

ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2015

№ 3 (32)

Свидетельство о регистрации: ПИ № ФС 77-29497 от 27 сентября 2007 г.



EDITORIAL BOARD

Alexander Gortsev – Editor-in-Chief, Doctor of Science, Prof., Head of the Operation Research Department, Dean of the Faculty of Applied Mathematics and Cybernetics.
Tel: +72822529599

Valery Smagin – Deputy Editor-in-Chief, Doctor of Science, Prof. of the Operation Research Department.
Tel: +72822529599

Lyudmila Nezhelskaya – Executive Editor, Cand. of Science, Associate Prof. of the Operation Research Department. E-mail: vestnik_uvti@mail.tsu.ru

Sergey Vorobeychikov – Doctor of Science, Prof. of the Higher Mathematics and Mathematical Modeling Department

Gennady Koshkin – Doctor of Science, Prof. of the Theoretical Cybernetics Department

Yury Kostyuk – Doctor of Science, Prof. of the Theoretical Informatics Department

Anjela Matrosova – Doctor of Science, Prof., Head of the Programming Department

Anatoly Nazarov – Doctor of Science, Prof., Head of the Probability Theory and Mathematical Statistics Department

Sergey Sushchenko – Doctor of Science, Prof., Head of the Applied of Information Department, Dean of the Faculty of Informatics

EDITORIAL COUNCIL

Ana Rosa Cavalli
PhD, Prof.
University VII
Paris, France

Vladimir Dombrovskii
Doctor of Science, Prof.
Tomsk State University
Russia

Alexander Dudin
Doctor of Science, Prof.
Belarusian State University
Minsk, Republic Belorussia

Enco Orzinger
PhD, Prof.
University of Rome
Italy

Paolo Prinetto
Prof.
Politechnic Institute
Torino, Italy

Gilbert Saporta
PhD, Prof.
Pierre and Marie Curie
University, Paris, France

Raimund Ubar
Doctor of Science, Prof.
University of Technology
Tallinn, Estonia

Nina Yevtushenko
Doctor of Science, Prof.
Tomsk State University
Russia

Yervant Zorian
PhD, Fellow & Chief Architect, Synopsys, Mountain View, CA, USA

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Горцев Александр Михайлович – гл. редактор, проф., д-р техн. наук, зав. каф. исследования операций, декан факультета прикладной математики и кибернетики ТГУ.
Тел. +72822529599

Смагин Валерий Иванович – зам. гл. редактора, проф., д-р техн. наук, проф. каф. исследования операций ТГУ.
Тел. +72822529599

Нежелская Людмила Алексеевна – отв. секретарь, доц., канд. техн. наук, доц. каф. исследования операций ТГУ.
E-mail: vestnik_uvti@mail.tsu.ru

Воробейчиков Сергей Эрикович – проф., д-р физ.-мат. наук, проф. каф. высшей математики и математического моделирования ТГУ

Косшкин Геннадий Михайлович – проф., д-р физ.-мат. наук, проф. каф. теоретической кибернетики ТГУ

Костюк Юрий Леонидович – проф., д-р техн. наук, проф. каф. теоретической информатики ТГУ

Матросова Анжела Юрьевна – проф., д-р техн. наук, зав. каф. программирования ТГУ

Назаров Анатолий Андреевич – проф., д-р техн. наук, зав. каф. теории вероятностей и математической статистики ТГУ

Сущенко Сергей Петрович – проф., д-р техн. наук, зав. каф. прикладной информатики, декан факультета информатики ТГУ

РЕДАКЦИОННЫЙ СОВЕТ

Ана Роза Ковали
д-р философии, проф.
Университет VII
Париж, Франция

Владимир Домбровский
д-р техн. наук, проф.
ТГУ, Томск, Россия

Александр Дудин
д-р физ.-мат. наук, проф.
БГУ, Минск,
Республика Беларусь

Енцо Орзингер
д-р философии, проф.
Римский университет
Италия

Паоло Принетто
проф.
Политехнический институт
Турин, Италия

Жильберт Сапорта
д-р философии, проф.
Университет им. Пьера и Марии
Кюри, Париж, Франция

Раймонд Убар
д-р техн. наук, проф.
Технологический университет,
Таллинн, Эстония

Нина Евтушенко
д-р техн. наук, проф.
ТГУ, Томск, Россия

Ервант Зорян
д-р философии,
гл. науч. сотр. фирмы
«Синописис», США

JORNAL INFO

Tomsk State University Journal of Control and Computer Science is an independent peer-reviewed research journal that welcomes submissions from across the world.

Tomsk State University Journal of Control and Computer Science is issued four times per year, and can be subscribed to in the Russian Press Joint Catalogue (Subscription Index 44031).

The publication in the journal is free of charge and may be in Russian or in English.

The topics of the journal are the following:

- *control of dynamical systems,*
- *mathematical modeling,*
- *data processing,*
- *informatics and programming,*
- *discrete function and automation,*
- *designing and diagnostics of computer systems.*

Rules of registration articles are given in a site:

<http://journals.tsu.ru/informatics/>
ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

О ЖУРНАЛЕ

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» выходит ежеквартально и распространяется по подписке, его подписной индекс 44031 в объединённом каталоге «Пресса России». Статьи публикуются на русском и английском языках.

Тематика публикаций журнала:

- *управление динамическими системами,*
- *математическое моделирование,*
- *обработка информации,*
- *информатика и программирование,*
- *дискретные функции и автоматы,*
- *проектирование и диагностика вычислительных систем.*

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» включен в Перечень изданий для публикации основных результатов кандидатских и докторских диссертаций, а также входит в систему Российского Индекса Научного Цитирования (РИНЦ). Правила оформления статей приведены на сайте:
<http://journals.tsu.ru/informatics/>

СОДЕРЖАНИЕ

ОБРАБОТКА ИНФОРМАЦИИ

Березин Д.В., Нежелская Л.А. Оптимальное оценивание состояний модулированного MAP-потока событий при его неполной наблюдаемости	4
Домбровский В.В., Обыедко Т.Ю., Самородова М.В. Управление с прогнозированием нелинейными стохастическими системами с марковскими скачками при ограничениях	14
Калягин А.А., Нежелская Л.А. Сравнение МП-и ММ-оценок длительности мертвого времени в обобщенном полусинхронном потоке событий	23
Kusainov M.I. Risk efficiency of adaptive one-step prediction of autoregression with parameter drift	33
Medvedev G.A. On the Nelson–Siegel–Svensson no-arbitrage yield curve models	44
Моисеев А.Н. Асимптотический анализ системы массового обслуживания MAP/GI/∞ с высокоинтенсивным входящим потоком	56
Сиренко М.А., Тарасенко П.Ф. Точные и асимптотические знаковые тесты для параметров нелинейной квантильной регрессии с Марковскими шумами	66
Тараник М.А., Копаница Г.Д. Анализ процесса контроля качества оказания медицинской помощи в рамках программы обязательного медицинского страхования	75

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

Головчинер М.Н., Юшин В.Ю. Автоматизированная система учета и контроля постановлений Ученого совета вуза	85
Змеев О.А., Политов А.М., Чайка Я.М. Концепция унифицированной модели сценария варианта использования для фиксации функциональных требований к программному продукту	91
Сведения об авторах	99

CONTENTS

DATA PROCESSING

Berezin D.V., Nezhel'skaya L.A. Modulated MAP states optimal estimation under conditions of its partial observability	4
Dombrovskii V.V., Obyedko T.Y., Samorodova M.V. Model Predictive Control for Nonlinear Stochastic Systems with Markovian Jumps under Constraints	14
Kalyagin A.A., Nezhel'skaya L.A. The comparison of maximum likelihood estimation and method of moments estimation of dead time value in a generalized semysynchronous flow of events	23
Kusainov M.I. Risk efficiency of adaptive one-step prediction of autoregression with parameter drift	33
Medvedev G.A. On the Nelson–Siegel–Svensson no-arbitrage yield curve models	44
Moiseev A.N. Asymptotic analysis of queueing system MAP/GI/∞ with high-rate arrivals	56
Sirenko M.A., Tarassenko P.F. Finite-sample and asymptotic sign-based tests for parameters of non-linear quantile regression with Markov noise	66
Taranik M.A., Kopanitsa G.D. Analysis of the process of control quality of medical service in the scope of compulsory health insurance program	75

INFORMATICS AND PROGRAMMING

Golovchiner M.N., Yushin V.Yu. Automated system of accounting and control of decisions of the Scientific Council of University	85
Zmeyev O.A., Politov A.M., Chayka Ya.M. Concept of unified use case model for software requirements specification	91
Brief information about the authors	99

ОБРАБОТКА ИНФОРМАЦИИ

УДК 519.21

DOI 10.17223/19988605/32/1

Д.В. Березин, Л.А. Нежелская

ОПТИМАЛЬНОЕ ОЦЕНИВАНИЕ СОСТОЯНИЙ МОДУЛИРОВАННОГО МАР-ПОТОКА СОБЫТИЙ ПРИ ЕГО НЕПОЛНОЙ НАБЛЮДАЕМОСТИ

Выведены явные выражения для апостериорных вероятностей состояний модулированного МАР-потока событий, учитывающие эффект непродлевающегося мертвого времени. Разработан алгоритм оптимального оценивания состояний модулированного МАР-потока событий. Проведены статистические эксперименты для установления качества оценивания, получены и проанализированы численные результаты.

Ключевые слова: модулированный МАР-поток событий; оптимальное оценивание состояний; метод максимума апостериорной вероятности; непродлевающееся мертвое время.

Интенсивное развитие компьютерной техники и информационных технологий послужило стимулом к созданию важной сферы приложений теории массового обслуживания – проектирования и создания информационно-вычислительных сетей, компьютерных сетей связи, спутниковых сетей, телекоммуникационных сетей и т.п. Интенсивность входящих потоков событий в системах и сетях массового обслуживания меняется со временем, как правило, случайно, что приводит к рассмотрению математических моделей дважды стохастических потоков событий. Дважды стохастические потоки можно разделить на два класса: к первому классу относятся потоки, интенсивность которых есть непрерывный случайный процесс [1]; ко второму классу относятся потоки с интенсивностью, представляющей собой кусочно-постоянный случайный процесс с конечным числом состояний [2–5]. Отметим, что МАР-потоки событий относятся ко второму классу дважды стохастических потоков и наиболее характерны для реальных телекоммуникационных сетей [6].

При исследовании потоков событий можно выделить два класса задач: 1) оценивание состояний потока событий [7–9]; 2) оценивание параметров потока [10–15].

В настоящей статье приведены аналитические и численные результаты оптимального оценивания состояний модулированного МАР-потока. Предлагается алгоритм оптимального оценивания состояний, когда решение о состоянии потока выносится по критерию максимума апостериорной вероятности, представляющей наиболее полную характеристику состояния потока, которую можно получить, располагая только выборкой наблюдений. Сам критерий минимизирует полную вероятность ошибки вынесения решения. Для получения численных результатов оценивания была построена имитационная модель потока, с помощью которой проведен ряд статистических экспериментов.

1. Постановка задачи

Рассматривается модулированный МАР-поток событий с интенсивностью, представляющей собой 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 -м состоянии, $i = 1, 2$, определяется двумя случайными величинами: первая случайная величина распределена по экспоненциальному закону $F_i^{(1)} = 1 - e^{-\alpha_i t}$, $i = 1, 2$; в момент окончания i -го состояния процесс $\lambda(t)$ переходит с вероятностью единица из i -го состояния в j -е, $i, j = 1, 2$ ($i \neq j$); вторая случайная величина распределена по экспоненциаль-

ному закону $F_i^{(2)} = 1 - e^{-\lambda_i t}$, $i = 1, 2$; в момент окончания i -го состояния процесс $\lambda(t)$ переходит с вероятностью $P_1(\lambda_j | \lambda_i)$ в j -е состояние ($i \neq j$) с наступлением события, либо с вероятностью $P_0(\lambda_j | \lambda_i)$ переходит в j -е состояние ($i \neq j$) без наступления события, либо с вероятностью $P_1(\lambda_i | \lambda_i)$ остается в i -м состоянии с наступлением события ($P_1(\lambda_j | \lambda_i) + P_0(\lambda_j | \lambda_i) + P_1(\lambda_i | \lambda_i) = 1$, $i, j = 1, 2$, $i \neq j$).

Первая и вторая случайные величины являются независимыми друг от друга. В сделанных предположениях $\lambda(t)$ – марковский процесс.

Блочная матрица инфинитезимальных характеристик процесса $\lambda(t)$ при этом примет вид

$$D = \begin{bmatrix} -(\alpha_1 + \lambda_1) & \alpha_1 + \lambda_1 P_0(\lambda_2 | \lambda_1) & \lambda_1 P_1(\lambda_1 | \lambda_1) & \lambda_1 P_1(\lambda_2 | \lambda_1) \\ \alpha_2 + \lambda_2 P_0(\lambda_1 | \lambda_2) & -(\alpha_2 + \lambda_2) & \lambda_2 P_1(\lambda_1 | \lambda_2) & \lambda_2 P_1(\lambda_2 | \lambda_2) \end{bmatrix} = \|D_0 | D_1\|.$$

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. Отметим, что если $P_0(\lambda_2 | \lambda_1) = P_0(\lambda_1 | \lambda_2) = 0$, то имеет место модулированный синхронный поток событий [16].

После каждого зарегистрированного в момент времени t_k события наступает время фиксированной длительности T_{dead} (мертвое время), в течение которого другие события исходного модулированного МАР-потока недоступны наблюдению. По окончании мертвого времени первое наступившее событие снова создает период мертвого времени длительности T_{dead} и т.д. Пример возникающей ситуации приведен на рис. 1, где 1 и 2 – состояния процесса $\lambda(t)$, $t_1, t_2, \dots$ – моменты наступления событий в наблюдаемом потоке; пунктиром обозначены длительности мертвого времени; черными кружками обозначены события модулированного МАР-потока, недоступные наблюдению.

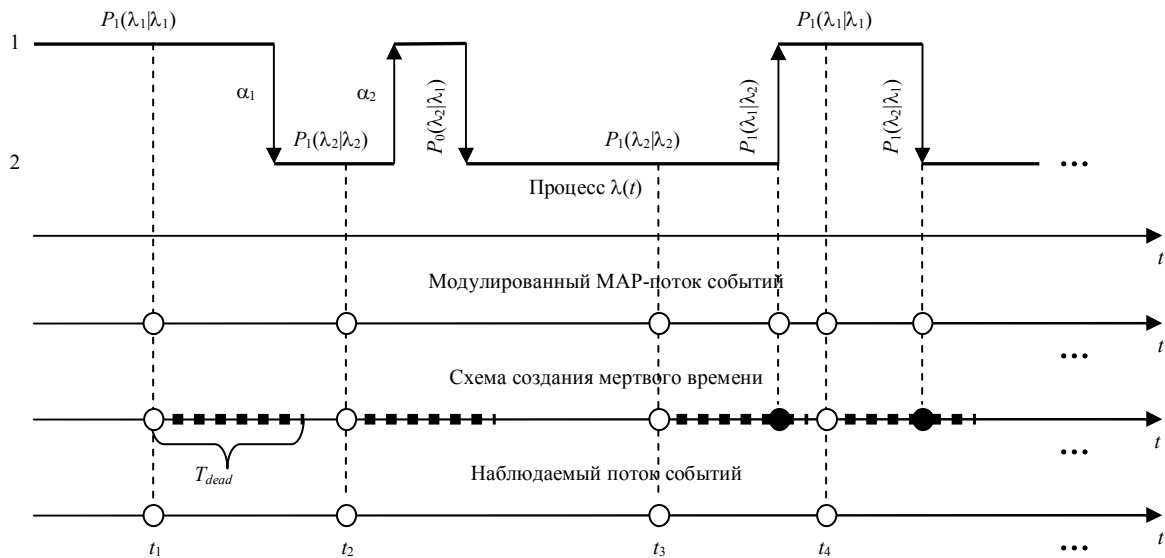


Рис. 1. Формирование наблюдаемого потока событий

Так как процесс $\lambda(t)$ принципиально ненаблюдаем (скрытый марковский процесс), а наблюдаемыми являются только временные моменты наступления событий потока $t_1, t_2, \dots$, то необходимо по этим наблюдениям оценить состояние процесса $\lambda(t)$ (или модулированного МАР-потока) в момент окончания наблюдения.

Рассматривается стационарный режим функционирования потока событий, поэтому переходными процессами на интервале наблюдения (t_0, t) , где t_0 – момент начала наблюдений, t – момент вынесения решения о состоянии процесса $\lambda(t)$, пренебрегаем. Тогда без потери общности можно положить $t_0 = 0$. Для вынесения решения о состоянии процесса $\lambda(t)$ в момент времени t необходимо определить апосте-

приорные вероятности $w(\lambda_i | t) = w(\lambda_i | t_1, \dots, t_m, t)$, $i = 1, 2$, того, что в момент времени t значение процесса $\lambda(t) = \lambda_i$ (m – количество наблюдаемых событий за время t), при этом $w(\lambda_1 | t) + w(\lambda_2 | t) = 1$. Решение о состоянии процесса $\lambda(t)$ выносится путем сравнения вероятностей: если $w(\lambda_i | t) \geq w(\lambda_j | t)$, $i, j = 1, 2$, $i \neq j$, то оценка состояния $\hat{\lambda}(t) = \lambda_i$, иначе $\hat{\lambda}(t) = \lambda_j$.

2. Алгоритм оптимального оценивания состояний модулированного МАР-потока событий

Рассмотрим интервал (t_k, t_{k+1}) , $k = 1, 2, \dots$, между соседними событиями рассматриваемого потока. Момент вынесения решения t будет принадлежать этому интервалу. При этом для начального интервала (t_0, t_1) момент t будет лежать между моментом начала наблюдения t_0 и моментом наступления первого события потока. Значение длительности интервала (t_k, t_{k+1}) есть $\tau_k = t_{k+1} - t_k$, $k = 1, 2, \dots$. С другой стороны, так как наблюдаемое в момент t_k событие порождает период мертвого времени длительности T , то $\tau_k = T + \eta_k$, где η_k – значение длительности интервала между моментом окончания периода мертвого времени $t_k + T_{\text{dead}}$ и моментом t_{k+1} , т.е. интервал (t_k, t_{k+1}) разбивается на два смежных: первый полуинтервал $(t_k, t_k + T_{\text{dead}}]$, второй – интервал $(t_k + T_{\text{dead}}, t_{k+1})$. Подчеркнем, что условия нахождения апостериорной вероятности $w(\lambda_i | t)$ на полуинтервале $(t_k, t_k + T_{\text{dead}}]$ и интервале $(t_k + T_{\text{dead}}, t_{k+1})$ принципиально разные. Кроме того, для нахождения вероятности $w(\lambda_i | t)$ необходимо точно знать значение T_{dead} либо, по крайней мере, предварительно осуществить оценку T_{dead} . В противном случае отсутствие такой информации делает попытку строгого нахождения вероятности $w(\lambda_i | t)$ невозможной. Здесь предполагается, что значение T_{dead} известно точно.

2.1. Выражения для апостериорной вероятности при отсутствии мертвого времени

Рассмотрим ситуацию, когда $T = 0$, т.е. мертвое время отсутствует. Для вывода формул апостериорной вероятности $w(\lambda_i | t)$ воспользуемся Δt -методом.

Лемма 1. На временных интервалах (t_0, t_1) и (t_k, t_{k+1}) , $k = 1, 2, \dots$, апостериорная вероятность $w(\lambda_i | t)$ удовлетворяет дифференциальному уравнению

$$\begin{aligned} w'(\lambda_1 | t) = & [\lambda_1 - \lambda_2 - \lambda_1 P_0(\lambda_2 | \lambda_1) + \lambda_2 P_0(\lambda_1 | \lambda_2)] w^2(\lambda_1 | t) - \\ & - [\alpha_1 + \alpha_2 + \lambda_1 - \lambda_2 + 2\lambda_2 P_0(\lambda_1 | \lambda_2)] w(\lambda_1 | t) + \alpha_2 + \lambda_2 P_0(\lambda_1 | \lambda_2); \end{aligned} \quad (1)$$

$$t_0 < t < t_1, \quad t_k < t < t_{k+1}, \quad k = 1, 2, \dots$$

Лемма 2. Апостериорная вероятность $w(\lambda_i | t)$ в момент t_k , $k = 1, 2, \dots$, наступления события модулированного МАР-потока определяется формулой пересчета

$$w(\lambda_1 | t_k + 0) = \frac{\lambda_2 P_1(\lambda_1 | \lambda_2) + [\lambda_1 P_1(\lambda_1 | \lambda_1) - \lambda_2 P_1(\lambda_1 | \lambda_2)] w(\lambda_1 | t_k - 0)}{\lambda_2 [1 - P_0(\lambda_1 | \lambda_2)] + [\lambda_1 - \lambda_2 - \lambda_1 P_0(\lambda_2 | \lambda_1) + \lambda_2 P_0(\lambda_1 | \lambda_2)] w(\lambda_1 | t_k - 0)}. \quad (2)$$

Замечание. В точке t_k вероятность $w(\lambda_i | t)$ претерпевает разрыв (имеет место конечный скачок). Вероятность $w(\lambda_i | t_k + 0)$ зависит от значения $w(\lambda_i | t_k - 0)$, где $w(\lambda_i | t_k - 0)$ – значение вероятности $w(\lambda_i | t)$ в момент времени t_k , когда $w(\lambda_i | t)$, определяемая в (1), изменяется на интервале (t_{k-1}, t_k) , соседнем с интервалом (t_k, t_{k+1}) , $k = 2, 3, \dots$. Таким образом, в значении $w(\lambda_i | t_k + 0)$ «сосредоточена» вся предыстория наблюдений за модулированным МАР-поток, начиная от момента $t_0 = 0$ до момента t_k . В качестве начального значения $w(\lambda_i | t_0 + 0) = w(\lambda_i | t_0 = 0)$ на полуинтервале $[t_0, t_1)$ выбирается априорная финальная вероятность первого состояния процесса $\lambda(t)$ [7]:

$$\pi_1 = \frac{\alpha_2 + \lambda_2 [1 - P_1(\lambda_2 | \lambda_2)]}{\alpha_1 + \alpha_2 + \lambda_1 [1 - P_1(\lambda_1 | \lambda_1)] + \lambda_2 [1 - P_1(\lambda_2 | \lambda_2)]}, \quad (3)$$

которая является решением дифференциального уравнения

$$\pi_1'(t | t_0) = (-\alpha_1 - \alpha_2 - \lambda_1 [1 - P_1(\lambda_1 | \lambda_1)] - \lambda_2 [1 - P_1(\lambda_2 | \lambda_2)]) \pi_1(t | t_0) + \alpha_2 + \lambda_2 [1 - P_1(\lambda_2 | \lambda_2)]$$

при $t \rightarrow \infty$.

Леммы 1, 2 позволяют сформулировать следующую теорему.

Теорема. Поведение апостериорной вероятности $w(\lambda_1 | t)$ на временных интервалах (t_0, t_1) и (t_k, t_{k+1}) , $k = 1, 2, \dots$, определяется явными формулами

$$w(\lambda_1 | t) = \frac{w_1[w_2 - w(\lambda_1 | t_k + 0)] - w_2[w_1 - w(\lambda_1 | t_k + 0)] \cdot e^{-A(w_2 - w_1)(t - t_k)}}{w_2 - w(\lambda_1 | t_k + 0) - [w_1 - w(\lambda_1 | t_k + 0)] \cdot e^{-A(w_2 - w_1)(t - t_k)}}, \quad (4)$$

$$w_1 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}, \quad w_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A},$$

$$A = \lambda_1 - \lambda_2 - \lambda_1 P_0(\lambda_2 | \lambda_1) + \lambda_2 P_0(\lambda_1 | \lambda_2) \neq 0, \quad B = -\alpha_1 - \alpha_2 - \lambda_1 + \lambda_2 - 2\lambda_2 P_0(\lambda_1 | \lambda_2),$$

$$C = \alpha_2 + \lambda_2 P_0(\lambda_1 | \lambda_2);$$

для $A = 0$

$$w(\lambda_1 | t) = \frac{(C + B \cdot w(\lambda_1 | t_k)) \cdot e^{B(t - t_k)} - C}{B}, \quad (5)$$

$t_0 < t < t_1$, $t_k < t < t_{k+1}$, $k = 1, 2, \dots$; $w(\lambda_1 | t_k + 0)$ определена формулой (2).

2.2. Выражение для апостериорной вероятности в условиях непродлевающегося мертвого времени

Возвратимся к ситуации, когда длительность мертвого времени $T_{\text{dead}} \neq 0$ (рис. 1). Тогда вычисление вероятности $w(\lambda_1 | t)$ по формуле (4) справедливо на интервале $(t_k + T_{\text{dead}}, t_{k+1})$. При этом начальное условие для $w(\lambda_1 | t)$ привязывается к моменту времени $t_k + T_{\text{dead}}$, т.е. в формуле (4) нужно заменить $w(\lambda_1 | t_k + 0)$ на $w(\lambda_1 | t_k + T_{\text{dead}})$ и $t_k + T_{\text{dead}} \leq t < t_{k+1}$, $k = 1, 2, \dots$. Формула (2) остается без изменения, так как предназначена для вычисления $w(\lambda_1 | t)$ в момент t_k наступления события, которое порождает мертвое время. Рассмотрим полуинтервал $(t_k, t_k + T_{\text{dead}}]$, $k = 1, 2, \dots$. На этом полуинтервале событие имеет место в граничной точке t_k , на самом полуинтервале события отсутствуют.

Утверждение. Поведение апостериорной вероятности $w(\lambda_1 | t)$ на временных полуинтервалах $(t_k, t_k + T_{\text{dead}}]$, $k = 1, 2, \dots$, определяется явной формулой

$$w(\lambda_1 | t) = \pi_1 + [w(\lambda_1 | t_k + 0) - \pi_1] e^{-B(t - t_k)}, \quad (6)$$

$$B = -\alpha_1 - \alpha_2 - \lambda_1 + \lambda_2 - 2\lambda_2 P_0(\lambda_1 | \lambda_2),$$

$t_k < t \leq t_k + T_{\text{dead}}$, $k = 1, 2, \dots$; $w(\lambda_1 | t_k + 0)$ определена формулой (2); π_1 определена в (3).

Полученные формулы позволяют сформулировать алгоритм расчета вероятности $w(\lambda_1 | t)$ и алгоритм принятия решения о состоянии процесса $\lambda(t)$ в любой момент времени t (алгоритм оптимального оценивания состояний модулированного МАР-потока):

1) вычисляется априорная вероятность π_1 нахождения процесса в состоянии λ_1 в момент начала наблюдения t_0 по формуле (3);

2) на промежутке (t_0, t_1) по формуле (4) или (5) в зависимости от значения коэффициента A вычисляется апостериорная вероятность $w(\lambda_1 | t)$, где в качестве $w(\lambda_1 | t_0 + 0)$, t_0 – момент наступления события, используется значение π_1 ;

3) в момент наступления события t_k апостериорная вероятность $w(\lambda_1 | t_k + 0)$ рассчитывается по формуле (2). Здесь в качестве $w(\lambda_1 | t_k - 0)$ используется значение, вычисленное по формуле (4) или (5) в момент времени t_k , $k = 1, 2, \dots$;

4) на промежутке $(t_k, t_k + T_{\text{dead}}]$ апостериорная вероятность $w(\lambda_1 | t)$ вычисляется по формуле (6), где $w(\lambda_1 | t_k + 0)$ – значение, вычисленное на предыдущем шаге алгоритма по формуле (2);

5) на интервале $(t_k + T_{\text{dead}}, t_{k+1})$, $k = 1, 2, \dots$, значение $w(\lambda_1 | t)$ рассчитывается по формуле (4) или (5), где в качестве $w(\lambda_1 | t_k + 0)$ используется значение $w(\lambda_1 | t_k + T_{\text{dead}})$, вычисленное на шаге 4. Далее осуществляется переход на шаг 3. Шаги 3–5 повторяются до момента t окончания наблюдения за потоком.

Параллельно по ходу вычисления $w(\lambda_1 | t)$ в любой момент времени выносится решение о состоянии процесса $\lambda(t)$: если $w(\lambda_1 | t) \geq w(\lambda_2 | t)$, то оценка состояния $\hat{\lambda}(t) = \lambda_1$, иначе $\hat{\lambda}(t) = \lambda_2$.

3. Результаты численных расчетов

Для получения численных результатов была построена имитационная модель модулированного МАР-потока событий с двумя состояниями и проведен статистический эксперимент. На первом этапе работы программы строится реализация потока. На втором этапе на основании полученной выборки $t_1, t_2, \dots, t_n$ моментов наступления событий в наблюдаемом потоке происходит оценивание состояний потока методом максимума апостериорной вероятности. На рис. 2 изображена реализация потока для $\lambda_1 = 4$, $\lambda_2 = 1$, $T = 5$, $T_{\text{dead}} = 1$, $\alpha_1 = 0,1$, $\alpha_2 = 0,1$, $P_1(\lambda_1 | \lambda_1) = 0,6$, $P_1(\lambda_2 | \lambda_1) = 0,3$, $P_0(\lambda_2 | \lambda_1) = 0,1$, $P_1(\lambda_2 | \lambda_2) = 0,4$, $P_1(\lambda_1 | \lambda_2) = 0,4$, $P_0(\lambda_1 | \lambda_2) = 0,2$.

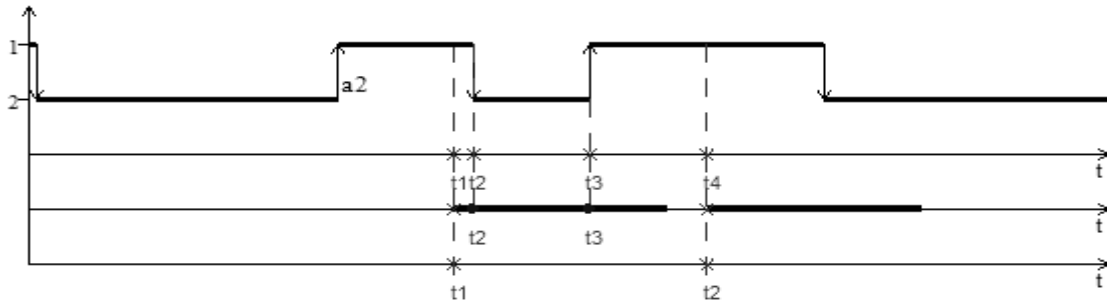


Рис. 2. Реализация модулированного МАР-потока событий

На рис. 3 приведено поведение апостериорной вероятности $w(\lambda_1 | t)$ для тех же значений параметров.

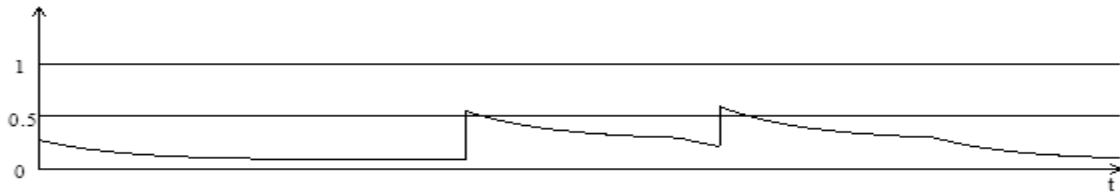


Рис. 3. Поведение апостериорной вероятности $w(\lambda_1 | t)$

На рис. 4 изображено поведение оценки $\hat{\lambda}(t)$ процесса $\lambda(t)$. Тонкая прямая линия проведена на тех временных участках, где поведение $\lambda(t)$ и $\hat{\lambda}(t)$ не совпадает.

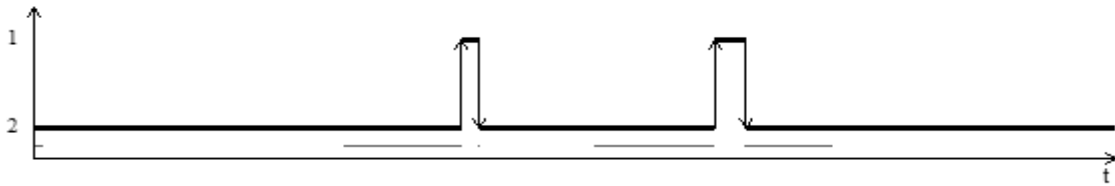


Рис. 4. Поведение $\hat{\lambda}(t)$

Чтобы установить частоту ошибочных решений о состоянии процесса $\lambda(t)$, были проведены статистические эксперименты, состоящие из следующих этапов:

- 1) для определенного набора параметров осуществляется моделирование потока (i -й опыт);
- 2) по формулам (2)–(6) рассчитывается апостериорная вероятность $w(\lambda_1 | t)$ и по методу максимума апостериорной вероятности выносится решение о значении процесса $\lambda(t)$ в любой момент времени t ;
- 3) определяется значение d_i – суммарная протяженность интервалов времени, на которых поведение $\lambda(t)$ и $\hat{\lambda}(t)$ не совпадает;

4) вычисляется доля ошибочных решений $\hat{p}_i = \frac{d_i}{T}$, где T – время моделирования;

5) шаги 1–4 повторяются N раз.

Результатом описанного алгоритма является выборка $\hat{p}_1, \dots, \hat{p}_N$ долей ошибочных решений в N испытаниях, на основании которой вычисляется выборочное среднее $\hat{\bar{p}} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i = \hat{P}_{\text{ош}}$ – оценка полной вероятности ошибки принятия решения $P_{\text{ош}}$, и выборочная дисперсия $S^2 = \frac{1}{N-1} \sum_{i=1}^N (\hat{p}_i - \hat{\bar{p}})^2$.

В первом эксперименте устанавливается зависимость оценки $\hat{P}_{\text{ош}}$ от значения параметра α_1 . Значения остальных параметров приведены в табл. 1. Результаты приведены в табл. 2.

Таблица 1

Данные для первого эксперимента

$T = 100$	$\lambda_1 = 10$	$\alpha_1 = 0; 0,1; \dots; 1$	$P_1(\lambda_1 \lambda_1) = 0,2$	$P_1(\lambda_2 \lambda_1) = 0,7$	$P_0(\lambda_2 \lambda_1) = 0,1$
$T_{\text{dead}} = 1$	$\lambda_2 = 1$	$\alpha_2 = 1$	$P_1(\lambda_2 \lambda_2) = 0,8$	$P_1(\lambda_1 \lambda_2) = 0,1$	$P_0(\lambda_1 \lambda_2) = 0,1$

Таблица 2

Результаты первого эксперимента

α_1	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$\hat{P}_{\text{ош}}$	0,025	0,037	0,049	0,059	0,071	0,079	0,090	0,094	0,104	0,112	0,121
$S^2 \times 10^{-4}$	0,63	0,72	0,81	0,93	0,89	1,37	1,26	1,51	1,83	1,92	2,04

Как видно, оценка $\hat{P}_{\text{ош}}$ возрастает по мере роста α_1 . Это объясняется тем, что с ростом α_1 переходы из первого состояния во второе происходят чаще, следовательно, событий в первом состоянии процесса $\lambda(t)$ наступает меньше (в вероятностном смысле) и, как следствие, событий в наблюдаемом потоке наступает меньше. Тем самым уменьшается количество информации, так как алгоритм в качестве исходных данных располагает только выборкой моментов наступления событий. Таким образом, частота принятия ошибочных решений о состоянии потока увеличивается.

Во втором эксперименте устанавливается зависимость оценки $\hat{P}_{\text{ош}}$ от значения параметра α_2 . Значения остальных параметров приведены в табл. 3. Результаты второго эксперимента продемонстрированы в табл. 4.

Таблица 3

Данные для второго эксперимента

$T = 100$	$\lambda_1 = 10$	$\alpha_1 = 1$	$P_1(\lambda_1 \lambda_1) = 0,2$	$P_1(\lambda_2 \lambda_1) = 0,7$	$P_0(\lambda_2 \lambda_1) = 0,1$
$T_{\text{dead}} = 1$	$\lambda_2 = 1$	$\alpha_2 = 0; 0,1; \dots; 1$	$P_1(\lambda_2 \lambda_2) = 0,8$	$P_1(\lambda_1 \lambda_2) = 0,1$	$P_0(\lambda_1 \lambda_2) = 0,1$

Таблица 4

Результаты второго эксперимента

α_2	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$\hat{P}_{\text{ош}}$	0,028	0,070	0,093	0,098	0,106	0,108	0,110	0,113	0,116	0,117	0,120
$S^2 \times 10^{-4}$	0,96	1,54	2,26	2,20	2,28	1,92	1,96	2,11	1,78	2,34	1,90

Как и в первом эксперименте, оценка $\hat{P}_{\text{ош}}$ возрастает с ростом α_2 . На рис. 5 изображено поведение оценки $\hat{P}_{\text{ош}}$ при изменении параметров α_1 и α_2 .

Заметим, что $\hat{P}_{\text{ош}}(\alpha_2) > \hat{P}_{\text{ош}}(\alpha_1)$ при равных значениях α_1 и α_2 . Это объясняется набором вероятностей, определяющих поток, в частности значениями $P_1(\lambda_2 | \lambda_1) = 0,7$ и $P_1(\lambda_2 | \lambda_2) = 0,8$. Последние заданы таким образом, что процесс $\lambda(t)$ во втором состоянии проводит больше времени. Поскольку $\lambda_1 = 10$, а $\lambda_2 = 1$ ($\lambda_1 \gg \lambda_2$), то алгоритм с большой апостериорной вероятностью $w(\lambda_1 | t_k + 0)$ выносит решение о том, что событие наступило в первом состоянии процесса $\lambda(t)$, тем самым совершая ошибку при оценке состояния.

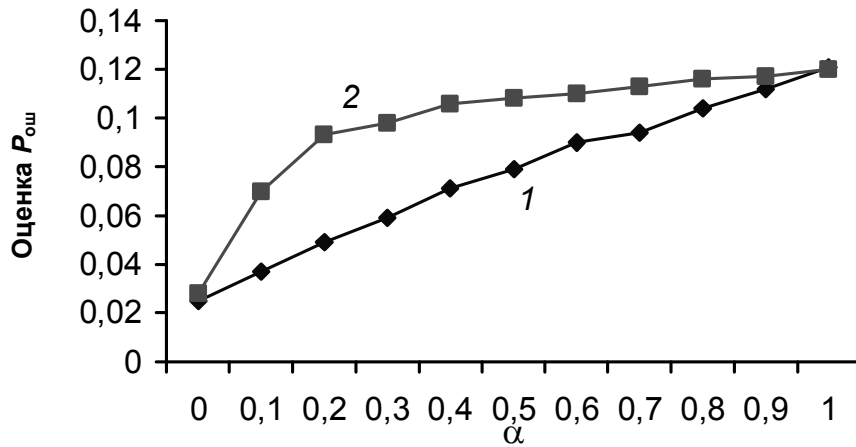


Рис. 5. Зависимость оценки $\hat{P}_{ош}$ от параметров α_i , $i = 1, 2$.

1 – поведение $\hat{P}_{ош}(\alpha_1)$; 2 – поведение $\hat{P}_{ош}(\alpha_2)$

В третьем эксперименте исследуется зависимость оценки $\hat{P}_{ош}$ от отношения λ_1/λ_2 . В табл. 5 приведены исходные данные эксперимента, результаты эксперимента продемонстрированы в табл. 6.

