тем самым, регрессия  $y_2$  по переменным  $x_i$ ,  $i = \overline{1,4}$  должна давать такие же коэффициенты и такие же остатки, которые имеют место, когда  $y_1$  регрессирует  $x_i$ , за исключением коэффициента  $x_1$ , который нужно увеличить на 1 [5].

### 2.1. Исследование для Томской области

Рассмотрим период с 1996 по 2013 гг.

Таблица 3

Значение коэффициентов и  $t$ -критерия для модели (2)

Параметр	Значение коэффициента	Значение $t$ -критерия
Свободный член	-1,36807	-1,09635
Ln дохода на душу населения	1,79276	7,68475
Темп роста дохода на душу населения	0,00581	1,84619
Ln процентного соотношения населения младше 14 лет	0,27138	1,05317
Ln процентного соотношения населения старше 65 лет	0,18835	0,33503

*Значимость коэффициентов.*

Для определения значимости воспользуемся  $t$ -критерием Стьюдента. Рассмотрим гипотезу  $H_0 : a_0, a_1, a_2, a_3, a_4 = 0$ , т.е. эти коэффициенты не являются статистически значимыми. Гипотеза будет принята в том случае, если  $|t| > |t_T|$ , где  $t_T = 1,78$  — табличное значение  $t$ -критерия с уровнем значимости  $\alpha = 0,05$  и числом степеней свободы равным 16;  $t$  — значение, вычисленное для каждого из коэффициентов.

Таким образом, незначимыми являются все коэффициенты, кроме  $a_1, a_2$  (Ln дохода на душу населения, темп роста дохода на душу населения).

*Общее качество уравнения регрессии.*

Коэффициент детерминации модели, который отражает долю изменения зависимой переменной, объясняемую рассмотренной моделью  $R^2 = 0,98$ ; скорректированный коэффициент детерминации  $R^2 = 0,9$ .

Далее воспользуемся  $F$ -критерием Фишера. Выдвигаем гипотезу  $H_0$  о неадекватности модели, которая выполняется в том случае, если  $F < F_{кр}$ , где  $F_{кр}(1;16) = 4,49$  — табличное значение  $F$ -критерия с уровнем значимости  $\alpha = 0,05$ ;  $F = 274,1249$  — значение  $F$ -критерия для рассматриваемой модели. Таким образом, гипотеза  $H_0$  отклоняется. Делаем вывод об адекватности модели.

*Результаты теста Уайта.*

$H_0$ : гомоскедастичность присутствует.  $H_1$ : гомоскедастичность отсутствует. Тестовая статистика = 16,78.  $\chi^2(0,05;14) = 23,685 > 16,78$ , отсюда следует, что выдвинутая гипотеза принимается, в модели отсутствует гетероскедастичность. В модели возможно присутствие эффекта мультиколлинеарности:  $r_{a_1, a_2} = 0,99$ ,  $r_{a_2, a_1} = 0,99$ .

*Результаты теста Харке-Бера для Томской области.*

$$H_0 : S = 0, K = 3$$

$$H_1 : S \neq 0, K \neq 0$$

Тестовая статистика = 1,034,  $p$ -value = 0,59638.  $\chi^2(0,05;16) = 26,296 > 1,034$ , отсюда следует, что выдвинутая гипотеза принимается, остатки в модели нормальные.

Для того, чтобы убедиться в наличии мультиколлинеарности, найдём определитель обратной матрицы произведения транспонированной матрицы независимых переменных  $a_1, a_2, a_3, a_4$  на исходную матрицу независимых переменных:

$$\det \left[ \left( X^T X \right)^{-1} \right] = 2,51281 \cdot 10^{-5}.$$

Полученный определитель близок к 0, что свидетельствует о наличии мультиколлинеарности.

## 2.2. Исследование для Новосибирской области

Рассмотрим период с 1996 по 2013 гг.

Таблица 4

Значение коэффициентов и  $t$ -критерия для модели (2)

Параметр	Значение коэффициента	Значение $t$ -критерия
Свободный член	-22,298	-2,8986
Ln дохода на душу населения	-0,9007	-9,5726
Темп роста дохода на душу населения	0,0108	1,01544
Ln процентного соотношения населения младше 14 лет	-2,9315	-4,88328
Ln процентного соотношения населения старше 65 лет	10,1674	5,3737

### Значимость коэффициентов.

Для определения значимости воспользуемся  $t$ -критерием Стьюдента. Рассмотрим гипотезу  $H_0: a_0, a_1, a_2, a_3, a_4 = 0$ , т.е. эти коэффициенты не являются статистически значимыми. Гипотеза будет принята в том случае, если  $|t| > |t_T|$ , где  $t_T = 1,78$  — табличное значение  $t$ -критерия с уровнем значимости  $\alpha = 0,05$  и числом степеней свободы равным 16;  $t$  — значение, вычисленное для каждого из коэффициентов.

Таким образом, значимыми являются все коэффициенты, кроме  $a_2$  (темп роста дохода на душу населения).

### Общее качество уравнения регрессии.

Коэффициент детерминации модели, который отражает долю изменения зависимой переменной, объясняемую рассмотренной моделью  $R^2 = 0,908$ ; скорректированный коэффициент детерминации  $R^2 = 0,88$ .

Далее воспользуемся  $F$ -критерием Фишера. Выдвигаем гипотезу  $H_0$  о неадекватности модели, которая выполняется в том случае, если  $F < F_{кр}$ , где  $F_{кр}(1;16) = 4,49$  — табличное значение  $F$ -критерия с уровнем значимости  $\alpha = 0,05$ ;  $F = 32,287$  — значение  $F$ -критерия для рассматриваемой модели. Таким образом, гипотеза  $H_0$  отклоняется. Делаем вывод об адекватности модели.

Также, подсчитав ошибку аппроксимации  $A = 9,58\%$ , можно сделать вывод об адекватности модели.

### Результаты теста Уайта.

$H_0$ : гомоскедастичность присутствует.  $H_1$ : гомоскедастичность отсутствует. Тестовая статистика = 15,3114.  $\chi^2(0,05;14) = 23,685 > 15,3114$ , отсюда следует, что выдвинутая гипотеза принимается, в модели отсутствует гетероскедастичность.

В модели возможно отсутствует эффект мультиколлинеарности, т.к. парные коэффициенты корреляции не превышают 0.57.

### Результаты теста Харке-Бера для Новосибирской области.

$$H_0: S = 0, K = 3$$

$$H_1: S \neq 0, K \neq 0$$

Тестовая статистика = 0,288, p-value = 0,86584.  $\chi^2(0,05;16) = 26,296 > 0,288$ , отсюда следует, что выдвинутая гипотеза принимается, остатки в модели нормальные.

Для того, чтобы убедиться в наличии мультиколлинеарности, найдём определитель обратной матрицы произведения транспонированной матрицы независимых переменных  $a_1, a_2, a_3, a_4$  на исходную матрицу независимых переменных:

$$\det \left[ (X^T X)^{-1} \right] = 3,90351 \cdot 10^{-7}.$$

Полученный определитель близок к 0, что свидетельствует о наличии мультиколлинеарности.

### **2.3. Выводы для логарифмической модели Н. Леффа**

При проверке значимости коэффициентов, общего качества уравнения регрессии и проверки на наличие гомоскедастичности, а также нормальности остатков в логарифмической модели Н. Леффа можно сделать вывод о том, что модель не подходит для конкретных наборов данных (Томская область и Новосибирская область).

Это свидетельствует о том, что данную модель нужно использовать с осторожностью, учитывая наличие мультиколлинеарности.

### **Заключение**

В настоящей работе рассмотрены регрессионные модели зависимости экономических факторов, положительно влияющих на рост валового регионального продукта и валовые внутренние сбережения таких регионов России, как Томская и Новосибирская области. Проведен анализ возрастных особенностей производства и потребления и выявлено их влияние на величину сбережений и темпы экономического роста.

### **ЛИТЕРАТУРА**

1. Саградов А.А. Экономическая демография/ М.: Изд-во ИНФРА-М, 2010. – с. 8.
2. Саградов А.А. Экономическая демография/ М.: Изд-во ИНФРА-М, 2010.– с. 164.
3. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс/ Учеб. – М.:Дело, 2004. – с. 177–178.
4. Википедия. Свободная энциклопедия. [Электронный ресурс]. URL: [https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D1%81%D1%82\\_%D0%A5%D0%B0%D1%80%D0%BA%D0%B5\\_%E2%80%94%D0%91%D0%B5%D1%80%D0%B0](https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D1%81%D1%82_%D0%A5%D0%B0%D1%80%D0%BA%D0%B5_%E2%80%94%D0%91%D0%B5%D1%80%D0%B0)
5. Arthur S. Golgberger. Dependency rates and savings rates: further comment// American Economic Review, 1969, 59(December), p.232–233.

## **ИСПОЛЬЗОВАНИЕ ДОПОЛНИТЕЛЬНОЙ ИНФОРМАЦИИ В ОЦЕНКЕ УДОВЛЕТВОРЕННОСТИ НАСЕЛЕНИЯ УСЛУГАМИ МЕТОДОМ ДЕТЕРМИНАЦИОННОГО АНАЛИЗА**

**А.Л. Сивушина**

Томский государственный университет  
ansivushina@gmail.com

### **Введение**

В современном и постоянно развивающемся обществе всё чаще появляется необходимость в качественной оценке степени удовлетворенности населения услугами, предоставляемыми различными организациями, будь то органы власти, частные фирмы и т.д. Оценивание производится по данным, полученным из проведённых обследований потребителей услуг. Однако существует некоторая сложность в анализе таких данных, которая заключается в том, что, как правило, данные являются номинальными, т.е. имеют качественную природу, и именно для таких данных применим детерминационный анализ [1].

В данной статье поставлена задача оценивания параметров детерминационного анализа на основании того, что имеется некоторая дополнительная информация о признаках имеющейся выборки. Строятся оценки с использованием этой информации и производится анализ влияния учёта дополнительной информации на качество оценивания признаков, приводятся примеры применения оценок в анализе реальных данных.

### 1. Постановка задачи

Необходимо определить по данным формализованного интервью «лицом к лицу» удовлетворенность населения качеством медицинских услуг, которые оказывают лечебные учреждения района. Данные получены из ответов респондентов на вопрос:

«Удовлетворяет ли Вас качество медицинской помощи, которую оказывают лечебные учреждения района / Вашего поселения?»

1. Удовлетворяет
2. Не удовлетворяет (почему?) \_\_\_\_\_
3. Затрудняюсь ответить
4. В моей семье никто не пользуется услугами этих лечебных учреждений

Для анализа была предоставлена выборка из 567 человек. Она была отсортирована по признакам: пол, возраст, оценка респондентом качества предоставляемой услуги.

### 2. Ключевые понятия ДА

Первичные эмпирические данные, получаемые в ходе социальных исследований, представляют собой веер отображений, т.е. совокупность отображений вида  $E \rightarrow X_i$ ,  $i = \overline{1, n}$ , где  $E$  — множество объектов,  $X_i$  — множество значений переменной  $x_i$ ,  $i$  — индекс, нумерующий переменные, участвующие в эмпирическом исследовании (см. рис. 1).

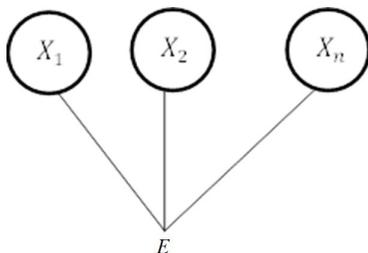


Рис. 1. Веер отображений

Метод ДА служит для поиска и интерпретации взаимосвязей между признаками или их группами на основе эмпирических данных исследований: опросов, анкет, и представляет собой вариант исчисления условных частот, или другими словами нахождения долей реализации объектов в выборке, которые содержатся в таблице сопряжённости.

Пусть у некоторой совокупности объектов выделены два признака  $X$  и  $Y$ , где у множества признаков  $X$  выделен признак  $a$ , а у множества  $Y$  — признак  $b$ . Тогда на языке детерминационного анализа детерминация  $a \rightarrow b$  с интенсивностью  $I(a \rightarrow b)$  — это не что иное, как высказывание «если  $a$ , то  $b$ », которому приписывается *интенсивность*, отражающая его *точность* или *истинность*.

Теперь рассмотрим случай, что среди респондентов, демонстрирующих тип поведения  $b$ , выделена доля тех, кто обладает некоторым свойством  $a$ . Тогда на языке детерминационного анализа это означает, что детерминация  $a \rightarrow b$  имеет ёмкость  $C(a \rightarrow b)$ . Она измеряет долю случаев реализации поведения  $b$ , которая «объясняется»

высказыванием «из  $a$  следует  $b$ ». Ёмкость  $C(a \rightarrow b)$  отражает полноту этой детерминации.

### 3. Оценивание интенсивности и ёмкости с учётом дополнительной информации

Рассмотрим задачу статистического оценивания характеристик точности и полноты в терминах событий  $A = \{X = a\}$ ,  $B = \{Y = b\}$ ,  $AB = \{X = a, Y = b\}$  и соответствующих вероятностей  $P(A)$ ,  $P(B)$ ,  $P(AB)$ . Тогда интенсивность и ёмкость детерминации  $A \rightarrow B$  примут вид:

$$I(A \rightarrow B) = P(B|A) = \frac{P(AB)}{P(A)}, \quad C(A \rightarrow B) = P(A|B) = \frac{P(AB)}{P(B)}. \quad (1)$$

Пусть случайной выборке объёма  $n$  соответствуют эмпирические вероятности:

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n I_i(A), \quad P_n(B) = \frac{1}{n} \sum_{i=1}^n I_i(B), \quad P_n(AB) = \frac{1}{n} \sum_{i=1}^n I_i(AB), \quad (2)$$

$I(\cdot)$  — индикаторная функция соответствующего события. Подставляя (1) в (2), получим статистические оценки функций интенсивности и ёмкости:

$$J_n = \frac{P_n(AB)}{P_n(A)}, \quad G_n = \frac{P_n(AB)}{P_n(B)}. \quad (3)$$

Объекты, попавшие в рассматриваемую выборку, могут быть иметь и другие наблюдаемые признаки. Допустим, что признак  $Z$  один из них. Тогда событие  $C$  принимает значение  $Z = c$ , для которого известна вероятность  $P(C)$ . Другими словами, пусть известна часть объектов генеральной совокупности, у которых  $Z$  имеет заданное значение  $c$ . Используем эту информацию для построения оценок вероятностей других рассматриваемых событий, а также для построения характеристик интенсивности и ёмкости. Рассмотрим полученные оценки, которые основаны на использовании формулы полной вероятности [2]:

$$\hat{P}_n(B) = P_n(B|C)P(C) + P_n(B|\bar{C})P(\bar{C}) = \frac{P_n(BC)}{P_n(C)}P(C) + \frac{P_n(B\bar{C})}{P_n(\bar{C})}P(\bar{C}), \quad (4)$$

где  $\bar{C}$  — событие, противоположное событию  $C$ . Применим формулу полной вероятности для оценивания  $P(A)$ :

$$\hat{P}_n(A) = P_n(A|C)P(C) + P_n(A|\bar{C})P(\bar{C}) = \frac{P_n(AC)}{P_n(C)}P(C) + \frac{P_n(A\bar{C})}{P_n(\bar{C})}P(\bar{C}). \quad (5)$$

Методом подстановки получим оценки характеристик интенсивности и ёмкости. Этот метод приводит к следующей оценке интенсивности с учётом дополнительной информации:

$$\hat{J}_n = J_n(\hat{P}_n(AB), \hat{P}_n(A)) = \frac{\hat{P}_n(AB)}{\hat{P}_n(A)} = \frac{\frac{P_n(ABC)}{P_n(C)}P(C) + \frac{P_n(ABC\bar{C})}{P_n(\bar{C})}P(\bar{C})}{\frac{P_n(AC)}{P_n(C)}P(C) + \frac{P_n(A\bar{C})}{P_n(\bar{C})}P(\bar{C})}. \quad (6)$$

Оценка ёмкости с учётом дополнительной информации:

$$\hat{G}_n = G_n(\hat{P}_n(AB), \hat{P}_n(B)) = \frac{\hat{P}_n(AB)}{\hat{P}_n(B)} = \frac{\frac{P_n(ABC)}{P_n(C)}P(C) + \frac{P_n(ABC\bar{C})}{P_n(\bar{C})}P(\bar{C})}{\frac{P_n(BC)}{P_n(C)}P(C) + \frac{P_n(B\bar{C})}{P_n(\bar{C})}P(\bar{C})}. \quad (7)$$

Асимптотическая дисперсия оценки интенсивности имеет вид:

$$\sigma_1^2 = \frac{P(AB)}{P^2(A)} \left( 1 - \frac{P(AB)}{P(A)} \right). \quad (8)$$

Исследуя асимптотические свойства оценок, можно прийти к выводу, что использование дополнительной информации уменьшает асимптотическую дисперсию на величину:

$$\delta_J = \frac{1}{P(C)(1-P(C))} \left[ \frac{P(ABC) - P(AB)P(C)}{P(A)} - \frac{P(AB)}{P^2(A)} P(AC) - P(A)P(C) \right]^2. \quad (8)$$

Выигрыш от включения в анализ дополнительной информации при оценивании точности выражается с помощью соотношения дисперсий:

$$W_1 = \frac{\sigma_1^2}{\sigma_2^2} = 1 - \frac{[P(A)(P(ABC) - P(AB)P(C)) - P(AB)(P(AC) - P(A)P(C))]^2}{P(A)P(C)P(\bar{C})P(AB)(P(A) - P(AB))}. \quad (9)$$

Таким образом, значение  $W_1$ , близкое к нулю, говорит о сильном влиянии дополнительной информации на точность оценки функции интенсивности; чем ближе значение к 1, тем это влияние слабее.

Аналогичным образом производится анализ асимптотических свойств оценки ёмкости, получим:

$$\bar{\sigma}_1^2 = \frac{P(AB)}{P^2(B)} \left( 1 - \frac{P(AB)}{P(B)} \right), \quad (10)$$

$$\delta_G = \frac{1}{P(C)(1-P(C))} \left[ \frac{P(ABC) - P(AB)P(C)}{P(B)} - \frac{P(AB)(P(AC) - P(A)P(C))}{P^2(B)} \right]^2.$$

Выигрыш от включения в анализ дополнительной информации при оценивании ёмкости выражается с помощью соотношения асимптотических дисперсий:

$$W_2 = \frac{\bar{\sigma}_1^2}{\bar{\sigma}_2^2} = 1 - \frac{[P(B)(P(ABC) - P(AB)P(C)) - P(AB)(P(BC) - P(B)P(C))]^2}{P(A)P(C)P(\bar{C})P(AB)(P(B) - P(AB))}. \quad (11)$$

Значение  $W_2$ , близкое к нулю, говорит о сильном влиянии дополнительной информации на точность оценки ёмкости; чем ближе значение к 1, тем это влияние слабее.

Также при оценивании функций истинности и ёмкости в случае независимости событий  $A, B$  и  $C$  величины  $W_1$  и  $W_2$  равны 1 и выигрыша в точности оценивания нет.

Теперь предположим, что событие  $A$  — достоверное, тогда положим  $P(A) = 1$ ,  $P(AB) = P(B)$ ,  $P(AC) = P(C)$ ,  $P(ABC) = P(BC)$ , имеем:

$$\sigma_2^2 = P(B)(1-P(B)) - \frac{[P(BC) - P(B)P(C)]^2}{P(C)(1-P(C))}, \quad \delta_1 = \frac{[P(BC) - P(B)P(C)]^2}{P(C)(1-P(C))}, \quad (11.1)$$

$$V_1 = 1 - \frac{[P(BC) - P(B)P(C)]^2}{P(C)(1-P(C))P(B)(1-P(B))}. \quad (11.2)$$

Полученные ниже соотношения применяются для анализа точности оценок характеристик.

#### 4. Использование дополнительной информации в оценке удовлетворённости населения услугами методом ДА

В данной работе нас интересует удовлетворенность населения Бакчарского района Томской области качеством предоставляемых медицинских услуг по месту проживания. Был произведён опрос населения, на основании которого были сделаны выводы и заключения. Объектом нашего наблюдения является респондент. Выделим два интересующих нас признака:  $X$  — пол респондента ( $X = a$  — респондент женщина,  $X = \bar{a}$  — мужчина),  $Y$  — оценка респондентом качества медицинского обслуживания ( $Y = b$  — респондент удовлетворён качеством предоставляемой услуги,  $Y = \bar{b}$  — респондент не удовлетворён качеством предоставляемой услуги).

Детерминация  $a \rightarrow b$  — это высказывание: «если респондент женщина, то она удовлетворена качеством медицинских услуг», данному высказыванию приписывается интенсивность  $I(a \rightarrow b)$ , отражающая его точность или истинность, а также ёмкость  $S(a \rightarrow b)$ , которая измеряет долю случаев реализации положительного ответа респондента, которая объясняется высказыванием «из  $a$  следует  $b$ », т.е. отражает полноту детерминации.

##### Пример расчётов реальных данных.

В одном из районов Томской области было опрошено 567 респондентов, и по результатам опроса была составлена сводная таблица (таблица сопряженности) двух признаков (табл. 1). Также априори известно, что численность всех избирателей составила 10 593, среди них женщин — 7 344, а мужчин — 3 249, таким образом, доля избирательниц —  $7\,344 / 10\,593 = 0,6933$ , мужчин —  $3\,249 / 10\,593 = 0,3067$ .

Таблица 1

Таблица сопряженности по переменным  $X$  и  $Y$  (объём выборки  $n$  — 567 человек)

$Y \setminus X$	женский ( $A$ )	мужской(не $A$ )	Общий итог
не удовлетворяет	171	187	358
удовлетворяет	121	88	209
общий итог	292	275	567

Из табл. 1 следует:

$$P_n(B) = \frac{209}{567} \approx 0.3686, \quad P_n(\bar{B}) = \frac{358}{567} \approx 0.6314,$$

$$P_n(A) = \frac{292}{567} \approx 0.515, \quad P_n(\bar{A}) = \frac{275}{567} \approx 0.485, \quad P_n(AB) = \frac{121}{567} \approx 0.2134, \quad (12)$$

$$J_n = \frac{P_n(AB)}{P_n(A)} = \frac{121}{292} \approx 0.4144, \quad G_n = \frac{P_n(AB)}{P_n(B)} = \frac{121}{209} \approx 0.5789. \quad (13)$$

Таким образом, удовлетворено 36,86% порошенных респондентов. Также эту величину можно уточнить, учитывая имеющуюся информацию о долях избирателей женщин и мужчин. В этом случае признак  $Z$  совпадает с признаком  $X (C = A)$ ,  $P(A) = 0,6933$ , значит  $P(A) = P(C) = 0,6933$ . Эту информацию учтём, положим в ней  $C = A$ . Получим уточнённую оценку  $\hat{P}_n(B) \approx 0.3626 = 36.26\%$ . Заметим, что уточнённая оценка практически не изменилась, также знание о долях избирателей по полу ничего не даст, т.к. при  $C = A$  из (9) следует, что  $\hat{J}_n = J_n$  и  $W_1 = 1$ .

Для вычисления выигрыша в точности воспользуемся формулами (8)–(11). В связи с тем, что в реальности далеко не все вероятности известны, рассмотрим два способа оценивания и подведем итоги.

1 способ.

На основании вычисленных значений долей (12), (13) и формул (11.1), (11.2) получим:

$$\delta_{n1} = 0.0027, V_{n1} = 0.988. \quad (14)$$

2 способ.

Основан на оценках вероятностей, построенных с учётом дополнительной информации:

$$\hat{P}_n(AB) = \frac{P_n(AB)}{P_n(A)} P(A) \approx 0.2873, \hat{P}_n(A) = P(A) \approx 0.6933, \hat{G}_n \approx 0.7923, \quad (15)$$

$$\hat{\delta}_{n1} = 0.0061, \hat{V}_{n1} = 0.9735, \hat{W}_{n2} = 0.7186. \quad (16)$$

Сравнивая результаты (12)–(16), заметим, что значения вероятностей немного изменились, и значение ёмкости с учётом дополнительной информации заметно изменилось. Однако значения характеристик точности практически совпадают. Значения величин  $V_{n1}, \hat{V}_{n1}, \hat{W}_{n2}$  говорят о том, что учёт знания долей избирателей по полу при оценивании вероятностей позволяет уменьшить объём выборки на 2–3% для достижения одинаковой точности с обычными оценками, а при оценивании ёмкости — на 28%.

Теперь, пусть признак  $Z$  означает возраст респондента, и значение  $C$  есть возрастной интервал (18–34) лет. Известно, что доля избирателей рассматриваемого поселения в таком возрастном интервале равна 0,327, т.е. число таких избирателей равно 3 464 и  $P(C) = P\{Z = c\} = 0,327$ . Рассмотрим таблицу сопряжённости для трёх признаков  $X, Y, Z$ :

Таблица 2  
Таблица сопряжённости по переменным  $X, Y$  и  $Z$  (объём выборки  $n = 567$  человек)

$Z \backslash XY$	женский	женский	мужской	мужской	Общий итог
	удовлетворяет	не удовлетворяет	удовлетворяет	не удовлетворяет	
18–34 лет	38	48	28	55	169
не 18–34 лет	83	123	60	132	398
общий итог	121	171	88	187	567

Применим вышеприведённые формулы для расчёта оценок вероятностей, ёмкости и интенсивности с учётом дополнительной информации о доли избирателей по возрасту (поемим их волной и рядом с оценкой дадим её точность). Имеем:

$$\begin{aligned} \tilde{P}_n(B) &\approx 0.3695 = 36.95\%, \tilde{V}_{n1} = 0.9993, \\ \tilde{P}_n(A) &\approx 0.5147 = 51.47\%, \tilde{V}_{n1} = 0.9994, \tilde{P}_n(AB) \approx 0.21327 = 21.32\%, \\ \tilde{J}_n &\approx 0.4143, \tilde{W}_{n1} = 0.9999, \tilde{G}_n \approx 0.5771, \tilde{W}_{n2} = 0.99999. \end{aligned} \quad (17)$$

Из полученных вычислений следует, что в данной реализации влияние возрастной структуры генеральной совокупности на точность оценивания очень слабое. Оценки вероятностей изменили свои значения, что даёт возможность строить другие доверительные интервалы с одинаковой вероятностью.

Детерминация  $c \rightarrow b$  есть не что иное, как высказывание «если возраст респондента в интервале 18–34 лет, то он удовлетворён качеством предоставляемой услуги». Интенсивность и ёмкость этой детерминации по данным табл. 2 равны соответственно:

$$J_n(c \rightarrow b) = \frac{P_n(CB)}{P_n(C)} = \frac{66}{169} \approx 0.3905, \quad (18)$$

$$G_n(c \rightarrow b) = \frac{P_n(CB)}{P_n(B)} = \frac{66}{209} \approx 0.3158. \quad (19)$$

Учёт информации по возрасту приводит к результату:

$$\bar{J}_n(c \rightarrow b) = \frac{\bar{P}_n(CB)}{\bar{P}_n(C)} = J_n(c \rightarrow b), \quad \bar{G}_n(c \rightarrow b) \approx 0.3456. \quad (20)$$

Таким образом, интенсивность не изменилась, а ёмкость изменилась в сторону увеличения.

### Заключение

В настоящее время наблюдается непрерывное изменение в изменении предпочтений и образа жизни людей. Поэтому, если организация желает закрепиться на долгосрочный период, не терять клиентов, предоставлять лучшее качество обслуживания, то ей необходимо постоянное исследование потребительских ожиданий и представлений.

Отметим, что оценка удовлетворенности при помощи детерминационного анализа применима во всех сферах социально-экономических услуг и при различных методах и способах сбора качественной информации [3,4].

В работе показано, каким образом можно привлекать дополнительную информацию о долях объектов генеральной совокупности по признакам и их значениям для повышения точности оценивания различных характеристик. Рассмотренный пример, основанный на реальных данных, позволил увидеть возможности подхода к построению оценок. Подход допускает обобщение на большее число значений признаков, чем рассмотрено в работе. Показано, что существенное повышение точности не всегда происходит. Формулы, характеризующие точность оценивания, дают возможность подбирать признаки, их комбинацию, конкретные значения признаков, которые приводят к существенному повышению точности при фиксированном объёме наблюдений или сокращению объёма выборки при заданной точности. Асимптотическая нормальность оценок с учётом информации позволяет строить доверительные интервалы, которые могут быть значительно уже по сравнению с обычными оценками.

Автор благодарит своего научного руководителя Ю.Г. Дмитриева за внимание к работе и ценные замечания.

### ЛИТЕРАТУРА

1. *Чесноков С.В.* Детерминационный анализ социально-экономических данных. М: Наука, 1982. 214 с.
2. *Дмитриев Ю.Г., Курицина С.В.* Об использовании дополнительной информации в статистическом оценивании параметров детерминационного анализа // Вестник ТГУ. Управление, вычислительная техника и информатика. 2012. №3(20) – С. 42–49.
3. *Victor Lorin Purcarea et al.* The Assessment of Perceived Service Quality of Public Health Care Services in Romania Using the SERVQUAL Scale // Procedia Economics and Finance 6 (2013) 573–585.
5. *S.E. Martynova, O.G. Maslennikova* The «service» model of the competences of the municipal employee as the basis of vocational training: russian experience// Procedia – Social and Behavioral Sciences 00 (2014) 000–000

# СЕКЦИЯ III. ПРИКЛАДНОЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ

## ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ДВУХФАЗНОЙ RQ-СИСТЕМЫ

А.А. Анисимова, А.Н. Моисеев

Томский государственный университет  
siberienne94@yandex.ru, moiseev.tsu@gmail.com

### Введение

В последнее время все большую роль стали играть компьютерные и телекоммуникационные системы, которые подходят под определение систем массового обслуживания [1]: они предназначены для удовлетворения массовых запросов на выполнение каких-либо услуг. Классические модели систем массового обслуживания подразумевают, что заявка, застав обслуживающий прибор занятым, встает в очередь или покидает систему. В данной статье рассматривается модель, в которой заявка в такой ситуации уходит в источник повторных вызовов, откуда через случайные моменты времени пытается вновь встать на обслуживание. Такие модели описываются в виде систем массового обслуживания с повторными вызовами или RQ-систем (Retrial queueing system) [2].

Одной из важнейших характеристик RQ-систем является распределение вероятностей числа заявок в источнике повторных вызовов. К сожалению, получить аналитическое решение относительно данного распределения слишком сложно, поэтому удобнее разработать программную модель, имитирующую поведение системы.

### 1. Описание модели

Рассмотрим двухфазную RQ-систему (рис. 1).

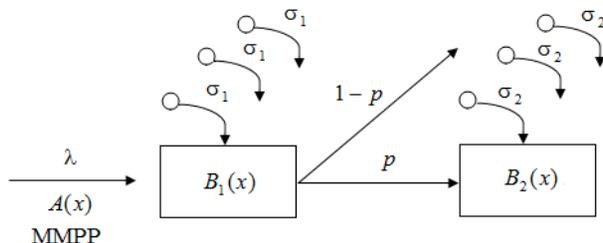


Рис. 1. Модель двухфазной RQ-системы

В данной работе рассматриваются модели с входящими потоками заявок трех видов: простейшим, рекуррентным или ММРР. Поступив в систему, заявка вначале проходит первую фазу, которая содержит прибор, обслуживающий заявку по заданному закону распределения  $B_1(x)$ , и источник повторных вызовов (ИПВ). Если заявка заставит прибор занятым, она отправляется в ИПВ, откуда через случайные моменты времени, имеющие экспоненциальное распределение с параметром  $\sigma_1$ , предпринимает попытки вновь обратиться за обслуживанием. После обслуживания на первой фазе заявка с вероятностью  $p$  отправляется на вторую фазу, которая содержит прибор с другим распределением времени обслуживания  $B_2(x)$  и ИПВ с параметром времени задержки  $\sigma_2$ . С вероятностью  $(1-p)$  заявка считается обработанной и покидает систему.

## 2. Имитационная модель

Имитационная модель данной системы строится на основе дискретно-событийного подхода, учитывающего, что система изменяет свои характеристики не непрерывно, а в определённые моменты времени, которые называются событиями. Основными типами событий являются:

- поступление заявки в систему;
- поступление заявки из ИПВ;
- завершение обслуживания заявки на приборе и ее передача на другой прибор, либо уход из системы;
- завершение моделирования.

Поступление заявки в систему моделируется в соответствии с выбранным пользователем типом потока. Для простейшего потока интенсивности  $\lambda$  генерируются независимые экспоненциально распределенные интервалы между событиями потока:

$\tau_1 = t_1 - t_0$ ,  $\tau_2 = t_2 - t_1, \dots, \tau_i = t_i - t_{i-1}, \dots$  по формуле  $\tau_i = -\frac{\ln(\alpha)}{\lambda}$ , где  $\alpha$  — случайная величина, равномерно распределенная на отрезке  $(0; 1)$ . Таким образом, каждое последующее событие потока генерируется по формуле:  $t_i = t_{i-1} - \frac{\ln(\alpha)}{\lambda}$ ,  $t_0 = 0$  [3].

ММРР-поток задается матрицей инфинитезимальных характеристик  $\mathbf{Q}$  с элементами  $\|q_{ij}\|_{n \times n}$  и вектором  $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ , содержащим значения интенсивности потока при каждом состоянии управляющей цепи Маркова. Для определения начального состояния вычисляется стационарное распределение вероятностей состояний цепи. Согласно этому распределению определяется начальное состояние  $i$ . По формуле

$T_i = \frac{\ln(\alpha)}{q_{ii}}$  рассчитывается время пребывания цепи Маркова в  $i$ -ом состоянии, в течение

которого генерируются события простейшего потока с интенсивностью  $\lambda_i$ . По истечении этого времени цепь Маркова должна перейти в следующее состояние. Для его на-

хождения вычисляются переходные вероятности  $p_{ij} = -\frac{q_{ij}}{q_{ii}}$ ,  $i \neq j$ , в соответствии с которыми определяется новая интенсивность наступления событий [3].

В рекуррентном потоке интервалы  $\tau_i$ ,  $i \geq 2$ , независимы и одинаково распределены с функцией распределения  $A(x)$ , а интервал  $\tau_1$  имеет распределение  $A_1(x)$ . Так как для получения адекватных результатов моделирования приходится генерировать очень большое число заявок во входящем потоке (1 млн. и более), этим условием можно пренебречь и моделировать все интервалы, используя функцию  $A(x)$ .

## 3. Приложение для имитационного моделирования

Приложение позволяет выбрать один из трех указанных выше типов входного потока заявок, причем для рекуррентного потока реализованы следующие распределения:

- двухпараметрическое гамма-распределение;
- двухпараметрическое бета-распределение;
- равномерное в заданном интервале;
- детерминированное.

Распределение времени обслуживания на каждой фазе также выбирает пользователь. Оно может быть экспоненциальным или иметь одно из вышеперечисленных распределений. Распределение времени между повторными вызовами имеет экспоненциальное распределение с заданным пользователем параметром.

Архитектура приложения реализована на основе объектной модели [4]. Для реализации дискретно-событийного механизма моделирования введен специальный класс Event, предназначенный для обработки событий внутри модели. Каждое событие связано с определенной заявкой. Класс Call реализует объект заявки, а процесс ее обработки заключается в генерации и последующей передаче элементам системы в зависимости от обрабатываемого события. Связь этих классов показана на рис. 2.

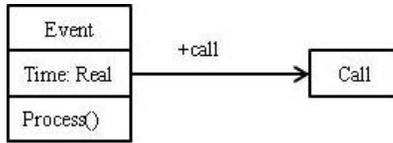


Рис.2. Модель событий

На рис. 3 представлены основные классы предметной области. Класс QueueSimulationModel отвечает за процесс моделирования. Он имеет поле Timer, реализующее таймер модельного времени, который после обработки очередного события сдвигается на момент следующего, и операцию DoStep(), выполняющую один шаг моделирования.

Все события заносятся в отсортированный по времени журнал событий (класс Journal), предназначенный для дискретно-событийного управления модельным временем.

При такой организации один шаг моделирования реализуется следующим образом: модель извлекает из журнала ближайшее событие (операция ReturnEvent()) и определяет его тип, в зависимости от которого переадресует обработку события одному из обрабатывающих элементов.

В общем случае, такие элементы имеют следующие операции:

- GenerateEvent() генерирует события в зависимости от заданного типа входящего потока или распределения времени обслуживания/времени между повторными вызовами;
- Accept() принимает на обслуживание заявку, соответствующую данному событию;
- ProcessEvent() производит обработку события.

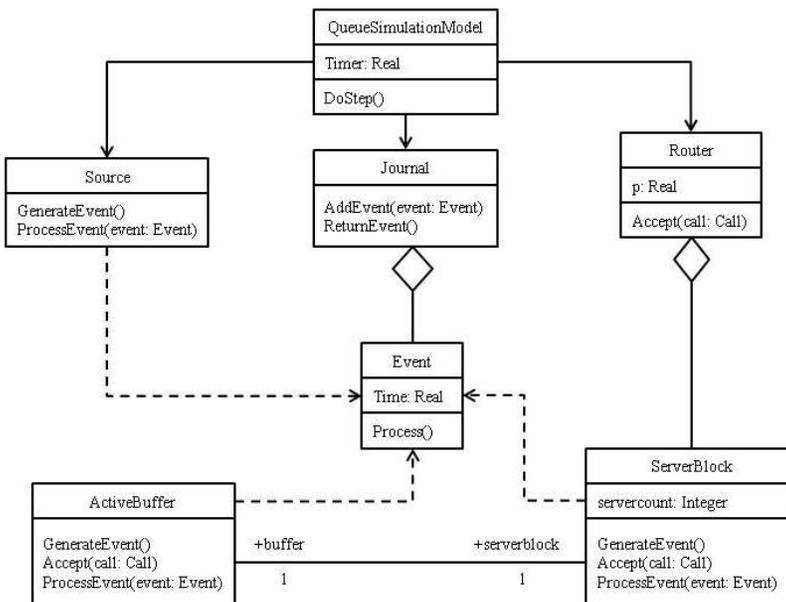


Рис.3. Иерархия элементов модели

Класс Source (источник заявок) генерирует входящий поток событий, а их обработка заключается в передаче соответствующей заявки маршрутизатору Router. Класс Router направляет заявку на очередную фазу, в зависимости от того, была ли она уже обработана, а также с учетом вероятности  $p$  продолжения обслуживания на второй фазе.

Класс ServerBlock реализует блок обслуживающих приборов, соответствующих одной фазе. После того, как заявка начала обслуживание, счетчик serverCount увеличивается на единицу (по завершении обслуживания уменьшается). Если заявка обращается за обслуживанием в тот момент, когда serverCount равен 1, она передается в ИПВ, реализованный посредством класса ActiveBuffer, откуда предпринимает дальнейшие попытки обслужиться.

На рис. 4 представлено главное окно программы. Пользователь задает число заявок, которые должны быть сгенерированы во входящем потоке, распределение входного потока, распределение времени обслуживания и параметры экспоненциального распределения времени ожидания в ИПВ на первой и второй фазе, а также вероятность, с которой заявка может покинуть систему после окончания обслуживания на первой фазе.

Результатом моделирования является двумерное распределение числа заявок в ИПВ первой и второй фазы, который сохраняется в файле. Также программа выводит на экран маргинальные распределения вероятностей числа заявок в ИПВ каждой фазы в виде полигона частот и сохраняет их в виде рядов распределений. Здесь же можно вывести полигон частот для выбранного распределения, чтобы провести сравнение с результатами имитационного моделирования. Программа позволяет рассчитывать вектор математических ожиданий и матрицу ковариаций этих распределений.

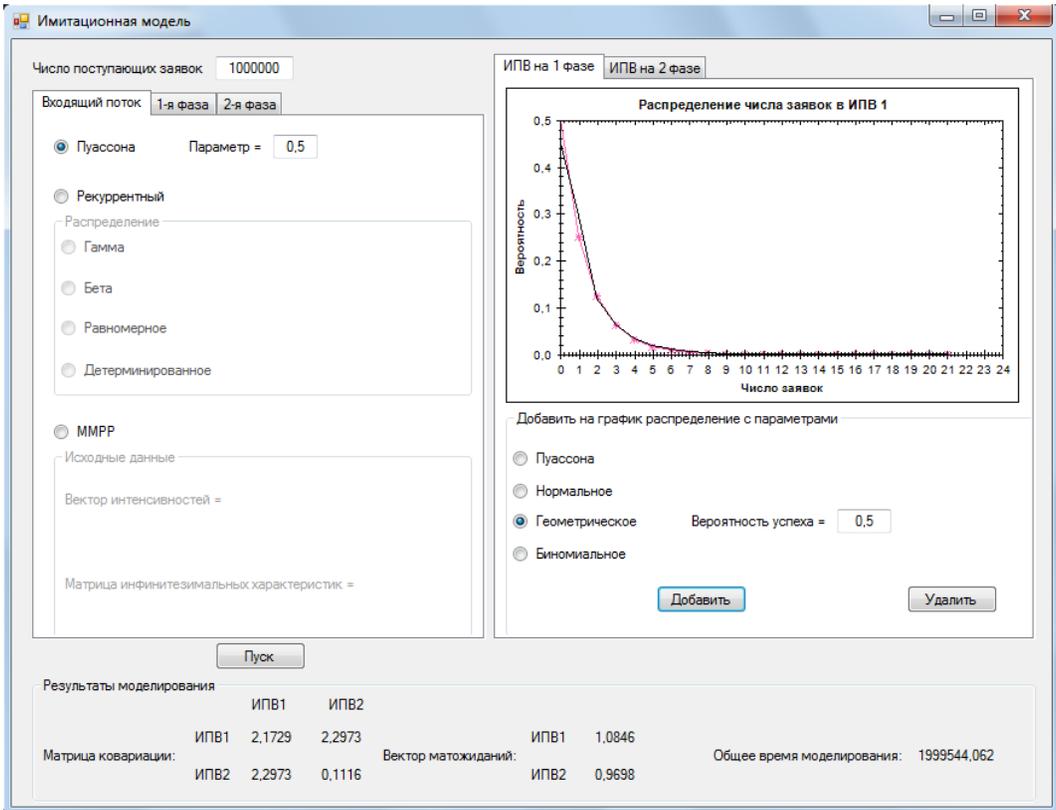


Рис.4. Главное окно программы

## Заключение

В работе представлено описание имитационной модели и приложения для моделирования двухфазной RQ-системы. Разработанное приложение позволяет получить основные вероятностные характеристики для сравнения с аналитическими результатами, полученными для данной системы.

## ЛИТЕРАТУРА

1. Гнеденко Б. В. Введение в теорию массового обслуживания. – 2-е изд., перераб. и доп. / Б.В. Гнеденко, И.Н. Коваленко. – М. : Наука. Гл. ред. физ.-мат. лит., 1987. – 336 с.
2. Artalejo J.R., Gomez-Corral A. Retrial Queueing Systems: A Computational Approach. – Berlin: Springer, 2008. – 267 p.
3. Даммер Д.Д. Имитационное моделирование: учебно-методическое пособие / Д.Д. Даммер, Н.Ю. Марголис, С.А. Цой. – Томск : ТГУ, 2010 – 32 с.
4. Мусеев А.Н., Синяков М.В. Разработка объектно-ориентированной модели системы имитационного моделирования процессов массового обслуживания // Вестник Том. гос. ун-т. Управление, вычислительная техника и информатика. 2010, №1, С. 89–93.

## ИССЛЕДОВАНИЕ СУММАРНОГО ОБЪЁМА ТРЕБОВАНИЙ В БЕСКОНЕЧНОЛИНЕЙНОЙ СМО С РЕКУРРЕНТНЫМ ВХОДЯЩИМ ПОТОКОМ

**В.А. Колбасова, Е.Ю. Лисовская**

*Томский государственный университет*

*vera-kolbasova@mail.ru*

Настоящая работа посвящена проблеме определения характеристик систем массового обслуживания (СМО) требований случайной длины, представляющих собой модели реальных информационных систем с ограниченным объёмом памяти, в которой хранится информация об обрабатываемых сообщениях. Более подробно задачи, возникающие при проектировании реальных информационных систем, моделируемых с помощью указанных СМО, а также примеры решения таких задач и классификация моделей рассмотрены в [1].

Рассмотрим систему массового обслуживания с неограниченным числом приборов, на вход которой поступает рекуррентный поток заявок, заданный функцией распределения вероятностей  $A(x)$  длин интервалов между наступлением событий. Продолжительности обслуживания заявок стохастически независимы, одинаково распределены и имеют экспоненциальное распределение с параметром  $\mu$ . Заявка, поступившая в систему, занимает любой из свободных приборов, завершив обслуживание, она покидает систему. Будем предполагать, что каждое требование характеризуется некоторым объёмом  $\xi_i > 0$ .

Введем обозначение  $V(t) = \sum_{i=0}^{i(t)} \xi_i$  — суммарный объём заявок, находящихся в системе. Объёмы различных требований представляют собой независимые случайные величины с функцией распределения вида  $F(x) = P\{\xi < x\}$ .

Ставится задача определения основных характеристик суммарного объёма заявок в системе.

Пусть  $i(t)$  — число заявок в системе, иначе говоря, число приборов, занятых в момент времени  $t$ . Для суммарного объёма находящихся в системе требований запишем характеристическую функцию в виде

$$H(u) = M \left\{ e^{ju \sum_{i=1}^{i(t)} \xi_i} \right\} = M \left\{ e^{u \sum_{i=1}^{i(t)} \xi_i} \mid P(i(t) = i) \right\} =$$

$$= M \left\{ e^{ju \sum_{i=1}^{i(t)} \xi_i} \right\} P\{i(t) = i\} = \sum_{i=0}^{\infty} M \{ (e^{ju\xi})^i \} P(i, t).$$

Введем следующее обозначение:  $\varphi(u) = M \{ e^{ju\xi} \}$ . Отсюда характеристическую функцию можно представить в виде  $H(u) = \sum_{i=0}^{\infty} (\varphi(u))^i P(i, t) = G(\varphi(u))$ .

Рассмотрим двумерный случайный процесс  $\{z(t), i(t)\}$ , который является марковским с непрерывным временем, где  $z(t)$  — длина интервала от момента времени  $t$  до момента наступления очередного события в рассматриваемом потоке. Для распределения вероятностей двумерного марковского случайного процесса  $\{z(t), i(t)\}$   $P(z, i, t) = P\{z(t) < z, i(t) = i\}$  составим систему дифференциальных уравнений Колмогорова в стационарном режиме [2]:

$$\frac{\partial P(z, i)}{\partial z} - i\mu P(z, i) + \mu(i+1)P(z, i+1) - \frac{\partial P(0, i)}{\partial z} + \frac{\partial P(0, i-1)}{\partial z} A(z) = 0.$$

Определим функции

$$H(z, u) = \sum_{i=0}^{\infty} e^{ju i} P(z, i), \quad (1)$$

для которых получим уравнение вида

$$\frac{\partial H(z, u)}{\partial z} + j\mu(1 - e^{-ju}) \frac{\partial H(z, u)}{\partial u} + \frac{\partial H(0, u)}{\partial z} (e^{ju} A(z) - 1) = 0. \quad (2)$$

Положим в уравнении (1)  $u = 0$ , получим  $H(z, 0) = \sum_{i=0}^{\infty} P(z, i) = P\{z(t) < z\} = R(z)$ , где  $R(z)$  — стационарное распределение вероятностей значений процесса  $\{z(t)\}$ . Стационарное распределение вероятностей  $R(z)$  является решением полученного из (2) уравнения  $\frac{\partial R(z)}{\partial z} + \frac{\partial R(0)}{\partial z} (A(z) - 1) = 0$ , которое, в свою очередь, можно представить в виде

$$R(z) = \int_0^z \frac{\partial R(0)}{\partial z} (1 - A(x)) dx. \quad (3)$$

Обозначим  $\frac{\partial R(0)}{\partial z} = \lambda$ , тогда уравнение переписывается в виде  $R(z) = \lambda \int_0^z (1 - A(x)) dx$ .

Найдем величину  $\frac{\partial R(0)}{\partial z} = \lambda$ . В силу того, что  $R(\infty) = 1$ , имеем  $1 = \lambda \int_0^{\infty} (1 - A(x)) dx$ , от-

куда  $\lambda = \frac{1}{\int_0^{\infty} (1 - A(x)) dx}$ . Обозначив  $\int_0^{\infty} (1 - A(x)) dx = a$ , окончательно получим  $\lambda = \frac{1}{a}$ .

Решение уравнения (2) будем искать с помощью метода асимптотического анализа в условии растущего времени обслуживания, то есть  $\mu \rightarrow 0$  [3,4].

### Асимптотика первого порядка

Для нахождения асимптотики первого порядка обозначим  $\mu = \varepsilon$  и выполним в уравнении (2) замены вида  $u = \varepsilon w$ ,  $H(z, u) = F_1(z, w, \varepsilon)$ . Для функции  $F_1(z, w, \varepsilon)$  получим уравнение

$$\frac{\partial F_1(z, w, \varepsilon)}{\partial z} + j(1 - e^{-j\varepsilon w}) \frac{\partial F_1(z, w, \varepsilon)}{\partial w} + \frac{\partial F_1(0, w, \varepsilon)}{\partial z} (e^{j\varepsilon w} A(z) - 1) = 0. \quad (4)$$

Выполним предельный переход  $\varepsilon \rightarrow 0$  в уравнении (4):

$$\frac{\partial F_1(z, w)}{\partial z} + \frac{\partial F_1(0, w)}{\partial z} (A(z) - 1) = 0.$$

Представим функцию  $F_1(z, w)$  в виде

$$F_1(z, w) = R(z) \Phi_1(w), \quad (5)$$

в уравнении (4) выполним предельный переход  $z \rightarrow \infty$ , в результате чего получим равенство вида  $j(1 - e^{-j\varepsilon w}) \frac{\partial \Phi_1(w)}{\partial w} + \lambda \Phi_1(w) (e^{j\varepsilon w} - 1) = 0$ , поделим левую и правую части этого равенства на  $\varepsilon$  и устремим его к нулю, получим

$$\frac{\partial \Phi_1(w)}{\partial w} = j\lambda \Phi_1(w). \quad (6)$$

Решение уравнения (6) удовлетворяет начальному условию  $\Phi_1(0) = 1$  и имеет вид  $\Phi_1(w) = e^{jw\lambda}$ .

В силу представления функции  $F_1(z, w)$  в виде (5), запишем  $F_1(z, w) = R(z) e^{jw\lambda}$ .