Таблица 5

Данные для третьего эксперимента

$T = 100$	$\lambda_1 = 10$	$\alpha_1 = 0,5$	$P_1(\lambda_1 \lambda_1) = 0,2$	$P_1(\lambda_2 \lambda_1) = 0,7$	$P_0(\lambda_2 \lambda_1) = 0,1$
$T_{dead} = 1$	$\lambda_1/\lambda_2 = 4; \dots; 200$	$\alpha_2 = 0,5$	$P_1(\lambda_2 \lambda_2) = 0,8$	$P_1(\lambda_1 \lambda_2) = 0,1$	$P_0(\lambda_1 \lambda_2) = 0,1$

Таблица 6

Результаты третьего эксперимента

λ_1/λ_2	4	5	10	20	50	100	200
$\hat{P}_{ош}$	0,108	0,097	0,076	0,068	0,061	0,057	0,057
$S^2 \times 10^{-4}$	1,94	1,51	1,27	1,73	1,66	1,47	1,16

Результаты показывают, что оценивание тем лучше, чем больше отношение λ_1/λ_2 . В этом случае состояния процесса $\lambda(t)$ становятся различимы лучше и частота принятия ошибочных решений значительно уменьшается.

В четвертом эксперименте устанавливается зависимость оценки $\hat{P}_{ош}$ от длительности мертвого времени T_{dead} . Данные для эксперимента приведены в табл. 7. Результаты продемонстрированы в табл. 8.

Таблица 7

Данные для четвертого эксперимента

$T = 100$	$\lambda_1 = 5$	$\alpha_1 = 0,2$	$P_1(\lambda_1 \lambda_1) = 0,6$	$P_1(\lambda_2 \lambda_1) = 0,3$	$P_0(\lambda_2 \lambda_1) = 0,1$
$T_{dead} = 0; 1; \dots; 6$	$\lambda_2 = 1$	$\alpha_2 = 0,2$	$P_1(\lambda_2 \lambda_2) = 0,1$	$P_1(\lambda_1 \lambda_2) = 0,7$	$P_0(\lambda_1 \lambda_2) = 0,2$

Таблица 8

Результаты четвертого эксперимента

T_{dead}	0	1	2	3	4	5	6
$\hat{P}_{ош}$	0,232	0,260	0,276	0,280	0,284	0,288	0,291
$S^2 \times 10^{-3}$	0,49	1,01	0,97	1,33	1,34	1,22	1,45

Результаты свидетельствуют о том, что оценка $\hat{P}_{\text{ош}}$ растет с увеличением значения T_{dead} , так как событий в наблюдаемом потоке наступает меньше, в результате чего уменьшается количество информации для алгоритма принятия решения о значении процесса $\lambda(t)$.

Заключение

В данной статье предложен алгоритм оптимального оценивания состояний модулированного МАР-потока при наличии непродлеваемого мертвого времени, приведены численные результаты ряда экспериментов по выявлению зависимости оценки $\hat{P}_{\text{ош}}$ от изменения различных параметров потока. Результаты экспериментов показывают, что оценка $\hat{P}_{\text{ош}}$ возрастает с увеличением параметров α_1 и α_2 , а с увеличением отношения λ_1 к λ_2 оценивание состояний процесса $\lambda(t)$ происходит значительно лучше.

ЛИТЕРАТУРА

1. Kingman J.F.C. On doubly stochastic Poisson process // Proc. of Cambridge Philosophical Society. 1964. V. 60, No. 4. P. 923–930.
2. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи // Изв. АН СССР. Техн. кибернетика. 1979. № 6. С. 92–99.
3. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи // Изв. АН СССР. Техн. кибернетика. 1980. № 1. С. 55–61.
4. Neuts M.F. A versatile Markov point process // Journal of Applied Probability. 1979. No. 16. P. 764–779.
5. Lucantoni D.M. New results on the single server queue with a batch Markovian arrival process // Communications in Statistics Stochastic Models. 1991. No. 7. P. 1–46.
6. Дудин А.Н., Клименок В.Н. Системы массового обслуживания с коррелированными потоками. Минск : Изд-во БГУ, 2000. 175 с.
7. Nezhe'skaya L.A. Optimal state estimation in modulated MAP event flows with unextendable dead time // Communications in Computers and Information Sciences: proceedings of the 13th International Scientific Conference ITMM 2014 named after A.F. Terpugov «Information Technologies and Mathematical Modeling» (November 20–22, 2014). Cham Heidelberg ; New York ; Dordrecht ; London : Springer, 2014. P. 342–350.
8. Gortsev A.M., Nezhe'skaya L.A., Solov'ev A.A. Optimal state estimation in MAP event flows with unextendable dead time // Automation and Remote Control. 2012. No. 8. P. 1316–1326.
9. Gortsev A.M., Nezhe'skaya L.A., Shevchenko T.I. Estimation of the states of an MC-stream of events in the presence of measurement errors // Russian Physics Journal. 1993. V. 36, No. 12. P. 1153–1167.
10. Gortsev A.M., Nezhe'skaya L.A. An asynchronous double stochastic flow with initiation of superfluous events // Discrete Mathematics and Applications. 2011. V. 21, No. 3. P. 283–290.
11. Bushlanov I.V., Gortsev A.M., Nezhe'skaya L.A. Estimating parameters of the synchronous twofold stochastic flow of events // Automation and Remote Control. 2008. V. 69, No. 9. P. 1517–1533.
12. Gortsev A.M., Nezhe'skaya L.A. Estimation of the dead time period and intensities of the synchronous double stochastic event flow // Радиотехника. 2004. No. 10. P. 8–16.
13. Gortsev A.M., Nezhe'skaya L.A. Estimation of the dead time period and parameters of a semi-synchronous double stochastic stream of events // Measurement Techniques. 2003. V. 46, No. 6. P. 536–545.
14. Gortsev A.M., Nezhe'skaya L.A. Estimation of parameters of synchronously alternating Poisson stream of events by the moment method // Telecommunications and Radio Engineering. 1996. V. 50, No. 1. P. 56–63.
15. Gortsev A.M., Nezhe'skaya L.A. Estimation of parameters of synchro-alternating Poisson event flow by the method of moments // Радиотехника. 1995. V. 40, No. 7–8. P. 6–10.
16. Gortsev A.M., Sirotina M. Joint probability density function of modulated synchronous flow interval duration // Communications in Computers and Information Sciences: proceedings of the 13th International Scientific Conference ITMM 2014 named after A.F. Terpugov «Information Technologies and Mathematical Modeling» (November 20–22, 2014). Cham Heidelberg ; New York ; Dordrecht ; London : Springer, 2014. P. 145–152.

Березин Дмитрий Владимирович. E-mail: berezin14@mail.ru

Нежелская Людмила Алексеевна, канд. техн. наук, доцент. E-mail: ludne@mail.tsu.ru
Томский государственный университет

Поступила в редакцию 14 мая 2015 г.

Berezin Dmitriy V., Nezhe'skaya Lyudmila A. (Tomsk State University, Russian Federation).

Modulated MAP states optimal estimation under conditions of its partial observability.

Keywords: Modulated MAP; optimal state estimation; method of a posteriori probability maximum; non-extendable dead time.

DOI 10.17223/19988605/32/1

This paper considers the modulated MAP, which rate is a piecewise constant random process $\lambda(t)$ with two states: $\lambda(t) = \lambda_1$ and $\lambda(t) = \lambda_2$ ($\lambda_1 > \lambda_2 \geq 0$). The time when the process $\lambda(t)$ remains at the i -th state, $i = 1, 2$, depends on two random values: 1) the first random value has the exponential distribution function $F_i^{(1)} = 1 - e^{-\alpha_i t}$, $i = 1, 2$; when the i -th state ends, the process $\lambda(t)$ transits with the probability equal to one from the i -th state to the j -th state, $i, j = 1, 2$ ($i \neq j$); 2) the second random value has the exponential distribution function $F_i^{(2)} = 1 - e^{-\lambda_i t}$, $i = 1, 2$; when the i -th state ends, the process $\lambda(t)$ transits with the probability $P_1(\lambda_j | \lambda_i)$ from the i -th state to the j -th state ($i \neq j$) and a flow event occurs or $\lambda(t)$ transits with the probability $P_0(\lambda_j | \lambda_i)$ from the i -th state to the j -th state ($i \neq j$) but an event does not occur, or the process $\lambda(t)$ transits with the probability $P_1(\lambda_i | \lambda_i)$ from the i -th state to the i -th state and a flow event occurs. Here $P_1(\lambda_j | \lambda_i) + P_0(\lambda_j | \lambda_i) + P_1(\lambda_i | \lambda_i) = 1$, $i, j = 1, 2$, $i \neq j$. The block transition rate matrix for the process $\lambda(t)$ is as follows:

$$D = \begin{bmatrix} -(\alpha_1 + \lambda_1) & \alpha_1 + \lambda_1 P_0(\lambda_2 | \lambda_1) & \lambda_1 P_1(\lambda_1 | \lambda_1) & \lambda_1 P_1(\lambda_2 | \lambda_1) \\ \alpha_2 + \lambda_2 P_0(\lambda_1 | \lambda_2) & -(\alpha_2 + \lambda_2) & \lambda_2 P_1(\lambda_1 | \lambda_2) & \lambda_2 P_1(\lambda_2 | \lambda_2) \end{bmatrix} = \|D_0 | D_1\|.$$

An event generates the period of time called the dead time, during which the flow cannot be observed. After this period ends, a new event also generates the dead time. Having only a sample $(t_1, \dots, t_n)$ of events moments, we need to estimate the flow states.

The optimal states estimation algorithm is as following:

- 1) at the initial moment t_0 a priori probability π_1 that $\lambda(t)$ is in λ_1 is calculated using the formula

$$\pi_1 = \frac{\alpha_2 + \lambda_2 [1 - P_1(\lambda_2 | \lambda_2)]}{\alpha_1 + \alpha_2 + \lambda_1 [1 - P_1(\lambda_1 | \lambda_1)] + \lambda_2 [1 - P_1(\lambda_2 | \lambda_2)]};$$

- 2) in the interval (t_0, t_1) a posteriori probability $w(\lambda_1 | t)$ is calculated using the formula

$$w(\lambda_1 | t) = \frac{w_1 [w_2 - w(\lambda_1 | t_0 + 0)] - w_2 [w_1 - w(\lambda_1 | t_0 + 0)] e^{-A(w_2 - w_1)(t - t_0)}}{w_2 - w(\lambda_1 | t_0 + 0) - [w_1 - w(\lambda_1 | t_0 + 0)] e^{-A(w_2 - w_1)(t - t_0)}}, \quad w(\lambda_1 | t_0 + 0) = \pi_1;$$

- 3) at event occurring moment t_k a posteriori probability $w(\lambda_1 | t_k + 0)$ is calculated using the formula

$$w(\lambda_1 | t_k + 0) = \frac{\lambda_2 P_1(\lambda_1 | \lambda_2) + [\lambda_1 P_1(\lambda_1 | \lambda_1) - \lambda_2 P_1(\lambda_1 | \lambda_2)] w(\lambda_1 | t_k - 0)}{\lambda_2 [1 - P_0(\lambda_1 | \lambda_2)] + [\lambda_1 - \lambda_2 - \lambda_1 P_0(\lambda_2 | \lambda_1) + \lambda_2 P_0(\lambda_1 | \lambda_2)] w(\lambda_1 | t_k - 0)},$$

where instead of $w(\lambda_1 | t_k - 0)$ the value calculated on the formula

$$w(\lambda_1 | t) = \frac{w_1 [w_2 - w(\lambda_1 | t_k + 0)] - w_2 [w_1 - w(\lambda_1 | t_k + 0)] \cdot e^{-A(w_2 - w_1)(t - t_k)}}{w_2 - w(\lambda_1 | t_k + 0) - [w_1 - w(\lambda_1 | t_k + 0)] \cdot e^{-A(w_2 - w_1)(t - t_k)}}$$

at t_k , $k = 1, 2, \dots$, is used;

- 4) in the interval $(t_k, t_k + T_{\text{dead}}]$ the probability $w(\lambda_1 | t)$ can be found as

$$w(\lambda_1 | t) = \pi_1 + [w(\lambda_1 | t_k + 0) - \pi_1] \cdot e^{-B(t - t_k)};$$

5) in the interval $(t_k + T_{\text{dead}}, t_{k+1})$, $k = 1, 2, \dots$, the value $w(\lambda_1 | t)$ can be calculated by formula used on the step 3, where instead of $w(\lambda_1 | t_k + 0)$ the value $w(\lambda_1 | t_k + T_{\text{dead}})$ calculated on the step 4 is used. Then go to the step 3. The steps 3-5 are reiterated during observing time.

Simultaneously with probability calculation we estimate the flow states: if $w(\lambda_1 | t) \geq w(\lambda_2 | t)$ then $\hat{\lambda}(t) = \lambda_1$, otherwise $\hat{\lambda}(t) = \lambda_2$.

The results show that the bigger α_i , $i=1, 2$, the higher estimation error, and that the bigger quotient λ_1/λ_2 , the better estimation.

REFERENCES

1. Kingman, J.F.C. (1964) On doubly stochastic Poisson process. *Proc. of Cambridge Philosophical Society*. 60 (4). pp. 923-930. DOI: 10.1017/S030500041000383X
2. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1979) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi [On the equivalent substitutions method for computing fragments of communication networks]. *Izv. Akad. Nauk USSR. Tekhn. Kibernet.* 6, pp. 92-99.
3. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1980) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi [On the equivalent substitutions method for computing fragments of communication networks]. *Izv. Akad. Nauk USSR. Tekhn. Kibernet.* 1, pp. 55-61.
4. Neuts, M.F. (1979) A versatile Markovian point process. *Journal of Applied Probability*. 16(4). pp. 764-779. DOI: 10.2307/3213143
5. Lucantoni, D.M. (1991) New results on the single server queue with a batch Markovian arrival process. *Communications in Statistics Stochastic Models*. 7. pp. 1-46.
6. Dudin, A.N. & Klimenok, V.I. (2000) *Queueing systems with correlated flows*. Minsk: BGU.
7. Nezhel'skaya, L.A. (2014) Optimal state estimation in modulated MAP event flows with unextendable dead time. *Communications in Computers and Information Sciences: proceedings of the 13th International Scientific Conference ITMM 2014 named after A.F. Terpigov "Information Technologies and Mathematical Modeling"*. 20th-22nd November. Cham Heidelberg, New York, Dordrecht, London: Springer. pp. 342-350.
8. Gortsev, A.M., Nezhel'skaya, L.A. & Solovyev, A.A. (2012) Optimal state estimation in MAP event flows with unextendable dead time. *Automation and Remote Control*. 73 (8). pp. 1316-1326. DOI: 10.1134/S000511791208005X

9. Gortsev, A.M., Nezhel'skaya, L.A. & Shevchenko, T.I. (1993) Estimation of the states of an MC-stream of events in the presence of measurement errors. *Russian Physics Journal*. 36 (12). pp. 1153-1167.
10. Gortsev, A.M. & Nezhel'skaya, L.A. (2011) An asynchronous double stochastic flow with initiation of superfluous events. *Discrete Mathematics and Applications*. 21 (3). pp. 283-290. DOI: 10.1515/dma.2011.017
11. Bushlanov, I.V., Gortsev, A.M. & Nezhel'skaya, L.A. (2008) Estimating parameters of the synchronous twofold stochastic flow of events. *Automation and Remote Control*. 69 (9). pp. 1517-1533. DOI: 10.1134/S0005117908090075
12. Gortsev, A.M. & Nezhel'skaya, L.A. (2014) Estimation of the dead time period and intensities of the synchronous double stochastic event flow. *Radiotekhnika*. 10. pp. 8-16.
13. Gortsev, A.M. & Nezhel'skaya, L.A. (2003) Estimation of the dead time period and parameters of a semi-synchronous double stochastic stream of events. *Measurement Techniques*. 46 (6). pp. 536-545. DOI: 10.1023/A:1025499509015
14. Gortsev, A.M. & Nezhel'skaya, L.A. (1996) Estimation of parameters of synchronously alternating Poisson stream of events by the moment method. *Telecommunications and Radio Engineering*. 50 (1). pp. 56-63.
15. Gortsev, A.M. & Nezhel'skaya, L.A. (1995) Estimation of parameters of synchro-alternating Poisson event flow by the method of moments. *Radiotekhnika*. 40 (7-8). pp. 6-10.
16. Gortsev A.M. & Sirotina, M. (2014) Joint probability density function of modulated synchronous flow interval duration. *Communications in Computers and Information Sciences: proceedings of the 13th International Scientific Conference ITMM 2014 named after A.F. Terpugov "Information Technologies and Mathematical Modeling"*. 20th to 22nd November, 2014. Cham Heidelberg, New York, Dordrecht, London: Springer. pp. 145-152.

В.В. Домбровский, Т.Ю. Объедко, М.В. Самородова

УПРАВЛЕНИЕ С ПРОГНОЗИРОВАНИЕМ НЕЛИНЕЙНЫМИ СТОХАСТИЧЕСКИМИ СИСТЕМАМИ С МАРКОВСКИМИ СКАЧКАМИ ПРИ ОГРАНИЧЕНИЯХ

Рассматривается задача управления с прогнозированием по квадратичному критерию для нелинейных дискретных систем со скачкообразно меняющимися параметрами. Синтезированы стратегии управления при наличии явных ограничений на управляющие воздействия. Алгоритм синтеза прогнозирующей стратегии сводится к решению последовательности задач квадратичного программирования.

Ключевые слова: нелинейные стохастические системы; прогнозирующее управление; марковские скачки; ограничения.

Моделями с марковскими скачкообразными параметрами описывается широкий класс реальных систем [1]. В этих моделях предполагается, что смена структуры системы осуществляется в соответствии с эволюцией марковской цепи с конечным пространством состояний. Решению различных задач управления и оценивания для таких систем посвящено значительное количество работ [2–10].

Эффективным подходом к синтезу систем управления с ограничениями, получившим широкое признание и применение в практике управления сложными технологическими процессами, является метод управления с прогнозирующей моделью (управление со скользящим горизонтом) [11]. Применению данного метода к управлению дискретными системами с марковскими скачками посвящены работы [3, 5]. В работе [3] рассматривается задача управления по квадратичному критерию линейными дискретными системами при «жестких» ограничениях на управляющие переменные.

Применению метода управления с прогнозирующей моделью к управлению нелинейными стохастическими системами посвящены работы [12–16]. В работах [12, 13] рассматривается задача управления на скользящем горизонте для нелинейных стохастических систем без учета ограничений. В работе [14] рассматривается задача прогнозирующего управления нелинейными системами, возмущенными белыми шумами, при наличии явных ограничений на управляющие переменные.

В настоящей работе рассматривается задача синтеза стратегий управления с прогнозированием для дискретных нелинейных систем с марковскими скачками. Получены уравнения синтеза оптимальных стратегий управления с учетом «жестких» ограничений на управляющие переменные.

1. Постановка задачи

Пусть объект управления описывается уравнением

$$x(k+1) = Ax(k) + B[\alpha(k+1), k+1]u(k) + f(x(k), u(k), w(k), \alpha(k+1)), \quad (1)$$

где $x(k)$ – n -мерный вектор состояния, $u(k)$ – n_u -мерный вектор управления, $w(k)$ – вектор белых шумов размерности n_w с нулевым средним и единичной матрицей ковариации, $\alpha(k)$ ($k = 0, 1, 2, \dots, v$) – однородная дискретная марковская цепь с конечным множеством состояний $\{1, 2, \dots, v\}$, известной матрицей переходных вероятностей

$$P = [P_{i,j}] \quad (i, j \in \{1, 2, \dots, v\}), P_{j,i} = P\{\alpha(k+1) = \alpha_j | \alpha(k) = \alpha_i\}, \sum_{j=1}^v P_{j,i} = 1,$$

и известным начальным распределением

$$p_i = P\{\alpha(0) = i\} \quad (i = 1, 2, \dots, v), \sum_{i=1}^v p_i = 1.$$

Предполагается, что состояние марковской цепи в момент времени k доступно наблюдению. Последовательности $w(k)$ и $\alpha(k)$ независимы. Характер нелинейной зависимости в функции f таков, что

$$M\{f(x(k), u(k), w(k), \alpha(k+1))/x(k), \alpha(k)=\alpha_j\} = 0 \quad (2)$$

для любых $x(k)$ и

$$M\{f(x(k), u(k), w(k), \alpha(k+1))f^T(x(k), u(k), w(k), \alpha(k+1))/x(k), \alpha(k)=\alpha_j\} = T^0[\alpha(k), k] + \sum_{i=1}^r T^i \left(x^T(k) W^i x(k) + u^T(k) M^i [\alpha(k), k] u(k) \right), \quad (3)$$

где $r = n(n+1)/2$, T^i , W^i и $M^i = \left(N^i \right)^T N^i$ ($i = \overline{1, r}$), $T^0 = \left(D^0 \right)^T D^0$ – неотрицательно определенные симметричные матрицы.

На управляющие воздействия наложены ограничения вида

$$u_{\min}(k) \leq S(k)u(k) \leq u_{\max}(k), \quad (4)$$

где $S(k)$ – матрица соответствующей размерности.

Необходимо определить закон управления системой (1) при ограничениях (4) из условия минимума критерия со скользящим горизонтом управления:

$$J(k+p/k) = M\left\{ \sum_{i=1}^p x^T(k+i) R_1(k, i) x(k+i) + u^T(k+i-1/k) R_2(k, i-1) u(k+i-1/k) / x(k), \alpha(k)=\alpha_j \right\}, \quad (5)$$

где $M\{\dots/\dots\}$ – оператор условного математического ожидания; p – горизонт прогноза; k – текущий момент времени; $R_1(k, i) \geq 0$, $R_2(k, i) \geq 0$ – весовые матрицы соответствующих размерностей.

2. Синтез стратегий прогнозирующего управления

Для решения сформулированной задачи используем методологию управления с прогнозирующей моделью. Данный подход позволяет получить стратегии управления с обратной связью с учетом явных ограничений на управляющие воздействия.

Стратегии управления с прогнозированием определяются по следующему правилу. На каждом шаге k минимизируем функционал (3) по последовательности прогнозирующих управлений $u(k/k), \dots, u(k+p-1/k)$, зависящих от состояния системы в момент времени k . В качестве управления в момент времени k берем $u(k) = u(k/k)$. Тем самым получаем управление $u(k)$ как функцию состояний $x(k)$ и $\alpha(k) = \alpha_j$, т.е. управление с обратной связью. Чтобы получить управление $u(k+1)$ на следующем шаге, процедура повторяется для следующего момента $k+1$ и т.д.

Цепь Маркова с дискретным временем допускает следующее представление в пространстве состояний [9]:

$$\theta(k+1) = P\theta(k) + v(k+1), \quad (6)$$

где $\theta(k) = [\delta(\alpha(k), 1), \dots, \delta(\alpha(k), v)]^T$, $\delta(\alpha(k), j)$ – функция Кронекера ($j = \overline{1, v}$); $v(k+1)$ – мартингал-разность с характеристиками

$$M\{v(k+1)/\theta(k)\} = 0, \quad (7)$$

$$C(k+1) = M\{v(k+1)v^T(k+1)/\theta(k)\} = \text{diag}\{P\theta(k)\} - P\text{diag}\{\theta(k)\}P^T. \quad (8)$$

С учетом (6) систему (1) можно представить в следующем виде:

$$x(k+1) = Ax(k) + B[\theta(k+1), k+1]u(k) + f(x(k), u(k), w(k), \theta(k+1)), \quad (9)$$

где

$$B[\theta(k), k] = \sum_{i=1}^v \theta_i(k) B^i(k), \quad (10)$$

здесь $\theta_i(k)$ ($i = 1, 2, \dots, v$) – компоненты вектора $\theta(k)$, $\{B^i(k)\}$ – множество значений матрицы $B[\theta(k), k]$.

Выражения (2) и (3) можно представить в виде

$$\begin{aligned}
M \{ f(x(k), u(k), w(k), \theta(k+1)) / x(k), \theta(k) \} &= 0, \\
M \{ f(x(k), u(k), w(k), \theta(k+1)) f^T(x(k), u(k), w(k), \theta(k+1)) / x(k), \theta(k) \} &= T^0[\theta(k), k] + \\
&+ \sum_{i=1}^r T^i \left(x^T(k) W^i x(k) + u^T(k) M^i[\theta(k), k] u(k) \right),
\end{aligned}$$

где, с учетом (6)–(8),

$$\begin{aligned}
T^0[\theta(k), k] &= \sum_{s=1}^v \sum_{n=1}^v \left(D_s^0(k) \right)^T [\theta^T(k) P^T E_s^T] [E_n P \theta(k)] D_n^0(k) + \sum_{s=1}^v \sum_{n=1}^v \left(D_s^0(k) \right)^T [E_n P C(k-1) P^T E_s^T] D_n^0(k), \\
M^j[\theta(k), k] &= \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k) \right)^T [\theta^T(k) P^T E_s^T] [E_n P \theta(k)] N_n^j(k) + \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k) \right)^T [E_n P C(k-1) P^T E_s^T] N_n^j(k),
\end{aligned}$$

здесь $\{D_i^0(k)\}, \{N_i^j(k)\}$ ($j=\overline{1, r}, i=\overline{1, v}$) – множество значений матриц $D^0(k)$ и $N^j(k)$ соответственно;

$E_i = [0, \dots, 0, 1, 0, \dots, 0]$.

Пусть $H(k)$ и $G(k)$ – блочные матрицы вида

$$H(k) = \begin{bmatrix} H_{00}(k) & H_{01}(k) & \dots & H_{0,p-1}(k) \\ H_{10}(k) & H_{11}(k) & \dots & H_{1,p-1}(k) \\ \dots & \dots & \dots & \dots \\ H_{p-1,0}(k) & H_{p-1,1}(k) & \dots & H_{p-1,p-1}(k) \end{bmatrix}, \quad (11)$$

$$G(k) = [G_0(k) \quad G_1(k) \quad \dots \quad G_{p-1}(k)], \quad (12)$$

блоки которых равны

$$H_{t,t}(k) = R_2(k, t) + \sum_{s=1}^v \sum_{n=1}^v \left[\left(B^{(s)}(k+t+1) \right)^T [\theta^T(k) (P^{t+1})^T E_s^T] Q(p-t-1) [E_n P^{t+1} \theta(k)] B^{(n)}(k+t+1) + \right. \quad (13)$$

$$+ \sum_{l=p-t}^p \left(B^{(s)}(k+t+1) \right)^T [E_n P^{l-p+t} C(k+p-l+1) (P^{l-p+t})^T E_s^T] Q(p-t-1) B^{(n)}(k+t+1) +$$

$$+ \sum_{j=1}^r \left(N_s^j(k+t) \right)^T \text{tr} \left\{ [\theta^T(k) (P^{t+1})^T E_s^T] Q(p-t-1) [E_n P^{t+1} \theta(k)] T^j \right\} N_n^j(k+t) +$$

$$+ \sum_{l=p-t}^p \sum_{j=1}^r \left(N_s^j(k+t) \right)^T \text{tr} \left\{ [E_n P^{l-p+t} C(k+p-l+1) (P^{l-p+t})^T E_s^T] Q(p-t-1) T^j \right\} N_n^j(k+t) \Big],$$

$$H_{t,f}(k) = \sum_{s=1}^v \sum_{n=1}^v \left[\left(B^{(s)}(k+t+1) \right)^T [\theta^T(k) (P^{t+1})^T E_s^T] Q(p-t-1) A^{t-f} [E_n P^{f+1} \theta(k)] B^{(n)}(k+f+1) + \right. \quad (14)$$

$$+ \sum_{l=p-f}^p \left(B^{(s)}(k+t+1) \right)^T [E_n P^{l-p+f} C(k+p-l+1) (P^{l-p+f})^T E_s^T] Q(p-t-1) A^{t-f} B^{(n)}(k+f+1), \quad t > f,$$

$$H_{t,f}(k) = H_{f,t}^T(k), \quad t < f, \quad (15)$$

$$G_t(k) = (A^{t+1})^T Q(p-t+1) \sum_{s=1}^v [E_s P^{t+1} \theta(k)] B^{(s)}(k+t+1), \quad (16)$$

где

$$Q(i) = A^T Q(i-1) A + \sum_{j=1}^r \text{tr} \{ Q(i-1) T^j \} W^j + R_1(k, p-i),$$

$$Q(0) = R_1(k, p),$$

$$C(k+i) = \text{diag} \{ P^i \theta(k) \} - P \text{diag} \{ P^{i-1} \theta(k) \} P^T.$$

Теорема 1. Вектор прогнозирующих управлений $U(k) = [u^T(k/k), \dots, u^T(k+p-1/k)]^T$, минимизирующий критерий (5) при ограничениях вида (4), на каждом шаге k определяется из решения задачи квадратичного программирования с критерием вида

$$Y(k+p/k) = 2x^T(k)G(k)U(k) + U^T(k)H(k)U(k)$$

при ограничениях

$$U_{\min}(k) \leq \bar{S}(k)U(k) \leq U_{\max}(k), \quad (17)$$

где

$$\bar{S}(k) = \text{diag}(S(k), \dots, S(k+p-1)),$$

$$U_{\min}(k) = \left[u_{\min}^T(k), \dots, u_{\min}^T(k+p-1) \right]^T, \quad U_{\max}(k) = \left[u_{\max}^T(k), \dots, u_{\max}^T(k+p-1) \right]^T,$$

$H(k)$ и $G(k)$ – блочные матрицы, определенные соотношениями (11)–(16).

Оптимальное управление равно

$$u(k) = \begin{bmatrix} I_{n_u} & 0_{n_u} & \dots & 0_{n_u} \end{bmatrix} U(k), \quad (18)$$

где I_{n_u} – единичная матрица размерности n_u , 0_{n_u} – квадратная нулевая матрица размерности n_u .

Оптимальная стратегия прогнозирующего управления системой (1) без учета ограничений определяется уравнением (18), где

$$U(k) = -H^{-1}(k)G^T(k)x(k). \quad (19)$$

При этом оптимальное значение критерия (5) определяется выражением

$$\begin{aligned} J^{\text{opt}}(k+p/k) = & x^T(k) \left[Q(p) - R_1(k,0) - G(k)H^{-1}(k)G^T(k) \right] x(k) + \\ & + \sum_{i=1}^p \sum_{s=1}^v \sum_{n=1}^v \left(D_s^0(k+t) \right)^T \text{tr} \left\{ \left[\theta^T(k) \left(P^{p-i+1} \right)^T E_s^T \right] Q(i-1) \left[E_n P^{p-i+1} \theta(k) \right] \right\} D_n^0(k+t) + \\ & + \sum_{i=1}^p \sum_{l=i}^p \sum_{s=1}^v \sum_{n=1}^v \left(D_s^0(k+t) \right)^T \text{tr} \left\{ \left[E_n P^{l-i} C(k+p-l+1) \left(P^{l-i} \right)^T E_s^T \right] Q(i-1) \right\} D_n^0(k+t). \end{aligned} \quad (20)$$

Доказательство. Критерий (5) можно записать следующим образом:

$$\begin{aligned} J(k+p/k) = & M \left\{ x^T(k+1)R_1(k,1)x(k+1) + u^T(k/k)R_2(k,0)u(k/k) + M \left\{ x^T(k+2)R_1(k,2)x(k+2) + \right. \right. \\ & + u^T(k+1/k)R_2(k,1)u(k+1/k) + \dots + M \left\{ x^T(k+p)R_1(k,p)x(k+p) + \right. \\ & \left. \left. + u^T(k+p-1/k)R_2(k,p-1)u(k+p-1/k) / x(k+p-1), \theta(k+p-1) \right\} \dots / x(k+1), \theta(k+1) \right\} / x(k), \theta(k) \left\} . \end{aligned}$$

Введем обозначение

$$\begin{aligned} J_{k+s} = & M \left\{ x^T(k+s+1)R_1(k,s+1)x(k+s+1) + u^T(k+s/k)R_2(k,s)u(k+s/k) + \right. \\ & + M \left\{ x^T(k+s+2)R_1(k,s+2)x(k+s+2) + u^T(k+s+1/k)R_2(k,s+1)u(k+s+1/k) + \dots \right. \\ & + M \left\{ x^T(k+p)R_1(k,p)x(k+p) + u^T(k+p-1/k)R_2(k,p-1)u(k+p-1/k) / x(k+p-1), \theta(k+p-1) \right\} \dots \\ & \left. \dots / x(k+s+1), \theta(k+s+1) \right\} / x(k+s), \theta(k+s) \left\} . \end{aligned}$$

Очевидно, что

$$J_{k+s} = M \left\{ x^T(k+s+1)R_1(k,s+1)x(k+s+1) + u^T(k+s/k)R_2(k,s)u(k+s/k) + J_{k+s+1} / x(k+s), \theta(k+s) \right\} \quad (21)$$

и

$$J(k+p/k) = J_k. \quad (22)$$

Рассмотрим

$$J_{k+p-1} = M \left\{ x^T(k+p)R_1(k,p)x(k+p) + u^T(k+p-1/k)R_2(k,p-1)u(k+p-1/k) / x(k+p-1), \theta(k+p-1) \right\}. \quad (23)$$

Выражая $x(k+p)$ через $x(k+p-1)$ с учетом (6) и (9), будем иметь

$$x(k+p) = Ax(k+p-1) + \sum_{i=1}^v [E_i P \theta(k+p-1)] B^{(i)}(k+p)u(k+p-1) +$$

$$+\sum_{i=1}^v [E_i v(k+p)] B^{(i)}(k+p) u(k+p-1) + f(x(k+p-1), u(k+p-1), w(k+p-1), \theta(k+p)). \quad (24)$$

Подставляя (24) в (23) и взяв условное математическое ожидание, получим:

$$\begin{aligned} J_{k+p-1} = & x^T(k+p-1) \left[A^T Q(0) A + \sum_{j=1}^r \text{tr}(Q(0) T^j) W^j \right] x(k+p-1) + 2x^T(k+p-1) A^T Q(0) \times \\ & \times \sum_{s=1}^v [E_s P \theta(k+p-1)] B^{(s)}(k+p) u(k+p-1/k) + u^T(k+p-1/k) \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p) \right)^T \times \right. \\ & \times [\theta^T(k+p-1) P^T E_s^T] Q(0) [E_n P \theta(k+p-1)] B^{(n)}(k+p) + \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p) \right)^T [E_n C(k+p) E_s^T] Q(0) \times \\ & \times B^{(n)}(k+p) + \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-1) \right)^T \text{tr} \left\{ [\theta^T(k+p-1) P^T E_s^T] Q(0) [E_n P \theta(k+p-1)] T^j \right\} N_n^j(k+p-1) + \\ & + \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-1) \right)^T \text{tr} \left\{ [E_n C(k+p) E_s^T] Q(0) T^j \right\} N_n^j(k+p-1) + R_2(k, p-1) \Big\} u(k+p-1/k) + \\ & + \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-1) \right)^T [\theta^T(k+p-1) P^T E_s^T] Q(0) [E_n P \theta(k+p-1)] D_n^0(k+p-1) \right\} + \\ & + \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-1) \right)^T [E_n C(k+p) E_s^T] Q(0) D_n^0(k+p-1) \right\}, \end{aligned}$$

где $Q(0) = R_1(k, p)$, $C(k+p) = M \{ v(k+p) v^T(k+p) / \theta(k) \}$.

Предположим далее, что для некоторого q верно

$$\begin{aligned} J_{k+p-q} = & x^T(k+p-q) \left[A^T Q(q-1) A + \sum_{j=1}^r \text{tr}(Q(q-1) T^j) W^j \right] x(k+p-q) + 2x^T(k+p-q) A^T \times \quad (25) \\ & \times \sum_{i=1}^q \left(A^{q-i} \right)^T Q(i-1) \sum_{s=1}^v [E_s P^{q-i+1} \theta(k+p-q)] B^{(s)}(k+p-i+1) u(k+p-i/k) + \sum_{i=1}^q u^T(k+p-i/k) \times \\ & \times \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T [\theta^T(k+p-q) (P^{q-i+1})^T E_s^T] Q(i-1) [E_n P^{q-i+1} \theta(k+p-q)] B^{(n)}(k+p-i+1) + \right. \\ & + \sum_{l=i}^q \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) B^{(n)}(k+p-i+1) + \\ & + \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-i) \right)^T \text{tr} \left\{ [\theta^T(k+p-q) (P^{q-i+1})^T E_s^T] Q(i-1) [E_n P^{q-i+1} \theta(k+p-q)] T^j \right\} \times \\ & \times N_n^j(k+p-i) + \sum_{l=i}^q \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-i) \right)^T \text{tr} \left\{ [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) T^j \right\} \times \\ & \times N_n^j(k+p-i) + R_2(k, p-i) \Big\} u(k+p-i/k) + 2 \sum_{i=1}^{q-1} \sum_{m=i+1}^q u^T(k+p-i/k) \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T \times \right. \\ & \times [\theta^T(k+p-q) (P^{q-i+1})^T E_s^T] Q(i-1) A^{m-i} [E_n P^{q-m+1} \theta(k+p-q)] B^{(n)}(k+p-m+1) + \\ & + \sum_{l=m}^q \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T [E_n P^{l-m} C(k+p-l+1) (P^{l-m})^T E_s^T] Q(i-1) A^{m-i} B^{(n)}(k+p-m+1) \Big\} \times \\ & \times u(k+p-m/k) + \sum_{i=1}^q \left\{ \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-i) \right)^T [\theta^T(k+p-q) (P^{q-i+1})^T E_s^T] Q(i-1) [E_n P^{q-i+1} \theta(k+p-q)] \times \right. \right. \\ & \times D_n^0(k+p-i) \Big\} + \sum_{l=i}^q \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-i) \right)^T [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) D_n^0(k+p-1) \right\} \Big\}, \end{aligned}$$

где $Q(i) = A^T Q(i-1)A + \sum_{j=1}^r \text{tr}(Q(i-1)T^j)W^j + R_1(k, p-i)$.