Запишем приближенное (асимптотическое) равенство для характеристической функции  $H(z, u)$ :

$$H(z, u) = F_1(z, w, \varepsilon) \approx F_1(z, w) = R(z) \cdot e^{jw\lambda} = R(z) \cdot e^{ju \frac{\lambda}{\mu}},$$

выполнив в котором предельный переход вида  $z \rightarrow \infty$ , получим функцию

$$h_1(u) = e^{ju \frac{\lambda}{\mu}}. \quad (7)$$

Полученную функцию  $h_1(u)$  и будем называть асимптотикой первого порядка характеристической функции числа занятых приборов в системе.

### Асимптотика второго порядка

Для нахождения асимптотики второго порядка в уравнении (2) выполним замену

$H(z, u) = H_2(z, u) e^{ju \frac{\lambda}{\mu}}$ . Тогда уравнение (2) можно переписать в виде

$$\frac{\partial H_2(z, u)}{\partial z} - \lambda(1 - e^{-ju}) H_2(z, u) + \frac{\partial H_2(0, u)}{\partial z} (e^{ju} A(z) - 1) + j\mu(1 - e^{-ju}) \frac{\partial H_2(z, u)}{\partial u} = 0. \quad (8)$$

Введем обозначения

$$u = \varepsilon w, \quad H_2(z, u) = F_2(z, w, \varepsilon). \quad (9)$$

Тогда (8) переписывается в виде

$$\begin{aligned} \frac{\partial F_2(z, w, \varepsilon)}{\partial z} - \lambda(1 - e^{-j\varepsilon w}) F_2(z, w, \varepsilon) + \frac{\partial F_2(0, w, \varepsilon)}{\partial z} (e^{j\varepsilon w} A(z) - 1) + \\ + j\varepsilon(1 - e^{-j\varepsilon w}) \frac{\partial F_2(z, w, \varepsilon)}{\partial w} = 0. \end{aligned} \quad (10)$$

Решение уравнения (10) будем искать в виде разложения

$$F_2(z, w, \varepsilon) = \Phi_2(w) \{R(z) + j\varepsilon w f_2(z)\} + O(\varepsilon^2).$$

В работе получено предельное при  $\varepsilon \rightarrow 0$  выражение для  $F_2(z, w, \varepsilon)$ , имеющее вид:

$$F_2(z, w) = \lim_{\varepsilon \rightarrow 0} F_2(z, w, \varepsilon) = R(z) \exp \left\{ \frac{(ju)^2}{2} \left[ \frac{\lambda + \kappa}{\mu} \right] \right\},$$

где  $\kappa = \frac{\lambda}{\mu} \int_0^{\infty} [A(z) - R(z)] dz$ . Следовательно, асимптотика второго порядка характеристической функции числа заявок в системе примет вид

$$h_2(u) = \exp \left\{ ju \frac{\lambda}{\mu} + \frac{(ju)^2}{2} \left[ \frac{\lambda + \kappa}{\mu} \right] \right\}.$$

Тогда для характеристической функции суммарного объема заявок в системе GI/M/∞ можно записать

$$h_2(u) = \exp \left\{ ju \frac{\lambda}{\mu} a_1 + \frac{(ju)^2}{2} \left[ \frac{\lambda + \kappa}{\mu} \right] a_2 \right\}. \quad (11)$$

Таким образом, асимптотическое приближение характеристической функции суммарного объема находившихся в системе заявок имеет нормальное распределение.

#### ЛИТЕРАТУРА

1. Тихоненко О.М. Моделирование процессов и систем обработки информации: курс лекций / О.М. Тихоненко. – Минск: БГУ, 2008. – 148 с.
2. Назаров А.А. Теория массового обслуживания: учебное пособие / А.А. Назаров, А.Ф. Терпугов. – 2-е изд., испр. – Томск: Изд-во НТЛ, 2010. – 228 с.
3. Назаров А.А. Методы асимптотического анализа в теории массового обслуживания / А.А. Назаров, С.П. Моисеева. – Томск: Изд-во НТЛ, 2006. 112 с.
4. Моисеева С.П. Разработка методов исследования математических моделей немарковских систем обслуживания с неограниченным числом приборов и непуассоновскими входящими потоками: Дис. ... докт. физ.-мат. наук: 05.13.18 / Светлана Петровна Моисеева. – Томск, 2014. – 280 с.

## АСИМПТОТИЧЕСКИЙ АНАЛИЗ СИСТЕМЫ MMPP/GI/∞ С ОБСЛУЖИВАНИЕМ ТРЕБОВАНИЙ СЛУЧАЙНОГО ОБЪЁМА

**Е.Ю. Лисовская, С.П. Моисеева**  
Томский государственный университет  
ekaterina\_lisovs@mail.ru, smoiseeva@mail.ru

### Введение

Системы массового обслуживания (СМО) требований случайного объема имеют свое применение в задачах проектирования информационных и телекоммуникационных систем, где информация передается порциями в виде сообщений случайного объема. Как было замечено в работах [1,4,8] задачи исследования таких систем играют большую роль при моделировании современных информационно-вычислительных систем. Однако аналитических решений при дисциплине обслуживания FIFO не было получено, поскольку для построения корректного марковского процесса необходимо учитывать объемы тех заявок, которые находятся в системе.

В работах [1,4–6] были исследованы СМО требований случайного объема с дисциплиной обслуживания LIFO и ограничением на суммарный объем требований. Оказалось, что в этом случае можно получить алгоритмы, пригодные для численных расчетов стационарных характеристик.

Целью данной работы является исследование многолинейной системы массового обслуживания ММРР/ГІ/∞ со случайным объёмом требований методом асимптотического анализа в условии высокой интенсивности входящего потока.

### Постановка задачи

Рассмотрим систему массового обслуживания (СМО) с неограниченным числом приборов, на вход которой поступает марковский модулированный поток (ММРР-поток), управляемый цепью Маркова  $k(t)$ , заданной матрицей инфинитезимальных характеристик  $Q = \|q_{ij}\|$  и диагональной матрицей условных интенсивностей  $\Lambda$ . Продолжительность обслуживания заявки имеет произвольную функцию распределения, одинаковую для всех приборов  $B(x)$ . Предполагаем, что каждое требование характеризуется некоторым случайным объёмом  $v > 0$ ,  $G(y) = P\{v < y\}$  — функция распределения случайной величины  $v$ . Объёмы различных требований независимы. По окончании обслуживания заявка покидает систему и «уносит» свой объём.

Пусть  $i(t)$  — число заявок, находящихся на обслуживании в системе в момент  $t$ ,  $V(t)$  — полная сумма объёмов требований, находящихся в системе в момент времени  $t$ . Поставим задачу нахождения характеристик двумерного случайного процесса  $\{i(t), V(t)\}$ . Отметим, что исследуемый процесс не является марковским, поэтому для его исследования будем использовать метод динамического просеивания (метод просеянного потока) [2,3].

Построим просеянный поток для рассматриваемой СМО ММРР/ГІ/∞. Для этого зафиксируем некоторый момент времени  $T$ . Полагаем, что заявка входящего потока, поступившая в систему в момент времени  $t < T$  с вероятностью  $S(t) = 1 - B(T - t)$  формирует событие просеянного потока, а с вероятностью  $1 - S(t)$  эта заявка не рассматривается.

Обозначим  $n(t)$  — число событий просеянного потока, наступивших до момента времени  $t$ . Тогда, если в начальный момент  $t_0 < T$  система была свободна, то для момента времени  $T$  для любых  $m$  выполняется равенство  $P\{i(T) = m\} = P\{n(T) = m\}$ .

Следует отметить, что использование метода просеянного потока позволяет более точно определить характеристики процесса  $V(t)$ , так как в просеянном потоке присутствуют только те заявки, которые не закончат обслуживание к моменту времени  $T$ .

### Дифференциальное уравнение Колмогорова в матричном виде

Введем обозначение  $P(k, n, z, t) = P\{k(t) = k, n(t) = n, V(t) < z\}$  — распределение вероятностей трехмерного Марковского процесса, где  $k(t)$  — состояние управляющей цепи Маркова в момент времени  $t$ ,  $n(t)$  — число событий просеянного потока, наступивших до момента времени  $t$ ,  $V(t)$  — суммарный объём требований, находящихся в просеянном потоке в момент времени  $t$ . Для этого распределения составим  $\Delta t$ -методом прямую систему дифференциальных уравнений Колмогорова. По формуле полной вероятности запишем равенство

$$P(k, n, z, t + \Delta t) = P(k, n, z, t)(1 - \lambda_k \Delta t)(1 - q_{kk} \Delta t) + P(k, n, z, t)\lambda_k \Delta t(1 - S(t)) + \\ + \lambda_k \Delta t S(t) \int_0^z P(k, n-1, z-y, t) dG(y) + \sum_{v \neq k} q_{vk} \Delta t P(v, n, z, t) + o(\Delta t),$$

$$k = \overline{1, K}, \quad n = 0, 1, 2, \dots, \quad z > 0,$$

откуда получаем систему дифференциальных уравнений Колмогорова

$$\frac{\partial P(k, n, z, t)}{\partial t} = \lambda_k S(t) \left[ \int_0^z P(k, n-1, z-y, t) dG(y) - P(k, n, z, t) \right] + \sum_v q_{vk} P(v, n, z, t), \quad (1)$$

$$k = \overline{1, K}, \quad n = 0, 1, 2, \dots, \quad z > 0.$$

Введем частичные характеристические функции вида:

$$H(k, u_1, u_2, t) = M \left\{ \exp(ju_1 n(t) + ju_2 V(t)) \right\} = \sum_{n=0}^{\infty} e^{ju_1 n} \int_0^{\infty} e^{ju_2 z} P(k, n, z, t) dz,$$

$$k = \overline{1, K}, \quad n = 0, 1, 2, \dots, \quad z > 0.$$

Учитывая, что

$$\begin{aligned} & \sum_{n=0}^{\infty} e^{ju_1 n} \int_0^{\infty} e^{ju_2 z} \int_0^z P(k, n-1, z-y, t) dG(y) dz = \\ & = e^{ju_1} \sum_{n=0}^{\infty} e^{ju_1 (n-1)} \int_0^{\infty} e^{ju_2 y} \cdot e^{ju_2 (z-y)} \int_0^z P(k, n-1, z-y, t) dG(y) dz = \\ & = e^{ju_1} \int_0^{\infty} e^{ju_2 y} \left[ \sum_{n=0}^{\infty} e^{ju_1 (n-1)} \int_0^{\infty} e^{ju_2 (z-y)} P(k, n-1, z-y, t) dz \right] dG(y) = \\ & = e^{ju_1} \int_0^{\infty} e^{ju_2 y} H(k, u_1, u_2, t) dG(y) = e^{ju_1} H(k, u_1, u_2, t) \int_0^{\infty} e^{ju_2 y} dG(y) = \\ & = e^{ju_1} H(k, u_1, u_2, t) G^*(u_2), \end{aligned}$$

где  $G^*(u_2)$  обозначено как  $G^*(u_2) = \int_0^{\infty} e^{ju_2 y} dG(y)$ , можно записать следующую систему уравнений:

$$\frac{\partial H(k, u_1, u_2, t)}{\partial t} = \lambda_k S(t) H(k, u_1, u_2, t) \left[ e^{ju_1} G^*(u_2) - 1 \right], \quad k = \overline{1, K}.$$

Запишем данную систему в виде матричного уравнения

$$\frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) \left[ \mathbf{\Lambda} S(t) \left( e^{ju_1} G^*(u_2) - 1 \right) + \mathbf{Q} \right] \quad (2)$$

с начальным условием

$$\mathbf{H}(u_1, u_2, t_0) = \mathbf{r}, \quad (3)$$

где

$$\mathbf{H}(u_1, u_2, t) = \left[ H(1, u_1, u_2, t), H(2, u_1, u_2, t), \dots, H(K, u_1, u_2, t) \right]$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_K \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1K} \\ q_{21} & q_{22} & \dots & q_{2K} \\ \dots & \dots & \dots & \dots \\ q_{K1} & q_{K2} & \dots & q_{KK} \end{bmatrix}, \quad \mathbf{r} = [r(1), r(2), \dots, r(K)].$$

Здесь и далее  $\mathbf{r}$  — вектор-строка стационарного распределения управляющей цепи Маркова, определяемая системой линейных уравнений

$$\begin{cases} \mathbf{r}\mathbf{Q} = 0 \\ \mathbf{r}\mathbf{e} = 1 \end{cases}, \quad (4)$$

$\mathbf{e}$  — единичный вектор-столбец.

### Метод асимптотического анализа

Так как прямое решение уравнения (2) не представляется возможным, то для решения задачи (2)–(3) воспользуемся методом асимптотического анализа [3] в условии неограниченно растущей интенсивности входящего потока [2]. Обозначим  $\Lambda = N\Lambda^1$ ,  $\mathbf{Q} = N\mathbf{Q}^1$ , где  $N \rightarrow \infty$ . Тогда уравнение (2) примет вид

$$\frac{1}{N} \frac{\partial \mathbf{H}(u_1, u_2, t)}{\partial t} = \mathbf{H}(u_1, u_2, t) \left[ \Lambda^1 S(t) (e^{ju_1} G^*(u_2) - 1) + \mathbf{Q}^1 \right] \quad (5)$$

с начальным условием

$$\mathbf{H}(u_1, u_2, t_0) = \mathbf{r}. \quad (6)$$

Асимптотический анализ первого порядка проведем в виде доказательства следующей теоремы.

**Теорема.** *Асимптотическая характеристическая функция распределения вероятностей процесса  $\{k(t), n(t), V(t)\}$  первого порядка имеет вид*

$$\mathbf{H}(u_1, u_2, t) = \mathbf{r} \exp \left\{ N\lambda [ju_1 + ju_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\}, \text{ где вектор-строка } r \text{ определяется систе-}$$

*мой линейных уравнений* **Ошибка! Источник ссылки не найден.**, а  $\lambda$  — выражением  $\lambda = \mathbf{r} \cdot \Lambda \cdot \mathbf{e}$ ,  $a_1$  — математическое ожидание случайной величины, определяемой функцией распределения объема требований  $G(y)$ .

Доказательство.

Выполним в выражениях (5) и (6) замены

$$\varepsilon = \frac{1}{N}, \quad u_1 = \varepsilon w_1, \quad u_2 = \varepsilon w_2, \quad \mathbf{H}(u_1, u_2, t) = \mathbf{F}_1(w_1, w_2, t, \varepsilon). \quad (7)$$

Тогда задача (5)–(6) примет вид

$$\varepsilon \frac{\partial \mathbf{F}_1(w_1, w_2, t, \varepsilon)}{\partial t} = \mathbf{F}_1(w_1, w_2, t, \varepsilon) \left[ \Lambda^1 S(t) (e^{ju_1} G^*(\varepsilon w_2) - 1) + \mathbf{Q}^1 \right], \quad (8)$$

с начальным условием

$$\mathbf{F}_1(w_1, w_2, t_0, \varepsilon) = \mathbf{r}, \quad (9)$$

Найдем асимптотическое при  $\varepsilon \rightarrow 0$  решение задачи (8)–(9), то есть  $\mathbf{F}_1(w_1, w_2, t) = \lim_{\varepsilon \rightarrow 0} \mathbf{F}_1(w_1, w_2, t, \varepsilon)$ .

*Этап 1.* Положим в (8)  $\varepsilon \rightarrow 0$ , получим  $\mathbf{F}_1(w_1, w_2, t) \mathbf{Q}^1 = 0$ . Сравнивая это равенство с первым уравнением **Ошибка! Источник ссылки не найден.**, можем сделать вывод, что  $\mathbf{F}_1(w_1, w_2, t)$  может быть представлена в виде

$$\mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \Phi_1(w_1, w_2, t), \quad (10)$$

где  $\Phi_1(w_1, w_2, t)$  — некоторая скалярная функция, в силу (9), удовлетворяющая условию  $\Phi_1(w_1, w_2, t_0) = 1$ .

Этап 2. Умножим (8) на вектор  $\mathbf{e}$ , подставим (10), поделим результат на  $\varepsilon$  и выполним предельный переход при  $\varepsilon \rightarrow 0$ . Тогда с учётом того, что  $\mathbf{Q}^1 \mathbf{e} = 0$  и  $\mathbf{r} \mathbf{e} = 1$ , получим следующее дифференциальное уравнение относительно функции  $\Phi_1(w_1, w_2, t)$ :

$$\frac{\partial \Phi_1(w_1, w_2, t)}{\partial t} = \Phi_1(w_1, w_2, t) [\lambda S(t)(jw_1 + jw_2 a_1)]. \quad (11)$$

Здесь и далее  $a_1 = \int_0^{\infty} y dG(y)$  — математическое ожидание случайной величины, определяемой функцией распределения объёма требований  $G(y)$ ,  $\lambda = \mathbf{r} \cdot \mathbf{\Lambda} \cdot \mathbf{e}$ . Решение (11) с учётом начального условия дает

$$\Phi_1(w_1, w_2, t) = \exp \left\{ \lambda (jw_1 + jw_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}.$$

Подставляя данное выражение в (10), получаем

$$\mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \exp \left\{ \lambda (jw_1 + jw_2 a_1) \int_{t_0}^t S(\tau) d\tau \right\}.$$

В силу замен (7) можно записать асимптотическое при  $\varepsilon \rightarrow 0$  равенство:

$$\begin{aligned} \mathbf{H}(u_1, u_2, t) &= \mathbf{F}_1(w_1, w_2, t, \varepsilon) \approx \mathbf{F}_1(w_1, w_2, t) = \mathbf{r} \Phi_1(w_1, w_2, t) = \\ &= \mathbf{r} \exp \left\{ \lambda \left[ j \frac{u_1}{\varepsilon} + j \frac{u_2}{\varepsilon} a_1 \right] \int_{t_0}^t S(\tau) d\tau \right\} = \mathbf{r} \exp \left\{ N \lambda [ju_1 + ju_2 a_1] \int_{t_0}^t S(\tau) d\tau \right\}. \end{aligned}$$

*Теорема доказана.*

Следствие. При  $t = T$  для характеристической функции процесса  $\{i(t), V(t)\}$  в стационарном режиме получим

$$H(u_1, u_2) = \exp \{ N \lambda b_1 [ju_1 + ju_2 a_1] \}.$$

Здесь и далее

$$b_1 = \int_{-\infty}^T S(\tau) d\tau = \int_{-\infty}^T (1 - B(T - \tau)) d\tau = \int_0^{\infty} (1 - B(\tau)) d\tau$$

определяет математическое ожидание случайной величины с функцией распределения  $B(x)$ .

### Заключение

Для бесконечнолинейной системы массового обслуживания требований случайного объёма была получена асимптотическая характеристическая функция первого порядка двумерного распределения вероятностей числа занятых приборов и суммарного объёма требований, находящихся в системе.

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 16-31-00292 мол\_а.

### ЛИТЕРАТУРА

1. Александров А.М., Кац Б.А. Обслуживание потоков неоднородных требований // Изв. АН СССР. Технич. Кибернетика, 1973. №2. С. 47–53.
2. Моисеев А.Н., Назаров А.А. Бесконечнолинейные системы и сети массового обслуживания. – Томск: Изд-во НТЛ, 2015. – 240с.
3. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. Томск: Изд-во НТЛ, 2006. – 112с.

4. Печинкин А.В., Печинкина О.А. Система  $M_k|G|1|n$  с дисциплиной обслуживания LIFO с прерыванием и ограничением на суммарный объём требований // Вестник Российского ун-та дружбы народов. Сер. Прикладная математика и информатика, 1996. №1. С. 86-93.

5. Печинкин А.В. Система  $M|G|1|n$  с дисциплиной обслуживания LIFO и ограничением на суммарный объём требований // Автоматика и телемеханика, 1998. №4. С. 106–116.

6. Печинкин А.В. Система обслуживания с дисциплиной LIFO и ограничением на суммарный объём требований // Вестник Российского ун-та дружбы народов. Сер. Прикладная математика и информатика, 1996. №2. С. 85-99.

7. Лисовская Е.Ю., Моисеева С.П. Исследование суммарного объёма требований в бесконечнолинейной системе массового обслуживания вида  $M|G|1|\infty$  // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем : материалы Российской конференции с международным участием. Москва, РУДН, 18–22 апреля 2016 г. – Москва : РУДН, 2016. – С. 28–30.

8. Тихоненко О.М. Модели массового обслуживания в системах обработки информации. – Минск: Университетское, 1990.

9. Колбасова В.А., Лисовская Е.Ю. Исследование суммарного объёма требований в СМО вида  $G|G|1|\infty$  методом асимптотического анализа // Научное творчество молодежи. Математика. Информатика : материалы XX Всероссийской научно-практической конференции (28-29 апреля 2016 г.) / сост. Ю.А. Наумкина. – Томск : Изд-во Том. ун-та, 2016. – Ч.1. – С. 97–101.

## ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ОБОБЩЁННОГО СИНХРОННОГО ПОТОКА ВТОРОГО ПОРЯДКА

**Л.А. Нежелская, Е.Ф. Сидорова**

*Томский государственный университет*

*ludne@mail.ru, katusha\_sidorova@mail.ru*

### Введение

В настоящее время теория массового обслуживания представляет собой одну из наиболее интенсивно развивающихся ветвей теории вероятностей и имеет широкое практическое применение в разнообразных направлениях исследований: в экономике, в планировании и организации процессов производства и потребления, в логистике, технике, транспорте, военном деле, а также в других сферах деятельности [1].

Бурное развитие информационных технологий и компьютерных наук привело к появлению важной сферы приложений теории массового обслуживания — проектированию и разработке информационно-вычислительных систем, компьютерных и телекоммуникационных сетей связи и т.д.

Зачастую работы по теории массового обслуживания направлены на определение стационарных характеристик систем обслуживания, для которых параметры входящих потоков заранее известны. В реальных ситуациях данные параметры изменяются с течением времени, и при этом в большинстве случаев подобные изменения носят случайный характер, что приводит к рассмотрению дважды стохастических потоков событий [2]. Двойная случайность (стохастика) подразумевает, во-первых, случайность моментов времени наступления событий потока, и, во-вторых, определение интенсивности потока как случайного процесса. Дважды стохастические потоки событий можно разделить на два класса, а именно:

- 1) потоки, интенсивность которых есть непрерывный случайный процесс;
- 2) потоки, для которых интенсивность — кусочно-постоянный случайный процесс с конечным числом состояний.

Потоки второго класса носят название МС-потоков (Markov chain) [3]. Заметим, что обобщённый синхронный поток второго порядка относится именно к МС-потокам и представляет собой актуальную математическую модель реальных потоков случайных событий.

В настоящей статье построена имитационная модель обобщённого синхронного потока второго порядка, а также проведен ряд экспериментов с целью получения численных результатов, подтверждающих работоспособность построенной модели.

# 1. Постановка задачи

Рассматривается обобщённый синхронный поток второго порядка (далее поток), интенсивность которого есть ненаблюдаемый кусочно-постоянный стационарный случайный процесс  $\lambda(t)$  с двумя состояниями:  $\lambda(t) = \lambda_1$  и  $\lambda(t) = \lambda_2$  ( $\lambda_1 > \lambda_2 \geq 0$ ). Длительность пребывания процесса  $\lambda(t)$  в  $i$ -ом состоянии ( $i = 1, 2$ ) определяется двумя случайными величинами. Первая случайная величина распределена по экспоненциальному закону с функцией распределения  $F_i^{(1)}(t) = 1 - e^{-\lambda_i t}$ ,  $i = 1, 2$ . Если процесс  $\lambda(t)$  в момент времени  $t$  находится в  $i$ -ом состоянии, т.е.  $\lambda(t) = \lambda_i$ , то на полуинтервале  $[t; t + \Delta t)$ , где  $\Delta t$  — достаточно малая величина, в момент окончания  $i$ -ого состояния происходит розыгрыш пар, т.е. включается вероятностный механизм: наступает событие потока и процесс  $\lambda(t)$  переходит из  $i$ -ого состояния в  $j$ -ое ( $i, j = 1, 2, i \neq j$ ) с вероятностью  $P_1^{(1)}(\lambda_j | \lambda_i)$  ( $i, j = 1, 2, i \neq j$ ) или наступает событие потока и процесс  $\lambda(t)$  переходит из  $i$ -ого состояния в  $i$ -ое ( $i = 1, 2$ ) с вероятностью  $P_1^{(1)}(\lambda_i | \lambda_i)$  ( $i = 1, 2$ ), причём  $P_1^{(1)}(\lambda_j | \lambda_i) + P_1^{(1)}(\lambda_i | \lambda_i) = 1$  ( $i, j = 1, 2, i \neq j$ ). Вторая случайная величина распределена по экспоненциальному закону с функцией распределения  $F_i^{(2)}(t) = 1 - e^{-\alpha_i t}$ ,  $i = 1, 2$ . В момент окончания  $i$ -ого состояния процесса  $\lambda(t)$  наступает событие потока и процесс  $\lambda(t)$  переходит из  $i$ -ого состояния в  $j$ -ое ( $i, j = 1, 2, i \neq j$ ) с вероятностью  $P_1^{(2)}(\lambda_j | \lambda_i)$  ( $i, j = 1, 2, i \neq j$ ) или наступает событие потока и процесс  $\lambda(t)$  переходит из  $i$ -ого состояния в  $i$ -ое ( $i = 1, 2$ ) с вероятностью  $P_1^{(2)}(\lambda_i | \lambda_i)$  ( $i = 1, 2$ ), причём  $P_1^{(2)}(\lambda_j | \lambda_i) + P_1^{(2)}(\lambda_i | \lambda_i) = 1$  ( $i, j = 1, 2, i \neq j$ ). Первая и вторая случайные величины являются независимыми друг от друга.

Пример описанной ситуации приведен на рис.1, где  $\lambda_1, \lambda_2$  — состояния процесса  $\lambda(t)$ ,  $t_1, t_2, \dots$  — моменты наступления событий в потоке.

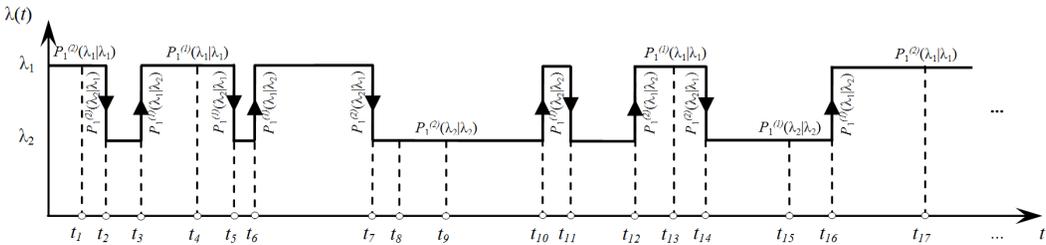


Рис.1. Одна из реализаций обобщённого синхронного потока событий второго порядка

В результате блочная матрица инфинитезимальных характеристик процесса  $\lambda(t)$  примет вид:  $D = \|D_0 | D_1\|$ , где

$$D_0 = \begin{vmatrix} -(\lambda_1 + \alpha_1) & 0 \\ 0 & -(\lambda_2 + \alpha_2) \end{vmatrix},$$

$$D_1 = \begin{vmatrix} \lambda_1 P_1^{(1)}(\lambda_1 | \lambda_1) + \alpha_1 P_1^{(2)}(\lambda_1 | \lambda_1) & \lambda_1 P_1^{(1)}(\lambda_2 | \lambda_1) + \alpha_1 P_1^{(2)}(\lambda_2 | \lambda_1) \\ \lambda_2 P_1^{(1)}(\lambda_1 | \lambda_2) + \alpha_2 P_1^{(2)}(\lambda_1 | \lambda_2) & \lambda_2 P_1^{(1)}(\lambda_2 | \lambda_2) + \alpha_2 P_1^{(2)}(\lambda_2 | \lambda_2) \end{vmatrix}.$$

Диагональные элементы матрицы  $D_0$  есть интенсивности выхода процесса  $\lambda(t)$  из своих состояний, взятые с противоположным знаком; недиагональные элементы — интенсивности переходов из состояния в состояние без наступления события. Эlemen-

тами матрицы  $D_1$  являются интенсивности переходов процесса  $\lambda(t)$  из состояния в состояние с наступлением события [3].

Задача настоящей работы заключается в построении имитационной модели обобщённого синхронного потока второго порядка.

### 2. Имитационное моделирование

Имитационное моделирование — метод исследования, позволяющий получить информацию об изучаемой системе в результате серий экспериментов над имитирующей системой, которая описывает реальную с достаточной степенью точности. Для имитационного моделирования как вида *компьютерного* моделирования характерно воспроизведение (имитация) на ЭВМ процесса функционирования исследуемой системы.

Модель исследуемого обобщённого синхронного потока событий второго порядка была построена с помощью метода обратных функций. Пусть  $\tau \geq 0$  — длительность пребывания процесса  $\lambda(t)$  в  $i$ -ом состоянии ( $i = 1, 2$ ). Случайная величина  $\tau$  имеет плотность распределения вероятностей  $p_i(\tau) = \beta_i e^{-\beta_i \tau}$ ,  $i = 1, 2$ . Тогда функция распределения имеет вид:  $F_i(\tau) = 1 - e^{-\beta_i \tau}$ ,  $i = 1, 2$ . Согласно методу обратных функций [4] обозначим  $F_i(\tau) = \gamma$ . Имеем  $\gamma = 1 - e^{-\beta_i \tau}$ . Тогда  $-\beta_i \tau = \ln(1 - \gamma)$ , откуда находим формулу моделирования длительностей интервалов между соседними событиями в обобщённом синхронном потоке второго порядка в следующем виде:  $\tau = -\frac{1}{\beta_i} \ln(1 - \gamma)$ , где  $\gamma$  — равномерно распределенная на интервале  $(0; 1)$  случайная величина,  $\beta_i = \{\lambda_i, \alpha_i\}$ ,  $i = 1, 2$ .

### 3. Результаты работы имитационной модели

Результаты испытаний, проделанных для обобщённого синхронного потока второго порядка, приведенные в табл. 1–4, отображают зависимость между параметрами потока (интенсивностями  $\lambda_1$  и  $\lambda_2$ ,  $\alpha_1$  и  $\alpha_2$ ) и количеством событий в исходном потоке, наступающих в процессе реализации испытания модели.

Зафиксируем значение времени моделирования  $T = 10$ .

Таблица 1

Исходные данные для первого испытания имитационной модели

$\lambda_1 = 3; 8$ ; $\lambda_2 = 1; 7$	$P_1^{(1)}(\lambda_1   \lambda_1) = 0,4$	$P_1^{(1)}(\lambda_2   \lambda_2) = 0,5$	$P_1^{(2)}(\lambda_1   \lambda_1) = 0,3$	$P_1^{(2)}(\lambda_2   \lambda_2) = 0,8$
$\alpha_1 = 0,5$ ; $\alpha_2 = 0,8$	$P_1^{(1)}(\lambda_2   \lambda_1) = 0,6$	$P_1^{(1)}(\lambda_1   \lambda_2) = 0,5$	$P_1^{(2)}(\lambda_2   \lambda_1) = 0,7$	$P_1^{(2)}(\lambda_1   \lambda_2) = 0,2$

Таблица 2

Результаты первого испытания

Моменты наступления событий	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	...	$t_{19}$
$\lambda_1 = 3$ $\lambda_2 = 1$	0,413	0,579	0,582	0,844	0,873	1,079	1,173	2,214		10,517
Моменты наступления событий	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	...	$t_{84}$
$\lambda_1 = 8$ $\lambda_2 = 7$	0,099	0,406	0,586	0,591	0,686	0,735	0,742	0,851		10,075

В результате анализа полученных в табл. 1, 2 данных делаем вывод, что с увеличением значений параметров потока  $\lambda_1$  и  $\lambda_2$  количество событий, наступивших в исходном потоке, увеличивается.

Исходные данные для второго испытания имитационной модели

$\lambda_1 = 3$ $\lambda_2 = 1$	$P_1^{(1)}(\lambda_1   \lambda_1) = 0,4$	$P_1^{(1)}(\lambda_2   \lambda_2) = 0,5$	$P_1^{(2)}(\lambda_1   \lambda_1) = 0,3$	$P_1^{(2)}(\lambda_2   \lambda_2) = 0,8$
$\alpha_1 = 0,5; 2,6$ $\alpha_2 = 0,8; 4,1$	$P_1^{(1)}(\lambda_2   \lambda_1) = 0,6$	$P_1^{(1)}(\lambda_1   \lambda_2) = 0,5$	$P_1^{(2)}(\lambda_2   \lambda_1) = 0,7$	$P_1^{(2)}(\lambda_1   \lambda_2) = 0,2$

Результаты второго испытания

Моменты наступления событий	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	...	$t_{27}$
$\alpha_1 = 0,5$ $\alpha_2 = 0,8$	0,159	0,249	0,683	0,949	1,081	1,445	3,182	3,239		10,840
Моменты наступления событий	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	...	$t_{55}$
$\alpha_1 = 2,6$ $\alpha_2 = 4,1$	0,206	0,426	0,722	1,283	1,404	1,528	1,547	1,637		10,082

Согласно данным табл. 3, 4 аналогичным образом получаем прямую зависимость между количеством наступивших в потоке событий и изменениями значений параметров потока  $\alpha_1$  и  $\alpha_2$ , т.е. при увеличении интенсивностей  $\alpha_1$  и  $\alpha_2$  количество событий в исходном потоке увеличивается.

В ходе экспериментов с различными значениями вероятностей  $P_1^{(1)}(\lambda_j | \lambda_i)$  ( $i, j = 1, 2$ ) и  $P_1^{(2)}(\lambda_j | \lambda_i)$  ( $i, j = 1, 2$ ) переходов процесса  $\lambda(t)$  из  $i$ -ого состояния в  $j$ -ое ( $i, j = 1, 2$ ) получены результаты для обобщённого синхронного потока второго порядка, иллюстрирующие уникальные свойства этого потока событий.

Зафиксируем значения следующих параметров для всех нижеописанных испытаний:  $\lambda_1 = 8, \lambda_2 = 7, \alpha_1 = 2,6, \alpha_2 = 4,1$ .

Исходные данные для третьего испытания имитационной модели

$P_1^{(1)}(\lambda_1   \lambda_1) = 1$	$P_1^{(1)}(\lambda_2   \lambda_1) = 0$	$P_1^{(1)}(\lambda_2   \lambda_2) = 1$	$P_1^{(1)}(\lambda_1   \lambda_2) = 0$
$P_1^{(2)}(\lambda_1   \lambda_1) = 1$	$P_1^{(2)}(\lambda_2   \lambda_1) = 0$	$P_1^{(2)}(\lambda_2   \lambda_2) = 1$	$P_1^{(2)}(\lambda_1   \lambda_2) = 0$

В результате реализации потока при исходных данных, соответствующих табл. 5, отмечаем, что процесс  $\lambda(t)$  устанавливается в одном из состояний ( $\lambda(t) = \lambda_1$  (рис. 2) или  $\lambda(t) = \lambda_2$  (рис. 3)) и не изменяет его с течением времени, т.е. процесс  $\lambda(t)$  останется в исходном состоянии с вероятностью, равной единице.

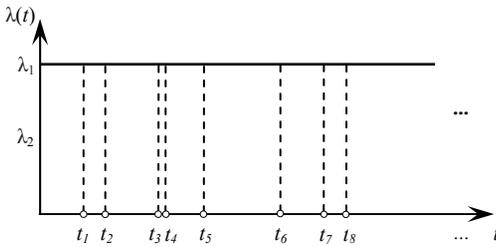


Рис. 2. Реализация потока

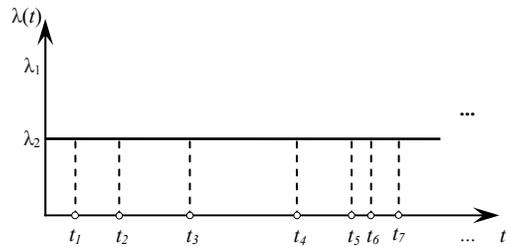


Рис. 3. Реализация потока

Исходные данные для четвертого испытания имитационной модели

$P_1^{(1)}(\lambda_1   \lambda_1) = 0$	$P_1^{(1)}(\lambda_2   \lambda_1) = 1$	$P_1^{(1)}(\lambda_2   \lambda_2) = 0$	$P_1^{(1)}(\lambda_1   \lambda_2) = 1$
$P_1^{(2)}(\lambda_1   \lambda_1) = 0$	$P_1^{(2)}(\lambda_2   \lambda_1) = 1$	$P_1^{(2)}(\lambda_2   \lambda_2) = 0$	$P_1^{(2)}(\lambda_1   \lambda_2) = 1$

Анализируя реализацию процесса, полученную при исходных данных табл. 6, замечаем, что в каждый момент наступления события процесс  $\lambda(t)$  переходит из  $i$ -ого состояния в  $j$ -ое ( $i, j = 1, 2, i \neq j$ ).

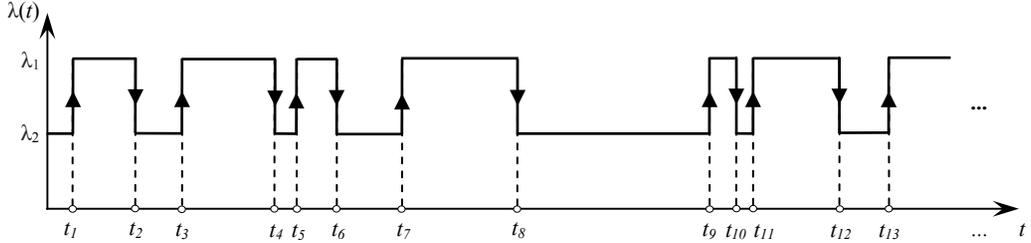


Рис. 4. Обобщённый поток событий второго порядка

Таким образом, анализируя результаты, полученные вследствие третьего и четвертого испытаний, делаем вывод, что для обобщённого синхронного потока второго порядка характерен переход из состояния в состояние *только* в момент наступления события.

В табл. 8, 10 величины  $\tau$  — длительности промежутков времени между двумя последовательными моментами наступления событий в потоке,  $t_1, t_2, \dots$  — моменты наступления события потока.

Таблица 7

Исходные данные для пятого испытания имитационной модели

$P_1^{(1)}(\lambda_1   \lambda_1)$	$P_1^{(1)}(\lambda_2   \lambda_1)$	$P_1^{(1)}(\lambda_2   \lambda_2)$	$P_1^{(1)}(\lambda_1   \lambda_2)$
0,97	0,03	0,1	0,9
$P_1^{(2)}(\lambda_1   \lambda_1)$	$P_1^{(2)}(\lambda_2   \lambda_1)$	$P_1^{(2)}(\lambda_2   \lambda_2)$	$P_1^{(2)}(\lambda_1   \lambda_2)$
0,88	0,12	0,07	0,93

Таблица 8

Результаты пятого испытания

$i$	1	2	3	4	5	6	7	8	9	10	...	108
$\tau$	0,064	0,021	0,043	0,073	0,089	0,037	0,139	0,054	0,026	0,112	...	0,015
$t_i$	0,064	0,085	0,128	0,201	0,290	0,327	0,466	0,520	0,546	0,658	...	10,215

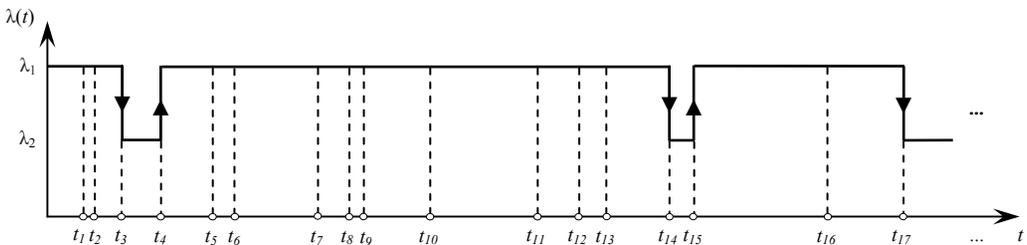


Рис. 5. Реализация обобщённого синхронного потока событий второго порядка

Исходные данные для шестого испытания имитационной модели

$P_1^{(1)}(\lambda_1   \lambda_1)$	$P_1^{(1)}(\lambda_2   \lambda_1)$	$P_1^{(1)}(\lambda_2   \lambda_2)$	$P_1^{(1)}(\lambda_1   \lambda_2)$
0,09	0,91	0	1
$P_1^{(2)}(\lambda_1   \lambda_1)$	$P_1^{(2)}(\lambda_2   \lambda_1)$	$P_1^{(2)}(\lambda_2   \lambda_2)$	$P_1^{(2)}(\lambda_1   \lambda_2)$
0,13	0,87	0,15	0,85

Таблица 10

Результаты шестого испытания

$i$	1	2	3	4	5	6	7	8	9	10	...	101
$\tau$	0,046	0,055	0,019	0,145	0,094	0,128	0,049	0,076	0,198	0,151	...	0,092
$t_i$	0,046	0,101	0,120	0,265	0,359	0,487	0,536	0,612	0,810	0,961	...	10,061

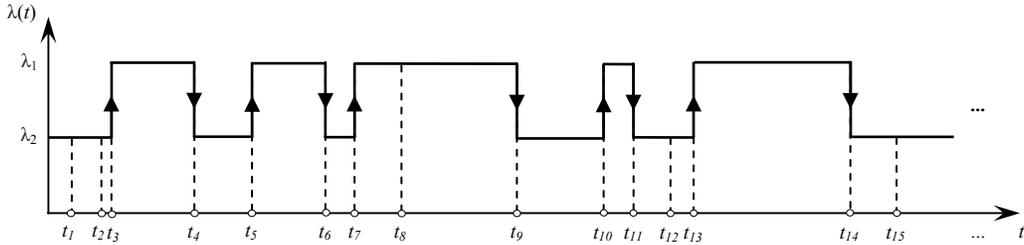


Рис. 6. Реализация обобщенного синхронного потока событий второго порядка

### Заключение

В данной статье был рассмотрен обобщенный синхронный поток событий второго порядка и получена матрица инфинитезимальных характеристик процесса  $\lambda(t)$ . Также была построена имитационная модель этого потока в виде программного кода на языке программирования C# в среде Visual Studio 2012. Анализ корректности результатов проведенных статистических испытаний дает возможность утверждать, что получена работоспособная модель, непротиворечащая входным данным модели.

### ЛИТЕРАТУРА

1. Гнеденко Б.В. Введение в теорию массового обслуживания // Б. В. Гнеденко, И. Н. Коваленко. – М.: Наука, 1987. – Гл. 1 : Задачи теории массового обслуживания в простейших предпосылках. – С. 7–87.
2. Горцев А.М., Голофастова А.М. Оптимальная оценка состояний модулированного синхронного дважды стохастического потока событий [Электронный ресурс] // Вестник Томского государственного университета. Управление, вычислительная техника и информатика, 2013 № 2 (23).
3. Горцев А.М., Нежелская Л.А. Математическое моделирование : о связи МС-потоков и МАР-потоков событий [Электронный ресурс] // Вестник Томского государственного университета. Управление, вычислительная техника и информатика, 2011 № 1 (14).
4. Соболев И.М. Численные методы Монте-Карло // И. М. Соболев. – М. : Наука. Гл. ред. физ.-мат. лит., 1973 – С. 210–247.

## ИМИТАЦИОННАЯ МОДЕЛЬ ПОЛУСИНХРОННОГО ПОТОКА ВТОРОГО ПОРЯДКА

Л.А. Нежелская, Д.А. Тумашкина  
Томский государственный университет  
ludne@mail.ru, diana1323@mail.ru

### Введение

При анализе экономических систем довольно часто возникает необходимость решать задачи массового обслуживания, которые рассматриваются в важнейшем из разделов математики — теории массового обслуживания. Основной задачей теории массо-

вого обслуживания является установление зависимости между вероятностными характеристиками, определяющими функциональные возможности системы массового обслуживания и эффективность ее функционирования. В свою очередь, случайные потоки событий являются основными элементами систем массового обслуживания и довольно широко применяются в качестве различных математических моделей, описывающих реальные физические, технические и другие процессы.

На практике довольно часто получается так, что параметры, определяющие случайный поток событий, неизвестны, либо изменяются со временем, причем данные изменения носят случайный характер. Таким образом, в наблюдаемом потоке событий случайны не только моменты наступления событий, но и интенсивность потока, т.е. имеется двойная стохастика. Это и приводит к изучению дважды стохастических потоков событий [1].

В данной работе рассмотрен один из типов дважды стохастических потоков — полусинхронный поток событий второго порядка, а также его основные свойства.

### 1. Постановка задачи

Рассматривается полусинхронный дважды стохастический поток событий второго порядка, интенсивность которого есть кусочно-постоянный случайный процесс  $\lambda(t)$  с двумя состояниями:  $\lambda(t) = \lambda_1$  и  $\lambda(t) = \lambda_2$  ( $\lambda_1 > \lambda_2 \geq 0$ ). Отметим, что процесс  $\lambda(t)$  является принципиально ненаблюдаемым. Длительность пребывания процесса  $\lambda(t)$  в  $i$ -ом состоянии,  $i = 1, 2$ , определяется двумя случайными величинами.

Первая случайная величина распределена по экспоненциальному закону с функцией распределения  $F_i^{(1)}(t) = 1 - e^{-\lambda_i t}$ ,  $i = 1, 2$ .

Пусть в момент времени  $t$  процесс  $\lambda(t)$  находится в первом состоянии, т.е.  $\lambda(t) = \lambda_1$ . В момент окончания первого состояния происходит розыгрыш пар:

- 1) наступает событие потока и процесс  $\lambda(t)$  переходит из первого состояния во второе, вероятность чего есть  $P_1^{(1)}(\lambda_2 | \lambda_1)$ ;
- 2) наступает событие потока и процесс  $\lambda(t)$  остается в первом состоянии, вероятность чего есть  $P_1^{(1)}(\lambda_1 | \lambda_1)$ ;

причем  $P_1^{(1)}(\lambda_2 | \lambda_1) + P_1^{(1)}(\lambda_1 | \lambda_1) = 1$ .

Пусть в момент времени  $t$  процесс  $\lambda(t)$  находится во втором состоянии, т.е.  $\lambda(t) = \lambda_2$ . В момент окончания второго состояния наступает событие потока и процесс  $\lambda(t)$  переходит из второго состояния во второе с вероятностью единица, т.е. имеет место пуассоновский поток событий с интенсивностью  $\lambda_2$ .

Вторая случайная величина распределена по экспоненциальному закону с функцией распределения  $F_i^{(2)}(t) = 1 - e^{-\lambda_i t}$ ,  $i = 1, 2$ .

В первом состоянии вторая случайная величина ведет себя аналогично первой случайной величине, т.е. если в момент времени  $t$  процесс  $\lambda(t)$  находится в первом состоянии, т.е.  $\lambda(t) = \lambda_1$ , то в момент окончания первого состояния наступает розыгрыш пар:

- 1) наступает событие потока и процесс  $\lambda(t)$  переходит из первого состояния во второе, вероятность чего есть  $P_1^{(2)}(\lambda_2 | \lambda_1)$ ;
- 2) наступает событие потока и процесс  $\lambda(t)$  остается в первом состоянии с вероятностью  $P_1^{(2)}(\lambda_1 | \lambda_1)$ ;

причем  $P_1^{(2)}(\lambda_2|\lambda_1) + P_1^{(2)}(\lambda_1|\lambda_1) = 1$ .

Пусть в момент времени  $t$  процесс  $\lambda(t)$  находится во втором состоянии, т.е.  $\lambda(t) = \lambda_2$ . В момент окончания второго состояния не наступает событие потока и процесс  $\lambda(t)$  переходит из второго состояния в первое с вероятностью единица.

Первая и вторая случайные величины являются независимыми друг от друга. Очевидно, что в сделанных предположениях  $\lambda(t)$  — марковский процесс.

На рис. 1 приведен пример одной из реализаций процесса  $\lambda(t)$ , где  $t_1, t_2, \dots, t_n$  — моменты наступления событий потока.

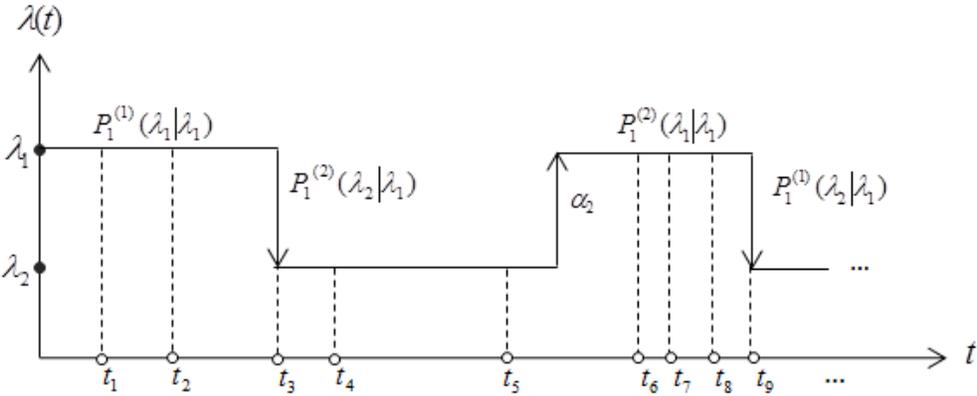


Рис. 1. Формирование полусинхронного потока событий второго порядка

Тогда блочная матрица инфинитезимальных характеристик процесса  $\lambda(t)$  примет вид:

$$D = \begin{vmatrix} -(\lambda_1 + \alpha_1) & 0 & \lambda_1 P_1^{(1)}(\lambda_1|\lambda_1) + \alpha_1 P_1^{(2)}(\lambda_1|\lambda_1) & \lambda_1 P_1^{(1)}(\lambda_2|\lambda_1) + \alpha_1 P_1^{(2)}(\lambda_2|\lambda_1) \\ \alpha_2 & -(\lambda_2 + \alpha_2) & 0 & \lambda_2 \\ & & & \lambda_2 \end{vmatrix} = \|D_0 | D_1\|.$$

Элементами матрицы  $D_1$  являются интенсивности переходов процесса  $\lambda(t)$  из состояния в состояние с наступлением события. Недиagonальные элементы матрицы  $D_0$  — интенсивности переходов из состояния в состояние без наступления события. Диагональные элементы матрицы  $D_0$  — интенсивности выхода процесса  $\lambda(t)$  из своих состояний, взятые с противоположным знаком [2].

Задача настоящей работы заключается в построении имитационной модели полусинхронного дважды стохастического потока второго порядка.