Покажем, что данная формула верна и для $q+1$. Действительно, из (21) следует, что

$$J_{k+p-(q+1)} = M \left\{ x^T(k+p-q)R_1(k, p-q)x(k+p-q) + \right. \\ \left. + u^T(k+p-(q+1)/k)R_2(k, p-(q+1))u(k+p-(q+1)/k) + J_{k+p-q}/x(k+p-(q+1)), \theta(k+p-(q+1)) \right\}. \quad (26)$$

Подставим в (26) вместо J_{k+p-q} его выражение из (25), вместо $x(k+p-q)$ – его выражение через $x(k+p-(q+1))$, используя (24), вместо $\theta(k+p-q)$ – его выражение через $\theta(k+p-(q+1))$, используя (6); возьмем условное математическое ожидание и, преобразовав выражение, получим, что

$$J_{k+p-(q+1)} = x^T(k+p-(q+1)) \left[A^T Q(q)A + \sum_{j=1}^r \text{tr}(Q(q)T^j)W^j \right] x(k+p-(q+1)) + \quad (27) \\ + 2x^T(k+p-(q+1))A^T \sum_{i=1}^{(q+1)} \left(A^{(q+1)-i} \right)^T Q(i-1) \sum_{s=1}^v [E_s P^{(q+1)-i+1} \theta(k+p-(q+1))] B^{(s)}(k+p-i+1) \times \\ \times u(k+p-i/k) + \sum_{i=1}^{(q+1)} u^T(k+p-i/k) \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T [\theta^T(k+p-(q+1)) (P^{(q+1)-i+1})^T E_s^T] \times \right. \\ \times Q(i-1) [E_n P^{(q+1)-i+1} \theta(k+p-(q+1))] B^{(n)}(k+p-i+1) + \sum_{l=i}^{(q+1)} \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T \times \\ \times [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) B^{(n)}(k+p-i+1) + \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-i) \right)^T \times \\ \times \text{tr} \left\{ [\theta^T(k+p-(q+1)) (P^{(q+1)-i+1})^T E_s^T] Q(i-1) [E_n P^{(q+1)-i+1} \theta(k+p-(q+1))] T^j \right\} N_n^j(k+p-i) + \\ + \sum_{l=i}^q \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-i) \right)^T \text{tr} \left\{ [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) T^j \right\} N_n^j(k+p-i) + \\ + R_2(k, p-i) \} u(k+p-i/k) + 2 \sum_{i=1}^{(q+1)-1} \sum_{m=i+1}^{(q+1)} u^T(k+p-i/k) \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T \times \right. \\ \times [\theta^T(k+p-(q+1)) (P^{(q+1)-i+1})^T E_s^T] Q(i-1) A^{m-i} [E_n P^{(q+1)-m+1} \theta(k+p-(q+1))] \times \\ \times B^{(n)}(k+p-m+1) + \sum_{l=m}^{(q+1)} \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T [E_n P^{l-m} C(k+p-l+1) (P^{l-i})^T E_s^T] \times \\ \times Q(i-1) A^{m-i} B^{(n)}(k+p-m+1) \} u(k+p-m/k) + \sum_{i=1}^{(q+1)} \left\{ \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-i) \right)^T \times \right. \right. \\ \times [\theta^T(k+p-(q+1)) (P^{(q+1)-i+1})^T E_s^T] Q(i-1) [E_n P^{(q+1)-i+1} \theta(k+p-(q+1))] D_n^0(k+p-i) \} + \\ \left. \left. + \sum_{l=i}^{(q+1)} \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-i) \right)^T [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) D_n^0(k+p-1) \right\} \right\} \right\}.$$

Формула (27) совпадает с (25), если в (25) q заменить на $q+1$, а значит, согласно принципу математической индукции формула (25) верна для всех $q = \overline{1, p}$.

Из (25) и (22) следует, что

$$J_k = x^T(k) \left[A^T Q(p-1)A + \sum_{j=1}^r \text{tr}(Q(p-1)T^j)W^j \right] x(k) + 2x^T(k)A^T \sum_{i=1}^p \left(A^{p-i} \right)^T Q(i-1) \times \quad (28) \\ \times \sum_{s=1}^v [E_s P^{p-i+1} \theta(k)] B^{(s)}(k+p-i+1) u(k+p-i/k) + \sum_{i=1}^p u^T(k+p-i/k) \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T \times \right.$$

$$\begin{aligned}
& \times [\theta^T(k) (P^{p-i+1})^T E_s^T] Q(i-1) [E_n P^{p-i+1} \theta(k)] B^{(n)}(k+p-i+1) + \sum_{l=i}^p \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T \times \\
& \times [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) B^{(n)}(k+p-i+1) + \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-i) \right)^T \times \\
& \times \text{tr} \left\{ [\theta^T(k) (P^{p-i+1})^T E_s^T] Q(i-1) [E_n P^{p-i+1} \theta(k)] T^j \right\} N_n^j(k+p-i) + \\
& + \sum_{l=i}^p \sum_{j=1}^r \sum_{s=1}^v \sum_{n=1}^v \left(N_s^j(k+p-i) \right)^T \text{tr} \left\{ [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) T^j \right\} N_n^j(k+p-i) + \\
& + R_2(k, p-i) \left\{ u(k+p-i/k) + 2 \sum_{i=1}^{p-1} \sum_{m=i+1}^p u^T(k+p-i/k) \left\{ \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T \times \right. \right. \\
& \left. \left. + \sum_{l=m}^p \sum_{s=1}^v \sum_{n=1}^v \left(B^{(s)}(k+p-i+1) \right)^T [E_n P^{l-m} C(k+p-l+1) (P^{l-i})^T E_s^T] \times \right. \right. \\
& \left. \left. \times Q(i-1) A^{m-i} B^{(n)}(k+p-m+1) \right\} u(k+p-m/k) \right\} + \sum_{i=1}^p \left\{ \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-i) \right)^T \times \right. \right. \\
& \left. \left. \times [\theta^T(k) (P^{p-i+1})^T E_s^T] Q(i-1) [E_n P^{p-i+1} \theta(k)] D_n^0(k+p-i) \right\} + \sum_{l=i}^{(q+1)} \sum_{s=1}^v \sum_{n=1}^v \text{tr} \left\{ \left(D_s^0(k+p-i) \right)^T \times \right. \right. \\
& \left. \left. \times [E_n P^{l-i} C(k+p-l+1) (P^{l-i})^T E_s^T] Q(i-1) D_n^0(k+p-1) \right\} \right\}.
\end{aligned}$$

Выражение (28) можно записать в матричном виде

$$J(k+p/k) = x^T(k) A^T Q(p-1) A x(k) + 2x^T(k) G(k) U(k) + U^T(k) H(k) U(k), \quad (29)$$

где матрицы $H(k)$, $G(k)$ имеют вид (11)–(16).

Таким образом, имеем задачу минимизации критерия (29) при ограничениях (17), которая эквивалентна задаче квадратичного программирования с критерием (5) при ограничениях (4).

Очевидно, что если ограничения на управляющие воздействия отсутствуют, то оптимальный вектор прогнозирующих управлений $U(k)$, минимизирующий критерий (29) на траекториях системы (1), определяется уравнением (19). Нетрудно показать, что при этом оптимальное значение критерия (29) имеет вид (20).

Заключение

В данной работе предложен метод синтеза стратегий прогнозирующего управления по квадратичному критерию для нелинейных дискретных систем со скачкообразно меняющимися параметрами. Данный подход позволяет в явном виде учесть ограничения на управления. Алгоритм синтеза прогнозирующей стратегии включает решение последовательности задач квадратичного программирования. Синтезированы стратегии управления с учетом явных ограничений на управляющие воздействия.

ЛИТЕРАТУРА

1. Пакишин П.В. Дискретные системы со случайными параметрами и структурой. М. : Физматлит, 1994.
2. Пакишин П.В., Ретинский Д.М. Робастная стабилизация систем случайной структуры с переключаемой статической обратной связью по выходу // Автоматика и телемеханика. 2005. № 7. С. 135–147.
3. Домбровский В.В., Обьедко Т.Ю. Управление с прогнозированием системами с марковскими скачками при ограничениях и применение к оптимизации инвестиционного портфеля // Автоматика и телемеханика. 2011. № 5. С. 96–112.
4. Смагин В.И., Поползухина Е.В. Синтез следящих систем управления для объектов со случайными скачкообразными параметрами и мультипликативными возмущениями // Вестник Томского государственного университета. 2000. № 271. С. 171–175.
5. Blackmore L., Bektassov A., Ono M., Williams B.C. Robust optimal predictive control of jump Markov linear systems using particles // Lecture Notes in Computer Science. 2007. V. 4416. P. 104–117.
6. Costa O.L.V., Okimura R.T. Discrete-time mean-variance optimal control of linear systems with Markovian jumps and multiplicative noise // International Journal of Control. 2009. V. 82, No. 2. P. 256–267.

7. Costa O.L.V., Oliveira A. Optimal mean-variance control for discrete-time linear systems with Markovian jumps and multiplicative noises // Automatica. 2012. V. 48, No. 2. P. 304–315.
8. Dragan V., Morozan T. The Linear Quadratic Optimization Problems for a Class of Linear Stochastic Systems With Multiplicative White Noise and Markovian Jumping // IEEE Transactions on Automatic Control. 2004. V. 49, No. 5. P. 665–675.
9. Elliott R.J., Aggoun L., Moore J.B. Hidden Markov Models: Estimation and Control. Berlin : Springer-Verlag, 1995.
10. Li X., Zhou X.Y. Indefinite stochastic LQ control with Markovian jumps in a finite time horizon // Communications in Information and Systems. 2002. No. 2. P. 265–282.
11. Rawlings J. Tutorial: Model Predictive Control Technology // Proc. Amer. Control Conf. San Diego. California. June 1999. P. 662–676.
12. Yaz E. A control scheme for a class of discrete nonlinear stochastic systems // IEEE Transactions on Automatic Control. 1987. V. AC-32, No. 1. P. 77–80.
13. Yaz E. Robust design of Stochastic controllers for nonlinear systems // IEEE Transactions on Automatic Control. 1989. V. AC-34, No. 3. P. 349–353.
14. Домбровский В.В., Домбровский Д.В., Ляшенко Е.А. Управление с прогнозированием нелинейными стохастическими системами при ограничениях // Вестник Томского государственного университета. 2006. № 18. С. 320–323.
15. Novara C., Fagiano L., Milanese M. Direct feedback control design for nonlinear systems // Automatica. 2013. V. 49. P. 849–860.
16. Mhaskar P., Nael H. El-Farra, Panagiotis D.C. Robust hybrid predictive control of nonlinear systems // Automatica. 2005. V. 41. P. 209–217.

Домбровский Владимир Валентинович, д-р техн. наук, профессор. E-mail: dombrovs@ef.tsu.ru

Объедко Татьяна Юрьевна, канд. физ.-мат. наук. E-mail: tatyana.obedko@mail.ru

Самородова Мария Владимировна. E-mail: samorodova21@gmail.com

Томский государственный университет

Поступила в редакцию 15 мая 2015 г.

Dombrovskii Vladimir V., Obyedko Tatyana Y., Samorodova Mariya V. (Tomsk State University, Russian Federation).

Model Predictive Control for Nonlinear Stochastic Systems with Markovian Jumps under Constraints.

Keywords: stochastic nonlinear systems; model predictive control; Markovian jumps, constrains.

DOI 10.17223/19988605/32/2

Let the control object is described by the equation

$$x(k+1) = Ax(k) + B[\alpha(k+1), k+1]u(k) + f(x(k), u(k), w(k), \alpha(k+1)), \quad (1)$$

where $x(k)$ is the n_x – dimensional vector of state, $u(k)$ is the n_u – dimensional vector of control; $w(k)$ is the n_w – dimensional vector of white noises with zero-mean and identity covariance matrix; $\alpha(k)$ ($k = 0, 1, 2, \dots$) denotes a time-invariant Markov chain taking values in a finite set of observable states $\{1, 2, \dots, v\}$ with the known transition probability matrix $P = [P_{i,j}]$ ($i, j \in \{1, 2, \dots, v\}$), $P_{j,i} = P\{\alpha(k+1) = \alpha_j | \alpha(k) = \alpha_i\}$,

$\sum_{j=1}^v P_{j,i} = 1$, and the initial distribution $p_i = P\{\alpha(0) = i\}$ ($i = 1, 2, \dots, v$), $\sum_{i=1}^v p_i = 1$.

It is assumed that the state of Markov chain is observable at time instant k , and $w(k)$ is independent of the Markov chain $\alpha(k)$ ($k = 0, 1, 2, \dots$). The function f is defined by its statistical properties as follows:

$$\begin{aligned} M\{f(x(k), u(k), w(k), \alpha(k+1)) / x(k), \alpha(k) = \alpha_j\} &= 0, \\ M\{f(x(k), u(k), w(k), \alpha(k+1)) f^T(x(k), u(k), w(k), \alpha(k+1)) / x(k), \alpha(k) = \alpha_j\} &= T^0[\alpha(k), k] + \\ &+ \sum_{i=1}^r T^i \left(x^T(k) W^i x(k) + u^T(k) M^i[\alpha(k), k] u(k) \right), \end{aligned}$$

for all $x(k)$, where $r = n(n+1)/2$; T^i , W^i , and $M^i = (C^i)^T C^i$ ($i = \overline{1, r}$), $T^0 = (D^0)^T D^0$ are positive semidefinite and symmetric matrices.

The following constraints are imposed on the control variables

$$u_{\min}(k) \leq S(k)u(k) \leq u_{\max}(k), \quad (2)$$

where $S(k)$ is a matrix of corresponding dimension.

For control of system (1) we synthesize the strategies with a predictive control model. At each step k we minimize the quadratic criterion with a receding horizon

$$J(k+p/k) = M \left\{ \sum_{i=1}^p x^T(k+i) R_1(k, i) x(k+i) + u^T(k+i-1/k) R_2(k, i-1) u(k+i-1/k) / x(k), \alpha(k) = \alpha_j \right\},$$

on trajectories of system (1) over the sequence of predictive controls $u(k/k), \dots, u(k+p-1/k)$, which depend on system's state and on the state of Markov chain at moment k , under constraints (2), where $R_1(k, i) \geq 0$, $R_2(k, i) \geq 0$ are weigh matrices of corresponding dimen-

sions; p is a prediction horizon, k is a current moment. The synthesis of predictive control strategies is reduced to the sequence of quadratic programming tasks.

REFERENCES

1. Pakshin, P.V. (1994) *Diskretnye sistemy so sluchaynymi parametrami i strukturoy* [Discrete-systems with stochastic parameters and structure]. Moscow: Fizmatlit.
2. Pakshin, P.V. & Retinskii, D.M. (2005) Robust Stabilization of Random-Structure Systems via Switchable Static Output Feedback. *Automation and Remote Control*. 66 (7). pp. 1153-1161. DOI: 10.1007/s10513-005-0155-5
3. Dombrovskii, V.V. & Obedko, T.Yu. (2011) Predictive control of systems with Markovian jumps under constraints and its application to the investment portfolio optimization. *Automation and Remote Control*. 72 (5). pp. 989-1003. DOI: 10.1134/S0005117911050079
4. Smagin, V.I. & Popolzhina, E.V. (2000) Sintez sledyashchikh sistem upravleniya dlya ob"ektov so sluchaynymi skachkoobraznymi parametrami i multiplikativnymi vozmushcheniyami [The synthesis of tracking control systems for objects with random switching parameters and multiplicative noises]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 271. pp. 171-175.
5. Blackmore, L., Bektassov, A., Ono, M. & Williams, B.C. (2007) Robust optimal predictive control of jump Markov linear systems using particles. *Lecture Notes in Computer Science*. 4416. pp. 104-117. DOI: 10.1007/978-3-540-71493-4_11
6. Costa, O.L.V. & Okimura, R.T. (2009) Discrete-time mean-variance optimal control of linear systems with Markovian jumps and multiplicative noise. *International Journal of Control*. 82 (2). pp. 256-267. DOI: 10.1080/00207170802050825
7. Costa, O.L.V. & Oliveira, A. (2012) Optimal mean-variance control for discrete-time linear systems with Markovian jumps and multiplicative noises. *Automatica*. 48 (2). pp. 304-315. DOI: 10.1080/00207170802050825
8. Dragan, V. & Morozan, T. (2004) The Linear Quadratic Optimization Problems for a Class of Linear Stochastic Systems With Multiplicative White Noise and Markovian Jumping. *IEEE Transactions on Automatic Control*. 49 (5). pp. 665-675.
9. Elliott, R.J., Aggoun, L. & Moore, J.B. (1995) *Hidden Markov Models: Estimation and Control*. Berlin: Springer-Verlag.
10. Li, X. & Zhou, X.Y. (2002) Indefinite stochastic LQ control with Markovian jumps in a finite time horizon. *Communications in Information and Systems*. 2. pp. 265-282.
11. Rawlings, J. (1999) Tutorial: Model Predictive Control Technology. *Proceedings American Control Conference*. San Diego. California. pp. 662–676. DOI: 10.1109/ACC.1999.782911
12. Yaz, E.A (1987) Control scheme for a class of discrete nonlinear stochastic systems. *IEEE Transactions on Automatic Control*. AC-32 (1). pp. 77-80. DOI: 10.1109/TAC.1987.1104428
13. Yaz, E. (1989) Robust design of Stochastic controllers for nonlinear systems. *IEEE Transactions on Automatic Control*. AC-34 (3). pp. 349-353. DOI: 10.1109/9.16432
14. Dombrovskii, V.V., Dombrovskii, D.V. & Lyashenko, E.A. (2006) Upravlenie s prognozirovaniem nelineynymi stokhasticheskimi sistemami pri ogranicheniyakh [Model predictive control for nonlinear stochastic systems under constraints]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 18. pp. 320-323.
15. Novara, C., Fagiano, L. & Milanese, M. (2013) Direct feedback control design for nonlinear systems. *Automatica*. 49. pp. 849-860. DOI: 10.1016/j.automatica.2013.01.002
16. Mhaskar, P., Nael, H. El-Farra & Panagiotis, D.C. (2005) Robust hybrid predictive control of nonlinear systems. *Automatica*. 41. pp. 209-217. DOI: 10.1016/j.automatica.2004.08.020

А.А. Калягин, Л.А. Нежелская

**СРАВНЕНИЕ МП- И ММ-ОЦЕНОК ДЛИТЕЛЬНОСТИ МЕРТВОГО ВРЕМЕНИ
В ОБОБЩЕННОМ ПОЛУСИНХРОННОМ ПОТОКЕ СОБЫТИЙ**

Работа выполнена в рамках государственного заказа Минобрнауки РФ на проведение научных исследований в Национальном исследовательском Томском государственном университете на 2014–2016 годы.

Изучается обобщенный полусинхронный поток событий, являющийся одной из адекватных математических моделей информационных потоков заявок в цифровых сетях интегрального обслуживания. Поток функционирует в условиях непродлевающегося мертвого времени, когда длительность мертвого времени – неизвестная фиксированная величина. Проводится сравнение качества получаемых (по наблюдениям за моментами наступления событий потока) оценок длительности мертвого времени методом максимального правдоподобия (МП-оценки) и методом моментов (ММ-оценки).

Ключевые слова: обобщенный полусинхронный поток событий; непродлевающееся мертвое время; МП-оценки; ММ-оценки; длительность мертвого времени.

Настоящая статья является непосредственным продолжением исследований обобщенного полусинхронного потока событий (далее – поток), начатых в статьях [1–5]. Изучаемый поток относится к классу дважды стохастических потоков событий и является одной из адекватных математических моделей информационных потоков сообщений, функционирующих в цифровых сетях интегрального обслуживания (ЦСИО) [6]. Дважды стохастические потоки делятся на два класса: к первому классу относятся потоки, интенсивность которых есть непрерывный случайный процесс; ко второму классу относятся потоки, интенсивность которых есть кусочно-постоянный случайный процесс с конечным числом состояний. Вторым класс потоков в настоящее время принято называть МС-потоками либо МАР-потоками событий. В [7] приведена классификация МС-потоков событий и установлена связь между МС-потоками и МАР-потоками событий. Наиболее полная литература по изучаемым типам МС-потоков приведена в [1, 8].

В реальных ситуациях параметры, задающие входящий поток событий, известны либо частично, либо вообще не известны, либо (что еще более ухудшает ситуацию) изменяются со временем. В подобных случаях наиболее рациональным является применение адаптивных систем массового обслуживания, которые в процессе функционирования оценивают неизвестные параметры либо состояния входящих потоков событий и изменяют дисциплины обслуживания в соответствии с полученными оценками [9]. Вследствие этого возникают задачи оценки состояний [10, 11] и оценки параметров [12, 13] потока по наблюдениям за моментами наступления событий.

Одним из искажающих факторов при оценке состояний и параметров потока событий выступает мертвое время регистрирующих приборов [14], которое порождается зарегистрированным событием. Другие же события, наступившие в течение периода мертвого времени, недоступны наблюдению (теряются). Можно считать, что этот период продолжается некоторое фиксированное время (непродлевающееся мертвое время). В качестве примера приведем CSMA/CD – протокол случайного множественного доступа с обнаружением конфликта, широко используемого в компьютерных сетях. В момент регистрации (обнаружения) конфликта на входе некоторого узла сети рассылается сигнал «заглушки» («пробки»); в течение времени рассылки сигнала «заглушки» заявки, поступившие в данный узел сети, получают отказ в обслуживании и направляются в источник повторных вызовов. Здесь время, в течение которого узел сети закрыт для обслуживания заявок, поступающих в него после обнаружения конфликта, можно трактовать как мертвое время прибора, регистрирующего конфликт в узле сети.

Для того чтобы оценить потери сообщений потока, возникающие из-за эффекта мертвого времени, необходимо оценить его длительность.

Подчеркнем, что задачи оценки параметров и длительности мертвого времени рассматривались в статьях [15–18] в рамках полусинхронного потока событий, который является частным случаем обобщенного полусинхронного потока событий, изучаемого в настоящей статье.

Для оценки длительности мертвого времени, как правило, используются два метода оценивания: метод максимального правдоподобия [19, 20] и метод моментов [16, 17]. На основании сравнения [21] по тем или иным критериям получаемых МП-оценок и ММ-оценок делается вывод о применимости в конкретных условиях упомянутых методов.

В настоящей статье производится сравнение оценок длительности мертвого времени в обобщенном полусинхронном потоке событий, полученных методом моментов и методом максимального правдоподобия.

1. Постановка задачи

Рассматривается обобщенный полусинхронный поток событий, интенсивность которого есть кусочно-постоянный случайный процесс $\lambda(t)$ с двумя состояниями λ_1 и λ_2 ($\lambda_1 > \lambda_2 \geq 0$). В течение временного интервала, когда $\lambda(t) = \lambda_i$, имеет место пуассоновский поток событий с интенсивностью λ_i , $i = 1, 2$. Переход из первого состояния процесса $\lambda(t)$ во второе возможен только в момент наступления события, при этом переход осуществляется с вероятностью p ($0 < p \leq 1$); с вероятностью $1 - p$ процесс $\lambda(t)$ остается в первом состоянии. Тогда длительность пребывания процесса $\lambda(t)$ в первом состоянии есть случайная величина с экспоненциальной функцией распределения $F_1(\tau) = 1 - e^{-p\lambda_1\tau}$. Переход из второго состояния процесса $\lambda(t)$ в первое может осуществляться в произвольный момент времени. При этом длительность пребывания процесса $\lambda(t)$ во втором состоянии распределена по экспоненциальному закону: $F_2(\tau) = 1 - e^{-\alpha\tau}$. При переходе процесса $\lambda(t)$ из второго состояния в первое инициируется с вероятностью δ ($0 \leq \delta \leq 1$) дополнительное событие. При этом блочная матрица инфинитезимальных коэффициентов примет вид

$$D = \begin{vmatrix} -\lambda_1 & 0 \\ (1-\delta)\alpha & -(\lambda_2 + \alpha) \end{vmatrix} \begin{vmatrix} (1-p)\lambda_1 & p\lambda_1 \\ \delta\alpha & \lambda_2 \end{vmatrix} = \|D_0 | D_1\|.$$

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 противоположенным знаком.



Рис. 1. Формирование наблюдаемого потока событий

В сделанных предположениях $\lambda(t)$ – скрытый марковский процесс. После каждого зарегистрированного в момент времени t_k события наступает время фиксированной длительности T (мертвое время), в

течение которого другие события исходного потока недоступны наблюдению. События, наступившие в течение мертвого времени, не вызывают продления его периода (непродлеваемое мертвое время). По окончании мертвого времени первое наступившее событие снова создает период мертвого времени T и т.д. Вариант возникающей ситуации показан на рис. 1, где 1, 2 – состояния случайного процесса $\lambda(t)$; дополнительные события, которые могут наступать в момент перехода процесса $\lambda(t)$ из второго состояния в первое, помечены буквой δ ; штриховка – периоды мертвого времени длительности T ; $t_1, t_2, \dots$ – моменты наступления событий в наблюдаемом потоке.

Подчеркнем, что если $\delta = 0$, то имеет место обычный полусинхронный поток событий [15–18]. Отметим также, что в соответствии с классификацией МАР-потоков событий, приведенной в [7], обобщенный полусинхронный поток относится к классу МАР-потоков событий второго порядка.

Процесс $\lambda(t)$ и типы событий (события пуассоновских потоков и дополнительные события) являются принципиально ненаблюдаемыми, а наблюдаемыми являются только временные моменты наступления событий $t_1, t_2, \dots$ наблюдаемого потока. Рассматривается установившийся (стационарный) режим функционирования наблюдаемого потока событий, поэтому переходными процессами на интервале наблюдения (t_0, t) , где t_0 – начало наблюдений, t – окончание наблюдений, пренебрегаем. Необходимо в момент окончания наблюдений (в момент времени t) осуществить методом максимального правдоподобия и методом моментов оценку $\hat{T}$ длительности мертвого времени и произвести сравнение полученных оценок.

2. МП-оценка длительности мертвого времени

Обозначим $\tau_k = t_{k+1} - t_k$ ($k = 1, 2, \dots$) – значение длительности k -го интервала между соседними событиями наблюдаемого потока ($\tau_k > 0$). Так как рассматривается стационарный режим, то плотность вероятностей значений длительности k -го интервала $p_T(\tau_k) = p_T(\tau)$, $\tau \geq 0$, для любого k (индекс T подчеркивает, что плотность вероятностей зависит от длительности мертвого времени). В силу этого момент t_k без потери общности можно положить равным нулю или, что то же самое, момент наступления события наблюдаемого потока есть $\tau = 0$. Тогда [4] плотность вероятностей примет вид

$$\begin{aligned} p_T(\tau) &= 0, \quad 0 \leq \tau < T; \quad p_T(\tau) = [1 - f(T)]\lambda_1 e^{-\lambda_1(\tau-T)} + f(T)(\alpha + \lambda_2)e^{-(\alpha + \lambda_2)(\tau-T)}, \\ f(T) &= \frac{p(\alpha + \lambda_2)(\lambda_1 - \lambda_2 - \alpha\delta)[\lambda_1 + (\alpha + p\lambda_1 - \lambda_1)e^{-(\alpha + p\lambda_1)T}]}{(\alpha + p\lambda_1)(\lambda_1 - \lambda_2 - \alpha)F(T)}, \\ F(T) &= (\alpha + \lambda_2) - [\lambda_2 - p(\lambda_2 + \alpha\delta)]e^{-(\alpha + p\lambda_1)T}, \quad 0 \leq T \leq \tau, \quad \tau \geq 0. \end{aligned} \quad (1)$$

В (1) функция $F(T) > 0$ для любых T ($0 \leq T \leq \tau$). Рассматривается общий случай, т.е. $(\lambda_1 - \lambda_2 - \alpha) \neq 0$. Подчеркнем, что (1) – одномерная плотность вероятностей.

Пусть $\tau_1 = t_2 - t_1$, $\tau_2 = t_3 - t_2$, ..., $\tau_k = t_{k+1} - t_k$ – последовательность измеренных (в результате наблюдения за потоком в течение интервала наблюдения $(0, t)$) значений длительностей интервалов между соседними событиями потока. Упорядочим величины $\tau_1, \dots, \tau_k$ по возрастанию: $\tau_{\min} = \tau^{(1)} < \tau^{(2)} < \dots < \tau^{(k)}$. В силу предпосылок наблюдаемый поток обладает марковским свойством, если его эволюцию рассматривать начиная с момента наступления события (с момента t_k , $k = 1, 2, \dots$). Тогда [22] функция правдоподобия, с учетом (1), запишется в виде

$$\begin{aligned} L(\lambda_i, p, \alpha, \delta, T | \tau^{(1)}, \dots, \tau^{(k)}) &= 0, \quad 0 \leq \tau_{\min} < T; \\ L(\lambda_i, p, \alpha, \delta, T | \tau^{(1)}, \dots, \tau^{(k)}) &= \prod_{j=1}^k p_T(\tau^{(j)}), \quad \tau_{\min} \geq T. \end{aligned}$$

Так как поставленная задача заключается в построении оценки $\hat{T}$ длительности мертвого времени (в предположении, что остальные параметры потока $\lambda_i, p, \alpha, \delta$ известны), то, согласно методу максимального правдоподобия, её реализация есть решение оптимизационной задачи

$$\begin{aligned} L(T | \tau^{(1)}, \dots, \tau^{(k)}) &= \prod_{j=1}^k p_T(\tau^{(j)}) = \prod_{j=1}^k \{ [1 - f(T)]\lambda_1 e^{-\lambda_1(\tau^{(j)}-T)} + \\ &+ f(T)(\alpha + \lambda_2)e^{-(\alpha + \lambda_2)(\tau^{(j)}-T)} \} \Rightarrow \max_T, \quad 0 \leq T \leq \tau_{\min}, \quad \tau_{\min} > 0, \end{aligned} \quad (2)$$

где $f(T)$ определена в (1).

Значение T , при котором (2) достигает своего глобального максимума, есть $\hat{T}_{МП}$ – оценка максимального правдоподобия длительности мертвого времени.

В [5] аналитически строго решена оптимизационная задача (2): при любых значениях параметров потока $\lambda_1 > 0$, $\lambda_2 \geq 0$ ($\lambda_1 > \lambda_2$), $0 < p \leq 1$, $0 \leq \delta \leq 1$, $\alpha > 0$ МП-оценка $\hat{T}_{МП} = \tau_{\min}$. Таким образом, в процессе наблюдения (в течение временного интервала (t_0, t)) потока событий вычисляются величины τ_k , $k = \overline{1, n}$, после чего находится $\tau_{\min} = \min \tau_k$ ($k = \overline{1, n}$) и полагается $\hat{T}_{МП} = \tau_{\min}$.

3. ММ-оценка длительности мертвого времени

В [4] показано, что обобщенный полусинхронный поток событий, функционирующий в условиях мертвого времени, в общем случае является коррелированным потоком. Только в частных случаях поток становится рекуррентным.

Пусть (t_k, t_{k+1}) , (t_{k+1}, t_{k+2}) – два смежных интервала в наблюдаемом потоке с соответствующими значениями длительностей: $\tau_k = t_{k+1} - t_k$, $\tau_{k+1} = t_{k+2} - t_{k+1}$; их расположение на временной оси, в силу стационарности потока, произвольно. Тогда можно положить $k = 1$ и рассматривать соседние интервалы (t_1, t_2) , (t_2, t_3) с соответствующими значениями длительностей: $\tau_1 = t_2 - t_1$, $\tau_2 = t_3 - t_2$; $\tau_1 \geq 0$, $\tau_2 \geq 0$. При этом $\tau_1 = 0$ соответствует моменту t_1 наступления события наблюдаемого потока, $\tau_2 = 0$ – моменту t_2 наступления следующего события наблюдаемого потока.

Соответствующая совместная плотность вероятностей при этом есть $p_T(\tau_1, \tau_2)$, $\tau_1 \geq 0$, $\tau_2 \geq 0$ [4]:

$$\begin{aligned} p_T(\tau_1, \tau_2) &= 0; \quad 0 \leq \tau_1 < T, \quad 0 \leq \tau_2 < T, \\ p_T(\tau_1, \tau_2) &= p_T(\tau_1)p_T(\tau_2) + e^{-(\alpha + p\lambda_1)T} f(T)[1 - f(T)] \frac{\lambda_2 - p(\lambda_2 + \alpha\delta)}{\alpha + \lambda_2} \times \\ &\times \left[\lambda_1 e^{-\lambda_1(\tau_1 - T)} - (\alpha + \lambda_2) e^{-(\alpha + \lambda_2)(\tau_1 - T)} \right] \left[\lambda_1 e^{-\lambda_1(\tau_2 - T)} - (\alpha + \lambda_2) e^{-(\alpha + \lambda_2)(\tau_2 - T)} \right], \quad \tau_1 \geq T, \quad \tau_2 \geq T, \end{aligned} \quad (3)$$

где $f(T)$, $p_T(\tau_k)$ определены в (1) для $\tau = \tau_k$, $k = 1, 2$.

Теоретическая ковариация значений τ_1 и τ_2 имеет вид

$$\text{cov}(\tau_1, \tau_2) = \int_T^\infty \int_T^\infty \tau_1 \tau_2 p_T(\tau_1, \tau_2) d\tau_1 d\tau_2 - \left[\int_T^\infty \tau p_T(\tau) d\tau \right]^2.$$

Подставляя в последнюю формулу $p_T(\tau_1, \tau_2)$ из (3), $p_T(\tau)$ из (1), находим явный вид теоретической ковариации:

$$\text{cov}(\tau_1, \tau_2) = f(T)[1 - f(T)](\lambda_1 - \lambda_2 - \alpha)^2 \frac{\lambda_2 - (\lambda_2 + \alpha\delta)p}{\lambda_1^2(\alpha + \lambda_2)^3} e^{-(\alpha + p\lambda_1)T}, \quad (4)$$

где $f(T)$ определена в (1).

Пусть за время наблюдения (в течение временного интервала (t_0, t)) реализовалось n интервалов (t_k, t_{k+1}) длительности τ_k , $k = \overline{1, n}$. Введем статистику

$$\widehat{\text{cov}}(\tau_1, \tau_2) = \frac{1}{n-1} \sum_{k=1}^{n-1} \tau_k \tau_{k+1} - \left(\frac{1}{n} \sum_{k=1}^n \tau_k \right)^2, \quad (5)$$

являющуюся оценкой теоретической ковариации (4). Тогда, согласно методу моментов [22], уравнение моментов, учитывающее коррелированность потока событий, запишется в виде

$$f(T)[1 - f(T)](\lambda_1 - \lambda_2 - \alpha)^2 \frac{\lambda_2 - (\lambda_2 + \alpha\delta)p}{\lambda_1^2(\alpha + \lambda_2)^3} e^{-(\alpha + p\lambda_1)T} = \widehat{\text{cov}}(\tau_1, \tau_2). \quad (6)$$

Подставляя в (6) выражение $f(T)$ из (1), вводя новую переменную $x = e^{-(\alpha + p\lambda_1)T}$ и проделывая при этом необходимые преобразования, находим (6) в виде

$$\begin{aligned} ax^3 + bx^2 + cx + d &= 0, \\ a &= h\{\lambda_1(\alpha + \lambda_2) - (\alpha + p\lambda_1)[\lambda_1(1 - p) + \lambda_2]\}; \end{aligned}$$

$$\begin{aligned}
b &= -\left\{ h[2\lambda_1(\alpha + \lambda_2) - (\alpha + p\lambda_1)(\lambda_1 + \lambda_2 + \alpha)] + [\lambda_2 - p(\lambda_2 + \alpha\delta)]^2 C_0 \right\}; \\
c &= (\alpha + \lambda_2) \left\{ \lambda_1 h + 2[\lambda_2 - p(\lambda_2 + \alpha\delta)] C_0 \right\}; \\
d &= C_0(\alpha + \lambda_2)^2; \\
h &= \frac{p\alpha(\lambda_1 - \lambda_2 - \alpha\delta)[\lambda_1(1 - p + p\delta) - (\alpha + \lambda_2)][\lambda_2 - p(\lambda_2 + \alpha\delta)]}{[\lambda_1(\alpha + \lambda_2)(\alpha + p\lambda_1)]}.
\end{aligned} \tag{7}$$

Решение уравнения (7) определит три корня x_i , $i = 1, 2, 3$, которые, в свою очередь, определяют три ММ-оценки длительности мертвого времени

$$\hat{T}_{MM}^{(i)} = -\frac{1}{\alpha + p\lambda_1} \ln x_i, \quad i = 1, 2, 3.$$

Алгоритм нахождения единственной оценки $\hat{T}_{MM}$ следующий [21]:

1) для определенного набора параметров λ_i , $i = 1, 2$, α , p , δ , T ед. времени осуществляется в течение T_m ед. времени имитационное моделирование наблюдаемого потока событий;

2) результатом работы имитационной модели является оценка теоретической ковариации (5), где n принимает одно из целых значений ($n \geq 2$);

3) решается кубическое уравнение (7), т.е. находятся корни x_1 , x_2 , x_3 ;

4) если все корни комплексные, то $\hat{T}_{MM} = \tau_{\min}$;

5) выделяются вещественные корни; здесь возможны три случая:

5.1) вещественный корень один – x_1 , тогда:

а) если $x_1 \leq 0$, то $\hat{T}_{MM} = \tau_{\min}$;

б) если $x_1 > 0$, то: б.1) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(1)} > \tau_{\min}$, б.2) $\hat{T}_{MM} = \hat{T}_{MM}^{(1)}$, если $0 < \hat{T}_{MM}^{(1)} \leq \tau_{\min}$,

б.3) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(1)} \leq 0$;

5.2) вещественных корней два – x_1 , x_2 ($x_1 < x_2$), тогда:

а) если $x_1 < x_2 \leq 0$, то $\hat{T}_{MM} = \tau_{\min}$;

б) если $x_1 \leq 0 < x_2$, то: б.1) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(2)} > \tau_{\min}$, б.2) $\hat{T}_{MM} = \hat{T}_{MM}^{(2)}$, если $0 < \hat{T}_{MM}^{(2)} \leq \tau_{\min}$,

б.3) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(2)} \leq 0$;

в) если $0 < x_1 < x_2$, то: в.1) $\hat{T}_{MM} = \tau_{\min}$, если $\tau_{\min} < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)}$, в.2) $\hat{T}_{MM} = \hat{T}_{MM}^{(2)}$, если $0 < \hat{T}_{MM}^{(2)} \leq \tau_{\min} < \hat{T}_{MM}^{(1)}$, в.3) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(2)} \leq 0 < \tau_{\min} \leq \hat{T}_{MM}^{(1)}$, в.4) $\hat{T}_{MM} = (\hat{T}_{MM}^{(1)} + \hat{T}_{MM}^{(2)})/2$, если $0 < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)} \leq \tau_{\min}$, в.5) $\hat{T}_{MM} = \hat{T}_{MM}^{(1)}$, если $\hat{T}_{MM}^{(2)} \leq 0 < \hat{T}_{MM}^{(1)} \leq \tau_{\min}$, в.6) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)} \leq 0$;

5.3) вещественных корней три – x_1 , x_2 , x_3 ($x_1 < x_2 < x_3$), тогда:

а) если $x_1 < x_2 < x_3 \leq 0$, то $\hat{T}_{MM} = \tau_{\min}$;

б) если $x_1 < x_2 \leq 0 < x_3$, то: б.1) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} > \tau_{\min}$, б.2) $\hat{T}_{MM} = \hat{T}_{MM}^{(3)}$, если $0 < \hat{T}_{MM}^{(3)} \leq \tau_{\min}$,

б.3) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} \leq 0$;

в) если $x_1 \leq 0 < x_2 < x_3$, то: в.1) $\hat{T}_{MM} = \tau_{\min}$, если $\tau_{\min} < \hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)}$, в.2) $\hat{T}_{MM} = \hat{T}_{MM}^{(3)}$, если $0 < \hat{T}_{MM}^{(3)} \leq \tau_{\min} < \hat{T}_{MM}^{(2)}$, в.3) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} \leq 0 < \tau_{\min} \leq \hat{T}_{MM}^{(2)}$, в.4) $\hat{T}_{MM} = (\hat{T}_{MM}^{(2)} + \hat{T}_{MM}^{(3)})/2$, если $0 < \hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} \leq \tau_{\min}$, в.5) $\hat{T}_{MM} = \hat{T}_{MM}^{(2)}$, если $\hat{T}_{MM}^{(3)} \leq 0 < \hat{T}_{MM}^{(2)} \leq \tau_{\min}$, в.6) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} \leq 0$;

г) если $0 < x_1 < x_2 < x_3$, то: г.1) $\hat{T}_{MM} = \tau_{\min}$, если $\tau_{\min} < \hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)}$, г.2) $\hat{T}_{MM} = \hat{T}_{MM}^{(3)}$, если $0 < \hat{T}_{MM}^{(3)} \leq \tau_{\min} < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)}$, г.3) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} \leq 0 < \tau_{\min} \leq \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)}$,

г.4) $\hat{T}_{MM} = (\hat{T}_{MM}^{(2)} + \hat{T}_{MM}^{(3)})/2$, если $0 < \hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} \leq \tau_{\min} < \hat{T}_{MM}^{(1)}$, г.5) $\hat{T}_{MM} = \hat{T}_{MM}^{(2)}$, если $\hat{T}_{MM}^{(3)} \leq 0 < \hat{T}_{MM}^{(2)} \leq \tau_{\min} < \hat{T}_{MM}^{(1)}$, г.6) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} \leq 0 < \tau_{\min} < \hat{T}_{MM}^{(1)}$, г.7) $\hat{T}_{MM} = (\hat{T}_{MM}^{(1)} + \hat{T}_{MM}^{(2)} + \hat{T}_{MM}^{(3)})/3$, если $0 < \hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)} \leq \tau_{\min}$, г.8) $\hat{T}_{MM} = (\hat{T}_{MM}^{(1)} + \hat{T}_{MM}^{(2)})/2$, если $\hat{T}_{MM}^{(3)} \leq 0 < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)} \leq \tau_{\min}$, г.9) $\hat{T}_{MM} = \hat{T}_{MM}^{(1)}$, если $\hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} \leq 0 < \hat{T}_{MM}^{(1)} \leq \tau_{\min}$, г.10) $\hat{T}_{MM} = \tau_{\min}$, если $\hat{T}_{MM}^{(3)} < \hat{T}_{MM}^{(2)} < \hat{T}_{MM}^{(1)} \leq 0 \leq \tau_{\min}$.