## 2. Результаты имитационного моделирования

Моделирование полусинхронного потока второго порядка осуществлено по формулам, которые получены методом обратных функций из экспоненциального распределения.

Пусть длительность пребывания процесса  $\lambda(t)$  в  $i$ -ом состоянии имеет плотность распределения вероятностей  $p_i(t) = \lambda e^{-\theta_i t}$  и функцию распределения  $F_i(t) = 1 - e^{-\theta_i t}$ ,  $i = 1, 2$ . В соответствии с методом обратных функций [3] обозначим  $F_i(t) = \gamma$ . Тогда имеем  $\gamma = 1 - e^{-\theta_i t}$ , откуда находим формулу моделирования длительности интервалов между соседними событиями в полусинхронном потоке второго порядка следующего вида:  $t = -\frac{1}{\theta_i} \ln(1 - \gamma)$ ,  $\theta_i \in \{\lambda_i, \alpha_i\}$ , где  $\gamma$  — равномерно распределенная на интервале

(0,1) случайная величина,  $\lambda_i$  — параметр экспоненциального распределения в  $i$ -ом состоянии первой случайной величины, а  $\lambda_i$  — параметр экспоненциального распределения в  $i$ -ом состоянии второй случайной величины,  $i = 1, 2$ .

Ниже приведены результаты экспериментов, заключающихся в сравнении наблюдаемых полусинхронных потоков событий второго порядка при различных наборах вероятностей перехода и интенсивностей.

**Эксперимент 1.**

Пусть  $\lambda_1 = 2$ ,  $\lambda_2 = 0,8$ ,  $\alpha_1 = 2$ ,  $\alpha_2 = 0,9$ ,  $T = 10$ . Зафиксируем данные значения и рассмотрим результаты моделирования при крайних значениях вероятностей перехода, т.е.

- 1) в первом случае положим  $P_1^{(1)}(\lambda_1|\lambda_1) = P_1^{(2)}(\lambda_1|\lambda_1) = 1$  и  $P_1^{(1)}(\lambda_2|\lambda_1) = P_1^{(2)}(\lambda_2|\lambda_1) = 0$ ;
- 2) во втором случае, наоборот, положим  $P_1^{(1)}(\lambda_2|\lambda_1) = P_1^{(2)}(\lambda_2|\lambda_1) = 1$  и  $P_1^{(1)}(\lambda_1|\lambda_1) = P_1^{(2)}(\lambda_1|\lambda_1) = 0$ ;

Результаты эксперимента приведены в табл. 1, 2. В строках приведенных таблиц указаны моменты наступления событий.

Таблица 1

**Результаты работы модели в первом случае эксперимента 1**

$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$
0,25790	2,65873	3,16783	4,18563	8,70645	9,02755

Таблица 2

**Результаты работы модели во втором случае первого эксперимента 1**

$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	$t_9$	$t_{10}$	$t_{11}$	$t_{12}$
1,17482	1,37835	1,96753	3,19753	3,56860	4,33565	4,99165	5,97628	6,25678	8,65742	8,96427	9,60826

На рис. 2, 3 приведены реализации случайного процесса  $\lambda(t)$  для эксперимента 1.

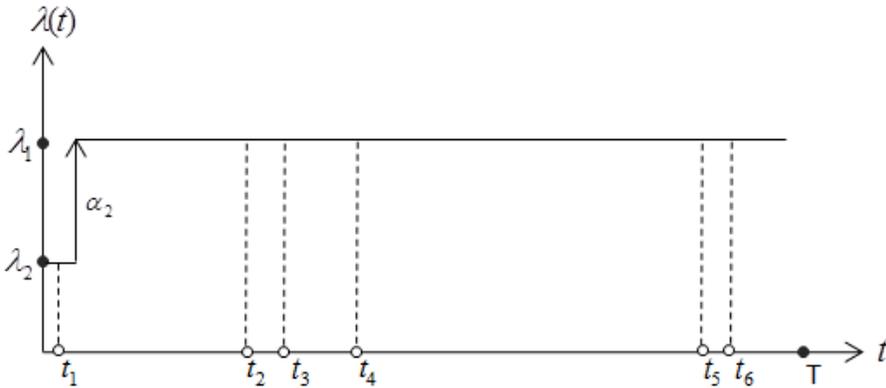


Рис. 2. Реализация полусинхронного потока событий второго порядка в первом случае

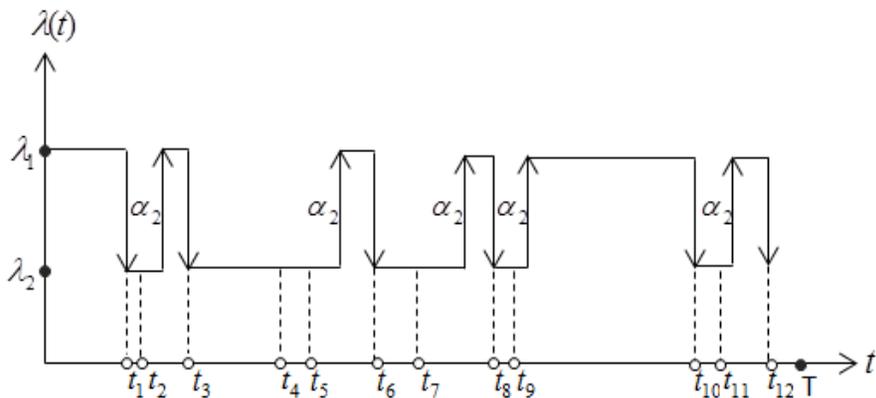


Рис. 3. Реализация полусинхронного потока событий второго порядка во втором случае

Как видно, в первом случае, т.е. при  $P_1^{(1)}(\lambda_1|\lambda_1) = P_1^{(2)}(\lambda_1|\lambda_1) = 1$ , процесс  $\lambda(t)$  находится в первом состоянии дольше, чем во втором, в отличие от второго случая, где процесс  $\lambda(t)$ , находясь в первом состоянии, при наступлении события переходит во второе состояние с вероятностью единица.

### Эксперимент 2.

Пусть  $\alpha_1 = 1$ ,  $\alpha_2 = 0,8$ ,  $P_1^{(1)}(\lambda_1|\lambda_1) = P_1^{(2)}(\lambda_1|\lambda_1) = 0,6$ ,  $P_1^{(1)}(\lambda_2|\lambda_1) = P_1^{(2)}(\lambda_2|\lambda_1) = 0,4$ ,  $T = 10$ . Рассмотрим результаты моделирования в двух случаях:

- 1)  $\lambda_1 = 1$ ,  $\lambda_2 = 0,5$ ;
- 2)  $\lambda_1 = 4$ ,  $\lambda_2 = 3,5$ .

Результаты данного эксперимента приведены в табл. 3, 4. В строках приведенных таблиц указаны моменты наступления событий.

Таблица 3

Результаты работы модели в первом случае эксперимента 2

$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$
1,48021	2,85762	3,51071	3,96251	6,13809	6,96072	8,59374

Таблица 4

Результаты работы модели во втором случае эксперимента 2

$t_1$	$t_2$	...	$t_{12}$	...	$t_{22}$	...	$t_{32}$	$t_{33}$
0,53709	1,27534		4,49031		6,14621		8,89473	9,27684

Как видно, с увеличением интенсивностей потока увеличилось количество событий, наступивших в потоке.

Таким образом, результаты моделирования демонстрируют работоспособность построенной модели.

### Заключение

В настоящей работе рассмотрен полусинхронный дважды стохастический поток событий второго порядка, получена матрица инфинитезимальных характеристик процесса  $\lambda(t)$ , построена имитационная модель данного потока на языке программирования C# в среде Visual Studio 2013, а также проведены статистические эксперименты; анализ результатов подтверждает работоспособность построенной модели.

## ЛИТЕРАТУРА

1. Горцев А.М., Калягин А.А., Нежелская Л.А. Оптимальная оценка состояний обобщенного полусинхронного потока событий // Вестник ТГУ. Управление, вычислительная техника и информатика. 2010. № 2(11). С. 66–81.
2. Горцев А.М., Калягин А.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном полусинхронном потоке // Вестник ТГУ. Управление, вычислительная техника и информатика. 2015. № 1(30). С. 27–37.
3. Соболев И.М. Численные методы Монте-Карло / И.М. Соболев. – М.: Наука, 1973. – 311 с.

# ИССЛЕДОВАНИЕ МНОГОФАЗНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ $SM|(M|\infty)^K$ С «РАСПАКОВКОЙ» ЗАЯВОК В УСЛОВИЯХ ПРЕДЕЛЬНО РЕДКИХ ИЗМЕНЕНИЙ СОСТОЯНИЙ И ВЫСОКОЙ ИНТЕНСИВНОСТИ ВХОДЯЩЕГО ПОТОКА

**М.В. Севостьянова, А.А. Назаров**

*Томский государственный университет  
sevast.rita@gmail.ru*

### 1. Постановка задачи

В данной работе рассматривается многофазная система массового обслуживания с «распаковкой» заявок, входящий поток полумарковский, имеется  $K$  фаз обслуживания с неограниченным числом приборов на каждой фазе, время обслуживания распределено по экспоненциальному закону с параметрами  $\mu_k$  ( $k = \overline{1, K}$ ) соответственно. Заявка, завершив обслуживание на  $k$ -ой фазе, «распаковывается» на  $n$  заявок  $n = \overline{0, \infty}$  с вероятностью  $p_k(n)$ , которые отправляются на  $(k + 1)$ -ую фазу для дальнейшего обслуживания. Таким образом, заявка может породить пачку заявок, либо перейти на следующую фазу в единственном числе или покинуть систему сразу после обслуживания на фазе с номером  $k$  (рис. 1).

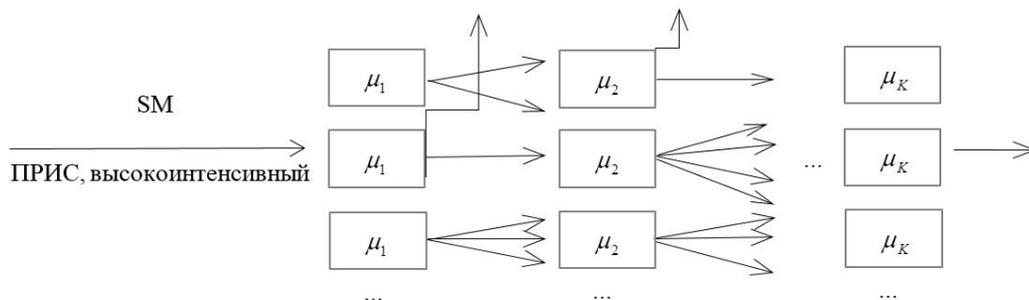


Рис. 1. Система массового обслуживания с «распаковкой» заявок

Полумарковский поток задается полумарковской матрицей  $\mathbf{A}(x)$ . Данную многофазную СМО мы рассматриваем в условии предельно редких изменений состояний входящего потока. Таким образом, запишем данное условие для полумарковского потока в следующем виде:

$$\mathbf{A}(x, \delta) = \mathbf{G}(x) * \mathbf{P}(\delta) = \mathbf{G}(x) * (\mathbf{I} + \delta \cdot \mathbf{Q}). \quad (1)$$

Оно представляет собой произведение Адамара. Здесь элементы матрицы  $\mathbf{G}(x)$  — это условные функции распределения длин интервалов полумарковского потока,  $\mathbf{P}(\delta)$  — матрица вероятностей переходов, имеющая вид  $\mathbf{P}(\delta) = \mathbf{I} + \delta \cdot \mathbf{Q}$ , где  $\delta$  — бесконечно

малая величина,  $\mathbf{I}$  — единичная матрица, матрица  $\mathbf{Q}$  — аналог матрицы инфинитезимальных характеристик.

Обозначим  $z(t)$  — остаточное время, т.е. промежуток от момента времени  $t$  до момента наступления в полумарковском потоке следующего события. Определим полумарковский процесс  $s(t)$  в виде:  $s(t) = \xi(n+1)$ ,  $t_n < t \leq t_{n+1}$ .

Пусть  $i_k(t)$  — число заявок на  $k$ -ой фазе системы в момент времени  $t$ . Запишем вектор  $\mathbf{i}(t) = \{i_1(t), i_2(t), \dots, i_K(t)\}$  — число заявок на фазах системы.

Таким образом, получаем многомерный марковский случайный процесс  $\{s(t), z(t), \mathbf{i}(t)\}$ .

Ставится задача нахождения стационарного многомерного распределения вероятностей случайного процесса  $\mathbf{i}(t)$  — числа заявок в рассматриваемой многофазной системе массового обслуживания.

## 2. Вывод уравнений Колмогорова

Найти решение нашей задачи в явном виде не представляется возможным, поэтому мы выполним исследование методом асимптотического анализа в условии высокой интенсивности входящего потока, т.е. при  $N \rightarrow \infty$ , где  $N$  — неограниченно большой параметр, обеспечивающий условие высокой интенсивности.

Для процесса  $\{s(t), z(t), \mathbf{i}(t)\}$  запишем систему дифференциальных уравнений

Колмогорова для распределения вероятностей  $P(s, z, \mathbf{i}, t) = P\left\{s(t) = s, z(t) < \frac{z}{N}, \mathbf{i}(t) = \mathbf{i}\right\}$

в стационарном виде:

$$\begin{aligned} & \frac{\partial P(s, z, \mathbf{i}, \delta)}{\partial z} - \frac{\partial P(s, 0, \mathbf{i}, \delta)}{\partial z} - \frac{1}{N} \sum_{k=1}^K P(s, z, \mathbf{i}, \delta) \lambda_k \mu_k + \\ & + \sum_{m=1}^M \frac{\partial P(m, 0, \mathbf{i} - \mathbf{e}_1, \delta)}{\partial z} A_{ms}(z, \delta) + \\ & + \frac{1}{N} \sum_{k=1}^{K-1} \sum_{n=0}^{\infty} P(s, z, \mathbf{i} + \mathbf{e}_k - n\mathbf{e}_{k+1}, \delta) p_k(n) (i_k + 1) \mu_k + \\ & + \frac{1}{N} P(s, z, \mathbf{i} + \mathbf{e}_K, \delta) (i_K + 1) \mu_K = 0, \end{aligned} \quad (2)$$

где  $\mathbf{e}_k = \left(0 \ 0 \dots 0 \ 1 \ 0 \dots 0\right)$ ,  $N \rightarrow \infty$ .

Для многомерных частных характеристических функции векторного аргумента  $\mathbf{u}^T = \{u_1, u_2, \dots, u_K\}$

$$H(s, z, \mathbf{u}, \delta) = \sum_{i_1=0}^{\infty} \dots \sum_{i_K=0}^{\infty} e^{ju_1 i_1 + \dots + ju_K i_K} P(s, z, i_1, \dots, i_K, \delta), \quad s = \overline{1, M},$$

система (2) примет вид:

$$\begin{aligned}
& \frac{\partial H(s, z, \mathbf{u}, \delta)}{\partial z} - \frac{\partial H(s, o, \mathbf{u}, \delta)}{\partial z} + e^{ju_1} \sum_{m=1}^M \frac{\partial H(m, o, \mathbf{u}, \delta)}{\partial z} A_{ms}(z, \delta) + \\
& + j \frac{1}{N} \sum_{k=1}^{K-1} \mu_k \frac{\partial H(s, z, \mathbf{u}, \delta)}{\partial u_k} \left( 1 - \sum_{n=0}^{\infty} e^{jnu_{k+1} - ju_k} p_k(n) \right) + \\
& + j \mu_K \frac{1}{N} \frac{\partial H(s, z, \mathbf{u}, \delta)}{\partial u_K} (1 - e^{-ju_K}) = 0,
\end{aligned} \tag{3}$$

Здесь, возвращаясь к условию предельно редких изменений состояний (1) входящего полумарковского потока, осуществим предельный переход при  $\delta \rightarrow 0$ , получим систему

$$\begin{aligned}
& \frac{\partial H(s, z, \mathbf{u})}{\partial z} + \frac{\partial H(s, o, \mathbf{u})}{\partial z} (e^{ju_1} G_{ss}(z) - 1) + \\
& + j \frac{1}{N} \sum_{k=1}^{K-1} \mu_k \frac{\partial H(s, z, \mathbf{u})}{\partial u_k} \left( 1 - \sum_{n=0}^{\infty} e^{jnu_{k+1} - ju_k} p_k(n) \right) + \\
& + j \mu_K \frac{1}{N} \frac{\partial H(s, z, \mathbf{u})}{\partial u_K} (1 - e^{-ju_K}) = 0,
\end{aligned} \tag{4}$$

при начальном условии  $H(s, z, \mathbf{0}) = R(s, z)$ , где  $R(s, z)$  определяется равенством

$$R(s, z) = \lambda_s \int_0^{\bar{z}} (1 - G_{ss}(x)) dx, \quad \lambda_s = \frac{1}{\int_0^{\bar{z}} (1 - G_{ss}(x)) dx}, \quad s = \overline{1, M}. \tag{5}$$

### 3. Метод асимптотического анализа

Запишем функцию  $H(s, z, \mathbf{u})$  в виде:

$$H(s, z, \mathbf{u}) = H_2(s, z, \mathbf{u}) \exp \left\{ \sum_{k=1}^K j \frac{\lambda_s N}{\mu_k} u_k \prod_{l=1}^{k-1} b_l \right\}, \quad b_l = \sum_{n=0}^{\infty} n p_l(n).$$

Далее, подставим данную функцию в (4):

$$\begin{aligned}
& \frac{\partial H_2(s, z, \mathbf{u})}{\partial z} + \frac{\partial H_2(s, 0, \mathbf{u})}{\partial z} (e^{ju_1} G_{ss}(z) - 1) + \\
& + j \frac{1}{N} \sum_{k=1}^{K-1} \mu_k \frac{\partial H_2(s, z, \mathbf{u})}{\partial u_k} \left( 1 - e^{-ju_k} \sum_{n=0}^{\infty} e^{jnu_{k+1}} p_k(n) \right) + \\
& + j \mu_K \frac{1}{N} \frac{\partial H_2(s, z, \mathbf{u})}{\partial u_K} (1 - e^{-ju_K}) - \\
& \lambda H_2(s, z, \mathbf{u}) \sum_{k=1}^{K-1} \left\{ \left( 1 - e^{-ju_k} \sum_{n=0}^{\infty} e^{jnu_{k+1}} p_k(n) \right) \prod_{l=1}^{k-1} b_l \right\} - \\
& - \lambda H_2(s, z, \mathbf{u}) (1 - e^{-ju_K}) \prod_{l=1}^{K-1} b_l = 0.
\end{aligned}$$

Здесь выполним следующие замены:  $\varepsilon^2 = \frac{1}{N}$ ,  $\mathbf{u} = \varepsilon \mathbf{w}$ ,  $H_2(s, z, \mathbf{u}) = F_2(s, z, \mathbf{w}, \varepsilon)$ . Получим задачу для функции  $F_2(s, z, \mathbf{w}, \varepsilon)$ :

$$\begin{aligned} & \frac{\partial F_2(s, z, \mathbf{w}, \varepsilon)}{\partial z} + \frac{\partial F_2(s, 0, \mathbf{w}, \varepsilon)}{\partial z} (e^{j\varepsilon w_1} G_{ss}(z) - 1) + \\ & + j\varepsilon \sum_{k=1}^{K-1} \mu_k \frac{\partial F_2(s, z, \mathbf{w}, \varepsilon)}{\partial w_k} \left( 1 - e^{-j\varepsilon w_k} \sum_{n=0}^{\infty} e^{jn\varepsilon w_{k+1}} p_k(n) \right) + \\ & + j\varepsilon \mu_K \frac{\partial F_2(s, z, \mathbf{w}, \varepsilon)}{\partial w_K} (1 - e^{-j\varepsilon w_K}) - \end{aligned} \quad (6)$$

$$-\lambda F_2(s, z, \mathbf{w}, \varepsilon) \left[ \left( 1 - e^{-j\varepsilon w_k} \sum_{n=0}^{\infty} e^{jn\varepsilon w_{k+1}} p_k(n) \right) \prod_{l=1}^{k-1} b_l \right] + (1 - e^{-j\varepsilon w_K}) \prod_{l=1}^{K-1} b_l = 0.$$

Докажем следующее утверждение.

**Теорема.** Асимптотическое (при  $\varepsilon \rightarrow 0$ ) решение  $F_2(s, z, \mathbf{w})$  задачи (6) имеет вид  $F_2(s, z, \mathbf{w}) = R(s, z)\Phi_2(\mathbf{w})$ , где  $R(s, z)$  определяется из (5), а функция  $\Phi_2(\mathbf{w})$ :

$$\Phi_2(\mathbf{w}) = \exp \left\{ -\frac{1}{2} \sum_{n,m=1}^K V_{nm} w_n w_m \right\}.$$

**Доказательство.**

*Этап 1.* В уравнении (6) выполним предельный переход при  $\varepsilon \rightarrow 0$ :

$$\frac{\partial F_2(s, z, \mathbf{w})}{\partial z} + \frac{\partial F_2(s, 0, \mathbf{w})}{\partial z} (G_{ss}(z) - 1) = 0.$$

Тогда функция  $F_2(s, z, \mathbf{w})$  имеет вид:

$$F_2(s, z, \mathbf{w}) = R(s, z)\Phi_2(\mathbf{w}). \quad (7)$$

*Этап 2.* Представим функцию  $F_2(s, z, \mathbf{w})$  в виде разложения

$$F_2(s, z, \mathbf{w}) = \Phi_2(\mathbf{w}) \left\{ R(s, z) + \sum_{k=1}^K j\varepsilon w_k f_k(s, z) \right\} + O(\varepsilon^2), \quad (8)$$

которое подставим в (6). В результате получим

$$\begin{aligned} & \Phi_2(\mathbf{w}) \left\{ R'(s, z) + \sum_{k=1}^K j\varepsilon w_k f'_k(s, z) \right\} + \\ & + (G_{ss}(z) + j\varepsilon w_1 G_{ss}(z) - 1) \Phi_2(\mathbf{w}) \left\{ R'(s, 0) + \sum_{k=1}^K j\varepsilon w_k f'_k(s, 0) \right\} - \\ & - \lambda \Phi_2(\mathbf{w}) \left\{ R(s, z) + \sum_{k=1}^K j\varepsilon w_k f_k(s, z) \right\} \times \\ & \times \left[ \sum_{k=1}^{K-1} \sum (j\varepsilon w_k - j\varepsilon n w_{k+1}) p_k(n) \prod_{l=1}^{k-1} b_l + j\varepsilon w_K \prod_{l=1}^{K-1} b_l \right] = 0. \end{aligned}$$

Получим выражения для неизвестных функций  $f_k(s, z)$ :

$$\begin{aligned} f'_1(s, z) - f'_1(s, 0)(1 - G_{ss}(z)) + \lambda_s (G_{ss}(z) - R(s, z)) &= 0, \\ f'_k(s, z) - f'_k(s, 0)(1 - G_{ss}(z)) &= 0, \quad \forall k = \overline{2, K}, \end{aligned}$$

интегрируя их по всем значениям  $z \in [0, \infty)$ , имеем:

$$f'_1(s, 0) - \lambda_s f_1(s, \infty) = \frac{\lambda_s^3}{2} (\sigma_s^2 - a_s^2) = \frac{\kappa_s}{2}, \quad s = \overline{1, M},$$

где  $a_s = \frac{1}{\lambda_s}$ ,  $\sigma_s^2 = \int_0^\infty (z^2 dG_{ss}(z) - (z dG_{ss}(z))^2)$ , и выражение для  $k = \overline{2, K}$  имеет вид:

$$f_k(s, z) = c_k R(s, z).$$

Этап 3. В (6) подставим разложение (8), устремим  $z \rightarrow \infty$ , получим следующее равенство:

$$\begin{aligned} & \Phi_2(\mathbf{w}) \left\{ R'(s, 0) + \sum_{k=1}^K j\varepsilon w_k f'_k(s, 0) \right\} \left( j\varepsilon w_1 + \frac{(j\varepsilon w_1)^2}{2} \right) + \\ & + j\varepsilon \sum_{k=1}^{K-1} \mu_k \frac{\partial \Phi_2(\mathbf{w})}{\partial w_k} R(s, \infty) \sum_{n=0}^\infty (j\varepsilon w_k - j\varepsilon n w_{k+1}) p_k(n) + \\ & + j^2 \varepsilon^2 w_K \mu_K \frac{\partial \Phi_2(\mathbf{w})}{\partial w_K} R(s, \infty) - \Phi_2(\mathbf{w}) \left\{ R(s, \infty) + \sum_{k=1}^K j\varepsilon w_k f_k(s, \infty) \right\} \times \\ & \times \left[ \sum_{k=1}^{K-1} \left\{ j\varepsilon w_k - j\varepsilon n w_{k+1} - \frac{(j\varepsilon w_k)^2}{2} - \frac{(j\varepsilon n w_{k+1})^2}{2} + j^2 \varepsilon^2 n w_k w_{k+1} \right\} p_k(n) \prod_{l=1}^{k-1} b_l \lambda_s + \right. \\ & \left. + \left( j\varepsilon w_K - \frac{(j\varepsilon w_K)^2}{2} \right) \prod_{l=1}^{K-1} b_l \lambda_s \right] = 0. \end{aligned}$$

Далее выполним преобразования и устремим  $\varepsilon \rightarrow 0$ :

$$\begin{aligned} & \sum_{k=1}^K \mu_k w_k \frac{\partial \Phi_2(\mathbf{w})}{\partial w_k} - \sum_{k=1}^{K-1} \mu_k w_{k+1} b_k \frac{\partial \Phi_2(\mathbf{w})}{\partial w_k} + \\ & + \Phi_2(\mathbf{w}) \left\{ \frac{w_1^2}{2} (\lambda_s + \kappa) + \lambda_s \left( \sum_{k=1}^K \frac{w_k^2}{2} \prod_{l=1}^{k-1} b_l - \sum_{k=1}^{K-1} w_k w_{k+1} \prod_{l=1}^k b_l + \sum_{k=1}^{K-1} \frac{w_{k+1}^2}{2} b_k^2 \prod_{l=1}^{k-1} b_l \right) \right\} = 0, \end{aligned}$$

откуда не трудно получить, что функция  $\Phi_2(\mathbf{w})$  принимает вид:

$$\Phi_2(\mathbf{w}) = \exp \left\{ -\frac{1}{2} \sum_{n,m=1}^K V_{nm} w_n w_m \right\}, \quad (10)$$

$V_{nm}$  — элементы матрицы ковариации компонент вектора  $\mathbf{w}$ , а функция  $\Phi_2(\mathbf{w})$  — характеристическая функция нормального распределения.