В результате работы алгоритма осуществляется один из описанных вариантов, тем самым определяется единственная ММ-оценка $\hat{T}_{MM}$ длительности мертвого времени.

4. Численное сравнение МП- и ММ-оценок

Для получения численных результатов разработан алгоритм вычисления МП- и ММ-оценок. Программа расчета реализована на языке программирования С++ в среде Builder 6. Первый этап расчета предполагает имитационное моделирование (при заданных значениях параметров λ_i , $i = 1, 2$, α , p , δ , T ед. времени и заданном времени моделирования T_m ед. времени) наблюдаемого потока событий. Описание алгоритма имитационного моделирования здесь не приводится, так как никаких принципиальных трудностей алгоритм не содержит. Результатом работы имитационной модели является последовательность значений длительностей временных интервалов $\tau_1, \tau_2, \dots, \tau_n$ ($n = 2, 3, \dots$). Второй этап расчета – непосредственное вычисление МП- и ММ-оценок. Коротко опишем второй этап: 1) находится оценка $\hat{T}_{МП} = \tau_{\min}(\tau_{\min} = \min \tau_k, k = \overline{1, n})$; 2) вычисляется оценка (5); 3) решается уравнение (7); 4) осуществляется алгоритм нахождения единственной оценки $\hat{T}_{MM}$; 5) вычисляются величины $\Delta\hat{T}_{МП} = (\hat{T}_{МП} - T)^2$, $\Delta\hat{T}_{MM} = (\hat{T}_{MM} - T)^2$, где T – истинное значение длительности мертвого времени, заданное на первом этапе расчета при осуществлении имитационного моделирования.

Для сравнения качества МП- и ММ-оценок проведен статистический эксперимент, состоящий из следующих этапов: 1) для заданного набора параметров λ_i , $i = 1, 2$, α , p , δ , T ед. времени осуществляется моделирование наблюдаемого потока событий для заданного T_m ед. времени (отдельный j -й эксперимент, $j = 1, 2, \dots$); 2) осуществляется расчет оценок $\hat{T}_{МП}^{(j)}$, $\hat{T}_{MM}^{(j)}$ для j -го эксперимента; 3) вычисляются величины $\Delta\hat{T}_{МП}^{(j)}$, $\Delta\hat{T}_{MM}^{(j)}$ для j -го эксперимента; 4) осуществляется повторение N раз ($j = \overline{1, N}$) шагов 1–3.

Результатом выполнения описанного алгоритма являются две выборки $(\hat{T}_{МП}^{(1)}, \hat{T}_{МП}^{(2)}, \dots, \hat{T}_{МП}^{(N)})$, $(\hat{T}_{MM}^{(1)}, \hat{T}_{MM}^{(2)}, \dots, \hat{T}_{MM}^{(N)})$, на основании которых вычисляются выборочные вариации полученных оценок:

$$\hat{V}_{МП} = (1/N) \sum_{j=1}^N \Delta\hat{T}_{МП}^{(j)}, \hat{V}_{MM} = (1/N) \sum_{j=1}^N \Delta\hat{T}_{MM}^{(j)}.$$

Путем сравнения значений выборочных вариаций устанавливается, какая из оценок при заданных параметрах лучше, какая хуже: если $\hat{V}_{МП} \leq \hat{V}_{MM}$, то МП-оценка лучше ММ-оценки, если наоборот, то ММ-оценка лучше МП-оценки. Отметим, что по определению МП-оценка при конечных T_m будет всегда смещенная ($\tau_{\min} > T$); ее несмещенность реализуется только в асимптотическом случае при $T_m \rightarrow \infty$.

Результаты статистического эксперимента приведены в табл. 1–8. В первой строке таблиц указана длительность имитационного моделирования T_m ($T_m = 10, 20, \dots, 50$ ед. времени в табл. 1–4; $T_m = 600, 700, \dots, 1000$ ед. времени в табл. 5–8). Во второй и третьей строках таблиц для каждой длительности имитационного моделирования T_m приведены численные значения для $\hat{V}_{МП}$ и $\hat{V}_{MM}$ соответственно. В четвертой строке таблиц для каждой длительности имитационного моделирования приведены численные значения разности $\hat{V}_{МП} - \hat{V}_{MM}$. Численные результаты во всех таблицах получены для $N = 100$.

Т а б л и ц а 1

Результаты статистического эксперимента
 $\lambda_1 = 1, \lambda_2 = 0,7, \alpha = 0,01, p = 0,02, \delta = 0,2, T = 0,4$

T_m	30	40	50	60	70
$\hat{V}_{MM}$	0,02186	0,00421	0,00369	0,00314	0,00253
$\hat{V}_{МП}$	0,02208	0,00447	0,00371	0,00314	0,00253
$\hat{V}_{МП} - \hat{V}_{MM}$	0,00022	0,00025	$2,62960 \cdot 10^{-5}$	0	0

Т а б л и ц а 2

Результаты статистического эксперимента
 $\lambda_1 = 0,7, \lambda_2 = 0,2, \alpha = 0,1, p = 0,2, \delta = 0,2, T = 1$

T_m	30	40	50	60	70
$\hat{V}_{MM}$	0,29872	0,09011	0,08394	0,07245	0,04156
$\hat{V}_{МП}$	0,29988	0,09965	0,10131	0,07568	0,04156
$\hat{V}_{МП} - \hat{V}_{MM}$	0,00115	0,00953	0,01737	0,00323	0

Т а б л и ц а 3

Результаты статистического эксперимента
 $\lambda_1 = 0,9, \lambda_2 = 0,2, \alpha = 0,1, p = 0,2, \delta = 0,2, T = 1$

T_m	30	40	50	60	70
$\hat{V}_{MM}$	0,47389	0,06118	0,05562	0,04330	0,04432
$\hat{V}_{МП}$	0,4759	0,06667	0,05884	0,04336	0,04432
$\hat{V}_{МП} - \hat{V}_{MM}$	0,00201	0,00549	0,00322	$6,38402 \cdot 10^{-5}$	0

Т а б л и ц а 4

Результаты статистического эксперимента
 $\lambda_1 = 0,7, \lambda_2 = 0,3, \alpha = 0,1, p = 0,2, \delta = 0,7, T = 0,4$

T_m	30	40	50	60	70
$\hat{V}_{MM}$	0,06944	0,05062	0,01856	0,01747	0,012561
$\hat{V}_{МП}$	0,06957	0,05288	0,01942	0,01747	0,012561
$\hat{V}_{МП} - \hat{V}_{MM}$	0,00012	0,00226	0,00085	0	0

Т а б л и ц а 5

Результаты статистического эксперимента
 $\lambda_1 = 1, \lambda_2 = 0,7, \alpha = 0,01, p = 0,02, \delta = 0,2, T = 0,4$

T_m	600	700	800	900	1 000
$\hat{V}_{MM}$	$3,4325 \cdot 10^{-5}$	0,0011	0,0017	$2,9796 \cdot 10^{-5}$	$9,2298 \cdot 10^{-5}$
$\hat{V}_{МП}$	$2,6118 \cdot 10^{-5}$	$2,0686 \cdot 10^{-5}$	$1,1477 \cdot 10^{-5}$	$1,0959 \cdot 10^{-5}$	$9,2298 \cdot 10^{-5}$
$\hat{V}_{МП} - \hat{V}_{MM}$	$-8,2068 \cdot 10^{-6}$	-0,001	-0,0017	$-1,8837 \cdot 10^{-5}$	0

Т а б л и ц а 6

Результаты статистического эксперимента
 $\lambda_1 = 0,7, \lambda_2 = 0,2, \alpha = 0,1, p = 0,2, \delta = 0,2, T = 1$

T_m	600	700	800	900	1 000
$\hat{V}_{MM}$	0,0094	0,0327	0,0116	0,0376	0,0238
$\hat{V}_{МП}$	0,0002	0,0002	0,0001	0,0001	0,0001
$\hat{V}_{МП} - \hat{V}_{MM}$	-0,0091	-0,0325	-0,0115	-0,0375	-0,0237

Результаты статистического эксперимента

$$\lambda_1 = 0,9, \lambda_2 = 0,2, \alpha = 0,1, p = 0,2, \delta = 0,2, T = 1$$

T_m	600	700	800	900	1 000
$\hat{V}_{MM}$	0,01542	0,0354	0,05089	0,03118	0,06874
$\hat{V}_{МП}$	0,00017	0,0002	0,00011	0,00013	0,0001
$\hat{V}_{МП} - \hat{V}_{MM}$	-0,01524	-0,0352	-0,05078	-0,03105	-0,0686

Результаты статистического эксперимента

$$\lambda_1 = 0,7, \lambda_2 = 0,3, \alpha = 0,1, p = 0,2, \delta = 0,7, T = 0,4$$

T_m	600	700	800	900	1 000
$\hat{V}_{MM}$	0,00093	$9,5221 \cdot 10^{-5}$	0,0007	0,00146	0,00481
$\hat{V}_{МП}$	0,00012	$7,9655 \cdot 10^{-5}$	0,0001	$4,6246 \cdot 10^{-5}$	$5,4977 \cdot 10^{-5}$
$\hat{V}_{МП} - \hat{V}_{MM}$	-0,00081	$-1,5566 \cdot 10^{-5}$	-0,0006	-0,0014	-0,0047

Анализ приведенных численных результатов показывает что: 1) при малых временах наблюдения за потоком (при малых $T_m = 30, 40, \dots, 70$ ед. времени) ММ-оценки лучше МП-оценок (табл. 1–4) либо, по крайней мере, не хуже МП-оценок (когда разность $\hat{V}_{МП} - \hat{V}_{MM}$ равна 0), что является вполне естественным, так как при малых временах наблюдения оценка $\hat{T}_{МП}$ может быть достаточно сильно смещенной относительно T ; 2) при больших временах наблюдения за потоком (при больших $T_m = 600, 700, \dots, 1\,000$ ед. времени) МП-оценки лучше ММ-оценок (табл. 5–8) либо не хуже ММ-оценок (когда разность $\hat{V}_{МП} - \hat{V}_{MM}$ равна 0), что также является естественным, так как при больших временах наблюдения смещение оценки $\hat{T}_{МП}$ относительно T уменьшается.

Заключение

Результаты проведенного исследования МП-оценок и ММ-оценок длительности мертвого времени T показывают общую тенденцию, что при малых временах наблюдения за потоком предпочтительнее применять оценку $\hat{T}_{MM}$, при больших временах наблюдения – оценку $\hat{T}_{МП}$. Границу применимости той или иной оценки (при заданных значениях параметров $\lambda_i, i = 1, 2, \alpha, p, \delta$) можно определить только численно путем имитационного моделирования.

ЛИТЕРАТУРА

1. Горцев А.М., Калягин А.А., Нежелская Л.А. Оптимальная оценка состояний обобщенного полусинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2(11). С. 66–81.
2. Горцев А.М., Калягин А.А. Оптимальная оценка состояний обобщенного полусинхронного потока событий в условиях непродлевающего мертвого времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 4(13). С. 50–60.
3. Горцев А.М., Калягин А.А. Совместная плотность вероятностей длительности интервалов обобщенного полусинхронного потока // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 2(19). С. 80–87.
4. Горцев А.М., Калягин А.А., Нежелская Л.А. Совместная плотность вероятностей длительности интервалов обобщенного полусинхронного потока событий при непродлевающем мертвом времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. № 2(27). С. 19–29.

- 5 Горцев А.М., Калягин А.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном полусинхронном потоке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 1(30). С. 27–37.
6. Дудин А.Н., Клименок В.И. Системы массового обслуживания с коррелированными потоками. Минск : Изд-во БГУ, 2000. 175 с.
7. Горцев А.М., Нежелская Л.А. О связи МС-потоков и МАР-потоков событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 1(14). С. 13–21.
- 8 Нежелская Л.А. Совместная плотность вероятностей длительности интервалов модулированного МАР-потока событий и условия рекуррентности потока // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 1(30). С. 57–67.
9. Горцев А.М., Назаров А.А., Терпугов А.Ф. Управление и адаптация в системах массового обслуживания. Томск : Изд-во ТГУ, 1978. 208 с.
10. Горцев А.М., Нежелская Л.А., Соловьев А.А. Оптимальная оценка состояний МАР-потока событий в условиях непродлевающегося мертвого времени // Автоматика и телемеханика. 2012. № 8. С. 49–63.
11. Горцев А.М., Нежелская Л.А., Шевченко Т.Н. Оценка состояний МС-потока событий при наличии ошибок измерений // Известия высших учебных заведений. Физика. 1993. № 12. С. 67–85.
12. Бушланов И.В., Горцев А.М., Нежелская Л.А. Оценка параметров синхронного дважды стохастического потока событий // Автоматика и телемеханика. 2008. № 9. С. 76–93.
13. Горцев А.М., Нежелская Л.А. Оценивание длительности «мертвого времени» и интенсивностей синхронного дважды стохастического потока событий // Радиотехника. 2004. №10. С. 8–16.
14. Апанасович В.В., Коляда А.А., Чернявский А.Ф. Статистический анализ случайных потоков в физическом эксперименте. Минск : Университетское, 1988. 254 с.
15. Горцев А.М., Нежелская Л.А. Полусинхронный дважды стохастический поток событий при продлеваемом мертвом времени // Вычислительные технологии. 2008. Т. 13, № 1. С. 31–41.
16. Горцев А.М., Нежелская Л.А. Оценивание периода мертвого времени и параметров полусинхронного дважды стохастического потока событий // Измерительная техника. 2003. № 6. С. 7–13.
17. Горцев А.М., Нежелская Л.А. Оценивание параметров полусинхронного дважды стохастического потока событий методом моментов // Вестник Томского государственного университета. 2002. № 1 (I). С. 18–23.
18. Нежелская Л.А. Оптимальное оценивание состояний полусинхронного потока событий в условиях его частичной наблюдаемости // Вестник Томского государственного университета. 2000. № 269. С. 95–98.
19. Леонова М.А., Нежелская Л.А. Оценка длительности непродлевающегося мертвого времени в обобщенном асинхронном потоке событий // Известия вузов. Физика. 2013. Т. 56, № 9/2. С. 220–222.
20. Леонова М.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2 (23). С. 54–63.
21. Горцев А.М., Леонова М.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 4 (25). С. 32–42.
22. Шулягин В.П. Математическая статистика. Часть 1. Томск: Изд-во НТЛ, 2012. 540 с.

Калягин Алексей Андреевич. E-mail: redall@inbox.ru

Нежелская Людмила Алексеевна, кандидат технических наук, доцент. E-mail: ludne@mail.ru
Томский государственный университет

Поступила в редакцию 12 мая 2015 г.

Kalyagin Aleksey A., Nezhel'skaya Lyudmila A. (Tomsk State University, Russian Federation).

The comparison of maximum likelihood method and moments method by estimation of dead time in a generalized semisynchronous flow of events.

Keywords: generalized semisynchronous flow of events; unprolonging dead time; maximum likelihood method; method of moments, dead time.

DOI 10.17223/19988605/32/3

We consider generalized semisynchronous flow of events, which intensity is piecewise constant stochastic process $\lambda(t)$ with two states λ_1 and λ_2 ($\lambda_1 > \lambda_2$) and unprolonging dead time. During the time interval when $\lambda(t) = \lambda_i$, the Poisson flow of events takes place with the intensity λ_i , $i = 1, 2$. The transition from the first state of process $\lambda(t)$ into the second one (from the second state into the first one) is carried out at any moment. The sojourn time in the i -th state is exponentially distributed with the parameter α_i , $i = 1, 2$. The transition of $\lambda(t)$ from the first state into the second one initiates with probability p ($0 \leq p \leq 1$) extra event in the second state. Also, the transition of $\lambda(t)$ from the second state into the first one initiates with probability q ($0 \leq q \leq 1$) extra event in the second state.

The flow is functioning in conditions of unprolonging dead time (the value of dead time is fixed). We solve the problem of estimation of dead time by using the likelihood function and the method of moments.

REFERENCES

1. Gortsev, A.M., Kalyagin, A.A. & Nezhel'skaya, L.A. (2011) Optimal states estimation of integrated semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2 (11). pp. 66-81. (In Russian).
2. Gortsev, A.M. & Kalyagin, A.A. (2010) Optimal states estimation of integrated semisynchronous flow of events in condition of constant dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4 (13). pp. 50-60. (In Russian).
3. Gortsev, A.M. & Kalyagin, A.A. (2012) The joint probability density of duration of the intervals in a generalized semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2 (19). pp. 80-87. (In Russian).
4. Gortsev, A.M., Kalyagin, A.A. & Nezhel'skaya, L.A. (2014) The joint probability density of duration of the intervals in a generalized semisynchronous flow of events with unprolonging dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2 (27). pp. 19-29. (In Russian).
5. Gortsev, A.M., Kalyagin, A.A. & Nezhel'skaya, L.A. (2015) Maximum likelihood estimation of dead time value at a generalized semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1 (30). pp. 27-37. (In Russian).
6. Dudin, A.N. & Klimenok, V.I. (2000) *Sistemy massovogo obsluzhivaniya s korrelirovannymi potokami* [Queueing systems with correlated flows]. Minsk: BGU.
7. Gortsev, A.M. & Nezhel'skaya, L.A. (2011) On connection of MC flows and MAP flows of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(14). pp. 13-21. (In Russian).
8. Nezhel'skaya, L.A. (2015) Joint probability density of the intervals duration in modulated MAP event flows and its recurrence conditions. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 57-67. (In Russian).
9. Gortsev, A.M., Nazarov, A.A. & Terpugov, A.F. (1978) *Upravlenie i adaptatsiya v sistemakh massovogo obsluzhivaniya* [Control and adaptation in queueing systems]. Tomsk: Tomsk State University.
10. Gortsev, A.M., Nezhel'skaya, L.A. & Solov'ev, A.A. (2012) Optimal'naya otsenka sostoyaniy MAR-potoka sobytiy v usloviyakh neprodle-vayushchegosya mertvogo vremeni [Optimal states estimation of MAP flows of events in the condition of constant dead time]. *Avtomatika i telemekhanika – Automation and Remote Control*. 8. pp. 49-63.
11. Gortsev, A.M., Nezhel'skaya, L.A. & Shevchenko, T.N. Otsenka sostoyaniy MS-potoka sobytiy pri nalichii oshibok izmereniy [State estimation MS – stream sobytiy pri presence of measurement error]. *Izvestiya vysshikh uchebnykh zavedeniy. Fizika – Russian Physics Journal*. 12. pp. 67-85.
12. Bushlanov, I.V., Gortsev, A.M. & Nezhel'skaya, L.A. (2008) Otsenka parametrov sinkhronnogo dvazhdy stokhasticheskogo potoka sobytiy [Estimation of parameters of synchronous double-stochastic event flow]. *Avtomatika i telemekhanika – Automation and Remote Control*. 9. pp. 76-93.
13. Gortsev, A.M. & Nezhel'skaya, L.A. (2004) Otsenivanie dlitel'nosti «mertvogo vremeni» i intensivnostey sinkhronnogo dvazhdy sto-khasticheskogo potoka sobytiy [Estimation of the dead time period and intensities of the synchronous double-stochastic event flow]. *Radiotekhnika*. 10. pp. 8-16.
14. Apanasovich, V.V., Kolyada, A.A. & Chernyavskiy, A.F. (1988) *Staticheskiy analiz sluchaynykh potokov v fizicheskoy eksperimente* [Statistical analysis of stochastic flows in physical experiment]. Minsk: Universitetskoe.
15. Gortsev, A.M. & Nezhel'skaya, L.A. (2008) Semisynchronous double stochastic flow of events when the dead time is prolonged. *Vychislitel'nye tekhnologii – Computational Technologies*. 13 (1). pp. 31-41. (In Russian).
16. Gortsev, A.M. & Nezhel'skaya, L.A. (2003) Otsenivanie perioda mertvogo vremeni i parametrov polusinkhronnogo dvazhdy stokhasticheskogo potoka sobytiy [Evaluation period of the dead time and the parameters of a semi-synchronous double stochastic flow of events]. *Izmeritel'naya tekhnika – Measurement Techniques*. 6. pp. 7-13.
17. Gortsev, A.M. & Nezhel'skaya, L.A. (2002) Otsenivanie parametrov polusinkhronnogo dvazhdy stokhasticheskogo potoka sobytiy metodom momentov [Estimating parameters of the semi-synchronous double stochastic flow of events by moment method]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 1 (I). pp. 18-23.
18. Nezhel'skaya, L.A. (2000) Optimal'noe otsenivanie sostoyaniy polusinkhronnogo potoka sobytiy v usloviyakh ego chastichnoy nablyu-daemosti [Optimal state estimation of the semi-synchronous flow of events under incomplete observability]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 269. pp. 95-98.
19. Leonova, M.A. & Nezhel'skaya, L.A. (2013) Otsenka dlitel'nosti neprodlevayushchegosya mertvogo vremeni v obobshchennom asinkhronnom potoke sobytiy [Estimating duration of non-continuous dead time in a generalized asynchronous event flow]. *Izvestiya vysshikh uchebnykh zavedeniy. Fizika – Russian Physics Journal*. 56 (9/2). pp. 220-222.
20. Leonova, M.A. & Nezhel'skaya, L.A. (2013) Maximum-likelihood estimation of the dead time in a generalized asynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2 (23). pp. 54-63. (In Russian).
21. Gortsev, A.M., Leonova, M.A. & Nezhel'skaya, L.A. (2013) Comparison of MP- and MM-grade duration of dead time in a generalized asynchronous event flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4 (25). pp. 32-42. (In Russian).
22. Shulenin, V.P. (2012) *Matematicheskaya statistika* [Mathematical statistics]. Tomsk: NTL.

M.I. Kusainov

**RISK EFFICIENCY OF ADAPTIVE ONE-STEP PREDICTION
OF AUTOREGRESSION WITH PARAMETER DRIFT**

A scalar stable autoregressive process with the dynamic parameter corrupted by additive noise is studied. The model parameters are assumed to be unknown. The truncated estimators of the parameters are used to build adaptive one-step predictors. The associated problem is to minimize a risk function of special form, defined to account for sample size and squared prediction error's sample mean. A sequential procedure is introduced to achieve the minimal risk.

Keywords: adaptive predictors; asymptotic risk efficiency; optimal sample size; scalar autoregression; stopping time; truncated parameter estimators.

When studying dynamic systems, the identification problem is often a major one to consider. A system's model is assumed to incorporate unknown parameters, estimation of which is vital for later research. Classic methods, such as maximum likelihood estimation, least squares fitting, etc., have known asymptotic properties. However, infinite samples do not occur in reality and efforts are being made to achieve non-asymptotic qualities of estimators. E.g., one may consider confidence regions for finite number of measurements (see [1, 2] among others).

Another approach are methods that employ sequential analysis. Among such methods is sequential estimation method (see, e.g., [3–12]), which has guaranteed accuracy by samples of random and finite, albeit unbounded, size. Its idea was further developed into the truncated sequential estimation method (see, e.g., [12–15]), which utilizes samples of bounded random size.

Recently, the truncated estimation method was suggested in [16] as a modification of the truncated sequential estimation method. Truncated estimators were constructed for ratio type functionals and only need samples of fixed non-random size to achieve guaranteed accuracy in the sense of the L_{2m} -norm, $m \geq 1$.

One possible use for estimated parameters is to predict future values of the modeled random process by existing observations. In order to control both the quality of predictions and the required sample size, a loss function dependent on the two is introduced. A risk efficiency problem arises, where the expected loss is minimized by choosing a certain duration of observations. Similar problems for autoregressive processes were examined in [17] and [18], the least squares estimators and sequential estimators of unknown parameters were used.

In this paper we consider a scalar stable AR(1) with parameter drift and construct real-time predictors based upon the truncated estimators of the parameters. All the parameters are assumed to be unknown. A similar model was studied in [14]. Sequential approach was, for the first time, applied to it to good effect. Resulting estimators were shown to have preassigned mean square accuracy and uniform in parameter asymptotic normality. These estimators, however, had different form than in this paper due to some parameters being known. We solve the optimization problem associated with the loss function of a special form. The proposed procedure is shown to be asymptotically risk efficient as the cost of prediction error tends to infinity. The simulation results confirm the result, but are not included for editorial reasons.

The scalar case without the drift was considered in [19], multivariate AR(1) in [20] and ARMA(1,1) in [21].

1. Problem statement

Consider the stable scalar autoregressive process satisfying the equation

$$x_k = \lambda_{k-1}x_{k-1} + \xi_k, \quad k \geq 1, \quad (1)$$

where

$$\lambda_k = \lambda + \eta_k, \quad k \geq 0,$$

the parameter λ is unknown, the condition $Ex_0^2 < \infty$ holds, ξ_k and η_k are sequences of independent identically distributed (i.i.d.) zero mean random variables, that are also independent of each other, with finite variances $\sigma_\xi^2 = E\xi_1^2, \sigma_\eta^2 = E\eta_1^2$. In addition, to guarantee stability of the process (1) we assume the following

$$\lambda^2 + \sigma_\eta^2 < 1. \quad (2)$$

It is known that the optimal in the mean square sense one-step predictor is the conditional expectation of the process with respect to its past, i.e.

$$x_k^{\text{opt}} = \lambda x_{k-1}, \quad k \geq 1.$$

Therefore, one needs an estimator $\hat{\lambda}_k$ for the unknown parameter λ to construct the adaptive predictors of the form

$$\hat{x}_k = \hat{\lambda}_{k-1} x_{k-1}, \quad k \geq 1. \quad (3)$$

Write the corresponding prediction errors

$$\hat{e}_k = x_k - \hat{x}_k = (\lambda - \hat{\lambda}_{k-1})x_{k-1} + \eta_{k-1}x_{k-1} + \xi_k.$$

Let e_n^2 denote the sample mean of squared prediction error

$$e_n^2 = \frac{1}{n} \sum_{k=1}^n \hat{e}_k^2. \quad (4)$$

Define the loss function

$$L_n = \frac{A}{n} e_n^2 + n.$$

One way the parameter $A(>0)$ can be interpreted is being the cost of prediction error.

The corresponding risk function

$$R_n = E_\theta L_n = \frac{A}{n} E_\theta e_n^2 + n, \quad (5)$$

E_θ denotes expectation under the distribution P_θ with the given parameter $\theta = (\lambda, \sigma_\xi^2, \sigma_\eta^2)$. Define

$\Theta = \{\theta : \lambda^2 + \sigma_\eta^2 < 1, \sigma_\xi^2 < \infty\}$ the process' stability parameter region.

The main aim is to minimize the risk R_n on the sample size n .

2. Main result

To solve the stated problem we shall use the truncated estimation method introduced in [16]. This method makes it possible to obtain the ratio type estimators with guaranteed accuracy using a sample of fixed size. According to the method, the truncated estimator of the autoregressive parameter λ is based on a ratio type estimator, the least-squares type in this case

$$\hat{\lambda}_k = \frac{\sum_{i=1}^k x_{i-1} x_i}{\sum_{i=1}^k x_{i-1}^2}, \quad k \geq 1, \quad (6)$$

and has the form

$$\tilde{\lambda}_k = \hat{\lambda}_k \chi(\bar{\Delta}_k \geq H_k), \quad k \geq 1, \quad (7)$$

where $\bar{\Delta}_k = \frac{1}{k} \sum_{i=1}^k x_{i-1}^2$, the notation $\chi(B)$ means the indicator function of the set B and

$$H_k = \log^{-1/2}(k+1). \quad (8)$$

It should be noted, that according to [16], H_k can be taken as any decreasing slowly changing positive function.

A model similar to (1) was studied in [14], but variances of the noises ξ_i and η_i , $i \geq 1$ were assumed to be known. This information allowed construction of true sequential least-squares estimators. However, it is absent in our case, forcing one to use estimators of the form (7).

Rewrite the formulae (3)–(5) with $\hat{\lambda}_k$ replaced by $\tilde{\lambda}_k$ as follows

$$\begin{aligned}\tilde{x}_k &= \tilde{\lambda}_{k-1} x_{k-1}, \\ \tilde{e}_k &= (\lambda - \tilde{\lambda}_{k-1}) x_{k-1} + \eta_{k-1} x_{k-1} + \xi_k, \\ \overline{e_n^2} &= \frac{1}{n} \sum_{k=1}^n \tilde{e}_k^2, \\ R_n &= \frac{A}{n} E_\theta \overline{e_n^2} + n.\end{aligned}\tag{9}$$

$$\tag{10}$$

To minimize the risk R_n we rewrite the risk function (10) using the definition of $\overline{e_n^2}$

$$R_n = \frac{A}{n} (\sigma_\xi^2 + \sigma_\eta^2 \sigma_x^2 + D_n) + n,\tag{11}$$

where

$$\sigma_x^2 = \frac{\sigma_\xi^2}{1 - (\lambda^2 + \sigma_\eta^2)}, \quad D_n = \frac{1}{n} \sum_{k=1}^n E_\theta (x_k^{\text{opt}} - \tilde{x}_k)^2 + \frac{\sigma_\eta^2}{n} \sum_{k=1}^n (\sigma_{x,k-1}^2 - \sigma_x^2), \quad \sigma_{x,k}^2 = E_\theta x_k^2.$$

From here on C denotes those non-negative constants, the values of which are not critical.

Further we show that

$$D_n = o(1) \quad \text{as } n \rightarrow \infty.\tag{12}$$

To this end we establish the two following estimates

$$\sum_{k=1}^n |\sigma_{x,k-1}^2 - \sigma_x^2| \leq C,\tag{13}$$

$$\sum_{k=1}^n E_\theta (x_k^{\text{opt}} - \tilde{x}_k)^2 \leq C \log^2 n.\tag{14}$$

In order to prove (13), write the solution of x_k using its definition (1) as follows

$$x_k = \sum_{i=0}^{k-1} \xi_{k-i} \prod_{j=1}^i (\lambda + \eta_{k-j}) + x_0 \prod_{j=1}^k (\lambda + \eta_{k-j}).$$

Squaring and taking the expectation of both sides, we use independence and zero mean of ξ_i and η_i , $i \geq 1$ to obtain

$$\begin{aligned}\sigma_{x,k}^2 &= \sum_{i=0}^{k-1} \sigma_\xi^2 \prod_{j=1}^i E(\lambda + \eta_{k-j})^2 + E x_0^2 \prod_{j=1}^k E(\lambda + \eta_{k-j})^2 = \\ &= \sigma_\xi^2 \sum_{i=0}^{k-1} (\lambda^2 + \sigma_\eta^2)^i + E x_0^2 (\lambda^2 + \sigma_\eta^2)^k.\end{aligned}$$

Using (2) one gets

$$\sigma_\xi^2 \sum_{i=0}^{k-1} (\lambda^2 + \sigma_\eta^2)^i = \frac{\sigma_\xi^2}{1 - (\lambda^2 + \sigma_\eta^2)} (1 - (\lambda^2 + \sigma_\eta^2)^k) = \sigma_x^2 (1 - (\lambda^2 + \sigma_\eta^2)^k)$$

and thus,

$$\sigma_{x,k}^2 - \sigma_x^2 = -\sigma_x^2 (\lambda^2 + \sigma_\eta^2)^{k-1} + E x_0^2 (\lambda^2 + \sigma_\eta^2)^k.$$

From (2) it follows, that

$$\sum_{k=1}^{\infty} |\sigma_{x,k-1}^2 - \sigma_x^2| \leq C,$$

hence (13).

To prove (14) rewrite its left-hand side

$$\sum_{k=1}^n E_{\theta}(x_k^{opt} - \tilde{x}_k)^2 = \sum_{k=1}^n E_{\theta}(\tilde{\lambda}_{k-1} - \lambda)^2 x_{k-1}^2. \quad (15)$$

We establish the properties of the estimators $\tilde{\lambda}_k$.

Define $k_0 = \max \left\{ 1, [e^{(\sigma_x^2)^{-2}}]_1 \right\}$, where $[a]_1$ denotes the integer part of a .

Lemma 1. Assume the model (1) and let for some integer $m \geq 1$ the conditions

$$E\xi_1^{4m} < \infty, \quad Ex_0^{4m} < \infty, \quad E(\lambda + \eta_1)^{4m} < 1 \quad (16)$$

be true. Then the truncated estimators $\tilde{\lambda}_k$ satisfy

(i) for $1 \leq k < k_0$

$$E_{\theta}(\tilde{\lambda}_k - \lambda)^{2m} \leq C; \quad (17)$$

(ii) for $k \geq k_0$

$$E_{\theta}(\tilde{\lambda}_k - \lambda)^{2m} \leq \frac{C \log^m k}{k^m}. \quad (18)$$

The proof of Lemma 1 is presented in Section 4.

Remark 1. The parameter H_k can be taken as a constant H , provided that $H \in (0, \sigma_{\xi}^2)$. The condition (2) then guarantees $H < \lim_{k \rightarrow \infty} \bar{\Delta}_k$, considering

$$\lim_{k \rightarrow \infty} \bar{\Delta}_k = \frac{\sigma_{\xi}^2}{1 - (\lambda^2 + \sigma_{\eta}^2)} \quad P_{\theta} - \text{a.s.}, \quad (19)$$

which follows from the ergodicity of the process $(x_k)_{k \geq 0}$ (see, e.g., [7]). Then the estimators $\tilde{\lambda}_k$ would satisfy

$$E_{\theta}(\tilde{\lambda}_k - \lambda)^{2m} \leq \frac{C}{k^m}$$

for every $k \geq 1$, which can be proved similarly to Theorem 2 of [16]. This, though, requires knowledge of σ_{ξ}^2 .

The Cauchy-Schwarz-Bunyakovsky inequality, (15) and (18) yield (14). The relation (12) follows directly from (13), (14).

Denote

$$\sigma^2 = \sigma_{\xi}^2 + \sigma_{\eta}^2 \sigma_x^2. \quad (20)$$

In view of (12) we minimize the principal term of R_n , analogously to [17]

$$R_n \approx \frac{A}{n} \sigma^2 + n \longrightarrow \min_n$$

to get the optimal sample size

$$n_A^o = A^{1/2} \sigma \quad (21)$$

and the corresponding approximate minimal risk value

$$R_{n_A^o} = 2A^{1/2} \sigma + O(\log^2 A) \quad \text{as } A \rightarrow \infty. \quad (22)$$

Similarly to [17, 18, 20], we introduce the stopping time T_A as an estimator of n_A^o , replacing σ^2 in its definition with an estimator $\tilde{\sigma}_n^2$

$$T_A = \inf_{n \geq n_A} \left\{ n \geq A^{1/2} \tilde{\sigma}_n \right\}, \quad (23)$$

where n_A is the initial sample size depending on A and specified below (see Theorem 1),

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \tilde{\lambda}_n x_{k-1})^2. \quad (24)$$

We formulate a theorem to prove the asymptotic equivalence of T_A and n_A^o in the sense of almost sure and mean convergences (see respectively (27), (28) below) and the optimality of the adaptive prediction procedure in the sense of equivalence of $R_{n_A^o}$ and the modified risk

$$R_A = E_\theta L_{T_A} = AE_\theta \frac{1}{T_A} \overline{e_{T_A}^2} + E_\theta T_A, \quad (25)$$

see (29).

Theorem 1. Assume that

$$E\xi_1^{16} < \infty, \quad Ex_0^{16} < \infty, \quad E(\lambda + \eta_1)^{16} < 1 \quad (26)$$

and n_A in (23) is such that

$$n_A \geq \max\{k_0, A^r \log^2 A\}, \quad n_A \cdot A^{-1/2} \xrightarrow{A \rightarrow \infty} 0$$

with $r \in (2/5, 1/2)$. Let the predictors $\tilde{x}_k$ be defined by (9) and the risk functions defined by (5), (25). Then for every $\theta \in \Theta$ and $\sigma^2 > 0$

$$\frac{T_A}{n_A^o} \xrightarrow{A \rightarrow \infty} 1 \quad P_\theta - \text{a.s.}, \quad (27)$$

$$\frac{E_\theta T_A}{n_A^o} \xrightarrow{A \rightarrow \infty} 1, \quad (28)$$

$$\frac{R_A}{R_{n_A^o}} \xrightarrow{A \rightarrow \infty} 1. \quad (29)$$

The proof of Theorem 1 is presented in Section 2.

3. Proofs

3.1. Proof of Lemma 1

It can be shown (see, e.g., [22], Lemma 1), that to guarantee $E_\theta x_k^{2m} \leq C$, $k, m \geq 1$, the following suffices

$$E\xi_1^{2m} < \infty, \quad Ex_0^{2m} < \infty, \quad E(\lambda + \eta_1)^{2m} < 1.$$

Thus, from the conditions of Lemma 1 on noise moments for $\theta \in \Theta$ it follows

$$\sup_{k \geq 0} E_\theta x_k^{4m} \leq C. \quad (30)$$

By the definition (7) of truncated estimators $\tilde{\lambda}_k$, their deviation has the form

$$\tilde{\lambda}_k - \lambda = (\hat{\lambda}_k - \lambda) \cdot \chi(\bar{\Delta}_k \geq H_k) - \lambda \cdot \chi(\bar{\Delta}_k < H_k). \quad (31)$$

The definitions (6) and (1) yield

$$\hat{\lambda}_k - \lambda = \frac{\sum_{i=1}^k x_{i-1} \xi_i + \sum_{i=1}^k \eta_{i-1} x_{i-1}^2}{\sum_{i=1}^k x_{i-1}^2},$$

Hence, from (31)

$$(\tilde{\lambda}_k - \lambda)^{2m} = \frac{\left(\sum_{i=1}^k x_{i-1} \xi_i + \sum_{i=1}^k \eta_{i-1} x_{i-1}^2 \right)^{2m}}{\left(\sum_{i=1}^k x_{i-1}^2 \right)^{2m}} \chi(\bar{\Delta}_k \geq H_k) + \lambda^{2m} \cdot \chi(\bar{\Delta}_k < H_k).$$

From the definition of $\bar{\Delta}_k$ (see (7)) and H_k (8) it follows

$$E_\theta(\tilde{\lambda}_k - \lambda)^{2m} \leq \frac{\log^m k}{k^{2m}} E_\theta \left(\sum_{i=1}^k x_{i-1} \xi_i + \sum_{i=1}^k \eta_{i-1} x_{i-1}^2 \right)^{2m} + \lambda^{2m} P_\theta(\bar{\Delta}_k < H_k). \quad (32)$$

It can be easily shown, that the sums $\sum_{i=1}^k x_{i-1} \xi_i$ and $\sum_{i=1}^k \eta_{i-1} x_{i-1}^2$ for $k \geq 1$ form martingales. Thus, by the Burkholder inequality and the Hölder inequality and (30), similarly to [16], Section 5.2, we get

$$\begin{aligned} \frac{\log^m k}{k^{2m}} E_\theta \left(\sum_{i=1}^k x_{i-1} \xi_i + \sum_{i=1}^k \eta_{i-1} x_{i-1}^2 \right)^{2m} &\leq \frac{2^{2m-1} \log^m k}{k^{2m}} \left(E_\theta \left(\sum_{i=1}^k x_{i-1} \xi_i \right)^{2m} + E_\theta \left(\sum_{i=1}^k \eta_{i-1} x_{i-1}^2 \right)^{2m} \right) \leq \\ &\leq \frac{C \log^m k}{k^{2m}} \left(E_\theta \left(\sum_{i=1}^k x_{i-1}^2 \xi_i^2 \right)^m + E_\theta \left(\sum_{i=1}^k \eta_{i-1}^2 x_{i-1}^4 \right)^m \right) \leq \frac{C \log^m k}{k^m}. \end{aligned} \quad (33)$$

The first assertion (17) of Lemma 1 follows from (32) and (33).

For the second summand of (6), using the Chebyshev inequality for $k \geq k_0$ one has

$$\lambda^{2m} P_\theta(\bar{\Delta}_k < H_k) \leq C P_\theta \left(\left| \bar{\Delta}_k - \sigma_x^2 \right| > \sigma_x^2 - H_k \right) \leq C \frac{E_\theta(\bar{\Delta}_k - \sigma_x^2)^{2m}}{(\sigma_x^2 - H_k)^{2m}}. \quad (34)$$

Note that for $k \geq k_0$,

$$H_k = \frac{1}{\sqrt{\log(k+1)}} < \frac{1}{\sqrt{\log e^{(\sigma_x^2)^{-2}}}} = \sigma_x^2$$

and hence the difference $\sigma_x^2 - H_k > 0$.

To estimate (34) we rewrite σ_x^2

$$\begin{aligned} \sigma_x^2 &= \frac{\sigma_\xi^2}{1 - (\lambda^2 + \sigma_\eta^2)} = \frac{\sigma_\xi^2(1 - \lambda^2)}{(1 - \lambda^2)(1 - (\lambda^2 + \sigma_\eta^2))} = \\ &= \frac{\sigma_\xi^2}{1 - \lambda^2} + \frac{\sigma_\xi^2 \sigma_\eta^2}{(1 - \lambda^2)(1 - (\lambda^2 + \sigma_\eta^2))} = \frac{1}{1 - \lambda^2} (\sigma_\xi^2 + \sigma_\eta^2 \sigma_x^2). \end{aligned}$$

Using this and the definition of the process (1), one can write $\bar{\Delta}_k - \sigma_x^2$ as follows

$$\begin{aligned} \bar{\Delta}_k - \sigma_x^2 &= \frac{1}{(1 - \lambda^2)} \left(\lambda^2 \cdot \frac{x_0^2 - x_k^2}{k} + \frac{2\lambda}{k} \sum_{i=1}^k \eta_{i-1} x_{i-1}^2 + \frac{2}{k} \sum_{i=1}^k \lambda_{i-1} \xi_i x_{i-1} + \right. \\ &\quad \left. + \frac{1}{k} \sum_{i=1}^k (\xi_i^2 - \sigma_\xi^2) + \frac{1}{k} \sum_{i=1}^k (\eta_{i-1}^2 x_{i-1}^2 - \sigma_\eta^2 \sigma_{x,i-1}^2) + \frac{1}{k} \sum_{i=1}^k \sigma_\eta^2 (\sigma_{x,i-1}^2 - \sigma_x^2) \right). \end{aligned}$$

Then (34), the Burkholder inequality, (30) and (13) yield

$$\lambda^{2m} P_\theta(\bar{\Delta}_k < H_k) \leq C \frac{1}{k^m}.$$

Together with (33) this proves the second assertion of Lemma 1.

3.1. Proof of Theorem 1

The conditions on noise moments (26) yield for $\theta \in \Theta$

$$\sup_{k \geq 0} E_\theta x_k^{16} \leq C. \quad (35)$$

Note that assuming the distribution of η_1 is symmetrical, the condition $E(\lambda + \eta_1)^{16} < 1$ reduces to

$$\lambda^{16} + 120(\lambda^{14} \sigma_\eta^2 + \lambda^2 \sigma_\eta^{14}) + 1820(\lambda^{12} \sigma_\eta^4 + \lambda^4 \sigma_\eta^{12}) + 8008(\lambda^{10} \sigma_\eta^6 + \lambda^6 \sigma_\eta^{10}) + 12870 \lambda^8 \sigma_\eta^8 + \sigma_\eta^{16} < 1,$$

where $\sigma_\eta^{2m} = E\eta_1^{2m}$, $m = \overline{2, 8}$.

Rewrite formula (24) for $\tilde{\sigma}_n^2$ using the definition of the process (1)

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k + \eta_{k-1}x_{k-1} + (\lambda - \tilde{\lambda}_n)x_{k-1})^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) + W_n + v_n, \quad (36)$$

where

$$W_n = \frac{(\tilde{\lambda}_n - \lambda)^2}{n} \sum_{k=1}^n x_{k-1}^2, \quad v_n = \frac{2}{n} \sum_{k=1}^n \xi_k \eta_{k-1} x_{k-1} - \frac{2}{n} \sum_{k=1}^n (\tilde{\lambda}_n - \lambda) x_{k-1} (\xi_k + \eta_{k-1} x_{k-1}).$$

Analogously to [17], we show that

$$\tilde{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{} \sigma^2 \quad P_\theta - \text{a.s.} \quad (37)$$

Consider W_n . Using the properties (18) and the Chebyshev inequality for any $\varepsilon > 0$ we get

$$P(|\tilde{\lambda}_n - \lambda| > \varepsilon) \leq \frac{1}{\varepsilon^4} E_\theta (\tilde{\lambda}_n - \lambda)^4 \leq C n^{-2} \log^2 n.$$

From the Borel-Cantelli lemma it follows that

$$\tilde{\lambda}_n \xrightarrow[n \rightarrow \infty]{} \lambda \quad P_\theta - \text{a.s.}$$

Together with (19) and (35) this yields

$$W_n \xrightarrow[n \rightarrow \infty]{} 0 \quad P_\theta - \text{a.s.} \quad (38)$$

Similar arguments can be used to show

$$v_n \xrightarrow[n \rightarrow \infty]{} 0 \quad P_\theta - \text{a.s.} \quad (39)$$

At the same time, strong law of large numbers, (13) and the Borel-Cantelli lemma yield

$$\frac{1}{n} \sum_{k=1}^n (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) \xrightarrow[n \rightarrow \infty]{} \sigma^2 \quad P_\theta - \text{a.s.} \quad (40)$$

Then (37) follows from the representation (36), (38)–(40).

From the definition (23) of T_A it follows that with P_θ -probability one $T_A \rightarrow \infty$ as $A \rightarrow \infty$. Therefore, by (11) we have $\tilde{\sigma}_{T_A}^2 \rightarrow \sigma^2$ P_θ -a.s. and hence

$$\frac{T_A}{A^{1/2} \sigma} \xrightarrow[A \rightarrow \infty]{} 1 \quad P_\theta - \text{a.s.}$$

To prove (28) we introduce for any positive A the auxiliary sequence of numbers $\gamma_{A,n}$

$$\gamma_{A,n} = n^2 A^{-1} \frac{1}{2 \log A}, \quad n \geq 1.$$

Denote

$$m_n = \frac{1}{n} \sum_{k=1}^n (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2 - \sigma^2).$$

Observe that m_n can be represented as a sum of two martingales and decaying to zero as $O(n^{-1})$ sequence

$$m_n = \frac{1}{n} \sum_{k=1}^n (\xi_k^2 - \sigma_\xi^2) + \frac{1}{n} \sum_{k=1}^n (\eta_{k-1}^2 x_{k-1}^2 - \sigma_\eta^2 \sigma_{x,k-1}^2) + \frac{1}{n} \sum_{k=1}^n \sigma_\eta^2 (\sigma_{x,k-1}^2 - \sigma_x^2).$$

By the definition of T_A and (36) we have

$$\begin{aligned} E_\theta T_A &\leq n_A + \sum_{n \geq n_A} P_\theta \left(n^2 A^{-1} \leq \frac{1}{n} \sum_{k=1}^n (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) + W_n + v_n \right) \leq \\ &\leq n_A + \sum_{n \geq n_A} \left\{ P_\theta \left(n^2 A^{-1} \leq \sigma^2 + 2\gamma_{A,n} \right) + P_\theta(|v_n| > \gamma_{A,n} / 2) + P_\theta(W_n > \gamma_{A,n} / 2) + P_\theta(|m_n| > \gamma_{A,n}) \right\}. \end{aligned} \quad (41)$$

It can be shown, analogously to [19], that for $n \geq n_A$

$$P_\theta(|v_n| > \gamma_{A,n} / 2) + P_\theta(W_n > \gamma_{A,n} / 2) + P_\theta(|m_n| > \gamma_{A,n}) \leq C \gamma_{A,n}^{-2} n^{-1} = 4CA^2 \log^2 A \cdot n^{-5}$$

and at the same time

$$\frac{n_A + \sum_{n \geq n_A} P_\theta(n^2 A^{-1} \leq \sigma^2 + 2\gamma_{A,n})}{A^{1/2} \sigma} \xrightarrow{A \rightarrow \infty} 1. \quad (42)$$

Therefore, by assumptions on n_A

$$\begin{aligned} & A^{-1/2} \sum_{n \geq n_A} \left\{ P_\theta(|v_n| > \gamma_{A,n} / 2) + P_\theta(W_n > \gamma_{A,n} / 2) + P_\theta(|m_n| > \gamma_{A,n}) \right\} \leq \\ & \leq C A^{3/2} \log^2 A \sum_{n \geq n_A} n^{-5} \leq C A^{3/2} \log^2 A \cdot n_A^{-4} \leq C A^{-\frac{8r-3}{2}} \log^{-6} A \xrightarrow{A \rightarrow \infty} 0. \end{aligned} \quad (43)$$

Then from (41) – (43) it follows that

$$\lim_{A \rightarrow \infty} \frac{E_\theta T_A}{A^{1/2} \sigma} \leq 1. \quad (44)$$

Analogously it can be shown that

$$\lim_{A \rightarrow \infty} \frac{E_\theta T_A}{A^{1/2} \sigma} \geq 1$$

and thus, in view of (44) the assertion (28) holds.

To prove (29) we need the following properties

$$\begin{aligned} P_\theta(T_A < N') &= O(A^{-r}), \quad P_\theta(T_A > N'') = O(A^{-1}), \\ N' &= [(\sigma - \varepsilon) A^{1/2}]_1, \quad N'' = [(\sigma + \varepsilon) A^{1/2}]_1 + 1, \quad 0 < \varepsilon < \sigma, \end{aligned} \quad (45)$$

which can be established similarly to (4.31) of [19].

Rewrite the left-hand side of (29) using (22) and (25)

$$\frac{R_A}{R_{n_A^o}} = \frac{A E_\theta \frac{1}{T_A} \overline{e_{T_A}^2} + E_\theta T_A}{2 A^{1/2} \sigma + O(\log^2 A)}. \quad (46)$$

From (28) and (46) it follows that to prove (29) it suffices to show the convergence

$$A^{1/2} E_\theta \frac{1}{T_A \sigma} \overline{e_{T_A}^2} \xrightarrow{A \rightarrow \infty} 1. \quad (47)$$

To this end we show that

$$A^{1/2} E_\theta \frac{1}{T_A} \overline{e_{T_A}^2} \chi(T_A < N') \xrightarrow{A \rightarrow \infty} 0, \quad A^{1/2} E_\theta \frac{1}{T_A} \overline{e_{T_A}^2} \chi(T_A > N'') \xrightarrow{A \rightarrow \infty} 0, \quad (48)$$

$$A^{1/2} E_\theta \frac{1}{T_A \sigma} \overline{e_{T_A}^2} \chi(N' \leq T_A \leq N'') \xrightarrow{A \rightarrow \infty} 1. \quad (49)$$

All three relations are proved analogously to (4.39)–(4.41) of [19] using the definition of $\overline{e_n^2}$. E.g., for (49) we write

$$\begin{aligned} & A^{1/2} E_\theta \frac{1}{T_A} \overline{e_{T_A}^2} \chi(N' \leq T_A \leq N'') = A^{1/2} E_\theta \frac{1}{T_A^2} \sum_{k=1}^{T_A} (\lambda - \tilde{\lambda}_{k-1})^2 x_{k-1}^2 \chi(N' \leq T_A \leq N'') + \\ & + 2 A^{1/2} E_\theta \frac{1}{T_A^2} \sum_{k=1}^{T_A} \left(\xi_k \eta_{k-1} x_{k-1} - (\tilde{\lambda}_{k-1} - \lambda) \xi_k x_{k-1} - (\tilde{\lambda}_{k-1} - \lambda) \eta_{k-1} x_{k-1}^2 \right) \chi(N' \leq T_A \leq N'') + \\ & + A^{1/2} E_\theta \frac{1}{T_A^2} \sum_{k=1}^{T_A} (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) \chi(N' \leq T_A \leq N''). \end{aligned} \quad (50)$$

By the definitions of N' and N'' , the Cauchy-Schwarz-Bunyakovsky inequality and Lemma 1, for the first summand one gets

$$A^{1/2} E_\theta \frac{1}{T_A^2} \sum_{k=1}^{T_A} (\lambda - \tilde{\lambda}_{k-1})^2 x_{k-1}^2 \chi(N' \leq T_A \leq N'') \leq C A^{-1/2} \sum_{k=1}^{N''} E_\theta (\lambda - \tilde{\lambda}_{k-1})^2 x_{k-1}^2 \leq C A^{-1/2} \log^2 A \xrightarrow{A \rightarrow \infty} 0.$$

For the second summand of (50) the Doob's maximal inequality for martingales (see, e.g., [3]) and the Cauchy-Schwarz-Bunyakovsky inequality can be used to show

$$A^{1/2} E_0 \frac{1}{T_A^2} \sum_{k=1}^{T_A} \left(\xi_k \eta_{k-1} x_{k-1} - (\tilde{\lambda}_{k-1} - \lambda) \xi_k x_{k-1} - (\tilde{\lambda}_{k-1} - \lambda) \eta_{k-1} x_{k-1}^2 \right) \chi(N' \leq T_A \leq N'') \leq CA^{-1/4} \xrightarrow{A \rightarrow \infty} 0.$$

Rewrite the left-hand side of the last summand of (50)

$$\begin{aligned} & A^{1/2} E_0 \frac{1}{T_A^2} \sum_{k=1}^{T_A} (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) \chi(N' \leq T_A \leq N'') = \\ & = A^{1/2} E_0 \frac{1}{T_A \sigma} m_{T_A} \chi(N' \leq T_A \leq N'') + A^{1/2} \sigma E_0 \frac{1}{T_A} \chi(N' \leq T_A \leq N''). \end{aligned}$$

We show that the first summand converges to 0 and the second one converges to 1. By (13) the Doob's maximal inequality and the Cauchy-Schwarz-Bunyakovsky inequality

$$\begin{aligned} & A^{1/2} E_0 \frac{1}{T_A \sigma} |m_{T_A}| \chi(N' \leq T_A \leq N'') \leq CA^{1/2} \frac{1}{(N')^2} \left(E_0 \max_{1 \leq n \leq N''} \left(\sum_{k=1}^n (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2 - (\sigma_\xi^2 + \sigma_\eta^2 \sigma_{x,k-1}^2)) \right)^2 \right)^{1/2} \leq \\ & \leq CA^{-1/2} \left(\sum_{k=1}^{N''} E_0 (\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2 - (\sigma_\xi^2 + \sigma_\eta^2 \sigma_{x,k-1}^2)) \right)^{1/2} \leq CA^{-1/4} \xrightarrow{A \rightarrow \infty} 0. \end{aligned}$$

The almost sure convergence of the second summand to 1 follows from (27), boundedness of the family $\left\{ A^{1/2} \frac{1}{T_A} \chi(N' \leq T_A \leq N'') \right\}_{A \geq 1}$ and bounded convergence theorem.

The assertion (29) follows from (48) and (49).

Summary

The problem of building optimal predictions for the values of scalar stable autoregressive process with parameter drift is considered. The predictors are constructed on the basis of the truncated estimators, which are shown to have prescribed mean-square accuracy on samples of fixed size. The optimal sample size is established both theoretically and as a stopping time defined on observational data. Asymptotic equivalence of the two is proved in the sense of almost sure convergence and convergence in mean, as well as asymptotic equivalence of the corresponding risk functions.

Acknowledgements

The author wishes to thank his scientific adviser professor Vasiliev V.A., who provided valuable comments and ideas regarding the research theme.

REFERENCES

1. Weyer E., Campi M. Non-Asymptotic Confidence Ellipsoids for the Least-Squares Estimate // *Automatica*. 2002. V. 38, Is. 9. P. 1539–1547.
2. Weyer E., Campi M. Guaranteed Non-Asymptotic Confidence Regions in System Identification // *Automatica*. 2005. V. 41, Is. 10. P. 1751–1764.
3. Liptser R., Shiryaev A. Statistics of Random Processes. N.Y. : Springer, 1977.
4. Konev V., Vorobeichikov S. On Sequential Identification of Stochastic Systems // *Izvestia of USSR Academy of Sciences, Technical Cybernetics*. 1980. Is. 4. P. 176–182.
5. Konev V. Sequential Parameter Estimation of Stochastic Dynamical Systems. Tomsk University Press, Tomsk, 1985.
6. Vasiliev V., Konev V. The Sequential Parameter Identification of the Dynamic Systems in the Presence of Multiplicative and Additive Noises in Observations // *Automation and Remote Control*. 1985. V. 46. P. 706–716.
7. Konev V., Pergamenschikov S. On the Duration of Sequential Estimation of Parameters of Stochastic Processes in Discrete Time // *Stochastics*. 1986. V. 18, Is. 2. P. 133–154.
8. Galtchouk L., Konev V. On Sequential Estimation of Parameters in Semimartingale Regression Models with Continuous Time Parameter // *Annals of Statistics*. 2001. V. 29, Is. 5. P. 1508–1536.
9. Kuchler U., Vasiliev V. On Sequential Parameter Estimation for Some Linear Stochastic Differential Equations with Time Delay // *Sequential Analysis*. 2001. V. 20, Is. 3. P. 117–146.

10. Küchler U., Vasiliev V. On Guaranteed Parameter Estimation of a Multiparameter Linear Regression Process // *Automatica*. 2010. V. 46, Is. 4. P. 637–646.
11. Malyarenko A., Vasiliev V. On Parameter Estimation of Partly Observed Bilinear Discrete-Time Stochastic Systems // *Metrika*. 2012. V. 75, Is. 3. P. 403–424.
12. Politis D., Vasiliev V. Sequential Kernel Estimation of a Multivariate Regression Function // *Proceedings of IX International Conference 'System Identification and Control Problems'*. Moscow, 2012. P. 996–1009.
13. Konev V., Pergamenshchikov S. Truncated Sequential Estimation of the Parameters in Random Regression // *Sequential Analysis*. 1990. V. 9, Is. 1. P. 19–41.
14. Konev V., Pergamenshchikov S. On Truncated Sequential Estimation of the Drifting Parametermean in the First Order Autoregressive Models // *Sequential Analysis*. 1990. V. 9, Is. 2. P. 193–216.
15. Fourdrinier D., Konev V., Pergamenshchikov S. Truncated Sequential Estimation of the Parameter of a First Order Autoregressive Process with Dependent Noises // *Mathematical Methods of Statistics*. 2008. V. 18, Is. 1. P. 43–58.
16. Vasiliev V. A Truncated Estimation Method with Guaranteed Accuracy // *Annals of Institute of Statistical Mathematics*. 2014. V. 66. P. 141–163.
17. Sriram T. Sequential Estimation of the Autoregressive Parameter in a First Order Autoregressive Process // *Sequential Analysis*. 1988. V. 7, Is. 1. P. 53–74.
18. Konev V., Lai T. Estimators with Prescribed Precision in Stochastic Regression Models // *Sequential Analysis*. 1995. V. 14, Is. 3. P. 179–192.
19. Vasiliev V., Kusainov M. Asymptotic Risk-Efficiency of One-Step Predictors of a Stable AR(1) // *Proceedings of XII All-Russian Conference on Control Problems*. Moscow, 2014. P. 2619–2627.
20. Kusainov M., Vasiliev V. On Optimal Adaptive Prediction of Multivariate Autoregression // *Sequential Analysis*. 2015. V. 34, Is. 2. 23 p.
21. Kusainov M. On Optimal Adaptive Prediction of Multivariate ARMA(1,1) Process // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2015. № 1(30). P. 44–57.
22. Malyarenko A. Estimating the Generalized Autoregression Model Parameters for Unknown Noise Distribution // *Automation and Remote Control*. 2012. V. 71, Is. 2. P. 291–302.

Kusainov Marat Islambekovich. E-mail: rjrltsk@gmail.com
Tomsk State University, Tomsk, Russian Federation

Поступила в редакцию 22 апреля 2015 г.

Кусаинов М.И. (Томский государственный университет. Россия).

Риск-эффективность адаптивных одношаговых прогнозов авторегрессии с шумящим параметром.

Ключевые слова: адаптивные прогнозы; асимптотическая риск-эффективность; оптимальный размер выборки; скалярная авторегрессия; момент остановки; усечённое оценивание.

DOI 10.17223/19988605/32/4

Изучается скалярный устойчивый процесс авторегрессии с аддитивным шумом в параметре динамики. Параметры модели предполагаются неизвестными. Усечённые оценки параметров используются для построения адаптивных одношаговых прогнозов. Сопутствующая проблема заключается в минимизации функции риска специального вида, описывающей размер выборки и выборочное среднее квадрата ошибки прогноза. Последовательная процедура вводится для достижения минимального риска.

REFERENCES

1. Weyer, E. & Campi, M. (2002) Non-Asymptotic Confidence Ellipsoids for the Least-Squares Estimate. *Automatica*. 38(9). pp. 1539–1547. DOI: 10.1109/CDC.2000.914211
2. Weyer, E. & Campi, M. (2005) Guaranteed Non-Asymptotic Confidence Regions in System Identification. *Automatica*. 41 (10). pp. 1751–1764.
3. Liptser, R. & Shiryaev, A. (1977) *Statistics of Random Processes*. New York: Springer.
4. Konev, V. & Vorobeichikov, S. (1980) On Sequential Identification of Stochastic Systems. *Izvestia of USSR Academy of Sciences, Technical Cybernetics*. 4. pp. 176–182.
5. Konev, V. (1985) *Sequential Parameter Estimation of Stochastic Dynamical Systems*. Tomsk: Tomsk State University Press.
6. Vasiliev, V. & Konev, V. (1985) The Sequential Parameter Identification of the Dynamic Systems in the Presence of Multiplicative and Additive Noises in Observations. *Automation and Remote Control*. 46. pp. 706–716. (In Russian).
7. Konev, V. & Pergamenshchikov, S. (1986) On the Duration of Sequential Estimation of Parameters of Stochastic Processes in Discrete Time. *Stochastics*. 18 (2). pp. 133–154. DOI: 10.1080/17442508608833405
8. Galtchouk, L. & Konev, V. (2001) On sequential estimation of parameters in semimartingale regression models with continuous time parameter. *Annals of Statistics*. 29 (5). pp. 1508–1536. DOI: 10.1214/aos/1013203463
9. Küchler, U. & Vasiliev, V. (2001) On Sequential Parameter Estimation for Some Linear Stochastic Differential Equations with Time Delay. *Sequential Analysis*. 20 (3). pp. 117–146. DOI: 10.1081/SQA-100106052

10. Küchler, U. & Vasiliev, V. (2001) On Guaranteed Parameter Estimation of a Multiparameter Linear Regression Process. *Automatica*. 46 (4). pp. 637-646. DOI: 10.1016/j.automatica.2010.01.003
11. Malyarenko, A. & Vasiliev, V. (2012) On Parameter Estimation of Partly Observed Bilinear Discrete-Time Stochastic Systems. *Metrika*. 75 (3). pp. 403-424. DOI: 10.1007/s00184-010-0333-5
12. Politis, D. & Vasiliev, V. (2012) Sequential Kernel Estimation of a Multivariate Regression Function. *Proceedings of IX International Conference 'System Identification and Control Problems'*. Moscow. pp. 996-1009.
13. Konev, V. & Pergamenshchikov, S. (1990) Truncated Sequential Estimation of the Parameters in Random Regression. *Sequential Analysis*. 9 (1). pp. 19-41. DOI: 10.1080/07474949008836194
14. Konev, V. & Pergamenshchikov, S. (1990) On Truncated Sequential Estimation of the Drifting Parametermean in the First Order Autoregressive Models. *Sequential Analysis*. 9 (2). pp. 193-216. DOI: 10.1080/07474949008836205
15. Fourdrinier, D., Konev, V. & Pergamenshchikov, S. (2008) Truncated Sequential Estimation of the Parameter of a First Order Autoregressive Process with Dependent Noises. *Mathematical Methods of Statistics*. 18 (1). pp. 43-58. DOI: 10.3103/S1066530709010037
16. Vasiliev, V. (2014) A Truncated Estimation Method with Guaranteed Accuracy. *Annals of Institute of Statistical Mathematics*. 66. pp. 141-163. DOI: 10.1007/s10463-013-0409-x.
17. Sriram, T.(1988) Sequential Estimation of the Autoregressive Parameter in a First Order Autoregressive Process. *Sequential Analysis*. 7 (1). pp. 53-74. DOI: 10.1080/07474948808836142
18. Konev, V. & Lai, T. (1995) Estimators with Prescribed Precision in Stochastic Regression Models. *Sequential Analysis*. 14 (3). pp. 179-192. DOI: 10.1080/07474949508836330
19. Vasiliev, V. & Kusainov, M. (2004) Asymptotic Risk-Efficiency of One-Step Predictors of a Stable AR(1). *Proceedings of XII All-Russian Conference on Control Problems*. Moscow. pp. 2619-2627. (In Russian).
20. Kusainov, M. & Vasiliev, V. (2015) On Optimal Adaptive Prediction of Multivariate Autoregression. *Sequential Analysis*. 34 (2). pp. 211-234. DOI: 10.1080/07474946.2015.1030977
21. Kusainov, M. (2015) On Optimal Adaptive Prediction of Multivariate ARMA (1,1) Process. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 44-57. (In Russian).
22. Malyarenko, A. (2012) Estimating the Generalized Autoregression Model Parameters for Unknown Noise Distribution. *Automation and Remote Control*. 71 (2). pp. 291-302.

G.A. Medvedev

ON THE NELSON–SIEGEL–SVENSSON NO-ARBITRAGE YIELD CURVE MODELS

It is shown that the requirement to satisfy the no-arbitrage conditions specifies the Nelson–Siegel–Svensson model in the sense that gives for coefficients of this model the obvious economic sense: the free coefficient should be function on term to maturity, and other coefficients should depend on the market state variables which, in turn, are selective values of stochastic processes at time point for which the time structure is designed. It is shown that the model is the member of family of affine yield models and is generated by two-dimensional model of a short-term interest rate for Nelson–Siegel model or four-dimensional model of a short-term interest rate for Nelson–Siegel–Svensson model.

Keywords: term structure; yield curve; factor model; affine no-arbitrage models; Nelson–Siegel yield model.

The term structure of yield interest rates is one of the most demanded characteristics of financial assets. At mathematical modeling of term structure are used no-arbitrage affine models as they imply possibility of derivation of decisions in an analytical form more often. For derivation of model of term structure is usual proceed from description as the financial market state variables evolve. It is accepted more often that the market state variables is a stochastic process of diffusion type.

Let's consider that for n -factor affine yield model it is supposed that the vector of a state variables of financial market $X(t) = (X_1, X_2, \dots, X_n)^T$ follows to homogeneous for time Markov process generated by the stochastic differential equation

$$dX(t) = \mu(X(t)) dt + \sigma(X(t)) dW(t) \quad (1)$$

with a drift n -vector $\mu(x)$, $(n \times m)$ -matrix volatility $\sigma(\xi)$, and m -vector $W(t)$ of independent standard Wiener processes.

In one-factor model the short-term interest rate is accepted that a market state variable $r(t)$ evolves as the stochastic process with property of returning to a stationary average θ . In this case there are three basic approaches for the description of stochastic process of the short-term interest rates, developed by the American authors:

1) the Vasicek model [1], when the rate volatility $\sigma = \sqrt{2kD}$ – the determined constant (D – a stationary variance of process $r(t)$)

$$dr(t) = k(\theta - r(t)) dt + \sqrt{2kD} dW(t); \quad (2)$$

2) the CIR model [2], in which the rate volatility – the non-negative stochastic process proportional to a square root of the interest rate

$$dr(t) = k(\theta - r(t)) dt + \sqrt{\frac{2kD}{\theta} r(t)} dW(t), r(0) > 0; \quad (3)$$

3) the Duffie–Kan model [3], in which frameworks the interest rate volatility – random, proportional to excess of the interest rate over the bottom (unattainable) border $r_{\inf}$:

$$dr(t) = k(\theta - r(t)) dt + \sqrt{\frac{2kD}{\theta - r_{\inf}} (r(t) - r_{\inf})} dW(t), r(0) > r_{\inf}. \quad (4)$$

Generally speaking, the Duffie–Kan model is more general model than early two because at zero bottom border of the interest rate it turns to CIR model, and at moving of the bottom border of the interest rate to a minus infinity (i.e. at $r_{\inf} \rightarrow -\infty$) model becomes the Vasicek model.

The term structure of yield interest rates is determined as dependence the yield interest rate (or the prices) zero coupon bonds at some currently time t on term to maturity of this bond.

For models (2)–(4) price of zero coupon bonds $P(t, r; T)$ with maturity date T provided that at time point t the short-term interest rate $r(t)$ has accepted value r , i.e. $r(t) = r$, is calculated by the formula

$$P(t, r; T) \equiv P(t, r; t + \tau) = P(t, r; T) \exp \{A(\tau) - rB(\tau)\}, \quad (5)$$

where $\tau = T - t$ – term to maturity, is usual is supposed that $P(T, r; T) = 1$ that does not limit an analysis generality. Functions $A(\tau)$ and $B(\tau)$ is usual are called as functions of term structure. At derivation of the formula (5) conditions of absence of arbitrage have been considered. For such function of the price the yield to maturity (or simply yield) $y(\tau, r)$ is affine function r and is calculated by the formula

$$y(\tau, r) = \frac{-\ln P(t, r; T)}{T - t} = \frac{rB(\tau) - A(\tau)}{\tau}. \quad (6)$$

Functions $A(\tau)$ and $B(\tau)$ are calculated in an explicit analytical form though for their derivation it is necessary to solve the differential equations of Riccati. Dependence $y(\tau, r)$ on τ also determines a term structure of yield.

The yield to maturity $y(\tau, r)$ is an *average* characteristic for the time period of duration τ . At the same time the practitioners are interested in the question what will be a *short-term rate* of yield in the end of this period. Such rate is named the instant forward rate $f(\tau, r)$ and it is functionally connected with yield to maturity $y(\tau, r)$ by relation

$$y(\tau, r) = \frac{1}{\tau} \int_0^\tau f(s, r) ds, \quad f(\tau, r) = y(\tau, r) + \tau \frac{\partial y(\tau, r)}{\partial \tau}. \quad (7)$$

In frameworks no-arbitrage affine models the forward rate $f(\tau, r)$ is calculated by the formula

$$f(\tau, r) = \frac{-\partial \ln P(t, r; T)}{\partial T} = r \frac{\partial B(\tau)}{\partial \tau} - \frac{\partial A(\tau)}{\partial \tau}. \quad (8)$$

Usually the analytical properties of function $f(\tau, r)$ turn out more simple, than properties $y(\tau, r)$. Both these functions – yield curve $y(\tau, r)$ and a forward curve $f(\tau, r)$ are equally interesting for financial analysts.

It appears that depending on a market variables (for one-factorial model it is values r) the function $y(\tau, r)$ can belong usually only to one of four types of curves: monotonously increasing to some final limiting value ("normal" yield curve), monotonously decreasing to some final limiting value ("inverse" yield curve), yield curve with a maximum (a "humped" curve) and flat yield curve.

Such classification of curves proves to be true on real financial markets. At the same time real market yield though have such functional appearance, but on size often enough strongly differ from received on the basis of the resulted models. Therefore there are various modifications described above models, in particular it is considered that the dimension increase, i.e. transition to two-factor, three-factor etc. models, can raise their accuracy of the description of market yield. In a multidimensional case when the financial market variables are determined by a vector $X(t)$ which evolution is described by diffusion process (1), formulas (6) and (8) for a yield curve and forward curve will be transformed to a form

$$y(\tau, r) = \frac{x B(\tau) - A(\tau)}{\tau}, \quad f(\tau, r) = x \frac{dB(\tau)}{d\tau} - \frac{dA(\tau)}{d\tau}, \quad (9)$$

provided that at point of time t the vector of a variables of financial market $X(t)$ has turned out equal to x , i.e. $X(t) = x$; in this case $B(\tau)$ – the vector, and $A(\tau)$ remains scalar function.

However the dimension increases essentially increases number of parameters of models, but accuracy of models increases poorly.

1. Nelson – Sigel – Svensson model

In this connection there was other method of improvement of model of the yield term structure, based on a small variety of types yield curve. As their only four, there was an idea – to enter some standard functional dependences and by them to build the combinations approximating yield curve. Nelson and Sigel [4] have offered as such standard functions for designing forward curve three simple functions: $\alpha_1(\tau) = 1$, $\alpha_2(\tau) = \exp(-\gamma\tau)$, $\alpha_3(\tau) = \gamma\tau \exp(-\gamma\tau)$. Function $\alpha_1(\tau)$ is intended for approximation of a long-term segment of a

curve, function $\alpha_2(\tau)$ – for approximation of a short-term segment and function $\alpha_3(\tau)$ should approximate medium-term of yield. Use of such approach leads to following result

$$f(\tau, r) = \beta_1 \alpha_{1f}(\tau) + \beta_2 \exp(-\gamma\tau) + \beta_3 \gamma\tau \exp(-\gamma\tau). \quad (10)$$

Evidently, function $f(\tau, r)$ is determined very simply, but for its final identification it is necessary to find four parameters $\gamma > 0$, $\beta_1 > 0$, β_2 and β_3 . By means of relations (7), (10) it is easy to receive and function $y(\tau, r)$ in a form

$$y(\tau, r) = \beta_1 \alpha_{1y}(\tau) + \beta_2 \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + \beta_3 \left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right). \quad (11)$$

Note that the assumption about $\alpha_1(\tau) = 1$, as it will be shown later, is not be agreed to the requirement of absence of arbitrage opportunities and therefore it is more correctly to write the first item in the right parts (10) and (11) as $\beta_1 \alpha_{1f}(\tau)$ and $\beta_1 \alpha_{1y}(\tau)$ where between $\alpha_{1f}(\tau)$ and $\alpha_{1y}(\tau)$ there is a mutual relation which is determined by equality (7):

$$\alpha_{1y}(\tau) = \frac{1}{\tau} \int_0^\tau \alpha_{1f}(s) ds.$$

To increase the flexibility and improve the fit models to empirical data Svensson [5] extend Nelson – Sigel function by adding a fourth function $\alpha_5(\tau) = \delta\tau \exp(-\delta\tau)$, $\delta > 0$, a second hump-shape with two additional parameters β_5 and $\delta \neq \gamma$. So

$$f(\tau, r) = \beta_1 + \beta_2 \exp(-\gamma\tau) + \beta_3 \gamma\tau \exp(-\gamma\tau) + \beta_4 \exp(-\delta\tau) + \beta_5 \delta\tau \exp(-\delta\tau). \quad (12)$$

So the number of parameters of model has increased by two and has reached six. In equality (12) for a generality is entered another function $\alpha_4(\tau) = \exp(-\delta\tau)$ but if to assume $\beta_4 = 0$, Svensson representation will turn out.

In June, 1996 in Bank for International Settlements (BIS, Basel) was accepted the agreement about that the central banks of Europe should to represent the data in BIS for calculations zero coupon yield curve and estimations of parameters of models. It was found out that the majority of banks of Europe is made use for modeling of yield curve the approach of Nelson – Sigel (Italy and Finland) or its Svensson modification (Belgium, Germany, Spain, Norway, France, Switzerland and Sweden) [6]. It, in particular, underlines importance of the analysis of this approach.

Unfortunately, the approach of Nelson – Sigel – Svensson (NSS) does not give recommendations how to determine parameters of model and in any way does not explain, whether such model is no-arbitrage. It makes sense to modify this model on purpose to enter it in the no-arbitrage class. We will notice in the beginning that no-arbitrage affine yields (6), (8) and (9) possess following general limit properties [7]:

$$\lim_{\tau \rightarrow 0} y_{NA}(\tau, r) = \lim_{\tau \rightarrow 0} f_{NA}(\tau, r) = x^T \phi = r, \quad \lim_{\tau \rightarrow \infty} y_{NA}(\tau, r) = \lim_{\tau \rightarrow \infty} f_{NA}(\tau, r) = y_\infty.$$

Here it is designated $B'(0) = \phi$; y_∞ – the limiting long-term yield, which determined only in parameters of model and not dependent on state variables of the market x . Since it is general properties of yields, yields of model of Nelson – Sigel (NS) should possess by them too:

$$\lim_{\tau \rightarrow 0} y_{NS}(\tau, r) = \lim_{\tau \rightarrow 0} f_{NS}(\tau, r) = \beta_1 + \beta_2, \quad \lim_{\tau \rightarrow \infty} y_{NS}(\tau, r) = \lim_{\tau \rightarrow \infty} f_{NS}(\tau, r) = \beta_1.$$

Thus, the sense of parameters β_1 and β_2 in Nelson – Sigel model is cleared up: $\beta_1 = y_\infty$, $\beta_2 = r - y_\infty$.