Теорема доказана.

### Заключение

Получили, что в условиях предельно редких изменений состояний и высокой интенсивности входящего полумарковского потока асимптотическое  $K$ -мерное распределение вероятностей числа занятых приборов на фазах рассматриваемой системы представляет собой  $K$ -мерное нормальное распределение с вектором средних значений с компонентами  $\frac{\lambda N}{\mu_k}$  ( $k = \overline{1, K}$ ) и матрицей ковариации  $\mathbf{V}$ .

### ЛИТЕРАТУРА

1. Назаров А.А., Тертугов А.Ф. Теория массового обслуживания: Учебное пособие. – Томск: Изд-во НТЛ, 2004. – 228 с.
2. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. – Томск: Изд-во НТЛ, 2006. – 112 с.
3. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. – 3-е изд, испр. и доп. – М.: КомКнига, 2005. – 408 с.

4. *Лопухова С.В.* Асимптотические и численные методы исследования специальных потоков однородных событий: дис. канд. ф.-м. наук/ С. В. Лопухова. – Томск, 2008. – 157 с.
5. *Moiseev A., Nazarov A.* Investigation of the High Intensive Markov-Modulated Poisson Process // Proc. Of The International Conference On Application Of Information And Communication Technology And Statistics In Economy And Education (ICAICTSEE-2012), Oct. 5-6, 2012, Sofia, Bulgaria. – Sofia: University Of National And World Economy, 2012, pp. 72–77.
6. *Jackson, J.R.*: Networks of the waiting lines. Oper. Res. 5(4), 518–521 (1957)

# СЕКЦИЯ IV. МАТЕМАТИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

## МОДЕЛИРОВАНИЕ И ОБНАРУЖЕНИЕ КАРДИОЛОГИЧЕСКИХ СОБЫТИЙ НА ОСНОВЕ РАЗМЕТКИ ВРЕМЕННОГО РЯДА

А.А. Бояркина, С.И. Колесникова  
Томский государственный университет  
boyarkina.93@inbox.ru, skolesnikova@yandex.ru

### Постановка задачи

Анализ состояний динамических объектов (нелинейных и плохо формализуемых), является сложной и не всегда решаемой (с приемлемой точностью) задачей. Основной проблемой анализа таких (сложных, по Л.А. Растригину) объектов является стохастичность реализаций изучаемого динамического объекта. Данное допущение делает невозможным применение стандартных методов анализа стохастических временных рядов (СВР), предполагающих стационарность основных характеристик. Очевидно, что возникает проблема нахождения фрагмента временного ряда, где изучаемый процесс испытывает изменения, необходимые для их скорейшего обнаружения с целью упреждения нежелательных (катастрофических в некотором смысле) событий. В связи с этим возникает задача определения оптимального объема выборки для распознавания состояний нестационарного нелинейного динамического объекта. Указанная задача имеет непосредственную связь с теорией распознавания образов и с современным ее ответвлением — машинным обучением, над созданием и проработкой алгоритмов трудятся в ВЦ РАН (Ю.И. Журавлев, К.В. Рудаков, К.В. Воронцов и др., [1]).

В данном докладе объектом исследования является СВР — множество выборочных значений случайного процесса  $\{y_k = y_{t_k}, t_k \in T\}$ , описание которого представлено в виде:

$$y_k = x_k + \xi_k, \quad (1)$$

где  $x_k$  — нелинейная детерминированная функция,  $\xi_k$  — случайная нормально распределенная составляющая  $N(0, \sigma)$ ,  $t_k = k\Delta$ ,  $\Delta > 0$ ,  $\Delta$  — интервал дискретизации.

Для решения задачи моделирования и обнаружения кардиологических событий будем использовать метод разметки СВР [1]. С этой целью будем полагать, что для каждого  $i$ -го фрагмента ряда процесс имеет описание вида:

$$y_k^{(i)} = x_k^{(i)} + \xi_k^{(i)}, \quad k \geq 0, \quad i = 1, I, \quad (2)$$

где  $x_k^{(i)} = f_k^{(i)} = f^{(i)}(k\Delta)$ ,  $\Delta > 0$ ,  $f^{(i)}(t) \in R$  — функция, описывающая поведение неслучайной составляющей СВР. Зафиксируем множество функциональных зависимостей  $\Phi = \{f_k(t)\}$ ,  $k = 1, n_f$  как основу алфавита  $M = \{l_1, \dots, l_m\}$  и системы аксиом. Поставим задачу выделения тренда как задачу классификации, в которой каждой точке ряда  $y$  сопоставлялся разметка из фиксированного словаря разметки  $M$ , и нахождения оптимального значения параметра  $k$ , с целью повышения достоверности распознавания состояний ДО.

### Полученные результаты

В докладе обсуждаются следующие полученные результаты. В процессе реализации алгоритма разметки было проведено исследование по обнаружению кардиологиче-

ского события как признака инфаркта миокарда (перегородочного). Обнаружение характерного признака наступления данного события производилось в наиболее информативных в данном случае отведениях  $V_2, V_3$  (рис. 1).

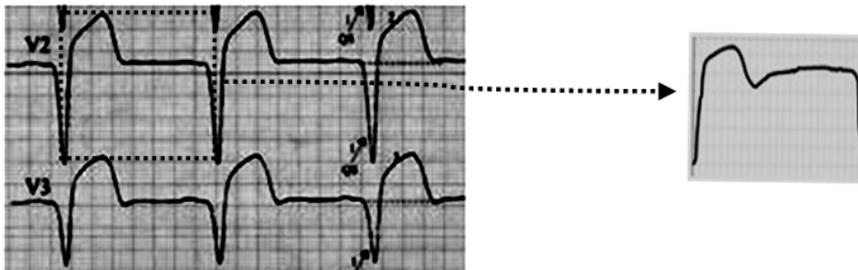


Рис.1. Трансмуральный передне-перегородочно-верхушечный инфаркт миокарда с переходом на боковую стенку ЛЖ

При решении задачи поиска оптимального размера окна для обнаружения признака кардиологического события была исследована зависимость оптимального размера окна от уровня шума (рис. 2).

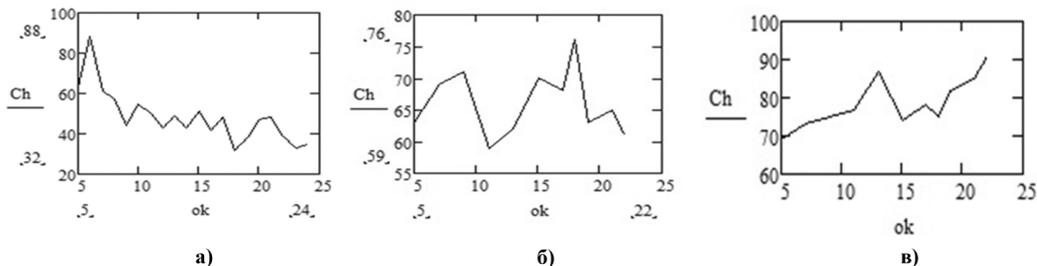


Рис. 2. Зависимость количества правильных ответов  $\nu$  в % от величины окна  $n$ :  
 а)  $N(0, 0.005)$ ; б)  $N(0, 0.06)$ ; в)  $N(0, 0.02)$

Для оценивания качества и корректности полученных результатов было проведено исследование по изучению эффективности работы алгоритма (рис. 3 б) и зависимости величины оптимального окна зависимости от уровня отношение сигнал  $\gamma = \frac{\delta(y + \xi)}{\delta(y)}$  (ОШС) (рис. 3 а).

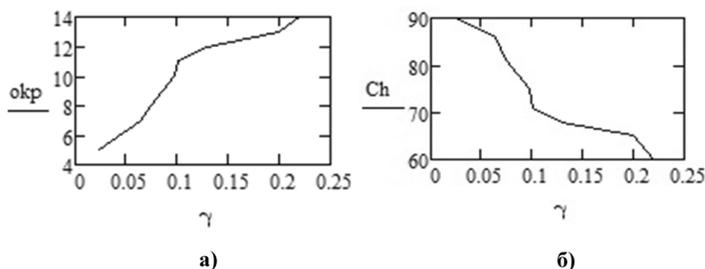


Рис. 3. а) величина оптимального окна  $n$  в зависимости от ОШС;  
 б) эффективность работы алгоритма в зависимости от ОШС

### Анализ полученных результатов

В ходе данного исследования был изучен алгоритм разметки временного ряда, позволяющий обнаруживать аномальные состояния динамических объектов любой при-

роды, по их реализациям в виде СВР. Алгоритм был применен для решения задачи моделирования кардиологических событий, в частности — для обнаружения характерных признаков перегородного инфаркта миокарда. На этапе решения данной задачи, с целью получения устойчивой работы алгоритма и повышения эффективности результатов его работы, был осуществлен поиск оптимального размера окна для обнаружения признака инфаркта миокарда по реализации ЭКГ сигнала в отведениях  $V_2$ ,  $V_3$ . Таким образом, был определен оптимальный размер окна для обнаружения кардиологического события, на этапе поиска которого было проведено исследование зависимости величины оптимального окна от шумовой составляющей в полезном сигнале. Оценивание результатов работы алгоритма было выполнено как изучение его эффективности в зависимости от уровня ОШС.

## ЛИТЕРАТУРА

1. Тырсин А.Н. Идентификация нестационарных экономических процессов на основе дискретно-совпадающих моделей авторегрессии // Известия Уральского государственного экономического университета. – 2004. – № 9. – С. 44–51.
2. Букреев В.Г., Колесникова С.И., Янковская А.Е. Выявление закономерностей во временных рядах в задачах распознавания состояний динамических объектов. – Томск: Изд-во ТПУ, 2010. – 254 с.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. Вып. 1.
4. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование, М.: Финансы и статистика, 2001.
5. Рудаков К.В., Чехович Ю.В. Алгебраический подход к проблеме синтеза обучаемых алгоритмов выделения трендов // ДАН. 2003. Т. 388. № 1. С. 33–36.
6. Ивахненко А.А., Каневский Д.Ю., Рудева А.В. и др. Выявление групп объектов, описанных набором многомерных временных рядов // Математические методы распознавания образов (ММРО\_13). М.: МАКС Пресс, 2007.
7. Васин Е.А., Костенко В.А., Коваленко Д.С. Автоматическое построение алгоритмов, основанных на алгебраическом подходе, для распознавания предаварийных ситуаций динамических систем // Искусственный интеллект. 2006. № 2. С. 130–134.
8. Коваленко Д.С., Костенко В.А., Васин Е.А. Исследование применимости алгебраического подхода к анализу временных рядов // Методы и средства обработки информации. М.: Изд. ВМиК МГУ, 2005. С. 553–559.
9. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. 1978. Т. 33. С. 5–68.
10. Рудаков К.В., Чехович Ю.В. Алгебраический подход к проблеме синтеза обучаемых алгоритмов выделения трендов // ДАН. 2003
11. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. 1978. Т. 33. С. 5–68.
12. Рудаков К.В., Чехович Ю.В. Алгебраический подход к проблеме синтеза обучаемых алгоритмов выделения трендов // ДАН. 2003.
13. Коваленко Д.С., Костенко В.А., Васин Е.А. Исследование применимости алгебраического подхода к анализу временных рядов // Методы и средства обработки информации.

## НЕЛИНЕЙНОЕ УПРАВЛЕНИЕ В БАЛАНСОВОЙ ТРЕХМЕРНОЙ МОДЕЛИ С ХАОТИЧЕСКИМИ РЕЖИМАМИ

**Н.Д. Дубина, С.И. Колесникова**

*Томский государственный университет*  
dubina-nina@mail.ru, skolesnikova@yandex.ru

### Введение

Решение вопроса синтеза устойчивых и робастных систем управления многомерными нелинейными динамическими объектами с возникновением методов управления в скользящем режиме (В.И. Уткин), бэкстепинг (Kokotovic P.V.), аналитического конструирования агрегированных регуляторов (АКАР, А.А. Колесников), принцип инвариантного погружения в задачах (immersion and invariance, Astolfi A., Ortega R.) вышло на новый уровень. Отметим, что все перечисленные методы используют аналитически обоснованный механизм, реализующий принцип самонастройки (адаптации) и инвариантных многообразий. Однако в силу наличия явления детерминированного хаоса

в нелинейных и многосвязных объектах общего универсального алгоритма синтеза управления многомерными с хаотическими режимами сложными объектами не существует, и вопрос управления такими объектами по-прежнему актуален.

В работе рассматривается применение метода аналитического конструирования агрегированных дискретных регуляторов совместно с методом адаптивного управления на многообразиях [1] для нелинейной балансовой модели третьего порядка, описывающей работу малой фирмы [2], при этом предполагается, что в модели описания присутствует внешнее возмущение с неизвестным законом описания, но с ограниченным диапазоном значений.

## 1. Постановка задачи

Рассматривается трехмерная нелинейная модель с хаотическими режимами, содержательно интерпретируемая как модель предприятия, осуществляющего деятельность в сфере компьютерных технологий, с соответствующей подгонкой параметров модели. Понятно, что параметры в системе, описывающие экономическую деятельность фирмы, в реальных условиях могут отклоняться от значений, гарантирующих устойчивое состояние [2].

С этой целью ставится задача такого управления объектом, которое обеспечивало бы асимптотически устойчивое движение изображающей точки объекта из любой начальной точки фазового пространства в окрестность аналитически заданного многообразия как модели целевого состояния  $\psi(X_1, X_2, X_3) = 0$ , где  $\psi(\cdot)$  — известная функция, а объект управления имеет следующее описание:

$$\begin{cases} \frac{dX_1}{dt} = \alpha X_2 X_3 - \gamma X_1, \\ \frac{dX_2}{dt} = \mu(X_2 + X_3) - \beta X_1 X_3, \\ \frac{dX_3}{dt} = \delta X_2 - \lambda X_3 + Z + u. \end{cases} \quad (1)$$

В описании (1) переменные  $X_1, X_2, X_3$  означают объем затрат на содержание сотрудников, величину капитала, объем затрат на использование новых технологий, соответственно;  $\alpha, \beta, \gamma, \delta, \lambda, \mu$  — некоторые коэффициенты пропорциональности,  $Z$  — внешнее неизвестное (ограниченное) возмущение, функция времени. Следует отметить, что объект (1) без двух последних слагаемых в последнем уравнении был исследован на устойчивость в работе [2].

Задача синтеза системы управления состоит в определении управления  $u(x)$  как функции координат состояния системы, обеспечивающего работу фирмы при заданных условиях и доставляющий глобальный минимум оптимизирующему функционалу качества управления:

$$J = \int_0^{\infty} (T_i \dot{\psi}_i^2(t) + \psi_i^2(t)) dt \rightarrow \min_{u(t)}, \quad 0 < T_i < \infty, \quad (2)$$

где величины  $T_i, i = 1, 2$  интерпретируются, как величины, пропорциональные времени движения изображающей точки системы управления до пересечения многообразий и являются параметрами настройки регулятора;  $\psi_i(t)$  — целевые макропеременные,  $i$  — при скалярном управлении номер этапа синтеза,  $i = 1, 2$ ,  $u(t)$  — управление. Как следует из вида (2), на каждом этапе синтеза управления задается свой оптимизирующий функционал.

## 2. Методика решения поставленной задачи управления

Решение задачи основано на методе аналитического конструирования агрегированных регуляторов совместно с методом адаптивного управления на многообразиях научной школы А.А. Колесникова [1].

Метод АКАР опирается на аппарат инвариантов и целевых многообразий. Основные положения алгоритма синтеза системы управления для объекта (1) реализуются следующими шагами.

1. Задается макропеременная  $\psi(x)$ ,  $X = (X_1, X_2, X_3)$ , равенство нулю которой ( $\psi(X) = 0$ ) задает целевое многообразие, содержательно означающее желаемые предельные соотношения между переменными объекта управления, реализующие технологические цели управления. Для объекта управления (1) в качестве примера вводится целевой инвариант

$$X_1 = X_{10}, \quad (3)$$

интерпретируемый как требование стабилизации затрат на сотрудников посредством управления затратами на новые технологии.

2. Согласно классическому методу АКАР производится расширение фазового пространства с целью перевода внешнего возмущения в фазовую переменную (система управления становится замкнутой):

$$\begin{cases} \frac{dX_1}{dt} = aX_2 X_3 - \gamma X_1, \\ \frac{dX_2}{dt} = \mu(X_2 + X_3) - \beta X_1 X_3, \\ \frac{dX_3}{dt} = \delta X_2 - \lambda X_3 + Z + u, \\ \frac{dZ}{dt} = \eta(X_1 - X_{10}). \end{cases} \quad (4)$$

3. Вводится вспомогательная макропеременная, где неизвестная на этом этапе функция  $\varphi(X)$ , интерпретируется как внутреннее управление:

$$\psi_1(t) = X_3 + \varphi(X_1, X_2, Z).$$

4. Определяется управление  $u$  с точностью до функции  $\varphi(X_1, X_2, Z)$  на основе метода Эйлера Лагранжа:

$$u = f\left(X_1, X_2, Z, \frac{\partial \varphi}{\partial X_1}, \frac{\partial \varphi}{\partial X_2}, \frac{\partial \varphi}{\partial Z}\right).$$

Управление  $u$  переводит изображающую точку системы (4) в окрестность многообразия  $\psi_1(t) = 0$ , движение вдоль которого описывается системой дифференциальных уравнений:

$$\begin{cases} \frac{dX_1}{dt} = aX_2 \varphi(X_1, X_2, Z) - \gamma X_1, \\ \frac{dX_2}{dt} = \mu(X_2 + \varphi(X_1, X_2, Z)) - \beta X_1 \varphi(X_1, X_2, Z), \\ \frac{dZ}{dt} = \eta(X_1 - X_{10}). \end{cases} \quad (5)$$

5. Для декомпозированной системы (5) задается вторая макропеременная:

$$\psi_2(t) = X_1 - X_{10} + \Delta Z = 0.$$

6. Определяются решения функционального уравнения, на которых выполняется условие глобального экстремума функционала (2):

$$T_2 \dot{\psi}_2^2(t) + \psi_2^2(t) = 0, \quad 0 < T_2 < \infty.$$

Заметим, что на инвариантном многообразии  $\psi_2(t) = 0$  выполняется соотношение:

$$X_1 - X_{10} = -\Delta Z,$$

которое гарантирует асимптотическую устойчивость решениям четвертого уравнения системы (5) при  $\eta\Delta > 0$ , в силу справедливости равенства:

$$\frac{\partial Z}{\partial t} = -\eta\Delta Z,$$

и, как следствие, достижение целевого инварианта (3) в установившемся режиме.

7. Определяется явный вид функции  $\varphi(X_1, X_2, Z)$  и явный вид выражения для  $u$ :

$$\begin{aligned} u = & \lambda X_3 - \delta X_2 - Z + \frac{1}{\alpha X_2} \left( \gamma - \Delta\eta - \frac{1}{T_2} \right) (\alpha X_2 X_3 - \gamma X_1) - \\ & - \frac{1}{\alpha X_2} \left( \gamma X_1 - \Delta\eta (X_1 - X_{10}) - \frac{X_1 - X_{10} + \mu Z}{T_2} \right) (\mu (X_2 + X_3) - \beta X_1 X_3) - \\ & - \frac{\mu}{\alpha X_2 T_2} \eta (X_1 - X_{10}) - \frac{\Psi_1}{T_1}. \end{aligned} \quad (6)$$

### 3. Численное моделирование

Проведено численное моделирование построенной системы управления с учетом влияния на систему случайных и гармонических помех, результаты которого свидетельствуют об асимптотически устойчивом стремлении координат объекта к окрестности заданного состояния. Таким образом, благодаря построенному управлению можно гарантировать вывод объекта-фирмы из неустойчивого состояния в заданное состояние (см. рис.1, 2).

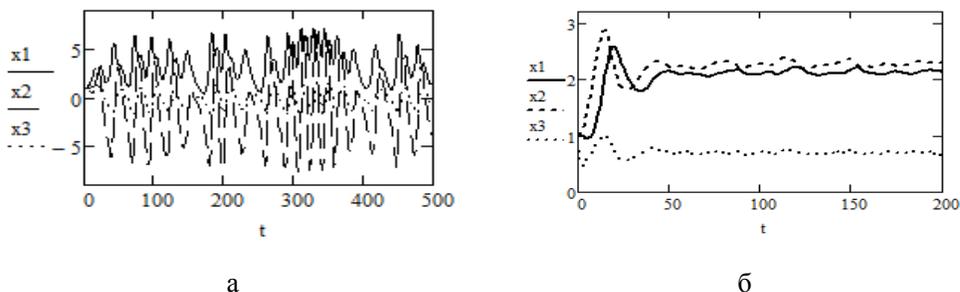


Рис. 1. Поведение объекта в нерасчетных условиях (случайные помехи) с исходными данными:  $\alpha = 1.3319$ ,  $\beta = 3.337$ ,  $\gamma = 0.99$ ,  $\delta = 1.047$ ,  $\lambda = 3.52$ ,  $\mu = 1.68$ ,  $\eta = -0.35$ ,  $\Delta = 0.1$ ;  
 начертание: сплошная линия — затраты на сотрудников, пунктир — величина капитала,  
 точки — затраты на обучение сотрудников;  
 а) — без управления; б) — с управлением

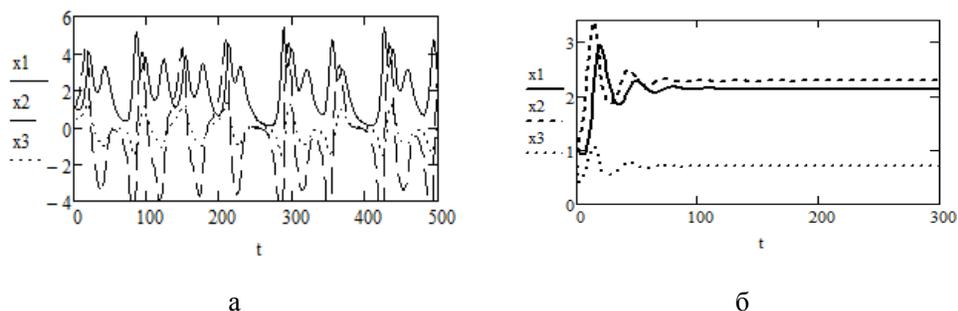


Рис. 2. Поведение объекта в расчетных условиях (гармонические помехи) с исходными данными:  $\alpha = 1.3319$ ,  $\beta = 3.337$ ,  $\gamma = 0.99$ ,  $\delta = 1.047$ ,  $\lambda = 3.52$ ,  $\mu = 1.68$ ,  $\eta = -0.35$ ,  $T_1 = 0.21$ ,  $T_2 = 0.06$ ,  $\Delta = 0.1$ ;  
 начертание: сплошная линия — затраты на сотрудников, пунктир — величина капитала,  
 точки — затраты на обучение сотрудников;  
 а) — без управления; б) — с управлением

Качество полученной системы управления (4), (6) в условиях внешних помех традиционно оценивается по параметрам переходного процесса (амплитуда, длительность) (табл. 1).

Таблица 1

Характеристики системы управления в условиях случайных / гармонических возмущений

Координаты системы (4)	Длительность переходного процесса (в днях)	Амплитуда переходного процесса	
		Максимальное значение отклонения	Среднее значение отклонения
Затраты на сотрудников	69 / 75	1.237 / 1.219	0.067 / 0.066
Капитал	78 / 87	1.344 / 1.29	0.058 / 0.068
Затраты на НИ	52 / 61	0.346 / 0.384	0.017 / 0.038

Результаты численного моделирования, приведенные в табл. 1, свидетельствуют о приемлемом качестве управления.

### Заключение

Для нелинейного трехмерного объекта с хаотическими режимами (1) сконструирована система управления (4), (6), гарантирующая асимптотически устойчивое движение изображающей точки объекта (1) из любого начального состояния в окрестность заданного состояния (3) и дальнейшее удержание объекта в этой окрестности. Полученное управление гарантирует глобальный минимум функционалу качества (2), который, как нетрудно убедиться, является обобщением квадратического функционала качества, традиционно используемого в классической теории адаптивного управления.