Comparing (10)–(12) with (9) it is possible to conclude that yields of model NSS are similar to no-arbitrage affine yields in case of four-factor model when the market condition is described by a four-dimensional vector $X(t) = (X_1(t), X_2(t), X_3(t), X_4(t))$. In this case it is possible to consider that the market state is described by a vector $(X_1(t) = \beta_2, X_2(t) = \beta_3, X_3(t) = \beta_4, X_4(t) = \beta_5)$ and in the following way compare details of these models:

$$\beta_1 \alpha_{1f}(\tau) \leftrightarrow -\frac{dA(\tau)}{d\tau}, \quad \alpha_2(\tau) \leftrightarrow \frac{dB_1(\tau)}{d\tau}, \quad \alpha_3(\tau) \leftrightarrow \frac{dB_2(\tau)}{d\tau}, \quad \alpha_4(\tau) \leftrightarrow \frac{dB_3(\tau)}{d\tau}, \quad \alpha_5(\tau) \leftrightarrow \frac{dB_4(\tau)}{d\tau}.$$

It means that parameters $\beta_2, \beta_3, \beta_4$ and β_5 are not constants, and are values of stochastic processes $X_1(t), X_2(t), X_3(t), X_4(t)$ at the point of time t for which the term structure is determined. Function $\beta_1 \alpha_{1f}(\tau)$ too not a constant and a determined function of τ .

The stochastic differential equation for $X(t) = (X_1(t), X_2(t), X_3(t), X_4(t))$ looks like (1). In order that yield was in frameworks of affine structure, components of the equation (1) should be specified as follows [8]:

$$\mu(x) = K(\theta - x), \sigma(x)\sigma(x)^T = \varphi + \sum_{i=1}^4 \omega_i x_i, \sigma(x)\lambda(x) = \xi + \sum_{i=1}^4 \eta_i x_i. \quad (13)$$

Here K, φ and ω_i – (4×4) -matrixes; θ, ξ and η_i – 4-vectors, x_i – vector components of x , the vector function $\lambda(x)$ determines the market prices of risk. It results [3] in the ordinary differential equations for function $A(\tau)$ and components of vector $B(\tau) = (B_1(\tau), B_2(\tau), B_3(\tau), B_4(\tau))$:

$$A'(\tau) = (\xi - K\theta)^T B(\tau) + B(\tau)^T \varphi B(\tau)/2, A(0) = 0, \quad (14)$$

$$B_i'(\tau) = \phi_i - B(\tau)^T (\eta_i + K_i) - B(\tau)^T \omega_i B(\tau)/2, B_i(0) = 0. \quad (15)$$

In the equation for $B_i(\tau)$ symbol K_i designates i -th column of matrix K , $1 \leq i \leq 4$.

In order that the NSS model was also affine no-arbitrage model, decisions of the equations (15) $B_1(\tau), B_2(\tau), B_3(\tau), B_4(\tau)$ should be agreed in appropriate way with functions $\alpha_2(\tau), \alpha_3(\tau), \alpha_4(\tau), \alpha_5(\tau)$.

By determination of function $\alpha_2(\tau), \alpha_3(\tau), \alpha_4(\tau), \alpha_5(\tau)$ are either exponents or their combinations. Such decisions of the equation (15) can be only when the equations (15) are linear. Hence, nonlinear items in the equations (15) should be absent, i.e. $\omega_i = 0, 1 \leq i \leq 4$. And it means that according to representation (13) volatility matrix $\sigma(\xi)$ should not depend on market state variables x , i.e. volatilities should be the determinate constants. From here also follows that function of market price of risk $\lambda(x)$ should be a vector consisting of constants, i.e. according to representation (13) $\eta_i = 0, 1 \leq i \leq 4$. This considerably simplifies a task of determination of vector $B(\tau) = (B_1(\tau), B_2(\tau), B_3(\tau), B_4(\tau))$ as instead of the equations (15) the linear system of the differential equations turns out

$$B'(\tau) = \phi - K^T B(\tau), B(0) = 0. \quad (16)$$

We define the matrix K in the form

$$K = \begin{pmatrix} \gamma & -\gamma & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & \delta & -\delta \\ 0 & 0 & 0 & \delta \end{pmatrix}. \quad (17)$$

Then the decision of system (16) will be the following

$$B_1(\tau) = \phi_1 \frac{1 - e^{-\gamma\tau}}{\gamma}, B_2(\tau) = (\phi_1 + \phi_2) \frac{1 - e^{-\gamma\tau}}{\gamma} - \phi_1 \tau e^{-\gamma\tau}, \quad (18)$$

$$B_3(\tau) = \phi_3 \frac{1 - e^{-\delta\tau}}{\delta}, B_4(\tau) = (\phi_3 + \phi_4) \frac{1 - e^{-\delta\tau}}{\delta} - \phi_3 \tau e^{-\delta\tau}.$$

It means that no-arbitrage yield curves $y_{NA}(\tau, r)$ and $f_{NA}(\tau, r)$ according to ratios (9) are determined by formulas

$$\begin{aligned} y_{NA}(\tau, r) &= x_1 \phi_1 \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + x_2 (\phi_1 + \phi_2) \frac{1 - e^{-\gamma\tau}}{\gamma\tau} - x_2 \phi_1 e^{-\gamma\tau} + \\ &+ x_3 \phi_3 \frac{1 - e^{-\delta\tau}}{\delta\tau} + x_4 (\phi_3 + \phi_4) \frac{1 - e^{-\delta\tau}}{\delta\tau} - x_4 \phi_3 e^{-\delta\tau} - \frac{A(\tau)}{\tau} = \\ &= (x_1 \phi_1 + x_2 (\phi_1 + \phi_2)) \frac{1 - e^{-\gamma\tau}}{\gamma\tau} - x_2 \phi_1 e^{-\gamma\tau} + (x_3 \phi_3 + x_4 (\phi_3 + \phi_4)) \frac{1 - e^{-\delta\tau}}{\delta\tau} - x_4 \phi_3 e^{-\delta\tau} - \frac{A(\tau)}{\tau}, \end{aligned} \quad (19)$$

$$\begin{aligned} f_{NA}(\tau, r) &= x_1 \phi_1 e^{-\gamma\tau} + x_2 \phi_2 e^{-\gamma\tau} + x_2 \phi_1 \gamma\tau e^{-\gamma\tau} + x_3 \phi_3 e^{-\delta\tau} + x_4 \phi_4 e^{-\delta\tau} + x_4 \phi_3 \delta\tau e^{-\delta\tau} - \frac{dA(\tau)}{d\tau} = \\ &= (x_1 \phi_1 + x_2 \phi_2) e^{-\gamma\tau} + x_2 \phi_1 \gamma\tau e^{-\gamma\tau} + (x_3 \phi_3 + x_4 \phi_4) e^{-\delta\tau} + x_4 \phi_3 \delta\tau e^{-\delta\tau} - \frac{dA(\tau)}{d\tau}. \end{aligned} \quad (20)$$

In these expressions the function $A(\tau)$ in an explicit form is not presented, as it has a cumbersome form though its computation by the formula (14) does not represent difficulties when functions $\{B_i(\tau)\}$ are already determined since the formula (14) is not the equation, and for determination $A(\tau)$ it is enough to integrate the right part of equality (14).

Let's notice that limiting values of functions $\{B_i(\tau)\}$ at $\tau \rightarrow \infty$ are finite quantities:

$$\lim_{\tau \rightarrow \infty} B_1(\tau) = \phi_1 / \gamma, \quad \lim_{\tau \rightarrow \infty} B_2(\tau) = (\phi_1 + \phi_2) / \gamma, \quad \lim_{\tau \rightarrow \infty} B_3(\tau) = \phi_3 / \delta, \quad \lim_{\tau \rightarrow \infty} B_4(\tau) = (\phi_3 + \phi_4) / \delta.$$

It allows by means of (9) and (14) to find limiting long-term yields in the explicit form

$$y_\infty = \lim_{\tau \rightarrow \infty} y_{NA}(\tau, r) = \lim_{\tau \rightarrow \infty} f_{NA}(\tau, r) = - \lim_{\tau \rightarrow \infty} \frac{dA(\tau)}{d\tau} = (\sigma\lambda - K\theta)^T B(\infty) + B(\infty)^T \sigma \sigma^T B(\infty) / 2.$$

Let's remind that vector components ϕ determine the weight coefficients of influence a component of a vector of the market states $X(t)$ on yield size (through a short-term rate r). If $\phi_3 = \phi_4 = 0$, then components $X_3(t)$ and $X_4(t)$ do not influence on yield, and considered NSS model turns to usual NS model. The affine no-arbitrage model in this case provides representations

$$f_{NA}(\tau, r) = - \frac{dA(\tau)}{d\tau} + (x_1\phi_1 + x_2\phi_2) e^{-\gamma\tau} + x_2\phi_1 \gamma \tau e^{-\gamma\tau}. \quad (21)$$

$$\begin{aligned} y_{NA}(\tau, r) &= - \frac{A(\tau)}{\tau} + (x_1\phi_1 + x_2(\phi_1 + \phi_2)) \frac{1 - e^{-\gamma\tau}}{\gamma\tau} - x_2\phi_1 e^{-\gamma\tau} = \\ &= - \frac{A(\tau)}{\tau} + (x_1\phi_1 + x_2\phi_2) \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + x_2\phi_1 \left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right). \end{aligned} \quad (22)$$

Comparison of representations (10) with (21) and (11) with (22) shows that no-arbitrage NS model is usual affine no-arbitrage model, coefficients $\beta_1, \beta_2, \beta_3$ of which are determined as follows

$$\beta_{1y}(\tau) = - \frac{A(\tau)}{\tau}, \quad \beta_2 = x_1\phi_1 + x_2\phi_2 = r, \quad \beta_3 = x_2\phi_1. \quad (23)$$

In contrast to determination of these coefficients by Nelson and Sigel here β_1, β_2 and β_3 are not constants: $\beta_{1y}(\tau) \equiv \beta_1\alpha_{1y}(\tau)$ – function τ , and β_2 and β_3 – functions of variables $x_1 = X_1(t)$ and $x_2 = X_2(t)$, the market state at point time t for which the term structure is designed, and consequently are essentially a sample of random variables. Note that by determination the coefficient β_2 is equal to value of a short-term interest rate at the specified point time t , $\beta_2 = r(t) = r$. Since for obtaining the affine no-arbitrage NS models in the equation (1) the function of drift $\mu(x)$ should be linear with respect to x , and a volatility matrix $\sigma(x)$ – a constant (not dependent on x) then the equation (1) generates stochastic process with normal distribution so the coefficients β_2 and β_3 in affine no-arbitrage NS models are normally distributed random variables correlated between themselves.

Let's consider, in what the modification of NS model offered by Svensson results. In this case $\phi_3 \neq 0, \phi_4 \neq 0$ and yield curves are determined by expressions (19) and (20). Comparison of representations (12) and (19) gives the following interpretation of coefficients of NSS model

$$\beta_{1y}(\tau) = - \frac{A(\tau)}{\tau}, \quad \beta_2 = x_1\phi_1 + x_2\phi_2, \quad \beta_3 = x_2\phi_1, \quad \beta_4 = x_3\phi_3 + x_4\phi_4, \quad \beta_5 = x_4\phi_3. \quad (24)$$

Unlike the NS model here $\beta_2 \neq r$, but $\beta_2 + \beta_4 = r$.

Thus, the affine no-arbitrage NSS model proves to be a special case of four-dimensional affine no-arbitrage model with a constant matrix of volatility and matrix K of coefficients of the influence, looking like (17). Accordingly, the NS model is a special case two-dimensional affine no-arbitrage models with a constant matrix of volatility and matrix K of coefficients of the influence, which has the form

$$K = \begin{pmatrix} \gamma & -\gamma \\ 0 & \gamma \end{pmatrix}. \quad (25)$$

In other words, models NS and NSS are special cases of multidimensional model of Vasiček (2) if by this to understand multidimensional model with a constant volatility matrix and linear drift. It is possible to assume that matrixes (17) and (25) serve as signs of models NS and NSS. Then it is possible to assume that if in model (1), (13) matrix K is specified in the form of (17) or (25) yield curve will have term structure NS or NSS. Whether so it?

Let's consider for an example two-dimensional model CIR (3) with a matrix of coefficients of influence (25). For this purpose we will set a matrix and a vector in relations (13) as follows

$$\sigma(x) = \begin{pmatrix} \sigma_1 \sqrt{x_1} & 0 \\ 0 & \sigma_2 \sqrt{x_2} \end{pmatrix}, \lambda(x) = \begin{pmatrix} \sigma_1 \lambda_1 \sqrt{x_1} \\ \sigma_2 \lambda_2 \sqrt{x_2} \end{pmatrix}.$$

Here it is for brevity designated $\sigma_i = \sqrt{2k_i D_i / \theta_i}$, $i = 1, 2$. It gives in expressions (13):

$$\varphi = 0, \omega_1 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}, \omega_2 = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \xi = 0, \eta_1 = \begin{pmatrix} \sigma_1^2 \lambda_1 \\ 0 \end{pmatrix}, \eta_2 = \begin{pmatrix} 0 \\ \sigma_2^2 \lambda_2 \end{pmatrix}.$$

The equations (14)–(15) become

$$\begin{aligned} A'(\tau) &= B(\tau)^T K \theta, \quad A(0) = 0, \\ B_1'(\tau) &= \phi_1 - (\lambda_1 \sigma_1^2 + \gamma) B_1(\tau) - \sigma_1^2 B_1(\tau)^2 / 2, \quad B_1(0) = 0, \\ B_2'(\tau) &= \phi_2 + \gamma B_1(\tau) - (\lambda_2 \sigma_2^2 + \gamma) B_2(\tau) - \sigma_2^2 B_2(\tau)^2 / 2, \quad B_2(0) = 0. \end{aligned}$$

In this case the equations have turned out nonlinear. The second and third equations are the equations of Riccati. And the equation for $B_1(\tau)$ – the equation with constant coefficients also can be solved in the explicit analytical form. The equation for $B_2(\tau)$ – the equation of Riccati with variable free coefficient and in the explicit analytical form does not be resolved.

Let's designate for brevity $g = \sqrt{(\gamma + \sigma_1^2 \lambda_1)^2 + 2\phi_1 \sigma_1^2}$. Then the solution of the equation for $B_1(\tau)$ has a form

$$B_1(\tau) = \frac{2\phi_1}{\gamma + \sigma_1^2 \lambda_1 + g + \frac{2g}{e^{g\tau} - 1}}, \quad (26)$$

that essentially differs from function $B_1(\tau)$ in NS model (18). The same it is possible to tell and about function $B_2(\tau)$ which can be found only numerically. Similar results are fair and for two-dimensional model of Duffie–Kan (4) in which function $B_1(\tau)$ has form like (26), but its parameters g and σ_1 are determined in another way:

$$g_{DK} = \sqrt{(\gamma + \sigma_{1,DK}^2 \lambda_1)^2 + 2\phi_1 \sigma_{1,DK}^2}, \quad \sigma_{1,DK} = \sqrt{2k_1 D_1 / (\theta_1 - x_{1,inf})}.$$

It turns out that the NS model – no-arbitrage only when in the equation (1) function of drift $\mu(x)$ is linear with regard to x , and the volatility matrix $\sigma(x)$ does not depend on x .

2. Latent factors

In [9] instead of the coefficients $\beta_1, \beta_2, \beta_3$ of NS model it is offered to use depending on the current time t «dynamic processes» L_t, S_t and C_t , respectively, which are interpreted as the level, the slope and the curvature factors. According to the above analysis (see Equation (23)), these processes are in accordance with the following state variables of the financial market

$$\beta_1 \Rightarrow L_t = -\frac{A(\tau)}{\tau} \Big|_{\tau=T-t}, \quad \beta_2 \Rightarrow S_t = \phi_1 X_1(t) + \phi_2 X_2(t) = r(t), \quad \beta_3 \Rightarrow C_t = \phi_1 X_2(t). \quad (27)$$

So the factor of level L_t in the explicit form does not depend on current time, and depends only on term to maturity and is not random. The slope and curvature factors $\{S_t, C_t\}$ are linear transformation of state variables $\{X_1(t), X_2(t)\}$

$$\begin{pmatrix} S_t \\ C_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_1 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}. \quad (28)$$

Therefore $\{S_t, C_t\}$ – two-dimensional normal random process with the mathematical expectations

$$\begin{pmatrix} E[S_t] \\ E[C_t] \end{pmatrix} = \begin{pmatrix} \phi_1 E[X_1(t)] + \phi_2 E[X_2(t)] \\ \phi_1 E[X_2(t)] \end{pmatrix}, \quad (29)$$

and the covariance matrix

$$\Sigma(t-t_0) = \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_1 \end{pmatrix} \begin{pmatrix} \Omega_{11}(t-t_0) & \Omega_{12}(t-t_0) \\ \Omega_{21}(t-t_0) & \Omega_{22}(t-t_0) \end{pmatrix} \begin{pmatrix} \phi_1 & 0 \\ \phi_2 & \phi_1 \end{pmatrix}. \quad (30)$$

Elements of expressions (29) and (30) are easily calculated on the basis of explicit formulas for $\{X_1(t), X_2(t)\}$ and for $\Omega(t-t_0)$.

Thus, at construction the forward curve and yield curve a problem of estimation of parameter γ , and values of processes L_t , S_t and C_t in current time arises.

As the process of state variables $X(t)$ satisfies to the stochastic differential equation (1) with drift and volatility (13) then process $\{S_t, C_t\}$ is generated by the stochastic differential equation

$$\begin{pmatrix} dS_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_1 \end{pmatrix} \begin{pmatrix} \gamma & -\gamma \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \theta_1 - (\phi_1 S_t - \phi_2 C_t) / \phi_1^2 \\ \theta_2 - C_t / \phi_1 \end{pmatrix} dt + \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} dW(t)$$

or

$$\begin{pmatrix} dS_t \\ dC_t \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} - \begin{pmatrix} S_t \\ C_t \end{pmatrix} dt + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} dW(t), \quad (31)$$

where

$$\begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \phi_1 \sigma_{11} + \phi_2 \sigma_{21} & \phi_1 \sigma_{12} + \phi_2 \sigma_{22} \\ \phi_1 \sigma_{21} & \phi_1 \sigma_{22} \end{pmatrix}.$$

This equation is linear and can be solved in an explicit form. Thus the fundamental matrix of solutions $\Theta(t-t_0)$ is the same as for process $\{X_1(t), X_2(t)\}$. Generally speaking, explicit dependence of factors $\{S_t, C_t\}$ on time may be found from equality (28) if there to substitute explicit expressions of state variables $\{X_1(t), X_2(t)\}$. Let s – some point time, such that $s < t$, and $\bar{S}$ and $\bar{C}$ – stationary averages of processes $\{S_t, C_t\}$, accordingly, $\bar{S} = \phi_1 \theta_1 + \phi_2 \theta_2$, $\bar{C} = \phi_1 \theta_2$. Then explicit expressions for these processes will be of the form

$$\begin{pmatrix} S_t \\ C_t \end{pmatrix} = \begin{pmatrix} S_s + \gamma(t-s)C_s \\ C_s \end{pmatrix} e^{-\gamma(t-s)} + \begin{pmatrix} (1 - e^{-(t-s)\gamma})\bar{S} - \gamma(t-s)e^{-(t-s)\gamma}\bar{C} \\ (1 - e^{-(t-s)\gamma})\bar{C} \end{pmatrix} + \zeta(t, s),$$

where $\zeta(t, s)$ – the two-dimensional stochastic process determined by the relation

$$\begin{aligned} \zeta(t, s) &= \Theta(t-s) \int_s^t \begin{pmatrix} e^{\gamma(u-s)} & -\gamma(u-s)e^{\gamma(u-s)} \\ 0 & e^{\gamma(u-s)} \end{pmatrix} \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} dW(u) = \\ &= \int_s^t \begin{pmatrix} \rho_{11} + \gamma(t-u)\rho_{21} & \rho_{12} + \gamma(t-u)\rho_{22} \\ \rho_{21} & \rho_{22} \end{pmatrix} e^{-\gamma(t-u)} dW(u). \end{aligned}$$

3. The no-arbitrage conditions for Nelson – Sigel – Svensson model

Until now, it was thought that the state variables form a two-dimensional process $\{X_1(t), X_2(t)\}$, and two-dimensional process $\{S_t, C_t\}$ is a process of the latent variables. Suppose now that the states variables are $\{S_t, C_t\}$, which are generated by the equation (31). The relationship of these variables with zero coupon yield $y_t(\tau)$ is established with help of state variables $\{X_1(t), X_2(t)\}$. In order to in the market the arbitrage opportunities are absent it is necessary that zero coupon bond price $P(t, T, S_t, C_t)$ satisfies the equation of term structure [1]

$$\begin{aligned} \frac{\partial P(t, T, x)}{\partial t} + \mu(x)^T \frac{\partial P(t, T, x)}{\partial x} + \frac{1}{2} \text{tr} \left(\sigma(x)^T \frac{\partial^2 P(t, T, x)}{\partial x^2} \sigma(x) \right) - r(t)P(t, T, x) = \\ = \lambda(t, x)^T \sigma(x)^T \frac{\partial P(t, T, x)}{\partial x}. \end{aligned}$$

This expression must be solved with the boundary condition $P(T, T, S_T, C_T) = 1$ for all states $\{S_T, C_T\}$. Here, for the sake of brevity we use the notation: $x \equiv \{S, C\}$; $\mu(x)$ and $\sigma(x)$ – accordingly a drift vector and a volatility matrix of the equations (31); $r(t)$ – a short-term interest rate; the vector function $\lambda(t, x)$ determines the market price of risk. As between the price of the bond and its yield there is a known relationship $\ln P(t, T, S_t, C_t) = -\tau y_t(\tau | S_t, C_t)$ then this equation is more convenient to write for yield, rather than the bond price:

$$\frac{\partial[\tau y(\tau|x)]}{\partial \tau} - \frac{\partial[\tau y(\tau|x)]}{\partial x} [\mu(x) - \sigma(x)\lambda(x)] + \frac{e^{\tau y(\tau|x)}}{2} \text{tr} \left(\sigma(x)^T \frac{\partial^2 e^{-\tau y(\tau|x)}}{\partial x^2} \sigma(x) \right) = r. \quad (32)$$

Substituting into this equation the function $y(\tau|S, C)$ in explicit form we can find conditions for the absence of arbitrage for Nelson–Siegel yield curve. Write down for the convenience the elements of the equation (32).

$$\begin{aligned} \frac{\partial[\tau y(\tau|x)]}{\partial \tau} &= \tau \frac{dL(\tau)}{d\tau} + L(\tau) + S e^{-\gamma\tau} + C \gamma \tau e^{-\gamma\tau}, \\ \frac{\partial[\tau y(\tau|x)]}{\partial x} &= \left(\frac{\partial[\tau y(\tau|S, C)]}{\partial S}, \frac{\partial[\tau y(\tau|S, C)]}{\partial C} \right) = \left(\frac{1 - e^{-\gamma\tau}}{\gamma}, \frac{1 - e^{-\gamma\tau}}{\gamma} - \tau e^{-\gamma\tau} \right), \\ \mu(x) - \sigma(x)\lambda(x) &= \begin{pmatrix} \gamma & -\gamma \\ 0 & \gamma \end{pmatrix} \begin{bmatrix} \bar{S} \\ \bar{C} \end{bmatrix} - \begin{pmatrix} S \\ C \end{pmatrix} - \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}. \end{aligned}$$

Here, through ρ_{ij} the elements of the volatility matrix of equation (31) are denoted. As has been obtained above to find the Nelson–Siegel yield curve it is necessary, that $\sigma(x)$ and $\lambda(x)$ were constants (did not depend on $\{S, C\}$).

$$\begin{aligned} \frac{\partial[\tau y(\tau|x)]}{\partial x} [\mu(x) - \sigma\lambda] &= S(1 - e^{-\gamma\tau}) - C\gamma\tau e^{-\gamma\tau} - \bar{S}(1 - e^{-\gamma\tau}) + \bar{C}\gamma\tau e^{-\gamma\tau} - \\ &- \frac{1}{\gamma} [(\rho_{11}\lambda_1 + \rho_{12}\lambda_2)(1 - e^{-\gamma\tau}) + (\rho_{21}\lambda_1 + \rho_{22}\lambda_2)(1 - e^{-\gamma\tau} - \gamma\tau e^{-\gamma\tau})]. \end{aligned}$$

As $y(\tau|S, C)$ linearly depends from S and C , then

$$e^{\tau y(\tau|x)} \frac{\partial^2 e^{-\tau y(\tau|x)}}{\partial x^2} = \left(\frac{\partial[\tau y(\tau|x)]}{\partial x} \right)^T \frac{\partial[\tau y(\tau|x)]}{\partial x}$$

does not depend on these variables, and the third item in the left side of the equation (32) can be presented in a form

$$\begin{aligned} \frac{1}{2} \text{tr} \left(\sigma(x)^T \frac{\partial^2 P(t, T, x)}{\partial x^2} \sigma(x) \right) &= \frac{1}{2} \frac{\partial[\tau y(\tau|x)]}{\partial x} \sigma \sigma^T \left(\frac{\partial[\tau y(\tau|x)]}{\partial x} \right)^T = \\ &= \frac{\tau^2}{2} \left[\left(\frac{(1 - e^{-\gamma\tau})}{\gamma\tau} \rho_{11} + \left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right) \rho_{21} \right)^2 + \left(\frac{(1 - e^{-\gamma\tau})}{\gamma\tau} \rho_{12} + \left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right) \rho_{22} \right)^2 \right]. \end{aligned}$$

Thus, substitution in the equation (32) yield $y(\tau|S, C)$, determined by expression (39), transforms this equation to a form

$$\begin{aligned} \tau \frac{dL(\tau)}{d\tau} + L(\tau) + S - \bar{S}(1 - e^{-\gamma\tau}) + \bar{C}\gamma\tau e^{-\gamma\tau} + \\ + \frac{1}{\gamma} [(\rho_{11}\lambda_1 + \rho_{12}\lambda_2)(1 - e^{-\gamma\tau}) + (\rho_{21}\lambda_1 + \rho_{22}\lambda_2)(1 - e^{-\gamma\tau} - \gamma\tau e^{-\gamma\tau})] + \\ + \frac{1}{2} \left[\left(\frac{(1 - e^{-\gamma\tau})}{\gamma} \rho_{11} + \left(\frac{1 - e^{-\gamma\tau}}{\gamma} - \tau e^{-\gamma\tau} \right) \rho_{21} \right)^2 + \left(\frac{(1 - e^{-\gamma\tau})}{\gamma} \rho_{12} + \left(\frac{1 - e^{-\gamma\tau}}{\gamma} - \tau e^{-\gamma\tau} \right) \rho_{22} \right)^2 \right] = r. \end{aligned} \quad (33)$$

Note that factor C has disappeared from the obtained expression, and it includes only average value $\bar{C}$ of this factor. Equality (33) should be satisfied for any values of factors $L(\tau)$ and S_t . In addition, note that only two values depend on the current time t in this expression: S_t and $r(t)$. Therefore, in order to the yield $y(\tau|S, C)$, which determined by expression (39), satisfies the equation (32) requires two conditions from which factors S_t and $L(\tau)$ are uniquely determined: 1) factor S_t is simply equal to a short-term interest rate

$$S_t = r(t), \quad (34)$$

that will completely be agreed equalities (27), and 2) factor $L(\tau)$ satisfies a following differential relation

$$\begin{aligned}
\frac{d[\tau L(\tau)]}{d\tau} &= \bar{S} (1 - e^{-\gamma\tau}) - \bar{C} \gamma \tau e^{-\gamma\tau} - \\
&- \frac{1}{\gamma} [(\rho_{11}\lambda_1 + \rho_{12}\lambda_2)(1 - e^{-\gamma\tau}) + (\rho_{21}\lambda_1 + \rho_{22}\lambda_2)(1 - e^{-\gamma\tau} - \gamma\tau e^{-\gamma\tau})] - \\
&- \frac{1}{2} \left[\left(\frac{(1 - e^{-\gamma\tau})}{\gamma} \rho_{11} + \left(\frac{1 - e^{-\gamma\tau}}{\gamma} - \tau e^{-\gamma\tau} \right) \rho_{21} \right)^2 + \left(\frac{(1 - e^{-\gamma\tau})}{\gamma} \rho_{12} + \left(\frac{1 - e^{-\gamma\tau}}{\gamma} - \tau e^{-\gamma\tau} \right) \rho_{22} \right)^2 \right],
\end{aligned} \tag{35}$$

which is substantially even is not equation with respect to $\tau L(\tau)$ and is simply an integral from it. The calculation of integral from (35) is simple, but leads to cumbersome result and is not included here.

So, conditions of absence of arbitrage in Nelson–Siegel model uniquely determine factors S_t and $L(\tau)$ by means of equalities (34) and (35). Factor C_t at first sight remained uncertain, but also on it there are restrictions. We will consider them.

Let's return at first to initial model (1) for which yield is determined by relation (9) using the functions of term structure $A(\tau)$, $B_1(\tau)$, $B_2(\tau)$ and factors $X_1(t)$, $X_2(t)$. Conditions of absence of arbitrage in this model are reduced to that functions $A(\tau)$, $B_1(\tau)$, $B_2(\tau)$ should satisfy to the equations (14)–(15) and equality $\phi_1 X_1(t) + \phi_2 X_2(t) = r(t)$, in which constant weight coefficients ϕ_1 and ϕ_2 carry matching factors $\{X_1(t), X_2(t)\}$ with a short-term rate $r(t)$. Factors $\{X_1(t), X_2(t)\}$ can be considered as some zero coupon yields [3], the combination of which is the rate $r(t)$. From relations (23) and (27) follows that $S_t = \phi_1 X_1(t) + \phi_2 X_2(t) = r(t)$, $C_t = \phi_1 X_2(t)$. Therefore factor C_t is also uniquely determined when conditions of no-arbitrage.

Let's pay attention that the Nelson–Siegel model is a special case of two-dimensional model (1), when there drift function $\mu(x)$ is affine concerning state variables (see (13)), and the volatility matrix is constant. If in this case to set $\phi_1 = 1$, and $\phi_2 = 0$ we will receive that $S_t = X_1(t)$, $C_t = X_2(t)$ and it is had complete coincidence of Nelson–Siegel model with the affine version of two-dimensional model (1).

4. The no-arbitrage conditions for the Svensson expansion

Let's consider now the no-arbitrage conditions for the Svensson expansion. As already it has been above told, in Svensson modification the forward curve can be calculated by the formula (12), which implies the following form of the yield curve

$$y(\tau) = \beta_0 + \beta_1 \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + \beta_2 \left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right) + \beta_3 \frac{1 - e^{-\delta\tau}}{\delta\tau} + \beta_4 \left(\frac{1 - e^{-\delta\tau}}{\delta\tau} - e^{-\delta\tau} \right). \tag{36}$$

As it has been shown above, from reasons of no arbitrage coefficients $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ cannot be constants. Accordingly redesignate them: $\beta_0 = L(\tau)$, $\beta_1 = S_t$, $\beta_2 = C_t$, $\beta_3 = G_t$, $\beta_4 = H_t$. So

$$y(\tau) = L(\tau) + S_t \frac{1 - e^{-\gamma\tau}}{\gamma\tau} + C_t \left(\frac{1 - e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \right) + G_t \frac{1 - e^{-\delta\tau}}{\delta\tau} + H_t \left(\frac{1 - e^{-\delta\tau}}{\delta\tau} - e^{-\delta\tau} \right). \tag{37}$$

We use this redesignation in representation (36) and we will substitute the obtained expression (37) in the equation (32), in which state vector x – four-dimensional, $x = (S_t, C_t, G_t, H_t)^T$. Function of drift

$$\mu(x) = K(\theta - x) = \begin{pmatrix} \gamma & -\gamma & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & \delta & -\delta \\ 0 & 0 & 0 & \delta \end{pmatrix} \begin{pmatrix} \bar{S} - S_t \\ \bar{C} - C_t \\ \bar{G} - G_t \\ \bar{H} - H_t \end{pmatrix},$$

where matrix K is determined by equality (17), and $(\bar{S}, \bar{C}, \bar{G}, \bar{H})^T$ – a vector of stationary expectations random coefficients (S_t, C_t, G_t, H_t) . A volatility matrix – (4×4) -matrix with constant elements, $\sigma(x) = \sigma$, and a vector of market prices of risk – a 4-vector with constant elements, $\lambda(x) = \lambda$.

Then the explicit form of elements of the equation (32) will have the form

$$\begin{aligned}
\frac{\partial[\tau y(\tau|x)]}{\partial \tau} &= \frac{d[\tau L(\tau)]}{d \tau} + S_t e^{-\gamma \tau} + C_t \gamma \tau e^{-\gamma \tau} + G_t e^{-\delta \tau} + H_t \delta \tau e^{-\delta \tau}, \\
\frac{\partial[\tau y(\tau|x)]}{\partial x} &= \left(\frac{1-e^{-\gamma \tau}}{\gamma} \frac{1-e^{-\gamma \tau}}{\gamma} - \tau e^{-\gamma \tau}, \frac{1-e^{-\delta \tau}}{\delta} \frac{1-e^{-\delta \tau}}{\delta} - \tau e^{-\delta \tau} \right), \\
\frac{\partial[\tau y(\tau|x)]}{\partial \tau} - \frac{\partial[\tau y(\tau|x)]}{\partial x} \mu(x) &= \\
= \frac{d[\tau L(\tau)]}{d \tau} - \bar{S}(1-e^{-\gamma \tau}) - \bar{G}(1-e^{-\delta \tau}) + \bar{C} \gamma \tau e^{-\gamma \tau} + \bar{H} \delta \tau e^{-\delta \tau} + S_t + G_t. \\
e^{\tau y(\tau|x)} \frac{\partial^2 e^{-\tau y(\tau|x)}}{\partial x^2} &= \left(\frac{\partial[\tau y(\tau|x)]}{\partial x} \right)^T \frac{\partial[\tau y(\tau|x)]}{\partial x}, \\
\frac{e^{\tau y(\tau|x)}}{2} \text{tr} \left(\sigma(x)^T \frac{\partial^2 e^{-\tau y(\tau|x)}}{\partial x^2} \sigma(x) \right) &= \frac{1}{2} \frac{\partial[\tau y(\tau|x)]}{\partial x} \sigma \sigma^T \left(\frac{\partial[\tau y(\tau|x)]}{\partial x} \right)^T.
\end{aligned} \tag{38}$$

Note that the vector $\frac{\partial[\tau y(\tau|x)]}{\partial x}$ in view of linearity $y(\tau|x)$ regarding the factors S_t, C_t, G_t, H_t does not

depend on these factors. In addition, the explicit form of equation (32) is in general not contain variables C_t and H_t . Thus as the equation should be carried out at any values of independent variables τ, S_t, C_t, G_t, H_t it breaks up on two parts – the equations regarding the factors S_t and G_t and the equation regarding the function of variable $\tau, L(\tau)$. On current time t are dependent only $r(t), S_t$ and G_t , other components of the equation (32) depend from τ . Therefore from the equation (32) it is obtained two relations, determining the no-arbitrage conditions

$$\begin{aligned}
S_t + G_t &= r(t), \\
\frac{d[\tau L(\tau)]}{d \tau} &= \bar{S}(1-e^{-\gamma \tau}) + \bar{G}(1-e^{-\delta \tau}) - \bar{C} \gamma \tau e^{-\gamma \tau} - \bar{H} \delta \tau e^{-\delta \tau} - \\
&\quad - \frac{\partial[\tau y(\tau|x)]}{\partial x} \sigma \lambda - \frac{1}{2} \frac{\partial[\tau y(\tau|x)]}{\partial x} \sigma \sigma^T \left(\frac{\partial[\tau y(\tau|x)]}{\partial x} \right)^T.
\end{aligned} \tag{39}$$

The vector $\frac{\partial[\tau y(\tau|x)]}{\partial x}$ depends only on a variable τ and parameters γ, δ . It is determined by equality

(38). Its explicit expression in the formula (40) is not used because of bulkiness.

The relation (39) is practically a condition of absence of arbitrage – the sum of factors S_t and G_t should be equal to a short-term interest rate $r(t)$ at any point t .

The relation (40) actually the equation is not, as function $L(\tau)$ is found from it by simple integration.

As to factors C_t and H_t to clarify their values it is necessary to address again to initial model (1), (13) in its four-dimensional version. Then according to relations (24) it is obtained

$$S_t = \phi_1 X_1(t) + \phi_2 X_2(t), C_t = \phi_1 X_2(t), G_t = \phi_3 X_3(t) + \phi_4 X_4(t), H_t = \phi_3 X_4(t).$$

Thus one of the basic properties of yield proves to be true

$$\lim_{T \rightarrow t} y(t, T, r) = \phi_1 X_1(t) + \phi_2 X_2(t) + \phi_3 X_3(t) + \phi_4 X_4(t) = r(t).$$

Strictly speaking, representation (37) does not coincide with the Svensson expansion as there is "extra" term (with factor G_t). Probably, availability of this term will improve approximation yield curve, it is a question opened. However, it should find out, what conditions of absence of arbitrage will be without this "extra" term, taking as an expression of the yield curve

$$y(\tau) = L(\tau) + S_t \frac{1-e^{-\gamma \tau}}{\gamma \tau} + C_t \left(\frac{1-e^{-\gamma \tau}}{\gamma \tau} - e^{-\gamma \tau} \right) + H_t \left(\frac{1-e^{-\delta \tau}}{\delta \tau} - e^{-\delta \tau} \right) \tag{41}$$

in accuracy corresponding to Svensson expansion. In this case we have three factors $X(t) = (S_t, C_t, H_t)$. Drift function will be such

$$\mu(x) = K(\theta - x) = \begin{pmatrix} \gamma & -\gamma & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \delta \end{pmatrix} \begin{pmatrix} \bar{S} - S_t \\ \bar{C} - C_t \\ \bar{H} - H_t \end{pmatrix} = \begin{pmatrix} \gamma(\bar{S} - \bar{C} - S_t + C_t) \\ \gamma(\bar{C} - C_t) \\ \delta(\bar{H} - H_t) \end{pmatrix}.$$

Let's substitute these expressions in the equation (32) and calculate its components.

$$\begin{aligned} \frac{\partial[\tau y(\tau|x)]}{\partial \tau} &= \frac{d[\tau L(\tau)]}{d\tau} + S_t e^{-\gamma\tau} + C_t \gamma \tau e^{-\gamma\tau} + H_t \delta \tau e^{-\delta\tau}, \\ \frac{\partial[\tau y(\tau|x)]}{\partial x} &= \left(\frac{1-e^{-\gamma\tau}}{\gamma}, \frac{1-e^{-\gamma\tau}}{\gamma} - \tau e^{-\gamma\tau}, \frac{1-e^{-\delta\tau}}{\delta} - \tau e^{-\delta\tau} \right), \\ \frac{\partial[\tau y(\tau|x)]}{\partial \tau} - \frac{\partial[\tau y(\tau|x)]}{\partial x} \mu(x) &= \frac{d[\tau L(\tau)]}{d\tau} - \\ &- \bar{S}(1-e^{-\gamma\tau}) + \bar{C} \gamma \tau e^{-\gamma\tau} - \bar{H}(1-e^{-\delta\tau}(1+\delta\tau)) + S_t + H_t(1-e^{-\delta\tau}). \end{aligned}$$

It turns out that in this case the no arbitrage conditions are reduced to the following

$$S_t + H_t(1-e^{-\delta\tau}) = r(t), \quad (42)$$

$$\begin{aligned} \frac{d[\tau L(\tau)]}{d\tau} &= \bar{S}(1-e^{-\gamma\tau}) - \bar{C} \gamma \tau e^{-\gamma\tau} + \bar{H}(1-e^{-\delta\tau}(1+\delta\tau)) - \\ &- \frac{\partial[\tau y(\tau|x)]}{\partial x} \sigma \lambda - \frac{1}{2} \frac{\partial[\tau y(\tau|x)]}{\partial x} \sigma \sigma^T \left(\frac{\partial[\tau y(\tau|x)]}{\partial x} \right)^T. \end{aligned} \quad (43)$$

The relation (43) is already known, as was met earlier, but equality (42) is unusual.

From it turns out that the short-term rate depends on the term to maturity, but it of course can't be. Therefore, the Svensson expansion in its original form (41) (without the "extra" term) leads to unrealizable no arbitrage conditions.

Conclusion

Thus, the requirement of accomplishment of conditions no-arbitrage specifies model of Nelson – Sigel – Svensson in the sense that gives to coefficients of this model the clear economic sense: the free coefficient should be function of term to maturity τ , and other coefficients should depend on market state variables $\{x_i\}$ which, in turn, are sample values of stochastic processes $\{X_i(t)\}$ at time point t for which the term structure is designed, i.e. random variables. We will notice that the description of stochastic processes $\{X_i(t)\}$ is produced under probability measure P , i.e. taking into account risk market prices $\lambda(x)$. The model is the representative of family of affine yield models and is generated by two-dimensional model of short-term interest rates for model of Nelson – Sigel or four-dimensional model of a short-term interest rates for model of Nelson – Sigel – Svensson. Stochastic processes $\{X_i(t)\}$, underlying NS- and NSS-models, are generated by the linear stochastic differential equations. In this connection the market state variables $\{x_i\}$ have normal distribution and can accept with positive probability negative values. That is a certain lack of models NS and NSS.

Let's notice that dependence of coefficients of NS-model on current time was discussed in [9]. Communication of NS-model with affine безарбитражными the models based on three-factorial casual process at a risk neutral Q -measure, is considered in [10]. Application affine no-arbitrage NS-models to real tasks of dynamics of an exchange rate of currencies is described in [11].

REFERENCES

1. Vasiček O.A. An Equilibrium characterisation of the term structure // J. of Financial Economics. 1977. V. 5. P. 177–188.
2. Cox J.C., Ingersoll J.E., Ross S.A. A Theory of the term structure of interest rate // Econometrica. 1985. V. 53. P. 385–467.
3. Duffie D., Kan R. A yield-factor model of interest rates // Mathematical Finance. 1996. V. 6. P. 379–406.
4. Nelson C.R., Siegel A.F. Parsimonious modelling of yield curves // Journal of Business. 1987. V. 60. P. 473–489.
5. Svensson L. Estimating forward interest rates with the extended Nelson-Siegel method // Quarterly Review. Sveriges Riksbank. 1995. No. 3. P. 13–26.

6. *Bank for International Settlements. Zero-Coupon Yield Curves: Technical Documentation* // BIS papers. 2005. No. 25. P. 1–55.
7. Siegel A., Nelson C.R. Long-term behavior of yield curves // *J. of Financial and Quantitative Analysis*. 1988. V. 23. P. 105–110.
8. Медведев Г.А. О временной структуре доходности. 1. Модель Васичека // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 1(18). С. 102–111.
9. Diebold F., Li C. Forecasting the term structure of government bond yields // *J. of Econometrics*. 2006. № 130. P. 337–364.
10. Christensen J., Diebold F., Rudebusch G. The affine arbitrage-free class of Nelson–Siegel term structure models // *PIER Working Paper*. 2008. No. 08-030. P. 1–38.
11. Yu Y. Modeling a two-currency affine arbitrage-free Nelson–Siegel term structure model // *M. Litt. in Finance*. 2012. P. 1–52.

Медведев Геннадий Алексеевич, д-р физ.-мат. наук, профессор. E-mail: MedvedevGA@cosmostv.by
Белорусский государственный университет (Республика Беларусь)

Поступила в редакцию 20 мая 2015 г.

Медведев Г.А. (Белорусский государственный университет. Республика Беларусь).

О безарбитражных моделях доходности Нельсона – Сигеля – Свенссона.

Ключевые слова: временная структура; кривая доходности; факторная модель; аффинные безарбитражные модели; модели доходности Нельсона–Сигеля.

DOI 10.17223/19988605/32/5

Показано, что требование выполнения условий отсутствия арбитража уточняет модель Нельсона–Сигеля–Свенссона в том смысле, что придает коэффициентам этой модели явный экономический смысл: свободный коэффициент должен быть функцией срока до погашения, а остальные коэффициенты должны зависеть от переменных состояния рынка, которые, в свою очередь, являются выборочными значениями случайных процессов в момент времени, для которого конструируется временная структура. Заметим, что описание случайных процессов производится при объективной вероятностной мере, т.е. с учетом рыночных цен риска. Показано, что сама модель является представителем семейства аффинных моделей доходности и порождается двухмерной моделью краткосрочной процентной ставки для модели Нельсона–Сигеля (NS) или четырехмерной моделью краткосрочной процентной ставки для модели Нельсона–Сигеля–Свенссона (NSS). Случайные процессы, лежащие в основе NS- и NSS-моделей, порождаются линейными стохастическими дифференциальными уравнениями, в связи с чем переменные состояния рынка имеют нормальное распределение и могут с положительной вероятностью принимать отрицательные значения, что является определенным недостатком моделей NS и NSS.

REFERENCES

1. Vasiček, O.A. (1977) An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*. 5. pp. 177–188.
2. Cox, J.C., Ingersoll, J.E. & Ross, J.E. (1985) A Theory of the Term Structure of Interest Rate. *Econometrica*. 53. pp. 385–467. DOI: 10.2307/1911242
3. Duffie, D. & Kan, R. (1996) A yield-factor model of interest rates. *Mathematical Finance*. 6. pp. 379–406. DOI: 10.1111/j.1467-9965.1996.tb00123.x
4. Nelson, C.R. & Siegel, A.F. (1987) Parsimonious modelling of yield curves. *Journal of Business*. 60. pp. 473–489.
5. Svensson, L. (1995) Estimating forward interest rates with the extended Nelson–Siegel method. *Quarterly Review. Sveriges Riksbank*. 3. pp. 13–26.
6. Bank for International Settlements. (2005) Zero-Coupon Yield Curves: Technical Documentation. *BIS papers*. 25. pp. 1–55.
7. Siegel, A. & Nelson, C.R. (1988) Long-term behavior of yield curves. *Journal of Financial and Quantitative Analysis*. 23. pp. 105–110. DOI: 10.3386/w1789
8. Medvedev, G.A. (2012) On term structure of yield rates. 1. Vasiček model. *Vestnik Tomskogo gosudarstvennogo universiteta. Informatika i vychislitel'naya tekhnika – Tomsk State University Journal of Control and Computer Science*. 1(18). pp. 102–111. (In Russian).
9. Diebold, F. & Li, C (2006) Forecasting the term structure of government bond yields. *Journal of Econometrics*. 130. pp. 337–364. DOI: 10.3386/w10048
10. Christensen, J., Diebold, F. & Rudebusch, G. (2008) The affine arbitrage-free class of Nelson–Siegel term structure models. *PIER Working Paper*. 08-030. pp. 1–38.
11. Yu, Y. (2012) Modeling a two-currency affine arbitrage-free Nelson–Siegel term structure model. *M. Litt. in Finance*. pp. 1–52.

А.Н. Моисеев

**АСИМПТОТИЧЕСКИЙ АНАЛИЗ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ
MAP/GI/∞ С ВЫСОКОИНТЕНСИВНЫМ ВХОДЯЩИМ ПОТОКОМ**

*Результаты получены в рамках выполнения государственного задания
Министерства образования и науки Российской Федерации № 1.511.2014/К.*

Представлен анализ системы массового обслуживания с неограниченным числом приборов, входящим MAP-поток и произвольным распределением времени обслуживания. Показано, что в условиях растущей интенсивности входящего потока стационарное распределение вероятностей числа заявок в системе является асимптотически гауссовским. Получены параметры соответствующей гауссовской аппроксимации. Кроме того, получена аппроксимация более высокого (третьего) порядка, установлены области применимости обеих аппроксимаций.

Ключевые слова: система массового обслуживания; MAP-поток; неограниченное число приборов; асимптотический анализ.

Системы массового обслуживания являются адекватными моделями современных систем передачи и обработки информации [1], социально-экономических систем [2, 3], систем управления транспортными потоками [4] и т.д. В научной литературе, как правило, исследуются системы с пуассоновским входящим потоком, однако имеются исследования [5], которые показывают, что, например, потоки сообщений в современных сетях передачи данных не являются пуассоновскими, а более адекватной моделью таких потоков является так называемый MAP-поток (Markovian Arrival Process) [6–8]. Таким образом, исследование моделей массового обслуживания с входящим MAP-поток является актуальной научной проблемой.

В настоящей работе представлен анализ системы массового обслуживания с входящим MAP-поток, неограниченным числом приборов и произвольным обслуживанием, которая в нотации Кендалла [9] обозначается как MAP/GI/∞. Исследование проведено в асимптотическом условии растущей интенсивности входящего потока, которое относится к классу широко используемых условий «heavy traffic» [10–12]. В результате получена гауссовская аппроксимация распределения вероятностей числа заявок в системе. Этот результат согласуется с известными результатами, связанными с диффузионной аппроксимацией [10, 13, 14], но предоставляет более удобную форму для использования. Кроме того, изложенная в настоящей работе методика исследования позволяет получить аппроксимацию более высокого порядка точности, которая обеспечивает достаточно малую погрешность и при небольшой интенсивности входящего потока.

С другими, близкими по тематике исследованиями в области анализа систем массового обслуживания и случайных потоков можно ознакомиться в работах [15–18].

1. Постановка задачи

Рассмотрим систему массового обслуживания с неограниченным числом приборов, на вход которой поступает высокоинтенсивный MAP-поток, время обслуживания заявки в системе является случайной величиной с функцией распределения $B(t)$. Входящий MAP-поток задан представлением [8] $(N\mathbf{D}_0, N\mathbf{D}_1)$, где $\mathbf{D}_0$ и $\mathbf{D}_1$ – квадратные матрицы порядка M , а скалярный мультипликатор N используется для того, чтобы выделить асимптотическую составляющую высокой интенсивности этого потока. Дело в том, что интенсивность MAP-потока, заданного таким образом, составляет λN , где величина λ определяется выражением

$$\lambda = \theta \mathbf{D}_1 \mathbf{e}. \quad (1)$$

Здесь $\mathbf{e}$ – вектор-столбец, состоящий из единиц, $\mathbf{\theta}$ – вектор-строка, элементы которой задают закон стационарного распределения вероятностей цепи Маркова $m(t)$, управляющей МАР-поток. Этот вектор удовлетворяет следующей системе уравнений:

$$\begin{cases} \mathbf{\theta}(\mathbf{D}_0 + \mathbf{D}_1) = \mathbf{0}, \\ \mathbf{\theta}\mathbf{e} = 1. \end{cases} \quad (2)$$

Таким образом, интенсивность описанного потока неограниченно возрастает при $N \rightarrow \infty$. Потоки такого рода называют высокоинтенсивными [18].

Обозначим $i(t)$ – количество заявок (или, что то же самое, количество занятых приборов) в системе в момент времени t . Ставится задача нахождения распределения вероятностей процесса $i(t)$ в произвольный момент времени в стационарном режиме функционирования системы.

2. Метод динамического просеивания. Уравнения Колмогорова

Для составления уравнений Колмогорова, которые являются основным механизмом исследования случайных процессов, необходимо провести так называемую марковизацию исследуемого процесса $i(t)$, т.е. преобразовать его таким образом, чтобы полученный в результате преобразований процесс был марковским. Стандартным инструментом такой марковизации является метод дополнительных компонент [9, 19, 20]. Для рассматриваемого процесса $i(t)$ этими дополнительными компонентами будут состояние $m(t)$ цепи Маркова, управляющей входящим МАР-поток, и значения остаточного времени обслуживания для каждой заявки, находящейся в системе в момент t . Таким образом, получаем многомерный марковский процесс с переменным числом компонент.

Прямое исследование такого процесса не представляется возможным ввиду переменного и потенциально неограниченного количества его компонент. В связи с этим в настоящей работе применяется метод динамического просеивания (просеянного потока) [21], который позволяет решить указанную задачу. Основная идея этого подхода заключается в том, чтобы заменить исследование многомерного процесса с переменным числом компонент на исследование двумерного марковского процесса $\{n(t), m(t)\}$, где $n(t)$ – считающий процесс просеянного потока, который конструируется следующим образом. Зафиксируем некоторый момент времени T . Будем считать, что заявка, 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. Через $n(t)$ обозначим число событий просеянного потока, наступивших до момента времени t . Тогда, если в начальный момент времени $t_0 < T$ система свободна, получаем, что для момента времени T имеет место равенство

$$P\{i(T) = k\} = P\{n(T) = k\}$$

для любых целых $k > 0$. Таким образом, найдя распределение вероятностей значений процесса $n(t)$ в момент времени t и полагая $t = T$, в силу произвольности выбора T получим искомое стационарное распределение вероятностей процесса $i(t)$.

Для распределения вероятностей $P(n, m, t) = P\{n(t) = n, m(t) = m\}$ значений марковского процесса $\{n(t), m(t)\}$ можно составить следующую систему дифференциальных уравнений Колмогорова:

$$\frac{1}{N} \frac{\partial P(n, m, t)}{\partial t} = \sum_{l=1}^M \left[P(n, l, t) (D_0)_{lm} + P(n, l, t) (D_1)_{lm} (1 - S(t)) + P(n-1, l, t) (D_1)_{lm} S(t) \right]$$

для $n \geq 0, m = \overline{1, M}$. Для частичной характеристической функции

$$H(u, m, t) = \sum_{n=0}^{\infty} e^{jun} P(n, m, t),$$

где $j = \sqrt{-1}$, эта система переписывается в виде

$$\frac{1}{N} \frac{\partial H(u, m, t)}{\partial t} = \sum_{l=1}^M \left[H(u, l, t) (D_0)_{lm} + H(u, l, t) (D_1)_{lm} (1 - S(t)) + H(u, l, t) (D_1)_{lm} S(t) e^{ju} \right].$$

Приводя подобные и используя обозначения $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$, $\mathbf{H}(u, t) = \{H(u, 1, t), \dots, H(u, M, t)\}$, получим матричное дифференциальное уравнение

$$\frac{1}{N} \frac{\partial \mathbf{H}(u, t)}{\partial t} = \mathbf{H}(u, t) [\mathbf{D} + \mathbf{D}_1 (e^{ju} - 1) S(t)] \quad (3)$$

с начальным условием

$$\mathbf{H}(u, t_0) = \mathbf{0}. \quad (4)$$

Так как прямое решение уравнения (3) не представляется возможным, для решения задачи (3)–(4) воспользуемся методом асимптотического анализа [21] в условии растущей интенсивности входящего потока, т.е. при $N \rightarrow \infty$.

3. Асимптотический анализ

3.1. Асимптотика первого порядка

Выполним в выражениях (3)–(4) следующие замены:

$$\frac{1}{N} = \varepsilon, \quad u = \varepsilon w, \quad \mathbf{H}(u, t) = \mathbf{F}_1(w, t, \varepsilon).$$

Тогда задача (3)–(4) примет вид

$$\varepsilon \frac{\partial \mathbf{F}_1(w, t, \varepsilon)}{\partial t} = \mathbf{F}_1(w, t, \varepsilon) [\mathbf{D} + \mathbf{D}_1(e^{j\varepsilon w} - 1)S(t)], \quad (5)$$

$$\mathbf{F}_1(w, t_0, \varepsilon) = \mathbf{0}. \quad (6)$$

Относительно асимптотического решения $\mathbf{F}_1(w, t) = \lim_{\varepsilon \rightarrow 0} \mathbf{F}_1(w, t, \varepsilon)$ этой задачи имеет место следующее утверждение.

Теорема 1. Асимптотическое решение задачи (5)–(6) имеет вид

$$\mathbf{F}_1(w, t) = \mathbf{0} \exp \left\{ jw\lambda \int_{t_0}^t S(\tau) d\tau \right\}, \quad (7)$$

где $\mathbf{0}$ и λ определяются выражениями (2) и (1) соответственно.

Доказательство выполним в два этапа.

Этап 1. Положим в (5) $\varepsilon \rightarrow 0$, получим

$$\mathbf{F}_1(w, t)\mathbf{D} = \mathbf{0}.$$

Сравнивая это равенство с первым уравнением (2), можем сделать вывод, что $\mathbf{F}_1(w, t)$ может быть представлена в виде

$$\mathbf{F}_1(w, t) = \mathbf{0}\Phi_1(w, t), \quad (8)$$

где $\Phi_1(w, t)$ – некоторая скалярная функция, в силу (6) удовлетворяющая условию $\Phi_1(w, t_0) = 1$.

Этап 2. Домножим (5) на вектор $\mathbf{e}$, подставим (8), поделим результат на ε и выполним предельный переход $\varepsilon \rightarrow 0$. Тогда с учетом того, что $\mathbf{D}\mathbf{e} = \mathbf{0}$ и $\mathbf{0}\mathbf{e} = 1$, получим дифференциальное уравнение относительно функции $\Phi_1(w, t)$

$$\frac{\partial \Phi_1(w, t)}{\partial t} = \Phi_1(w, t) jw\lambda S(t),$$

решение которого с учетом начального условия дает

$$\Phi_1(w, t) = \exp \left\{ jw\lambda \int_{t_0}^t S(\tau) d\tau \right\}.$$

Подставляя данное выражение в (8), получаем (7). Теорема доказана.

3.2. Асимптотика второго порядка

Обозначим через $\mathbf{H}_2(u, t)$ многомерную функцию, удовлетворяющую выражению

$$\mathbf{H}(u, t) = \mathbf{H}_2(u, t) \exp \left\{ juN\lambda \int_{t_0}^t S(\tau) d\tau \right\}. \quad (9)$$

Подставляя это выражение в (3)–(4), получим следующую задачу Коши:

$$\begin{cases} \frac{1}{N} \frac{\partial \mathbf{H}_2(u, t)}{\partial t} + \mathbf{H}_2(u, t) ju\lambda S(t) = \mathbf{H}_2(u, t) [\mathbf{D} + \mathbf{D}_1(e^{ju} - 1)S(t)], \\ \mathbf{H}_2(u, t_0) = \mathbf{0}. \end{cases} \quad (10)$$

Выполним в (10) следующие замены:

$$\frac{1}{N} = \varepsilon^2, \quad u = \varepsilon w, \quad \mathbf{H}_2(u, t) = \mathbf{F}_2(w, t, \varepsilon). \quad (11)$$

С использованием этих обозначений (10) перепишется в виде

$$\begin{cases} \varepsilon^2 \frac{\partial \mathbf{F}_2(w, t, \varepsilon)}{\partial t} + \mathbf{F}_2(w, t, \varepsilon) j \varepsilon w \lambda S(t) = \mathbf{F}_2(w, t, \varepsilon) [\mathbf{D} + \mathbf{D}_1 (e^{j \varepsilon w} - 1) S(t)], \\ \mathbf{F}_2(w, t_0, \varepsilon) = \mathbf{0}. \end{cases} \quad (12)$$

Относительно асимптотического решения $\mathbf{F}_2(w, t) = \lim_{\varepsilon \rightarrow 0} \mathbf{F}_2(w, t, \varepsilon)$ этой задачи имеет место следующее утверждение.

Теорема 2. Асимптотическое решение задачи (12) имеет вид

$$\mathbf{F}_2(w, t) = \mathbf{0} \exp \left\{ \frac{(jw)^2}{2} \left[\lambda \int_{t_0}^t S(\tau) d\tau + \kappa \int_{t_0}^t S^2(\tau) d\tau \right] \right\}, \quad (13)$$

где

$$\kappa = 2\mathbf{g}(\mathbf{D}_1 - \lambda \mathbf{I})\mathbf{e}, \quad (14)$$

а вектор-строка $\mathbf{g}$ удовлетворяет линейному матричному уравнению

$$\mathbf{gD} = \mathbf{0}(\lambda \mathbf{I} - \mathbf{D}_1), \quad (15)$$

здесь $\mathbf{I}$ – единичная матрица.

Доказательство этой теоремы выполним в три этапа.

Этап 1. В (12) выполним предельный переход $\varepsilon \rightarrow 0$, получим следующую систему уравнений:

$$\begin{cases} \mathbf{F}_2(w, t)\mathbf{D} = \mathbf{0}, \\ \mathbf{F}_2(w, t_0) = \mathbf{0}. \end{cases}$$

Отсюда с учетом (2) следует, что $\mathbf{F}_2(w, t)$ можно записать в виде

$$\mathbf{F}_2(w, t) = \mathbf{0} \Phi_2(w, t), \quad (16)$$

где $\Phi_2(w, t)$ – некоторая скалярная функция, удовлетворяющая условию $\Phi_2(w, t_0) = 1$.

Этап 2. С учетом (16) функцию $\mathbf{F}_2(w, t, \varepsilon)$ представим в виде разложения

$$\mathbf{F}_2(w, t, \varepsilon) = \Phi_2(w, t) [\mathbf{0} + \mathbf{g} \cdot j \varepsilon w S(t)] + \mathbf{O}(\varepsilon^2), \quad (17)$$

где $\mathbf{g}$ – некоторая вектор-строка, а $\mathbf{O}(\varepsilon^2)$ – вектор-строка, состоящая из величин порядка малости ε^2 и выше. Подставим (17) и разложение $e^{j \varepsilon w} = 1 + j \varepsilon w + \mathbf{O}(\varepsilon^2)$ в первую строку (12), получим

$$\mathbf{0} \cdot j \varepsilon w \lambda S(t) = \mathbf{0D} + \mathbf{0D}_1 \cdot j \varepsilon w S(t) + \mathbf{gD} \cdot j \varepsilon w S(t) + \mathbf{O}(\varepsilon^2).$$

Выполняя в этом равенстве предельный переход $\varepsilon \rightarrow 0$, получаем матричное уравнение относительно вектора $\mathbf{g}$

$$\mathbf{gD} = \mathbf{0}(\lambda \mathbf{I} - \mathbf{D}_1),$$

которое совпадает с (15).

Этап 3. Умножая обе части уравнения (12) на вектор $\mathbf{e}$, используя (17) и разложение

$$e^{j \varepsilon w} = 1 + j \varepsilon w + \frac{(j \varepsilon w)^2}{2} + \mathbf{O}(\varepsilon^3), \quad (18)$$

в результате несложных преобразований с учетом обозначения (14) получаем следующее дифференциальное уравнение относительно функции $\Phi_2(w, t)$:

$$\frac{\partial \Phi_2(w, t)}{\partial t} = \Phi_2(w, t) \left\{ \frac{(jw)^2}{2} [\lambda S(t) + \kappa S^2(t)] \right\}.$$

Решение этого уравнения при имеющемся начальном условии $\Phi_2(w, t_0) = 1$ дает выражение

$$\Phi_2(w, t) = \exp \left\{ \frac{(jw)^2}{2} \left[\lambda \int_{t_0}^t S(\tau) d\tau + \kappa \int_{t_0}^t S^2(\tau) d\tau \right] \right\},$$

подставляя которое в (16), получаем (13). Теорема доказана.

3.3. Гауссовская аппроксимация стационарного распределения вероятностей числа заявок в системе

Выполнив в выражении (13) замены, обратные к (11) и (9), получим следующее выражение для частичной характеристической функции $\mathbf{H}(u, t)$ числа событий, наступивших в просеянном потоке до момента времени t :

$$\mathbf{H}(u, t) = \theta \exp \left\{ juN\lambda \int_{t_0}^t S(\tau) d\tau + \frac{(ju)^2}{2} \left[N\lambda \int_{t_0}^t S(\tau) d\tau + N\kappa \int_{t_0}^t S^2(\tau) d\tau \right] \right\}.$$

Умножим обе части этого равенства на e . Тогда, возвращаясь к исходному исследуемому процессу $i(t)$ числа заявок в системе, подставим $t = T$, а в качестве начального момента (когда система свободна) выберем $t_0 = -\infty$, получим следующее выражение для характеристической функции $h(u)$ числа заявок в системе в стационарном режиме:

$$h(u) = \exp \left\{ juN\lambda b + \frac{(ju)^2}{2} [N\lambda b + N\kappa\beta] \right\}, \quad (19)$$

где $b = \int_0^\infty (1 - B(t)) dt$ есть среднее время обслуживания, а $\beta = \int_0^\infty (1 - B(t))^2 dt$.

Таким образом, стационарное распределение числа заявок в системе MAP/GI/ ∞ является асимптотически гауссовским с математическим ожиданием $N\lambda b$ и дисперсией $[N\lambda b + N\kappa\beta]$.

3.4. Асимптотика третьего порядка

Применение гауссовской аппроксимации (19) достаточно удобно, однако, как будет показано в п. 4, допустимая погрешность достигается при больших значениях N (30 и больше). Для меньших значений интенсивности входящего потока можно получить более точную аппроксимацию, используя асимптотику третьего порядка.

Итак, введем функцию $\mathbf{H}_3(u, t)$, удовлетворяющую выражению

$$\mathbf{H}_2(u, t) = \mathbf{H}_3(u, t) \exp \left\{ \frac{(ju)^2}{2} \left[N\lambda \int_{t_0}^t S(\tau) d\tau + N\kappa \int_{t_0}^t S^2(\tau) d\tau \right] \right\}, \quad (20)$$

подставим (20) в (10), получим следующую задачу Коши:

$$\begin{cases} \frac{1}{N} \frac{\partial \mathbf{H}_3(u, t)}{\partial t} + \mathbf{H}_3(u, t) \left\{ ju\lambda S(t) + \frac{(ju)^2}{2} [\lambda S(t) + \kappa S^2(t)] \right\} = \mathbf{H}_3(u, t) [\mathbf{D} + \mathbf{D}_1(e^{ju} - 1)S(t)], \\ \mathbf{H}_3(u, t_0) = \theta. \end{cases} \quad (21)$$

В (21) выполним замены

$$\frac{1}{N} = \varepsilon^3, \quad u = \varepsilon w, \quad \mathbf{H}_3(u, t) = \mathbf{F}_3(w, t, \varepsilon),$$

получим

$$\begin{cases} \varepsilon^3 \frac{\partial \mathbf{F}_3(w, t, \varepsilon)}{\partial t} + \mathbf{F}_3(w, t, \varepsilon) \left\{ j\varepsilon w\lambda S(t) + \frac{(j\varepsilon w)^2}{2} [\lambda S(t) + \kappa S^2(t)] \right\} = \mathbf{F}_3(w, t, \varepsilon) [\mathbf{D} + \mathbf{D}_1(e^{j\varepsilon w} - 1)S(t)], \\ \mathbf{F}_3(w, t_0, \varepsilon) = \theta. \end{cases} \quad (22)$$

Обозначим $\mathbf{F}_3(w, t) = \lim_{\varepsilon \rightarrow 0} \mathbf{F}_3(w, t, \varepsilon)$. Тогда имеет место следующая теорема.

Теорема 3. Асимптотическое решение $\mathbf{F}_3(w, t)$ задачи (22) имеет вид

$$\mathbf{F}_3(w, t) = \theta \exp \left\{ \frac{(jw)^3}{6} \left[\lambda \int_{t_0}^t S(\tau) d\tau + 3\kappa \int_{t_0}^t S^2(\tau) d\tau + \mu \int_{t_0}^t S^3(\tau) d\tau \right] \right\}, \quad (23)$$

где

$$\mu = 6\mathbf{g}_2(\mathbf{D}_1 - \lambda\mathbf{I})\mathbf{e} - 3\kappa\mathbf{g}, \quad (24)$$

$\mathbf{g}$ определяется выражением (15), а вектор-строка $\mathbf{g}_2$ удовлетворяет линейному матричному уравнению

$$\mathbf{g}_2\mathbf{D} = \mathbf{0} \left[\left(\frac{\kappa}{2} - \lambda \right) \mathbf{I} - \mathbf{D}_1 \right] + \lambda\mathbf{g}. \quad (25)$$

Доказательство этой теоремы выполним в три этапа.

Этап 1. Выполним в (22) предельный переход $\varepsilon \rightarrow 0$, получим

$$\begin{cases} \mathbf{F}_3(w, t)\mathbf{D} = \mathbf{0}, \\ \mathbf{F}_3(w, t_0) = \mathbf{0}. \end{cases}$$

Отсюда с учетом (2) делаем вывод, что $\mathbf{F}_3(w, t)$ можно представить в виде

$$\mathbf{F}_3(w, t) = \mathbf{0}\Phi_3(w, t), \quad (26)$$

где $\Phi_3(w, t)$ – некоторая скалярная функция.

Этап 2. Решение задачи (22) будем искать в виде разложения

$$\mathbf{F}_3(w, t, \varepsilon) = \Phi_3(w, t) \left[\mathbf{0} + \mathbf{g}_1 \cdot j\varepsilon w S(t) + (\mathbf{g}_1 + 2\mathbf{g}_2) \frac{(j\varepsilon w)^2}{2} S^2(t) \right] + \mathbf{O}(\varepsilon^3), \quad (27)$$

где $\mathbf{g}_1$ и $\mathbf{g}_2$ – некоторые вектор-строки. Подставим разложения (27) и (18) в (22). После несложных преобразований получаем равенство

$$\begin{aligned} j\varepsilon w \lambda S(t)\mathbf{0} + \frac{(j\varepsilon w)^2}{2} \lambda S(t)\mathbf{0} + (j\varepsilon w)^2 S^2(t) \left[\frac{\kappa}{2} \mathbf{0} + \lambda \mathbf{g}_1 \right] = \\ = j\varepsilon w S(t) [\mathbf{0}\mathbf{D}_1 + \mathbf{g}_1\mathbf{D}] + \frac{(j\varepsilon w)^2}{2} S(t) [\mathbf{0}\mathbf{D}_1 + \mathbf{g}_1\mathbf{D}] + (j\varepsilon w)^2 S^2(t) [\mathbf{g}_1\mathbf{D}_1 + \mathbf{g}_2\mathbf{D}] + \mathbf{O}(\varepsilon^2). \end{aligned} \quad (28)$$

Приравнявая выражения при одинаковых степенях ε , получаем, во-первых, $\lambda\mathbf{0} = \mathbf{0}\mathbf{D}_1 + \mathbf{g}_1\mathbf{D}$, откуда следует $\mathbf{g}_1\mathbf{D} = \mathbf{0}(\lambda\mathbf{I} - \mathbf{D}_1)$, что совпадает с (15). С учетом этого вторые слагаемые левой и правой частей (28) уничтожаются, и мы получаем равенство

$$\frac{\kappa}{2} \mathbf{0} + \lambda \mathbf{g} = \mathbf{g}\mathbf{D}_1 + \mathbf{g}_2\mathbf{D},$$

откуда следует выражение (25).

Этап 3. Умножим каждую часть (22) на $\mathbf{e}$, тогда с использованием обозначения $f(w, t, \varepsilon) = \mathbf{F}_3(w, t, \varepsilon)\mathbf{e}$ получаем

$$\begin{cases} \varepsilon^3 \frac{\partial f(w, t, \varepsilon)}{\partial t} + f(w, t, \varepsilon) \left\{ j\varepsilon w \lambda S(t) + \frac{(j\varepsilon w)^2}{2} [\lambda S(t) + \kappa S^2(t)] \right\} = f(w, t, \varepsilon) \mathbf{D}_1 \mathbf{e} (e^{j\varepsilon w} - 1) S(t), \\ f(w, t_0, \varepsilon) = 1. \end{cases}$$

Подставляя сюда (27) и разложение

$$e^{j\varepsilon w} = 1 + j\varepsilon w + \frac{(j\varepsilon w)^2}{2} + \frac{(j\varepsilon w)^3}{6} + O(\varepsilon^4),$$

получаем

$$\begin{aligned} \varepsilon^3 \frac{\partial \Phi_3(w, t)}{\partial t} + \Phi_3(w, t) \left\{ 1 + \mathbf{g}\mathbf{e} \left[j\varepsilon w + \frac{(j\varepsilon w)^2}{2} \right] S(t) + \mathbf{g}_2 \mathbf{e} (j\varepsilon w)^2 S^2(t) \right\} \left\{ \lambda \left[j\varepsilon w + \frac{(j\varepsilon w)^2}{2} \right] S(t) + \frac{\kappa}{2} (j\varepsilon w)^2 S^2(t) \right\} = \\ = \Phi_3(w, t) \left\{ \mathbf{0} + \mathbf{g} \left[j\varepsilon w + \frac{(j\varepsilon w)^2}{2} \right] S(t) + \mathbf{g}_2 (j\varepsilon w)^2 S^2(t) \right\} \mathbf{D}_1 \mathbf{e} \left[j\varepsilon w + \frac{(j\varepsilon w)^2}{2} + \frac{(j\varepsilon w)^3}{6} \right] S(t) + O(\varepsilon^4). \end{aligned}$$

После несложных преобразований с учетом (15), (25), (14) и с использованием обозначения (24) получаем дифференциальное уравнение относительно неизвестной функции $\Phi_3(w, t)$

$$\frac{\partial \Phi_3(w, t)}{\partial t} = \Phi_3(w, t) \left\{ \frac{(jw)^3}{6} [\lambda S(t) + 3\kappa S^2(t) + \mu S^3(t)] \right\},$$

решение которого при имеющемся начальном условии $\Phi_3(w, t_0) = 1$ дает

$$\Phi_3(w, t) = \exp \left\{ \frac{(jw)^3}{6} \left[\lambda \int_{t_0}^t S(\tau) d\tau + 3\kappa \int_{t_0}^t S^2(\tau) d\tau + \mu \int_{t_0}^t S^3(\tau) d\tau \right] \right\}.$$

Подставляя это выражение в (26), получаем (23). Теорема доказана.

Делая обратные замены и аналогично п. 3.3 возвращаясь к характеристической функции числа заявок в системе в стационарном режиме, получим

$$h(u) = \exp \left\{ juN\lambda b + \frac{(ju)^2}{2} [N\lambda b + N\kappa\beta] + \frac{(ju)^3}{6} [N\lambda b + 3N\kappa\beta + N\mu\gamma] \right\}, \quad (29)$$

где $\gamma = \int_0^\infty (1 - B(t))^3 dt$.

К сожалению, пользоваться аппроксимацией (29) не так удобно, как гауссовской (19) (для получения ряда распределения придется выполнить обратное преобразование Фурье от этого выражения), однако, как будет показано ниже, выражение (29) обеспечивает меньшую погрешность и более широкую область применимости асимптотических результатов.

4. Численные результаты

Демонстрацию точности аппроксимаций (19) и (29) проведем на конкретном численном примере. Пусть входящий МАР-поток задан представлением (ND_0, ND_1) , где

$$D_0 = \begin{pmatrix} -6,50 & 0,50 & 0,97 \\ 0,36 & -1,90 & 0,96 \\ 0,19 & 0,54 & -0,85 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 5,00 & 0,00 & 0,03 \\ 0,04 & 0,50 & 0,04 \\ 0,01 & 0,06 & 0,05 \end{pmatrix}.$$

Время обслуживания имеет гамма-распределение с параметрами масштаба и формы, равными 0,5. Пример составлен таким образом, что параметр λ приблизительно равен 1, среднее время обслуживания также равно 1, и, таким образом, интенсивность входящего потока превышает интенсивность обслуживания в N раз.

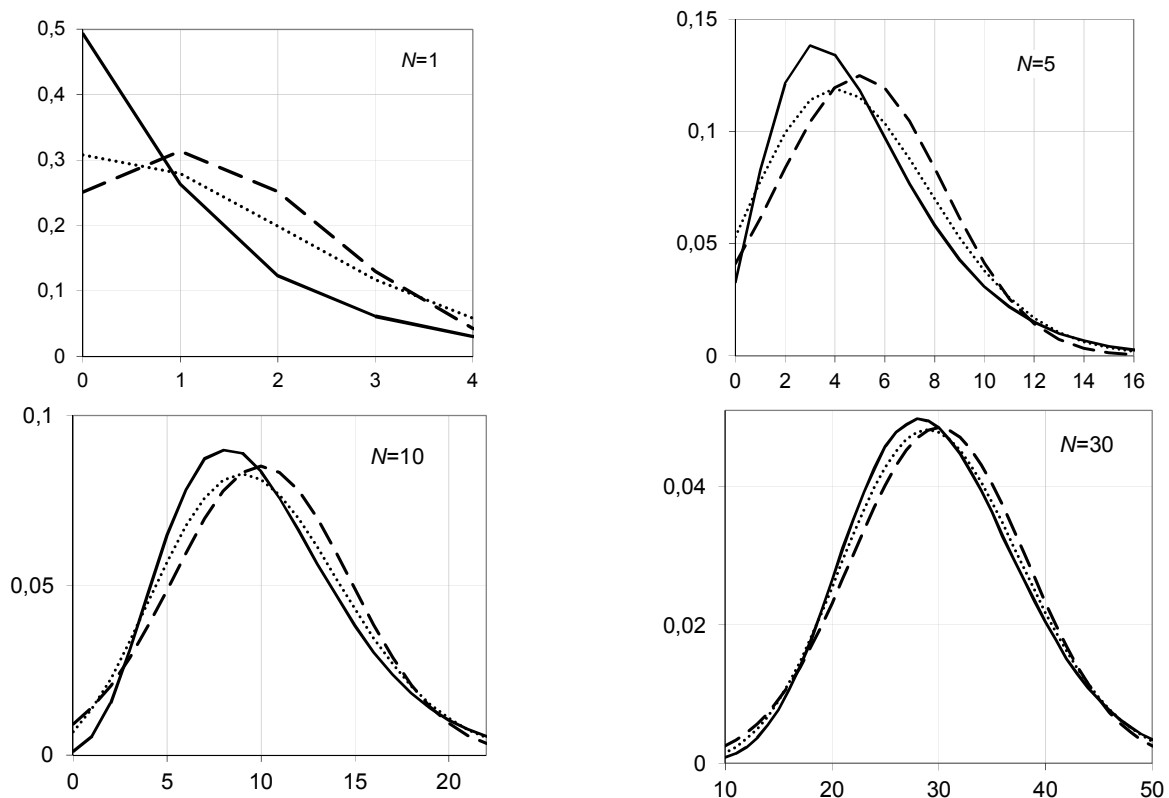


Рис. 1. Сравнение законов распределения вероятностей для эмпирического распределения (сплошная линия), гауссовской аппроксимации (пунктирная линия) и аппроксимации третьего порядка (точечная линия)

На рис. 1 представлены графики законов распределения вероятностей для эмпирического распределения, составленного на основе результатов имитационного моделирования, и распределений, полученных на основе аппроксимаций (19) и (29). Из графиков видно, что при достаточно больших значениях N ($N \geq 10$) распределение по форме близко к нормальному.

Для числовой оценки точности аппроксимаций воспользуемся расстоянием Колмогорова [22]. В таблице приведены значения расстояния Колмогорова для различных значений величины N . Если в качестве допустимой погрешности выбрать расстояние Колмогорова, равное 0,03, то из таблицы видно, что третья асимптотика дает достаточно точную аппроксимацию уже при $N = 10$, в то время как гауссовская аппроксимация дает такую же точность только при $N = 30$.