Результаты работы могут быть актуальны в системах поддержки принятия решений, используемых в экономических прикладных задачах, в задачах перераспределения финансовых потоков для обеспечения стабильной деятельности малых предприятий, поскольку найденный закон управления представляет собой способ организации обратной связи (упреждающего воздействия), позволяющий контролировать поведение системы.

### ЛИТЕРАТУРА

1. Колесников А.А. Синергетика и проблемы теории управления: сборник научных трудов / Под ред. А.А. Колесникова. М.: ФИЗМАТЛИТ, 2004.
2. Шаповалов В.И. Моделирование синергетических систем: Метод пропорций и другие математические методы / В.И. Шаповалов. – Проспект, 2015. – 136 с.

# УПРАВЛЕНИЕ ДИСКРЕТНЫМИ СИСТЕМАМИ С ЗАПАЗДЫВАНИЕМ ПРИ НЕПОЛНОЙ ИНФОРМАЦИИ О СОСТОЯНИИ И ВОЗМУЩЕНИЯХ С ИСПОЛЬЗОВАНИЕМ НЕПАРАМЕТРИЧЕСКИХ АЛГОРИТМОВ

К.С. Ким, В.И. Смагин

Томский государственный университет  
kks93@rambler.ru, vsm@mail.tsu.ru

## Введение

В работе рассматривается решение задачи управления в условиях неполной информации о состоянии объекта и модели возмущений с учётом запаздывания по управлению. Предполагается, что модель объекта содержит неизвестные параметры, которые входят в модель возмущений. Для определения оптимального управления применяется метод локально-оптимального слежения, выполненный с использованием калмановской фильтрации и экстраполяции с неизвестным входом [1,2]. Выполнено сравнение качества управления и оценивания в зависимости от применяемого метода оценки неизвестного входа, которые реализуются двумя алгоритмами: МНК [1,2] и непараметрическим [3,4].

## 1. Модель объекта

Пусть модель нашего объекта с запаздыванием по управлению описывается дискретным уравнением:

$$\begin{aligned}x(k+1) &= Ax(k) + Bu(k-h) + Fs(k) + \eta(k), \\x(0) &= x_0, u(j) = \psi(j), j = -h, -h+1, \dots, -1,\end{aligned}\quad (1)$$

где  $x(k) \in \mathbb{R}^n$  — вектор состояния;  $u(k-h) \in \mathbb{R}^m$  — вектор управления;  $h$  — количество тактов запаздывания;  $s(k) \in \mathbb{R}^n$  — вектор возмущений,  $x_0$  и  $\psi(j)$  ( $j = -h, -h+1, \dots, -1$ ) — заданные векторы;  $A$ ,  $B$  и  $F$  — заданные постоянные матрицы;  $\eta(k)$  — случайные ошибки наблюдений со следующими характеристиками  $M\{\eta(k)\} = 0$ ,  $M\{\eta(k)\eta^T(k)\} = V\delta_{kj}$ .

Модель канала наблюдений имеет вид:

$$y(k) = H_x x(k) + \eta_x(k),$$

где  $\eta_x(k)$  — независимые гауссовские случайные последовательности с характеристиками

$$M\{\eta_x(k)\} = 0, M\{\eta_x(k)\eta_x^T(k)\} = V_{xkj}.$$

## 2. Модель возмущений

Модель возмущений содержит неизвестные параметры и определяется разностным уравнением:

$$s(k+1) = (R + \Delta R)s(k) + f + \Delta f + q(k), s(0) = s_0, \quad (2)$$

где  $R$  — известная матрица,  $f$  — известный вектор (матрица  $R$  и вектор  $f$  — номинальные параметры модели возмущения),  $\Delta R$  и  $\Delta f$  — некоторые неизвестные матрицы и вектор, которые можно интерпретировать как ошибки определения параметров модели (2). Модель возмущений (2) можно интерпретировать как динамическую модель с неизвестным входом

$$s(k+1) = Rs(k) + f + r(k) + q(k), s(0) = s_0, \quad (3)$$

где  $r(k) = \Delta Rs(k) + \Delta f$  — неизвестный вход.

Предполагается также, что косвенные наблюдения за вектором возмущений описываются следующей моделью:

$$\omega(k) = \Phi s(k) + \tau(k). \quad (4)$$

Здесь  $\omega(k) \in R^{m_1}$  — вектор наблюдений;  $\Phi$  —  $m_1 \times n$ -матрица;  $\tau(k)$  — случайные ошибки наблюдений; в (2)  $s_0$  — случайный вектор начальных условий, независимый от  $q(k)$ ,  $\tau(k)$  и  $M\{s_0\} = \bar{s}_0$ ,  $M\{(s_0 - \bar{s}_0)(s_0 - \bar{s}_0)^T\} = P_0$ ;  $q(k)$ ,  $\tau(k)$  — независимые гауссовские случайные последовательности с характеристиками:

$$M\{q(k)\} = 0, \quad M\{\tau(k)\} = 0, \quad M\{q(k)q^T(j)\} = Q\delta_{kj}, \quad M\{\tau(k)\tau(j)^T\} = T\delta_{kj}. \quad (5)$$

Будем также предполагать, что вектор состояния объекта  $x(k)$  полностью доступен измерению.

Требуется построить управление такое, чтобы вектор выхода системы  $w(k) \in R^{n_2}$

$$w(k) = Hx(k),$$

отслеживал значение заданного вектора  $z(k) \in R^n$ .

### 3. Локально оптимальное управление с запаздыванием

Определим управление, отслеживающее заданный вектор  $z$ , на остове оптимизации следующего локального критерия [5]:

$$I(k) = M\left\{\left(w(k+1) - z(k)\right)^T C \left(w(k+1) - z(k)\right) + u^T(k-h) Du(k-h) / S_0^k, X_0^k\right\}, \quad (6)$$

где  $C > 0$ ,  $D \geq 0$  — весовые матрицы;  $z(k)$  — заданный отслеживаемый вектор;  $S_0^k = \{s(0), s(1), \dots, s(k)\}$  и  $X_0^k = \{x(0), x(1), \dots, x(k)\}$ .

Сначала будем предполагать, что все компоненты вектора  $x(k)$  и  $s(k)$  измеряются точно. Вычислим значения критерия (6):

$$I(k) = u^T(k-h) \left( B^T H^T CHB + D \right) u(k-h) + u^T(k-h) B^T H^T C \left( HAx(k) + HF_s(k) - z(k) \right) + \left( HAx(k) + HF_s(k) - z(k) \right)^T CHBu(k-h). \quad (7)$$

Оптимальное управление определим из условия:

$$\frac{dI(k)}{du(k-h)} = 0. \quad (8)$$

Тогда, в силу (8), получим уравнение:

$$\left( B^T H^T CHB + D \right) u(k-h) + B^T H^T C \left( HAx(k) + HF_s(k) - z(k) \right) = 0. \quad (9)$$

Выражая  $u(k-h)$  из (9), получаем управление в следующем виде:

$$u(k-h) = - \left( B^T H^T CHB + D \right)^{-1} B^T H^T C \left( HAx(k) + HF_s(k) - z(k) \right). \quad (10)$$

Далее, учитывая (1), имеем следующие равенства

$$\begin{aligned} x(k) &= Ax(k-1) + Bu(k-h-1) + Fs(k-1) + \eta(k-1), \\ x(k-1) &= Ax(k-2) + Bu(k-h-2) + Fs(k-2) + \eta(k-2), \\ &\vdots \\ x(k-h+1) &= Ax(k-h) + Bu(k-2h) + Fs(k-h) + \eta(k-h), \end{aligned} \quad (11)$$

тогда, учитывая (11), локально-оптимальное управление (10) представляется в виде

$$u(k-h) = - \left( B^T H^T CHB + D \right)^{-1} B^T H^T C \left( HA^{h+1} x(k-h) + \right.$$

$$+ \sum_{i=1}^h HA^i Bu(k-h-i) + \sum_{i=0}^h HA^i Fs(k-i) - z(k). \quad (12)$$

Управление (12) формируется в момент времени  $k-h$  и для его реализации необходимо знать состояние  $x(k-h)$ , возмущение  $s(k-h)$  и прошлые значения управлений  $u(k-h-i)$ , а также необходимо осуществлять прогноз возмущений для моментов времени  $k, k-1, \dots, k-h+1$ .

Построим управление для случая неполной информации об аддитивном возмущении  $s(k)$  (3). В этом случае оптимизируемый локальный критерий примет вид

$$I(k) = M \left\{ (w(k+1) - z(k))^T C (w(k+1) - z(k)) + u^T(k-h) Du(k-h) / \Omega_0^k, X_0^k \right\}, \quad (13)$$

где  $\Omega_0^k = \{\omega(0), \omega(1), \dots, \omega(k)\}$ ,  $X_0^k = \{x(0), x(1), \dots, x(k)\}$ .

Управление определим на основе принципа разделения, используя оценки фильтрации и прогноза. Тогда получим

$$u(k-h) = - \left( B^T H^T CHB + D \right)^{-1} B^T H^T C \left( HA^{h+1} \hat{x}(k-h) + \sum_{i=1}^h HA^i Bu(k-h-i) + HA^h F \hat{s}_f(k-h) + \sum_{i=0}^{h-1} HA^i F \hat{s}_p(k-i) - z(k) \right), \quad (14)$$

где  $\hat{s}_f(k-h)$  — оценка фильтрации, которая определяется с помощью алгоритма оптимальной калмановской фильтрации:

$$\hat{s}_f(k-h) = R \hat{s}_f(k-h-1) + f + \hat{r}(k-h-1) + K_f(k-h) [\omega(k-h) - \Phi (R \hat{s}_f(k-h-1) + f + \hat{r}(k-h-1))], \quad \hat{s}_f(0) = \bar{s}_0, \quad (15)$$

$$K_f(k-h) = P(k-h/k-h-1) \Phi^T (\Phi P(k-h/k-h-1) \Phi^T + \Gamma)^{-1}, \quad (16)$$

$$P(k-h/k-h-1) = RP(k-h-1)R^T + Q, \quad (17)$$

$$P(k-h) = (E_n - K_f(k-h)\Phi)P(k-h/k-h-1), \quad P(0) = P_0. \quad (18)$$

Фильтр (15) использует информацию, поступившую из канала измерений в момент  $k-h$ . В (14) требуется вычислять также оценки и в моменты большие, чем  $k-h$  (оценки прогноза), поэтому здесь воспользуемся экстраполятором, который позволит вычислить оценку возмущения с прогнозом на 1 такт  $\hat{s}_p(k-h+1)$ :

$$\hat{s}_p(k-h+1) = R \hat{s}_p(k-h) + f + \hat{r}(k-h) + K_p(k-h) (\omega(k-h) - \Phi \hat{s}_p(k-h)), \quad \hat{s}_p(0) = \bar{s}_0, \quad (19)$$

$$K_p(k-h) = RP_{pr}(k-h)\Phi^T (\Phi P_{pr}(k-h)\Phi^T + \Gamma)^{-1}, \quad (20)$$

$$P_{pr}(k-h+1) = (R - K_p(k-h)\Phi)P_{pr}(k-h)(R - K_p(k-h)\Phi)^T + Q + K_p(k-h)TK_p^T(k-h), \quad P_{pr}(0) = P_0, \quad (21)$$

а оценки прогнозов  $\hat{s}_p(k-h+j)$  для  $j \geq 2$  определяются по формулам

$$\hat{s}_p(k-h+j) = R \hat{s}_p(k-h+j-1) + f + \hat{r}(k-h+j-1). \quad (22)$$

Отметим, что в (22) оценки  $\hat{r}(k-h+j-1)$  для значений  $j \geq 2$  могут быть вычислены с использованием методов прогнозирования временных рядов.

#### 4. Методы вычисления оценки $\hat{r}$

В (15) и (19) оценку  $\hat{r}$  можно вычислить несколькими способами. Рассмотрим два метода: метод наименьших квадратов (МНК) и непараметрический метод (использование ядерной оценки плотности).

##### 4.1. Метод наименьших квадратов

Используем метод наименьших квадратов на основе минимизации следующего критерия:

$$J = \sum_{i=1}^k \left\{ \|\chi(i)\|_V^2 + \|r(i-1)\|_W^2 \right\}, \quad (23)$$

где  $\chi(i) = \omega(i) - \Phi\tilde{s}(i)$  ( $\tilde{s}(i) = R\hat{s}(i-1) + f$ );  $V > 0$ ,  $W \geq 0$  — весовые матрицы соответствующих размерностей;  $\|\chi(i)\|_V^2 = \chi^T(i)V\chi(i)$ . В результате имеем оценку:

$$\hat{r}(k) = \left[ \Phi^T V \Phi + W \right]^{-1} \Phi^T V \left\{ \omega(k+1) - \Phi [R\hat{s}(k) + f] \right\}, \quad (24)$$

которая учитывается при определении оценок  $\hat{s}_f$  и  $\hat{s}_p$ , вычисляемых по формулам (15), (19).

##### 4.2. Непараметрический алгоритм

В этом случае получим следующую оценку  $\hat{r}$ :

$$\hat{r}(k) = \left[ \Phi^T V \Phi + W \right]^{-1} \Phi^T V \tilde{w}(k). \quad (25)$$

В (25)  $j$ -ая компонента вектора  $\tilde{w}(k)$  определяется по формуле

$$\tilde{w}_j(k) = \frac{\sum_{i=1}^k w_j(i) K\left(\frac{k-i+1}{h_j}\right)}{\sum_{i=1}^k K\left(\frac{k-i+1}{h_j}\right)}, \quad (26)$$

где  $K(\alpha)$  — функция плотности,  $h$  — коэффициент размытости,

$$K(\alpha) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2}\right). \quad (27)$$

#### 5. Вычисление оценок вектора состояния

В (14) оценки вектора состояния вычисляются с помощью фильтра:

$$\begin{aligned} \hat{x}(k) = & A\hat{x}(k-1) + Bu(k-h-1) + \hat{s}_f(k-1) + \hat{r}_x(k) + K_x(k) \left\{ w_x(k) - \right. \\ & \left. - H \left[ \hat{x}(k-1) + Bu(k-h-1) + \hat{s}_f(k-1) + \hat{r}_x(k) \right] \right\}, \quad \hat{x}(0) = \bar{x}_0, \end{aligned} \quad (28)$$

где  $K_x$  и  $\hat{r}_x(k)$  рассчитываются по аналогии с (16) и (25):

$$\begin{aligned} K_x(k) = & P_x(k/k-1)H^T \left( HP_x(k/k-1)H^T + V \right)^{-1}, \\ P_x(k/k-1) = & AP_x(k-1)A^T + Q_x, \\ P_x(k) = & \left( E_{m_1} - K_x(k)H \right) P_x(k/k-1), \quad P_x(0) = P_{x,0}, \\ w_x(k) = & Hx(k) + \eta(k), \end{aligned} \quad (29)$$

$$\hat{r}_x(k) = \left[ H^T V H + W \right]^{-1} H^T V \left\{ y(k) - H \left[ \hat{x}(k-1) + Bu(k-h-1) + \hat{s}_f(k-1) \right] \right\}$$

и являются оценкой неизвестного входа для модели объекта. Здесь  $\bar{x}_0 = M \{x(0)\}$ ,

$$P_{x,0} = M \left\{ (x(0) - \bar{x}_0)(x(0) - \bar{x}_0)^T \right\}.$$

## 6. Сравнение численных результатов

Проведем сравнение полученных результатов для следующих начальных данных

$$A = \begin{pmatrix} 0,97 & 0 \\ 0 & 0,85 \end{pmatrix}, R = \begin{pmatrix} 0 & 1 \\ 0,1 & 0,58 \end{pmatrix},$$

$$Q = \text{diag}\{0,05 \quad 0,02\},$$

$$T = \text{diag}\{0,05 \quad 0,05\}, D = W = 0, B = H = F = C = V = P_0 = E_2;$$

$$r_1(k) = \begin{cases} 2,5, & \text{если } 0 \leq k < 50, \\ 0,4, & \text{если } 50 \leq k < 100, \\ 1,4, & \text{если } 100 \leq k \leq 150, \end{cases}$$

$$r_2(k) = \begin{cases} 2,4, & \text{если } 0 \leq k < 50, \\ 0,5, & \text{если } 50 \leq k < 100, \\ 1,4, & \text{если } 100 \leq k \leq 150. \end{cases}$$

Результаты моделирования представлены в виде графиков и таблиц. Усреднение производилось по 50 реализациям. Формулы расчета среднеквадратических ошибок представлены ниже.

$$\sigma_{r,i} = \sqrt{\frac{\sum_{k=1}^N (r_i(k) - \hat{r}_i(k))^2}{N-1}}, \quad \sigma_{x,i} = \sqrt{\frac{\sum_{k=1}^N (x_i(k) - z_i(k))^2}{N-1}},$$

$$\sigma_{s,i} = \sqrt{\frac{\sum_{k=1}^N (s_i(k) - \hat{s}_{f,i}(k))^2}{N-1}}, \quad \sigma_{s,p,i} = \sqrt{\frac{\sum_{k=1}^N (s_i(k) - \hat{s}_{p,i}(k))^2}{N-1}}.$$

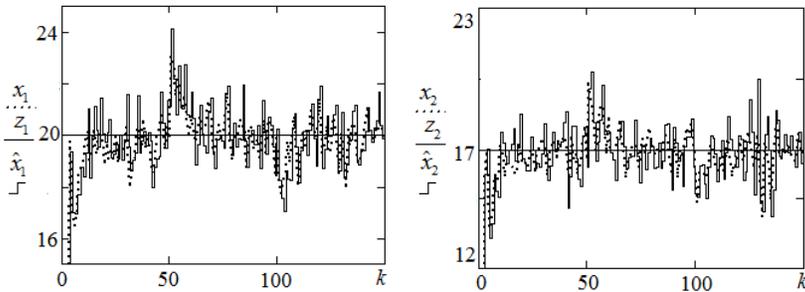


Рис. 1. Реализации компонент векторов  $x(k)$ ,  $\hat{x}(k)$  и  $z(k)$

(оценки объекта построены с использованием метода МНК для оценивания модели возмущений)

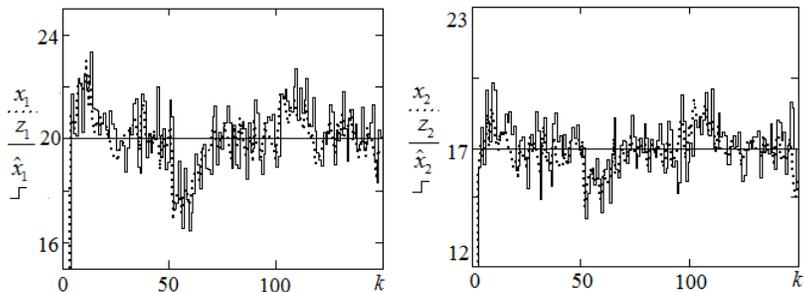
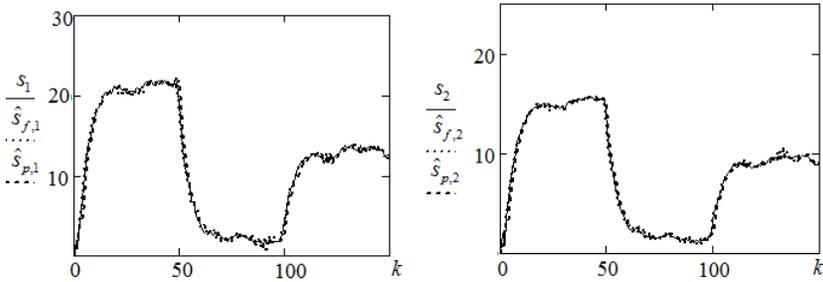


Рис. 2. Реализации компонент векторов  $x(k)$ ,  $\hat{x}(k)$  и  $z(k)$

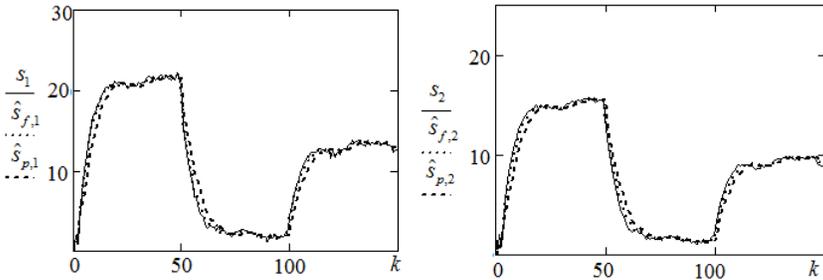
(оценки объекта построены с использованием непараметрического метода для оценивания модели возмущений)

## Среднеквадратические ошибки точности слежения и оценивания вектора состояния объекта

	МНК		Непараметрическая оценка	
	1-ая компонента	2-ая компонента	1-ая компонента	2-ая компонента
Ошибка отслеживания $z$	0,645	0,611	0,447	0,455
Ошибка оценивания $\hat{x}$	0,776	0,743	0,743	0,717

Рис. 3. Реализации компонент векторов  $s(k)$ ,  $\hat{s}_f(k)$ ,  $\hat{s}_p(k)$ 

(оценки построены с использованием метода МНК для оценивания модели возмущений)

Рис. 4. Реализации компонент векторов  $s(k)$ ,  $\hat{s}_f(k)$ ,  $\hat{s}_p(k)$ 

(оценки построены с использованием непараметрического метода для оценивания модели возмущений)

## Среднеквадратические ошибки значений вектора возмущений

	МНК		Непараметрическая оценка	
	1-ая компонента	2-ая компонента	1-ая компонента	2-ая компонента
Оценка фильтрации $\hat{s}_f$	0,222	0,22	0,185	0,169
Оценка прогноза $\hat{s}_p$	0,615	0,55	0,594	0,533

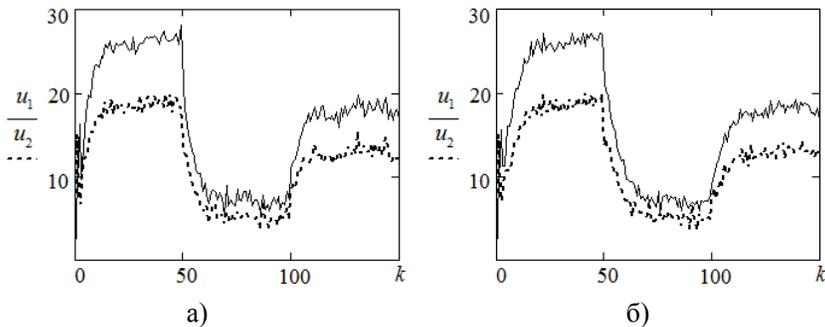


Рис. 5. Реализации вектора управления при использовании а) МНК и б) непараметрического метода

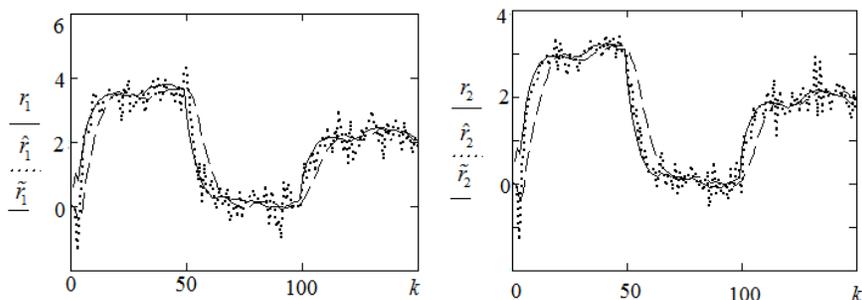


Рис. 6. Реализации компонент вектора неизвестного входа  $r(k)$  и его оценок  $\hat{r}(k)$ , построенной с помощью МНК и  $\tilde{r}(k)$ , полученной с использованием непараметрического алгоритма

Таблица 3

Среднеквадратические ошибки оценивания неизвестного входа

МНК		Непараметрическая оценка	
1-ая компонента	2-ая компонента	1-ая компонента	2-ая компонента
0,518	0,427	0,316	0,273

### Заключение

Результаты моделирования показали, что применение методов непараметрической статистики в задаче управления дискретным объектом с запаздыванием по управлению позволяет повысить точности слежения за заданным вектором  $z(k)$ .

### ЛИТЕРАТУРА

1. Hsieh C.-S. On the optimality of two-stage Kalman filtering for systems with unknown inputs // Asian Journal of Control. 2010. 12(4). P.510–520.
2. Witczak M. Fault diagnosis and fault-tolerant control strategies for non- systems. Chapter 2. Unknown input observers and filters. Lecture Notes in Electrical Engineering. Springer International Publishing, Switzerland, 2014. P. 19–56.
3. Koshkin G.M. Filtering and prediction for discrete systems with unknown input using nonparametric algorithms / G.M. Koshkin, V.I.Smagin // Proc. 10th International Conference on Digital Technologies. Žilina, Slovakia. July 9–11. 2014. pp. 120–124.
4. G.M. Koshkin and V.I. Smagin, “Filtering and prediction for discrete systems with unknown input using non-parametric algorithms”. Proc. 10th International Conference on Digital Technologies. Žilina, Slovakia. July 9–11, 2014. pp. 120–124.
5. Смагин В.И. Адаптивное управление запасами с учетом ограничений и транспортных запаздываний / В.И. Смагин, С.В. Смагин // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – №3(4). – 2008. – С. 19–26.

## ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННО-ТОРГОВОЙ СИСТЕМЫ QUIK ДЛЯ ФОРМИРОВАНИЯ ОПЦИОННОГО ПОРТФЕЛЯ

**М.Э. Фатьянова, М.Е. Семенов**

Томский политехнический университет  
fmare13@gmail.com, sme@tpu.ru

### Введение

Интерес к рынку финансовых продуктов неуклонно растет. При этом брокерские компании стремятся учитывать инвестиционные цели клиентов — получить максимальный доход при заранее определенной величине убытков. Наиболее часто брокеры создают финансовые продукты со стандартными опционными стратегиями. Однако в этом случае бывает сложно реализовать различные запросы инвестора. В работе описан подход конструирования сложных диверсифицированных портфелей биржевых

опционов [1,2]. Целью данной работы является построение сложных опционных стратегий с использованием информационно-торговой системы Quik.

Банк взаимодействует с клиентами (инвесторами), удовлетворяя их пожелания относительно различных финансовых продуктов на российском рынке. Предполагается, что он обеспечивает клиента всеми необходимыми аналитическими материалами с прогнозами интересующих цен активов. При этом инвестор стремится максимизировать будущую прибыль, допуская определенный уровень риска.

Процесс формирования портфеля сводится к нахождению оптимального числа опционных контрактов колл и пут, путем решения задачи линейной оптимизации (линейного программирования, ЗЛП), при заданном ряде ограничений в виде неравенств и равенств.

## 1. Основные определения

Опционный продукт (опционный портфель, ОП) — инвестиционная стратегия, упакованная индивидуально для клиента исходя из его целей и запросов, сформированная посредством купли/продажи опционных контрактов [3,4].

Опцион «call» (колл)/«put» (пут) дает право покупателю опциона купить/продать базисный актив у продавца опциона по цене исполнения в установленные сроки или отказаться от этой покупки. Фьючерс — производный финансовый инструмент, стандартный срочный биржевой контракт купли-продажи базового актива, при заключении которого стороны (продавец и покупатель) договариваются только об уровне цены и сроке поставки. Страйк — фиксированная в опционном контракте цена (цена исполнения), по которой может быть куплен или продан базовый актив в случае исполнения опциона [5]. Цена ASK (BID) цена продажи (покупки) базового актива. При этом спредом называется разность цен ASK и BID [6–8].

В данном исследовании имеется четыре фактора, влияющих на выбор того или иного финансового продукта:

1) Стоимость опционного продукта: отрицательная, положительная и нулевая. Отрицательная стоимость — продукт бесплатен для клиента и предполагает первоначальную денежную выплату в момент создания инвестиционного портфеля (монетизация). Положительная — стоимость продукта, установленная банком, оплачивается инвестором в момент формирования портфеля (демонетизация). Нулевая — продукт является полностью бесплатным для клиента [6–8].

2) Сценарий движения цены базового актива опционного портфеля (например, «бычий» или «медвежий» спреда).

3) Портфель должен иметь ограниченный максимальный уровень убытка в случае нереализации прогноза движения цены актива инвестора.

4) Если цена базового актива в момент экспирации опционного продукта совпадет с прогнозируемым значением цены, портфель должен принести максимальную прибыль.

Введём предположения и условные обозначения.

1. Предполагается, что банк имеет доступ к срочному рынку Московской биржи FORTS, на котором заключаются сделки купли-продажи различных деривативов с соответствующими сроками экспирации.

2. Базовый актив (БА) — обыкновенные акции ПАО «Газпром». Тогда в качестве деривативов будет выступать маржируемый опцион конкретного вида на фьючерсный контракт на акции ПАО «Газпром».

Торговлю маржируемыми опционами запустила биржа РТС в 2009 г. Основное отличие маржируемых от немаржируемых опционов состоит в том, что покупатель сразу не платит весь размер премии продавцу, т.е. в момент заключения сделки на счетах не происходит движения реальных денежных средств. Вместо этого с обеих сторон (покупателя и продавца) удерживается гарантийное обеспечение (ГО). При этом по итогам

изменения цены в дневной (14:00 мск.) и вечерний (18:45 мск.) клиринг им начисляется положительная или отрицательная разница (вариационная маржа) [5].

Гарантийное обеспечение (ГО) — сумма, составляющая некоторый процент от стоимости базового актива, которая должна находиться на счете клиента. Гарантийное обеспечение (ГО) обычно варьируется от 10 до 25 процентов от стоимости базового актива и указывается в спецификации фьючерсного контракта. Размер ГО тесно связан с волатильностью (динамикой изменения) цены на рынке.

Гарантийное обеспечение определяет депозитную маржу (возвращаемая страховая сумма, которая взимается биржей с обеих сторон (покупателя и продавца)) при покупке/продаже опционных контрактов.

Рассмотрим формирование портфеля в системе Quik.

## 2. Формирование портфеля в Quik

Пусть инвестор выдвигает прогноз движения цен индекса РТС от текущего значения  $M_{\text{now}} = 89470$  руб. до ожидаемого значения  $M_E = 92500$  руб., в котором он желает получить максимальный доход. Динамика изменения индекса представлена на рис. 1. При этом инвестор хочет получить 1000 руб. наличными в момент приобретения продукта и ограничить максимальный убыток величиной  $L = 460$  руб.



Рис. 1. График изменения значения индекса от времени

Для удовлетворения предпочтений инвестора следует сформировать портфель из 6 опционов «call» (колл) на фьючерсный контракт на индекс РТС и 6 опционов «put» (пут) с одним сроком исполнения и различными страйками.

Процесс формирования портфеля сводится к нахождению оптимального числа опционных контрактов колл и пут  $(X_{\text{opt}}, Y_{\text{opt}})$ , путем решения задачи линейной оптимизации (линейного программирования, ЗЛП), состоящей из 12 переменных (по 6 опционов колл и пут с различными страйками) при заданном ряде ограничений в виде неравенств и равенств. В результате решения задачи находится вектор оптимальных долей опционов колл и пут:  $(X_{\text{opt}}, Y_{\text{opt}})(X_{\text{opt}(1)}, \dots, X_{\text{opt}(n)}, Y_{\text{opt}(1)}, \dots, Y_{\text{opt}(n)})$ .

Таким образом, полученное оптимальное количество опционов  $(X_{\text{opt}}, Y_{\text{opt}})$  необходимо купить (продать) при формировании продукта для достижения максимального значения функции прибыли  $F_{\text{opt}}(P, Q, X_{\text{opt}}, Y_{\text{opt}}, M)$  в момент экспирации.

В соответствии с методикой конструирования опционного продукта работы [9] сформирован портфель, представленный на рис. 2. В соответствии с конкретным страйком и типом опциона, рассчитан оптимальный план.

Инструмент	Тип	Страйк	Дата исп.	Контракт	К-во	Поз. откр. по
✓ Опцион	Call	87 500	21.04.16	RI87500BD6	4	2 840,00
✓ Опцион	Call	90 000	21.04.16	RI90000BD6	4	1 350,00
✓ Опцион	Call	92 500	21.04.16	RI92500BD6	-8	500,00
✓ Опцион	Put	90 000	21.04.16	RI90000BP6	-4	1 820,00
✓ Опцион	Put	92 500	21.04.16	RI92500BP6	10	3 600,00
✓ Опцион	Put	95 000	21.04.16	RI95000BP6	-6	7 670,00

Рис. 2. Портфель опционного продукта

Функция прибыли опционного продукта представлена на рис. 3.

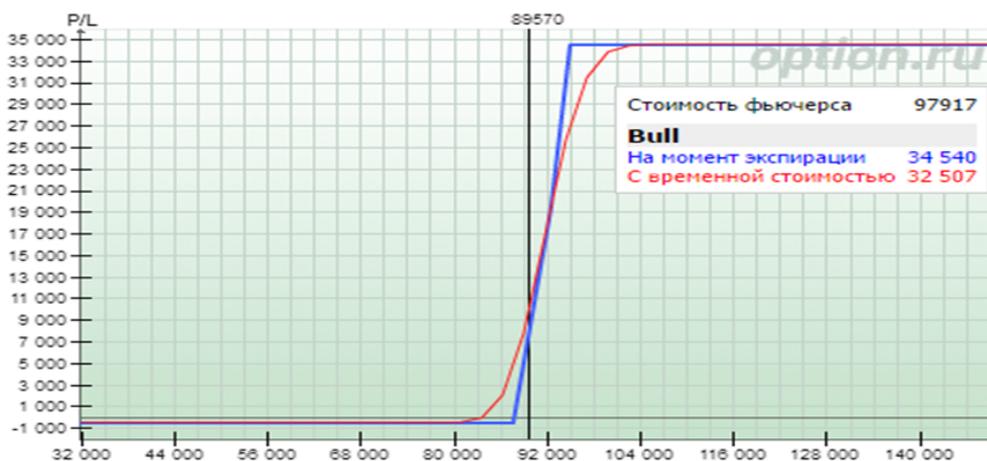


Рис. 3. График изменения величины прибыли в зависимости от страйка

## Заключение

В данном исследовании ЗЛП была решена симплекс-методом и методом Монте-Карло. Идея симплекс-метода состоит в монотонном изменении величины целевой функции при переходе к следующему плану. Основным недостатком симплекс-метода — оптимальный план является нецелочисленным и приходится прибегать к округлению. Для устранения указанного недостатка использован метод Монте-Карло, который позволяет получить целочисленный оптимальный план.

## ЛИТЕРАТУРА

1. Вайн С. Опционы: Полный курс для профессионалов / С. Вайн. – М.: Альпина Бизнес Букс, 2008. – 466 с.
2. Буренин А.Н. Форварды, фьючерсы, опционы, экзотические производные. – М.:НТО – 2008.– 512 с.
3. Фельдман А.Б. Производные финансовые и товарные инструменты: Учебник. – Финансы и статистика, 2005. – 304 с.
4. Натенберг Шелдон, Ник Антилл. Опционы: Волатильность и оценка стоимости. Стратегии и методы опционной торговли. / Пер. с англ. – М.: Альпина Паблишерз, 2011. – 546 с.
5. Буренин А.Н. Рынок ценных бумаг и производных финансовых инструментов. – М.: Научно-техническое общество имени академика С. И. Вавилова, 2011. – 394 с.
6. Курочкин С.В., Пичугин И.С. Структурированный коллар: построение сложных опционных продуктов // Рынок Ценных Бумаг. 2005. № 14 (293). С. 64–68
7. Курочкин С.В. Функции выплат, реализуемые с помощью опционных стратегий // Экономика и математические методы. 2005. Т. 41, № 3.
8. Пичугин И.С. Структурирование опционных продуктов на основе метода оптимизации конечных денежных выплат. – Диссертация на соискание ученой степени кандидата экономических наук / ГУ ВШЭ. – М., 2007. – 154 с.
9. Фатьянова М.Э. Конструирование сложных портфелей биржевых опционов [Электронный ресурс] // Перспективы развития фундаментальных наук: сборник научных трудов XIII Международной конференции студентов и молодых учёных, Томск, 26–29 Апреля 2016. – Томск: ТПУ, 2016 – Т. 3. Математика – С. 114–116 – [http://science-persp.tpu.ru/Arch/Proceedings\\_2016\\_vol\\_3.pdf](http://science-persp.tpu.ru/Arch/Proceedings_2016_vol_3.pdf)

## НАШИ АВТОРЫ

Анисимова Анна Александровна	бакалавр 4 курса ФПМК ТГУ siberienne94@yandex.ru
Белая Елена Владимировна	магистрант 2 курса ФПМК ТГУ elena-belaya-fpmk@ya.ru
Белошапко Георгий Алексеевич	аспирант ИСИ СО РАН ordcoder@gmail.com
Боронихина Елена Александровна	магистрант 2 курса ФПМК ТГУ lin_san@mail.ru
Бояркина Анна Александровна	магистрант 1 курса ФПМК ТГУ boyarkina.93@inbox.ru
Бударина Евгения Александровна	бакалавр 3 курса ФПМК ТГУ budarinajane@gmail.com
Бычков Игорь Валерьевич	зав. каф. РиЭ ФФ ЧелГУ, проф., д.ф.-м.н. bychkov@csu.ru
Веденеев Виктор Сергеевич	преподаватель каф. РиЭ ФФ ЧелГУ ingafen@gmail.com
Дарибаева Нурзия Туребековна	бакалавр 4 курса филиала «Восход» Московско- го авиационного института nurz_dar7@mail.ru
Дмитриев Юрий Глебович	зав. каф. ТК ФПМК ТГУ, доц., д.ф.-м.н. dmitr@fpmk.tsu.ru
Дубина Нина Дмитриевна	магистрант 1 курса ФПМК ТГУ dubina-nina@mail.ru
Дудкина Мария Владимировна	бакалавр 4 курса ФПМК ТГУ masha_gold@mail.ru
Жукова Екатерина Викторовна	магистрант 1 курса ФПМК ТГУ water_melon93@mail.ru
Журко Елена Сергеевна	магистрант 1 курса ФПМК ТГУ helenal993@mail.ru
Замщикова Анастасия Александровна	бакалавр 4 курса ФПМК ТГУ zamashikova.n@mail.ru
Зенкова Жанна Николаевна	доцент каф. ТК ФПМК ТГУ, доц., к.ф.-м.н. thankoffjean@mail.ru
Калякин Сергей Николаевич	с.н.с. ИХХТ СО РАН, к.х.н. snk@icct.ru
Ким Константин Станиславович	магистрант 1 курса ФПМК ТГУ kks93@rambler.ru
Колбасова Вера Александровна	магистрант 2 курса ФПМК ТГУ vera-kolbasova@mail.ru
Колесникова Светлана Ивановна	проф. каф. ВМиММ ФПМК ТГУ, д.т.н. skolesnikova@yandex.ru
Конев Виктор Васильевич	зав. каф. ВМиММ ФПМК ТГУ, проф., д.ф.-м.н. vvkonev@mail.tsu.ru
Коробейников Артем Игоревич	аспирант ИХХТ СО РАН artkor777@rambler.ru
Кошевая Тамара Олеговна	магистрант 1 курса ФПМК ТГУ tomakoshevaya@gmail.com
Кошкин Геннадий Михайлович	проф. каф. ТК ФПМК ТГУ, проф., д.ф.-м.н. kgm@mail.tsu.ru
Лисовская Екатерина Юрьевна	аспирант ФПМК ТГУ ekaterina_lisovs@mail.ru
Луков Вадим Юрьевич	магистрант 1 курса ФПМК ТГУ lukov_vadim@rambler.ru
Макеева Оксана Борисовна	магистрант 1 курса ФПМК ТГУ oksi-mak-tsu@mail.ru
Мизюркина Ольга Игоревна	бакалавр 4 курса ФПМК ТГУ olga.mizyurkina@mail.ru
Моисеев Александр Николаевич	доц. каф. ПИ ФИнФ ТГУ, доц., к.т.н. moiseev.tsu@gmail.com

Моисеева Светлана Петровна	проф. каф. ТВиМС ФПМК ТГУ, доц., д.ф.-м.н. smoiseeva@mail.ru
Назаренко Богдан Николаевич	бакалавр 4 курса ФПМК ТГУ nazarenkob@sibmail.com
Назаров Анатолий Андреевич	зав. каф. ТВиМС ФПМК ТГУ, проф., д.т.н. nazarov.tsu@gmail.com
Нежельская Людмила Алексеевна	доц. каф. ИО ФПМК ТГУ, доц., к.т.н. ludne@mail.ru
Непомнящий Валерий Александрович	зав. лаб. ТП ИСИ СО РАН, доц., к.ф.-м.н. vner@iis.nsk.su
Овчинникова Татьяна Сергеевна	магистрант 2 курса ФПМК ТГУ tanyalastochkina@mail.ru
Пальянова Юлия Евгеньевна	бакалавр 3 курса ФПМК ТГУ pje0360@gmail.com
Савченко Полина Алексеевна	бакалавр 3 курса ФПМК ТГУ sapi95@rambler.ru
Севостьянова Маргарита Владимировна	аспирант ФПМК ТГУ sevost.rita@gmail.com
Семенов Михаил Евгеньевич	доц. каф. ВМиМФ ФТИ ТПУ, доц., к.ф.-м.н. sme@tpu.ru
Сибирякова Валентина Александровна	ст. преп. каф. Пр. ФПМК ТГУ val349@mail.ru
Сивушина Анастасия Леонидовна	бакалавр 4 курса ФПМК ТГУ ansivushina@gmail.com
Сидорова Екатерина Филипповна	бакалавр 3 курса ФПМК ТГУ katusha_sidorova@mail.ru
Смагин Валерий Иванович	проф. каф. ИО ФПМК ТГУ, проф., д.т.н. vsm@mail.tsu.ru
Тумашкина Диана Александровна	бакалавр 3 курса ФПМК ТГУ diana1323@mail.ru
Фатьянова Маргарита Эдуардовна	магистрант 2 курса ФПМК ТГУ fmare13@gmail.com
Цыбулькикова Нина Романовна	бакалавр 4 курса ФПМК ТГУ tsybulnikova94@gmail.com

## АВТОРСКИЙ УКАЗАТЕЛЬ

Анисимова Анна Александровна .....	92
Белая Елена Владимировна.....	34
Белошапко Георгий Алексеевич.....	3
Боронихина Елена Александровна.....	7
Бояркина Анна Александровна .....	120
Бударина Евгения Александровна .....	37
Бычков Игорь Валерьевич .....	10
Веденеев Виктор Сергеевич .....	10
Дарибаева Нурзия Туребековна .....	15
Дмитриев Юрий Глебович .....	39
Дубина Нина Дмитриевна.....	122
Дудкина Мария Владимировна .....	19
Жукова Екатерина Викторовна .....	43
Журко Елена Сергеевна .....	47
Замщикова Анастасия Александровна.....	52
Зенкова Жанна Николаевна .....	47, 52, 57, 62
Калякин Сергей Николаевич .....	22
Ким Константин Станиславович .....	127
Колбасова Вера Александровна .....	96
Колесникова Светлана Ивановна .....	120, 122
Конев Виктор Васильевич.....	65
Коробейников Артем Игоревич.....	22
Кошечкина Тамара Олеговна .....	71
Кошкин Геннадий Михайлович .....	43
Лисовская Екатерина Юрьевна .....	96, 99
Луков Вадим Юрьевич.....	74
Макеева Оксана Борисовна.....	57
Мизюркина Ольга Игоревна .....	25
Моисеев Александр Николаевич .....	92
Моисеева Светлана Петровна.....	99
Назаренко Богдан Николаевич .....	65
Назаров Анатолий Андреевич .....	114
Нежелская Людмила Алексеевна .....	104, 109
Непомнящий Валерий Александрович .....	3
Овчинникова Татьяна Сергеевна.....	29

Пальянова Юлия Евгеньевна .....	79
Савченко Полина Алексеевна .....	39
Севостьянова Маргарита Владимировна .....	114
Семенов Михаил Евгеньевич .....	133
Сибирякова Валентина Александровна .....	7
Сивушина Анастасия Леонидовна .....	85
Сидорова Екатерина Филипповна .....	104
Смагин Валерий Иванович .....	127
Тумашкина Диана Александровна .....	109
Фатьянова Маргарита Эдуардовна .....	133
Цыбульникова Нина Романовна .....	62

# СОДЕРЖАНИЕ

<b>СЕКЦИЯ I. МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН И КОМПЬЮТЕРНЫХ СЕТЕЙ .....</b>	<b>3</b>
<b>Белашапко Г.А., Непомнящий В.А.</b> Анализ и верификация иерархических временных типизированных сетей Петри .....	3
<b>Боронихина Е.А., Сибирякова В.А.</b> Исследование гибридных эвристических методов решения задачи коммивояжера .....	7
<b>Веденеев В.С., Бычков И.В.</b> Конфигурирование нейронных сетей Кохонена .....	10
<b>Дарибаева Н.Т.</b> Комплексная система обработки метеорологической и аэрологической информации при обеспечении подготовки стартов и пусков ракет космического назначения .....	15
<b>Дудкина М.В.</b> Использование графа для визуального задания базовых правил документооборота .....	19
<b>Коробейников А.И., Калякин С.Н.</b> Создание автоматизированного электронного справочника для расчета термодинамических и физико-химических свойств сложных химических систем .....	22
<b>Мизюркина О.И.</b> Проектирование и разработка информационного хранилища для автоматизированной диалоговой системы медицинской диагностики .....	25
<b>Овчинникова Т.С.</b> Обучающая программа на тему «Дерево ван Эмде Боаса» .....	29
<b>СЕКЦИЯ II. ПРИКЛАДНОЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ.....</b>	<b>34</b>
<b>Белая Е.В.</b> Оценивание вероятностных характеристик с учетом информации о качественном признаке.....	34
<b>Бударина Е.А.</b> Оценивание симметричных распределений с неизвестным центром .....	37
<b>Дмитриев Ю.Г., Савченко П.А.</b> Регрессионные модели заработной платы в Российской Федерации .....	39
<b>Жукова Е.В., Кошкин Г.М.</b> Исследование тарифных ставок при страховании автотранспорта: статистическое моделирование .....	43
<b>Журко Е.С., Зенкова Ж.Н.</b> Метод ценообразования PSM для случая цензурированных выборок малого объема.....	47
<b>Замщикова А.А., Зенкова Ж.Н.</b> Прогнозирование временных рядов на основе сингулярного спектрального анализа .....	52
<b>Зенкова Ж.Н., Макеева О.Б.</b> Влияние вида распределения стоимости оборотных средств на показатели оборачиваемости по интервальным данным.....	57
<b>Зенкова Ж.Н., Цыбульникова Н.Р.</b> Исследование цены спроса на продукцию ЗАО «Натуральное мыло» методом Штоцеля по цензурированным данным .....	62
<b>Конев В.В., Назаренко Б.Н.</b> Экспериментальное исследование непараметрической процедуры CUSUM для обнаружения разладки в авторегрессионных моделях с неизвестным конечным параметром .....	65
<b>Кошечкина Т.О.</b> Улучшенные оценки доли в выборочном контроле качества.....	71
<b>Луков В.Ю.</b> Непараметрические алгоритмы идентификации, прогнозирования и управления в ARX-процессах.....	74

<b>Пальянова Ю.Е.</b> Математические модели зависимости производства и потребления от возрастной структуры населения .....	79
<b>Сивушина А.Л.</b> Использование дополнительной информации в оценке удовлетворенности населения услугами методом детерминационного анализа .....	85
<b>СЕКЦИЯ III. ПРИКЛАДНОЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ.....</b>	<b>92</b>
<b>Анисимова А.А., Моисеев А.Н.</b> Имитационное моделирование двухфазной RQ-системы	92
<b>Колбасова В.А., Лисовская Е.Ю.</b> Исследование суммарного объема требований в бесконечнолинейной СМО с рекуррентным входящим потоком .....	92
<b>Лисовская Е.Ю., Моисеева С.П.</b> Асимптотический анализ системы MMPP/GI/∞ с обслуживанием требований случайного объема.....	99
<b>Нежелская Л.А., Сидорова Е.Ф.</b> Имитационное моделирование обобщенного синхронного потока второго порядка.....	104
<b>Нежелская Л.А., Тумашкина Д.А.</b> Имитационная модель полусинхронного потока второго порядка .....	109
<b>Севостьянова М.В., Назаров А.А.</b> Исследование многофазной системы массового обслуживания SM (M ∞)K с «распаковкой» заявок в условиях предельно редких изменений состояний и высокой интенсивности входящего потока.....	114
<b>СЕКЦИЯ VI. МАТЕМАТИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ .....</b>	<b>120</b>
<b>Бояркина А.А., Колесникова С.И.</b> Моделирование и обнаружение кардиологических событий на основе разметки временного ряда .....	120
<b>Дубина Н.Д., Колесникова С.И.</b> Нелинейное управление в балансовой трехмерной модели с хаотическими режимами .....	122
<b>Ким К.С., Смагин В.И.</b> Управление дискретными системами с запаздыванием при неполной информации о состоянии и возмущениях с использованием непараметрических алгоритмов.....	127
<b>Фатьянова М.Э., Семенов М.Е.</b> Использование информационно-торговой системы Quik для формирования опционного портфеля.....	133
<b>НАШИ АВТОРЫ .....</b>	<b>137</b>
<b>АВТОРСКИЙ УКАЗАТЕЛЬ .....</b>	<b>139</b>

Научное издание

**МАТЕРИАЛЫ  
IV Всероссийской молодежной  
научной конференции  
«МАТЕМАТИЧЕСКОЕ  
И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ  
ИНФОРМАЦИОННЫХ, ТЕХНИЧЕСКИХ  
И ЭКОНОМИЧЕСКИХ СИСТЕМ»**

**Томск, 20–21 мая 2016 г.**

*Под общей редакцией  
кандидата технических наук И.С. Шмырина*

*Издание подготовлено в авторской редакции*

Подписано в печать 11.10.2016 г. Формат 70×108 <sup>1</sup>/<sub>16</sub>

Печ. л. 9; усл. печ. л. 12,6.

Тираж 30 экз. Заказ № 2092.

Отпечатано на оборудовании  
Издательского Дома  
Томского государственного университета  
634050, г. Томск, пр. Ленина, 36  
Тел. 8+(382-2) 53-15-28  
Сайт: <http://publish.tsu.ru>; E-mail: [rio.tsu@mail.ru](mailto:rio.tsu@mail.ru)

ISBN 978-5-94621-570-1