Расстояние Колмогорова между эмпирическим распределением и аппроксимациями второго и третьего порядка для различных значений интенсивности входящего потока N

Интенсивность входящего потока N	Расстояние Колмогорова	
	для аппроксимации второго порядка (гауссовской)	для аппроксимации третьего порядка
1	0,242	0,186
5	0,099	0,048
10	0,060	0,027
30	0,029	0,013
100	0,017	0,007

Заключение

В работе представлено исследование системы массового обслуживания с неограниченным числом приборов, входящим МАР-поток и произвольным распределением времени обслуживания. Исследование выполнено в асимптотическом условии растущей интенсивности входящего потока. Показано, что для достаточно больших значений параметра N интенсивности входящего потока стационарное распределение вероятностей числа заявок в системе аппроксимируется гауссовским (нормальным) распределением, параметры которого получены в работе. Для повышения степени точности выполнен асимптотический анализ третьего порядка и получены параметры аппроксимации третьего порядка для стационарного распределения вероятностей числа заявок в системе.

Числовые примеры показывают, что гауссовская аппроксимация дает достаточно малую погрешность при $N \geq 30$, где N – параметр, определяющий, во сколько раз интенсивность входящего потока превышает интенсивность обслуживания. Аппроксимация третьего порядка расширяет область применимости до значений $N \geq 10$, что существенно увеличивает точность полученных результатов. Однако при использовании полученных в работе аппроксимационных формул следует учитывать, что гауссовская аппроксимация более проста в применении. Таким образом, при $N \geq 30$ целесообразно использовать более удобную гауссовскую аппроксимацию, а для меньших значений N следует воспользоваться аппроксимацией третьего порядка.

Результаты могут быть использованы для анализа систем массового обслуживания вида МАР/GI/ ∞ , а методика исследования – для выполнения исследования многофазных систем и сетей массового обслуживания с входящим высокоинтенсивным МАР-поток.

ЛИТЕРАТУРА

1. Грачев В.В., Моисеев А.Н., Назаров А.А., Ямпольский В.З. Многофазная модель массового обслуживания системы распределенной обработки данных // Доклады ТУСУР. 2012. № 2(26), ч. 2. С. 248–251.
2. Матальцкий М.А., Статкевич С.Э. НМ-сети как новые стохастические модели прогнозирования доходов различных объектов // Вестник ГрГУ. Сер. 5. Экономика. 2009. № 1. С. 107–115.
3. Гарайшина И.Р. Исследование математических моделей процессов государственного пенсионного страхования : дис. ... канд. физ.-мат. наук. Томск : ТГУ, 2005. 148 с.
4. Гасников А.В. и др. Введение в математическое моделирование транспортных потоков : учеб. пособие / под ред. А.В. Гасникова. М. : МФТИ, 2010. 362 с.

5. Heyman D.P., Lucantoni D. Modelling multiple IP traffic streams with rate limits // IEEE/ACM Trans. on Networking. 2003. V. 11, No. 6. P. 948–958.
6. Neuts M.F. A versatile Markovian arrival process // J. of Appl. Prob. 1979. V. 16. P. 764–779.
7. Lucantoni D.M. New results on the single server queue with a batch Markovian arrival process // Stochastic Models. 1991. V. 7. P. 1–46.
8. Breuer L., Baum D. An introduction to queueing theory and matrix-analytic methods. Springer, 2005. 271 p.
9. Kendall D.G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chains // Ann. Math. Statist. 1953. V. 24(3). P. 338–354.
10. Whitt W. On the heavy-traffic limit theorem for GI/G/∞ queues // J. Adv. Appl. Prob. 1982. V. 14. P. 171–190.
11. Kingman J.F.C. On queues in heavy traffic // J. of the Royal Statistical Society. 1962. Series B 24(2). P. 383–392.
12. Iglehart D.L., Whitt W. Multiple channel queues in heavy traffic // Adv. Appl. Prob. 1970. V. 2. P. 150–172.
13. Iglehart D.L. Limit diffusion approximations for the many server queue and the repairman problem // J. of Appl. Prob. 1965. V. 2. P. 429–441.
14. Боровков А.А. О предельных законах для процессов обслуживания в многоканальных системах // Сибирский математический журнал. 1967. Т. 8, № 5. С. 983–1004.
15. Горбатенко А.Е., Назаров А.А. Исследование МАР-потока в условиях растущей интенсивности // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2008. № 3(4). С. 66–70.
16. Лопухова С.В. Исследование ММР-потока асимптотическим методом m -го порядка // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2008. № 3(4). С. 71–76.
17. Назаров А.А., Семенова И.А. Исследование системы ММР|GI|∞ методом просеянного потока // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 4(17). С. 74–84.
18. Моисеев А.Н., Назаров А.А. Исследование системы массового обслуживания HIGI|GI|∞ // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 2(23). С. 75–83.
19. Cox D.R. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables // Proc. Camb. Phil. Soc. 1955. V. 51. P. 433–441.
20. Henderson W. Alternative approaches to the analysis of M/G/1 and G/M/1 queue // J. Oper. Res. Soc. Japan. 1972. V. 15. P. 92–101.
21. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. Томск : Изд-во НТЛ, 2006. 112 с.
22. Рыков В.В., Иткин В.Ю. Математическая статистика и планирование эксперимента : учеб. пособие. М. : МАКС Пресс, 2010. 308 с.

Моисеев Александр Николаевич, канд. техн. наук, доцент. E-mail: moiseev.tsu@gmail.com
Томский государственный университет

Поступила в редакцию 20 апреля 2015 г.

Moiseev Alexander N. (Tomsk State University, Russian Federation).

Asymptotic analysis of queueing system MAP/GI/∞ with high-rate arrivals

Keywords: queueing system; Markovian arrival process; infinite-server system; asymptotic analysis.

DOI 10.17223/19988605/32/6

In the paper, the infinite-server queueing system with the Markovian arrival process and with the general form of service time distribution is considered. The analysis is performed under the asymptotic condition of arrivals' rate infinite growth. The high intensive Markovian arrival process is defined by representation $(N\mathbf{D}_0, N\mathbf{D}_1)$, where N is the scalar that characterizes the high rate of arrivals. Let $\mathbf{D}_0$ and $\mathbf{D}_1$ be square matrices. A rate of the arrivals is equal to $N\lambda$. Here $\lambda = \boldsymbol{\theta}\mathbf{D}_1\mathbf{e}$, $\boldsymbol{\theta}$ is a row vector of stationary probability distribution of the embedded Markov chain of the MAP, $\mathbf{e}$ is the column vector with entries all equal to 1. The service time for each customer is a random variable with the cumulative distribution function $B(t)$.

The approximations of the 2nd and 3rd orders for the characteristic function of number of customers in the stationary regime are obtained. The approximation of the second order is the Gaussian distribution with mean $N\lambda b$ and variance $[N\lambda b + N\kappa\beta]$. Here

$b = \int_0^\infty (1 - B(t))dt$ is the average service time, $\beta = \int_0^\infty (1 - B(t))^2 dt$, κ is defined by the expression $\kappa = 2\mathbf{g}(\mathbf{D}_1 - \lambda\mathbf{I})\mathbf{e}$, and the row

vector $\mathbf{g}$ satisfies the linear matrix equation $\mathbf{g}\mathbf{D} = \boldsymbol{\theta}(\lambda\mathbf{I} - \mathbf{D}_1)$, where $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$ and $\mathbf{I}$ is the identity matrix.

The approximation of the third order is equal to

$$h(u) = \exp \left\{ juN\lambda b + \frac{(ju)^2}{2} [N\lambda b + N\kappa\beta] + \frac{(ju)^3}{6} [N\lambda b + 3N\kappa\beta + N\mu\gamma] \right\},$$

where $j = \sqrt{-1}$, $\gamma = \int_0^\infty (1 - B(t))^3 dt$, $\mu = 6\mathbf{g}_2(\mathbf{D}_1 - \lambda\mathbf{I})\mathbf{e} - 3\kappa\mathbf{g}$, the row vector $\mathbf{g}_2$ satisfies the following linear matrix equation

$$\mathbf{g}_2 \mathbf{D} = \theta \left[\left(\frac{\kappa}{2} - \lambda \right) \mathbf{I} - \mathbf{D}_1 \right] + \lambda \mathbf{g}.$$

The numerical analysis of the applicability of the obtained asymptotic results is also performed. It is shown that the asymptotic approximation of the second order (Gaussian approximation) provides enough accuracy when the arrivals rate is greater than rate of the servicing by 30 times or more ($N \geq 30$). On other hand, the third order approximation provides the similar accuracy for values $N \geq 10$. Therefore, it is reasonable to use the obtained Gaussian approximation as a more comfortable form for usage when $N \geq 30$, and we need take the approximation of the third order when a value of the parameter N is less than 30.

REFERENCES

1. Grachev, V.V., Moiseev, A.N., Nazarov, A.A. & Yampol'skiy, V.Z. (2012) Multi-stage queueing model of distributed data processing system. *Doklady TUSUR – Proceedings of TUSUR*. 2(26/2). pp. 248-251. (In Russian).
2. Matalytskiy, M.A. & Statkevich, S.E. (2009) NM-seti kak novye stokhasticheskie modeli prognozirovaniya dokhodov razlichnykh ob"ek-tov [HM-networks as new stochastic models for forecasting of revenues of different objects]. *Vestnik GrGU. Ser. 5. Ekonomika*. 1. pp. 107-115.
3. Garayshina, I.R. (2005) *Issledovanie matematicheskikh modeley protsessov gosudarstvennogo pensionnogo strakhovaniya* [Investigation of mathematical models of the State pension insurance processes]. Physics and Mathematics Cand. Dis. Tomsk: Tomsk State University.
4. Gasnikov, A.V. et al. (2010) *Vvedenie v matematicheskoe modelirovanie transportnykh potokov* [Introduction into mathematical modeling of traffic flows]. Moscow: MFTI.
5. Heyman, D.P. & Lucantoni, D. (2003) Modelling multiple IP traffic streams with rate limits. *IEEE/ACM Trans. on Networking*. 11 (6). pp. 948-958.
6. Neuts, M.F. (1979) A versatile Markovian arrival process. *Journal of Applied Prob.* 16. pp. 764-779.
7. Lucantoni, D.M. (1991) New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*. 7. pp. 1-46. DOI: 10.1080/15326349108807174
8. Breuer, L. & Baum, D. (2005) *An introduction to queueing theory and matrix-analytic methods*. Springer.
9. Kendall, D.G. (1953) Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chains. *The Annals of Mathematical Statistics*. 24(3). pp. 338-354.
10. Whitt, W. (1982) On the heavy-traffic limit theorem for GI/G/∞ queues. *Journal of Advances in Applied Probability*. 14. p. 171-190. DOI: 10.2307/1426738
11. Kingman, J.F.C. (1962) On queues in heavy traffic. *Journal of the Royal Statistical Society. B* 24(2). pp. 383-392.
12. Iglehart, D.L. & Whitt, W. (1970) Multiple channel queues in heavy traffic. *Journal of Advances in Applied Probability*. 2. pp. 150-172.
13. Iglehart, D.L. (1965) Limit diffusion approximations for the many server queue and the repairman problem. *Journal of Applied Probability*. 2. pp. 429-441. DOI: 10.2307/3212203
14. Borovkov, A.A. (1967) O predel'nykh zakonakh dlya protsessov obsluzhivaniya v mnogokanal'nykh sistemakh [On limit laws for service processes in multi-channel systems]. *Sibirskiy matematicheskiy zhurnal – Siberian Mathematical Journal*. 8 (5). pp. 983-1004.
15. Gorbatenko, A.E. & Nazarov, A.A. (2008) Issledovanie MAR-potoka v usloviyakh rastushchey intensivnosti [The research of MAP in conditions of growing intensity]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(4). pp. 66-70.
16. Lopukhova, S.V. (2008) Issledovanie MMR-potoka asimptoticheskim metodom m-go poryadka [Research of MMP-process by the asymptotical analysis method of the m-th order]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(4). pp. 71-76.
17. Nazarov, A.A. & Semenova, I.A. (2011) Investigation of non-Markovian system of mass service with unlimited number of servers and input MMP-flow by the method of sifted flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(17). pp. 74-84. (In Russian).
18. Moiseev, A.N. & Nazarov, A.A. (2013) Investigation of queueing system HIGI|GI|∞. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 75-83. (In Russian).
19. Cox, D.R. (1955) The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Mathematical Proceedings of the Cambridge Philosophical Society*. 51. pp. 433-441. DOI: 10.1017/S0305004100030437
20. Henderson, W. (1972) Alternative approaches to the analysis of M/G/1 and G/M/1 queue. *J. Oper. Res. Soc.* 15. pp. 92-101.
21. Nazarov, A.A. & Moiseeva, S.P. (2006) *Metod asimptoticheskogo analiza v teorii massovogo obsluzhivaniya* [Asymptotic analysis method in queueing theory]. Tomsk: NTL.
22. Rykov, V.V. & Itkin, V.Yu. (2010) *Matematicheskaya statistika i planirovanie eksperimenta* [Mathematical statistics and design of experiments]. Moscow: MAKSS Press.

М.А. Сиренко, П.Ф. Тарасенко

ТОЧНЫЕ И АСИМПТОТИЧЕСКИЕ ЗНАКОВЫЕ ТЕСТЫ ДЛЯ ПАРАМЕТРОВ НЕЛИНЕЙНОЙ КВАНТИЛЬНОЙ РЕГРЕССИИ С МАРКОВСКИМИ ШУМАМИ

Одним из привлекательных свойств знаковых статистических процедур является возможность строить точные тесты для проверки простых гипотез о параметрах регрессионных моделей. В данной работе этот подход распространяется на случай нелинейной модели с зависимыми шумами. Рассматривается модель многоквантильной регрессии, что позволяет совместно проверять гипотезы как о параметрах регрессии, так и о параметрах масштаба шума.

Ключевые слова: знаковые статистические процедуры; нелинейная регрессия; квантильная регрессия; Марковский процесс; проверка простых гипотез; точные тесты.

Знаковые статистические процедуры позволяют строить тесты для проверки простых гипотез о параметрах линейных моделей в непараметрической постановке задачи. Это объясняется тем, что при гипотезе распределение знаков невязок известно, поэтому такие тесты позволяют точно контролировать уровень значимости. Знаковые методы сначала были разработаны для меданной регрессии [1], т.е. для частного случая квантильной регрессии, для которой намного раньше были развиты методы, основанные на минимизации взвешенных модулей невязок (см. основополагающую работу [2]). В частности, на основе метода наименьших модулей в [3] были предложены методы анализа симметричной двухквантильной регрессии. Позднее были предложены знаковые процедуры для квантильной регрессии [4, 5], в том числе для зависимых шумов [6].

В данной работе рассматривается ряд обобщений знакового подхода к проверке простых гипотез для параметров квантильной регрессии. Во-первых, мы рассматриваем случай нелинейной модели. Во-вторых, рассматривается многоквантильная регрессия, поэтому от знаков невязок мы переходим к индикаторам их попадания в межквантильные интервалы, как это предлагается в [7]. В-третьих, рассматриваются зависимые наблюдения, когда ошибки образуют стационарный Марковский процесс с дискретным временем.

Рассмотрение методов проверки простых гипотез является первым шагом для построения процедур оценивания параметров на основе подхода, который, как принято считать, впервые был использован в статье [8], а также для построения процедур проверки линейных гипотез, которые более интересны для практических приложений.

1. Постановка задачи

Рассмотрим нелинейную модель регрессии с зависимыми случайными ошибками вида

$$y_t = g_t(\theta) + \varepsilon_t, \quad t = \overline{1, n}, \quad (1)$$

где $g_t(\theta)$ – заданная при каждом t непрерывно дифференцируемая функция от параметров $\theta = (\theta_1, \dots, \theta_T)' \in \mathbb{R}^T$. Случайные отклонения ε_t образуют стационарный Марковский процесс $(r-1)$ -го порядка. Одномерные функции распределения $P\{\varepsilon_t < x\}$ неизвестны и не обязательно совпадают при разных t , но имеют несколько совпадающих квантилей заданных уровней.

Подробнее: пусть конечный набор смежных интервалов $C_1(\mu), \dots, C_K(\mu)$ образует разбиение $\mathbb{R}^1$, при этом $P\{\varepsilon_t \in C_k(\mu)\} = p_k$, $k = \overline{1, K}$, где вероятности p_k заданы. Параметры μ определяют масштаб одномерного распределения шумов, так как влияют на ширину межквантильных интервалов. Границы

интервалов $C_k(\mu) = \langle c_{k-1}(\mu), c_k(\mu) \rangle$ (угловая скобка может означать либо открытую, либо закрытую границу) зависят от неизвестных параметров μ линейно $c_k(\mu) = a_k + d'_k \mu$, $k = \overline{1, K-1}$, $c_0(\mu) = -\infty$, $c_K(\mu) = +\infty$. При этом a_k и d_k заданы, а допустимые параметры образуют множество $\left\{ \mu : (d_k - d_{k-1})' \mu + (a_k - a_{k-1}) > 0, k = \overline{2, K-1} \right\}$.

Наиболее интересными являются самые простые случаи симметричной двухквантильной и трехквантильной регрессии, когда единственный параметр μ равен половине интерквантильного размаха. Для двухквантильной регрессии $K = 3$, $c_1(\mu) = -\mu$, $c_2(\mu) = \mu$, $p_1 = p_3 = p$, $p_2 = 1-2p$. Для трехквантильной регрессии $K = 4$, $c_1(\mu) = -\mu$, $c_2(\mu) = 0$, $c_3(\mu) = \mu$, $p_1 = p_4 = p$, $p_2 = p_3 = (1-2p)/2$. В эту же модель укладывается одноквантильная регрессия, при этом параметры μ отсутствуют, $K = 2$, $c_1(\mu) = 0$, $p_1 = p$, $p_2 = 1-p$. Во всех трех моделях вероятность p задана.

Введем обозначения для совместных вероятностей:

$$P^{(l)}(k_1, \dots, k_l) = P\{\varepsilon_{t-l+1} \in C_{k_1}(\mu), \dots, \varepsilon_t \in C_{k_l}(\mu)\}, \quad l = \overline{1, r}, \quad k_1, \dots, k_r = \overline{1, K}, \quad (2)$$

$$\tilde{P}^{(l)} = \{P^{(l)}(k_1, \dots, k_l) : k_1, \dots, k_l = \overline{1, K}\}, \quad l = \overline{1, r}.$$

В частности, $\tilde{P}^{(1)} = \{p_1, \dots, p_K\} = \{P^{(1)}(1), \dots, P^{(1)}(K)\}$.

Неизвестными параметрами являются параметры θ , μ и набор $\tilde{P}^{(r)}$, причем в составе $\tilde{P}^{(r)}$ есть линейно зависимые вероятности. Обозначим через Q вектор, образованный некоторым набором линейно независимых вероятностей из состава $\tilde{P}^{(r)}$.

Рассмотрим структурные преобразования $j_l(i_1, \dots, i_l) = 1 + \sum_{j=1}^l (i_j - 1)K^{l-j}$, $l = \overline{1, r}$, которые осуществляют развертку r -мерного набора вероятностей $\tilde{P}^{(l)}$ в одномерный вектор $P^{(l)}$ по правилу $[P^{(l)}]_{j_l(i_1, \dots, i_l)} = P^{(l)}(i_1, \dots, i_l)$. Здесь и далее $[A]_j$ означает j -ю строку матрицы или j -й элемент вектора. С помощью структурной матрицы G и вектора D можно осуществить переход от свободных вероятностей Q к $P^{(r)} = D + GQ$. Это преобразование может учитывать не только условие нормировки, заданные одномерные вероятности, условия стационарности, но и симметрию конечномерных распределений, если это необходимо. Кроме того, далее будем использовать структурные матрицы F_i , $i = \overline{1, r-1}$, которые обеспечивают переходы к вероятностям меньшего порядка $P^{(i)} = F_i P^{(i+1)}$.

В этих условиях, обозначив истинные параметры через $v = (\theta', \mu', Q')$, а гипотетические через $v_0 = (\theta'_0, \mu'_0, Q'_0)$, сформулируем задачу проверки простой гипотезы H_0 о параметрах моделей (1) и (2) против сложной альтернативы H_1 :

$$H_0 : v = v_0, \quad H_1 : v \neq v_0. \quad (3)$$

В качестве признаков для построения статистических процедур будем использовать индикаторы $\bar{s} = (s_1, \dots, s_n)$ – номера интервалов, в которые попадают невязки $y_t - g_t(\theta_0)$, т.е. $s_t = s_t(\theta_0, \mu_0) = s(y_t - g_t(\theta_0), \mu_0)$, где $s(u, \mu_0) = k$ для $u \in C_k(\mu_0)$.

Тест для проверки гипотез (3) построим по принципу максимума отношения правдоподобия, отбирая в критическую область те параметры, которые доставляют наибольшие значения нормы его градиента в гипотетической точке:

$$\nabla_v L(\bar{s} | v, v_0) \Big|_{v=v_0} = \frac{\nabla_v P(\bar{s} | v)}{P(\bar{s} | v_0)} \Big|_{v=v_0}. \quad (4)$$

Здесь $L(\bar{s} | v, v_0) = P(\bar{s} | v) / P(\bar{s} | v_0)$ – функция отношения правдоподобия индикаторных признаков.

2. Градиент правдоподобия индикаторных признаков

При поиске выражений для правой части в равенстве (4) воспользуемся представлением

$$P(s_1, \dots, s_n | v) = P(s_1 | v) P(s_2 | s_1, v) \dots P(s_{r-1} | s_1, s_2, \dots, s_{r-2}, v) \prod_{t=r}^n P(s_t | s_{t-r+1}, \dots, s_{t-1}, v), \quad (5)$$

чтобы получить разложение Тейлора $P(\bar{s} | v) = P(\bar{s} | v_0) + \nabla'_v P(\bar{s} | v_0)(v - v_0) + o(\|v - v_0\|)$. Для данного разложения требуется непрерывно дифференцируемая параметризация семейства конечномерных распределений по μ и Q . Для любой такой параметризации имеем

$$\begin{aligned} P(s_1 | v) &= P(s_1 | v_0) + \nabla'_v P(s_1 | v_0)(v - v_0) + o(\|v - v_0\|), \\ P(s_2 | s_1, v) &= P(s_2 | s_1, v_0) + \nabla'_v P(s_2 | s_1, v_0)(v - v_0) + o(\|v - v_0\|), \dots \\ \dots P(s_t | s_{t-r+1}, \dots, s_{t-1}, v) &= P(s_t | s_{t-r+1}, \dots, s_{t-1}, v_0) + \nabla'_v P(s_t | s_{t-r+1}, \dots, s_{t-1}, v_0) \times \\ &\quad \times (v - v_0) + o(\|v - v_0\|), \quad t = \overline{r, n}. \end{aligned} \quad (6)$$

Подставив (6) в (5), после перегруппировки получим

$$\begin{aligned} \frac{P(\bar{s} | v)}{P(\bar{s} | v_0)} &= 1 + \left(\frac{\nabla'_v P(s_1 | v_0)}{P(s_1 | v_0)} + \frac{\nabla'_v P(s_2 | s_1, v_0)}{P(s_2 | s_1, v_0)} + \dots + \frac{\nabla'_v P(s_{r-1} | s_1, \dots, s_{r-2}, v_0)}{P(s_{r-1} | s_1, \dots, s_{r-2}, v_0)} + \right. \\ &\quad \left. + \sum_{t=r}^n \frac{\nabla'_v P(s_t | s_{t-r+1}, \dots, s_{t-1}, v_0)}{P(s_t | s_{t-r+1}, \dots, s_{t-1}, v_0)} \right) \times (v - v_0) + o(\|v - v_0\|). \end{aligned}$$

Отсюда

$$\nabla_v L(\bar{s} | v, v_0) \Big|_{v=v_0} = \frac{\nabla_v P(s_1 | v_0)}{P(s_1 | v_0)} + \frac{\nabla_v P(s_2 | s_1, v_0)}{P(s_2 | s_1, v_0)} + \dots + \frac{\nabla_v P(s_{r-1} | s_1, \dots, s_{r-2}, v_0)}{P(s_{r-1} | s_1, \dots, s_{r-2}, v_0)} + \sum_{t=r}^n \frac{\nabla_v P(s_t | s_{t-r+1}, \dots, s_{t-1}, v_0)}{P(s_t | s_{t-r+1}, \dots, s_{t-1}, v_0)}.$$

Последнее выражение можно преобразовать таким образом, чтобы градиент отношения правдоподобия был выражен через совместные вероятности порядков r и $r-1$:

$$\nabla_v L(\bar{s} | v, v_0) \Big|_{v=v_0} = \sum_{t=r}^n \frac{\nabla_v P(s_{t-r+1}, \dots, s_t | v_0)}{P(s_{t-r+1}, \dots, s_t | v_0)} - \sum_{t=r+1}^n \frac{\nabla_v P(s_{t-r+1}, \dots, s_{t-1} | v_0)}{P(s_{t-r+1}, \dots, s_{t-1} | v_0)}. \quad (7)$$

Перейдем к задаче определения градиентов $\nabla_v P(s_{t-r+1}, \dots, s_t | v) \Big|_{v=v_0}$ и $\nabla_v P(s_{t-r+1}, \dots, s_{t-1} | v) \Big|_{v=v_0}$. Для этого необходимо получить производные по каждому из входящих в v параметров. Несмотря на то что заданным значениям v соответствует целое множество распределений на $P(\bar{s} | v)$, следующая теорема показывает, что для произвольной непрерывно дифференцируемой параметризации распределения $P(\bar{s} | v)$ по параметрам μ и Q вид градиентов не зависит от способа параметризации. Поэтому они могут использоваться для дальнейших суждений о локальных изменениях функции правдоподобия.

Теорема 1. Пусть существует r -мерная непрерывная плотность распределения $f_t(x_1, \dots, x_r)$ случайных величин $\varepsilon_{t-r+1}, \dots, \varepsilon_t$, $t = \overline{r, n}$. Тогда для произвольной непрерывно дифференцируемой параметризации распределения $P(\bar{s} | v)$ по параметрам μ и Q имеет место

$$\begin{aligned} \nabla_\theta P(s_{t-r+1}, \dots, s_t | v) \Big|_{v=v_0} &= \sum_{i=1}^r [P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_{s_{t-r+i-1}}) f_{t-r+i}(c_{s_{t-r+i-1}}) - \\ &\quad - P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_{s_{t-r+i}}) f_{t-r+i}(c_{s_{t-r+i}})] \cdot \nabla_\theta g_{t-r+i}(\theta_0). \end{aligned} \quad (8)$$

$$\begin{aligned} \nabla_\theta P(s_{t-r+1}, \dots, s_{t-1} | v) \Big|_{v=v_0} &= \sum_{i=1}^{r-1} [P_{r-1,i}(C_{s_{t-r+1}}, \dots, C_{s_{t-1}} | c_{s_{t-r+i-1}}) f_{t-r+i}(c_{s_{t-r+i-1}}) - \\ &\quad - P_{r-1,i}(C_{s_{t-r+1}}, \dots, C_{s_{t-1}} | c_{s_{t-r+i}}) f_{t-r+i}(c_{s_{t-r+i}})] \cdot \nabla_\theta g_{t-r+i}(\theta_0). \end{aligned} \quad (9)$$

$$\nabla_{\mu} P(s_{t-r+1}(\theta_0, \mu_0), \dots, s_t(\theta_0, \mu_0) | v) \Big|_{v=v_0} = \sum_{i=1}^r \left[P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_{s_{t-r+1}-1}) f_{t-r+i}(c_{s_{t-r+1}-1}) d_{s_{t-r+1}-1} - \right. \\ \left. - P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_{s_{t-r+1}}) f_{t-r+i}(c_{s_{t-r+1}}) d_{s_{t-r+1}} \right]. \quad (10)$$

$$\nabla_{\mu} P(s_{t-r+1}(\theta_0, \mu_0), \dots, s_{t-1}(\theta_0, \mu_0) | v) \Big|_{v=v_0} = \sum_{i=1}^{r-1} \left[P_{r-1,i}(C_{s_{t-r+1}}, \dots, C_{s_{t-1}} | c_{s_{t-r+1}-1}) f_{t-r+i}(c_{s_{t-r+1}-1}) d_{s_{t-r+1}-1} - \right. \\ \left. - P_{r-1,i}(C_{s_{t-r+1}}, \dots, C_{s_{t-1}} | c_{s_{t-r+1}}) f_{t-r+i}(c_{s_{t-r+1}}) d_{s_{t-r+1}} \right]. \quad (11)$$

$$\nabla_Q P(s_{t-r+1}(\theta_0, \mu_0), \dots, s_t(\theta_0, \mu_0) | v) \Big|_{v=v_0} = [G]_{j_r(s_{t-r+1}, \dots, s_t)}', \quad (12)$$

$$\nabla_Q P(s_{t-r+1}(\theta_0, \mu_0), \dots, s_{t-1}(\theta_0, \mu_0) | v) \Big|_{v=v_0} = [F_{r-1} G]_{j_{r-1}(s_{t-r+1}, \dots, s_{t-1})}'. \quad (13)$$

Здесь $f_t(\cdot)$ – плотность распределения ε_t , $P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_k)$ – условная вероятность попадания вектора $\|\varepsilon_s : s = \overline{t-r+1, t}; s \neq t-r+i\|$ в параллелепипед, образованный интервалами $\{C_{s_j} : j = \overline{t-r+1, t}; j \neq t-r+i\}$ при условии $\varepsilon_{t-r+i} = c_k$.

Доказательство. Для доказательства (8) обозначим $c_t(k | \mu_0, \theta_0, \theta) = c_k + d'_k \mu_0 - g_t(\theta) + g_t(\theta_0)$. Тогда

$$P(s_{t-r+1}, \dots, s_t | v) = \frac{c_t(s_t | \mu_0, \theta_0, \theta)}{c_t(s_t - 1 | \mu_0, \theta_0, \theta)} \dots \frac{c_{t-r+1}(s_{t-r+1} | \mu_0, \theta_0, \theta)}{c_{t-r+1}(s_{t-r+1} - 1 | \mu_0, \theta_0, \theta)} f_t(x_1, \dots, x_r) dx_1 \dots dx_r.$$

Из этого следует (8). Аналогично получаем (9).

Для получения градиентов по параметрам μ воспользуемся тем, что совместные вероятности полностью определяются параметрами Q . Поэтому если рассмотреть произвольную непрерывно дифференцируемую параметризацию конечномерных распределений по μ , то при любых v выполняется

$$\nabla_{\mu} P(s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu) | v) = \nabla_{\mu} P^{(r)}_{s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu)} = 0. \quad (14)$$

С другой стороны,

$$\nabla_{\mu} P(s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu) | v) \Big|_{v=v_0} = P'_{\mu}(s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu) | v) \Big|_{v=v_0} - \\ - \sum_{i=1}^r \left[P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_{s_{t-r+1}-1}) f_{t-r+i}(c_{s_{t-r+1}-1}) d_{s_{t-r+1}-1} - \right. \\ \left. - P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_{s_{t-r+1}}) f_{t-r+i}(c_{s_{t-r+1}}) d_{s_{t-r+1}} \right], \quad (15)$$

где P'_{μ} – градиент по параметризации самого распределения без учета зависимости аргументов от μ . Нас интересует градиент при гипотетических параметрах $\nabla_{\mu} P(s_{t-r+1}(\theta_0, \mu_0), \dots, s_t(\theta_0, \mu_0) | v) \Big|_{v=v_0} = P'_{\mu}(s_{t-r+1}(\theta_0, \mu_0), \dots, s_t(\theta_0, \mu_0) | v) \Big|_{v=v_0}$, который благодаря непрерывной дифференцируемости параметризации семейства распределений равен $P'_{\mu}(s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu) | v) \Big|_{v=v_0}$. Поэтому из (14) и (15) получаем (10). Аналогично можно получить (11) для градиента правдоподобия меньшего порядка.

Воспользуемся теперь произвольной параметризацией семейства распределений по Q :

$$\nabla_Q P(s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu) | v) \Big|_{v=v_0} = \nabla_Q P^{(r)}_{s_{t-r+1}(\theta, \mu), \dots, s_t(\theta, \mu)} \Big|_{v=v_0} = [G]_{j_r(s_{t-r+1}, \dots, s_t)}'. \quad (16)$$

Нас интересует градиент $\nabla_Q P(s_{t-r+1}(\theta_0, \mu_0), \dots, s_t(\theta_0, \mu_0) | v) \Big|_{v=v_0}$, который совпадает с (16), поскольку параметризация непрерывно дифференцируема. Отсюда имеем (12) и аналогично получаем (13).

Теорема доказана.

3. Знаковые тесты

Формально искомый тест для проверки гипотез (3) может иметь вид

$$\left\| \nabla_v L(\bar{s} | v, v_0) \Big|_{v=v_0} \right\|^2 > \text{const}, \quad (17)$$

где $\|\cdot\|$ – некоторая подходящая норма вектора, а формулы (7)–(13) дают выражение для тестовой статистики. Нетрудно убедиться, что если выбрать $p_1 = \dots = p_K = 1/K$ и гипотетические параметры Q_0 определяют равномерное r -мерное распределение (т.е. $[Q_0]_i = K^{-r}$ и гипотеза состоит в том, что зависимость отсутствует), то тест (17) будет локально наиболее мощным против любой линейной одномерной односторонней альтернативы, поскольку знаменатель отношения правдоподобия $P(\bar{s} | v_0)$ превращается в константу. В остальных случаях приходится опираться на логическую обоснованность принципа максимума отношения правдоподобия.

В качестве альтернативы для (17) можно рассматривать тест вида

$$\left\| \nabla_v P(\bar{s} | v) \Big|_{v=v_0} \right\|^2 = \left\| P(\bar{s} | v_0) \cdot \nabla_v L(\bar{s} | v, v_0) \Big|_{v=v_0} \right\|^2 > \text{const}, \quad (18)$$

который является локально наиболее мощным против любой линейной одномерной односторонней альтернативы.

Относительно тестов (17) и (18) остаются вопросы, на которые необходимо дать ответ. Во-первых, в (8)–(11) присутствуют неизвестные величины, которые требуют обоснованной замены. Во-вторых, необходимо указать, каким образом можно определять критические значения и какие нормы вектора при этом использовать.

Начнем с ответа на первый вопрос. Чтобы избавиться от неизвестных величин, входящих в состав (17), используем метод, аналогичный [7], – заменим их доступными нам величинами. Так, игнорируя некоторые эффекты зависимости, можно заменить

$$P_{r,i}(C_{s_{t-r+1}}, \dots, C_{s_t} | c_k) \text{ на } P^{(r,i)}(k_1, \dots, k_r) = \sum_{s=1}^K P^{(r)}(k_1, \dots, k_{i-1}, s, k_{i+1}, \dots, k_r).$$

После этого неизвестными останутся величины

$$(f_t(c_{k-1}) - f_t(c_k)) / p_k \text{ и } (f_t(c_{k-1})d_{k-1} - f_t(c_k)d_k) / p_k,$$

замена которых специальными весами $B_1(k)$ и $B_2(k)$ подробно обсуждается в [7]. Например, для $K = 2$ (квантильная регрессия) $B_1 = \{-1/p, 1/(1-p)\}$. При $K = 3$ (симметричная двухквантильная регрессия) $B_1 = \{-1, 0, 1\}$, $B_2 = \{1, -2p/(1-2p), 1\}$. Для $K = 3$ (симметричная трехквантильная регрессия) $B_1 = \{-A, -\alpha, \alpha, A\}$, $B_2 = \{A, -1, -1, A\}$, где $A = (1-2p)/2p$, α – априорная догадка о величине $(f_t(0) - f_t(\mu)) / f_t(\mu)$.

Заметим, что похожие «разумные» замены часто встречаются при конструировании непараметрических процедур. В ранговом анализе полученные похожим образом веса рангов называют метками (scores). Важными свойствами меток являются их нулевые средние $\sum_{k=1}^K B_1(k) p_k = 0$, $\sum_{k=1}^K B_2(k) p_k = 0$ и нулевая ковариация $\sum_{k=1}^K B_1(k) B_2(k) p_k = 0$.

В результате замен из (17) и (18) получаем тесты вида

$$\left\| \xi_n(\bar{s} | v_0) \right\|^2 > \text{const}, \quad (19)$$

$$\left\| P(\bar{s} | v_0) \cdot \xi_n(\bar{s} | v_0) \right\|^2 > \text{const}, \quad (20)$$

где $\xi_n(\bar{s} | v_0)$ – векторная статистика, которая является модифицированным и нормированным градиентом отношения правдоподобия и задается выражениями

$$\xi_n(\bar{s} | v) = n^{-1/2} \sum_{t=r}^n \sum_{i=1}^r W_{t,i}(\bar{s} | v),$$

$$W_{t,i}(\bar{s}|\mathbf{v}) = \begin{bmatrix} \nabla_{\theta} g_{t-r+i}(\theta) R_{t,1}^{(r,i)}(\mathbf{v}) \\ R_{t,2}^{(r,i)}(\mathbf{v}) \\ \frac{[G]_{j_r(s_{t-r+1}(\mathbf{v}), \dots, s_t(\mathbf{v}))}' - (1-\delta_{t,r}) \frac{[F_{r-1}G]_{j_{r-1}(s_{t-r+1}(\mathbf{v}), \dots, s_{t-1}(\mathbf{v}))}' }{P^{(r-1)}(s_{t-r+1}(\mathbf{v}), \dots, s_{t-1}(\mathbf{v}))}}{P^{(r)}(s_{t-r+1}(\mathbf{v}), \dots, s_t(\mathbf{v}))} \end{bmatrix}, \quad (21)$$

$$R_{t,l}^{(r,i)}(\mathbf{v}) = B_l(s_{t-r+i}) \left(L_t^{(r,i)}(\mathbf{v}) - (1-\delta_{t,r})(1-\delta_{i,r}) L_{t-1}^{(r-1,i)}(\mathbf{v}) \right),$$

$$L_t^{(r,i)}(\mathbf{v}) = \frac{P^{(r,i)}(s_{t-r+1}(\mathbf{v}), \dots, s_t(\mathbf{v})) P^{(1)}(s_{t-r+i}(\mathbf{v}))}{P^{(r)}(s_{t-r+1}(\mathbf{v}), \dots, s_t(\mathbf{v}))}.$$

Заметим, что тест (20), даже после замены неизвестных величин весами, остается локально наиболее мощным против линейных односторонних одномерных альтернатив, если $[Q_0]_i = K^{-r}$, но при дополнительном условии, что эти альтернативы отличаются от гипотезы только по параметрам θ , только по μ или только по Q . Однако имитационные эксперименты показывают, что поведение этого теста при альтернативах не всегда удовлетворительно. Кроме того, возникают трудности с поиском его асимптотических критических значений. Поэтому для использования мы рекомендуем тест (19)

Перейдем к ответу на второй из поставленных выше вопросов – о выборе критических значений и способов определения нормы векторных статистик. Как и для случая независимых наблюдений [1, 7], для проверки гипотез (3) можно построить тесты вида (19) и (20) с точным уровнем значимости, поскольку при гипотезе распределение статистик $\|\xi_n(\bar{s}|\mathbf{v}_0)\|^2$ и $\|P(\bar{s}|\mathbf{v}_0) \cdot \xi_n(\bar{s}|\mathbf{v}_0)\|^2$ совпадает с распределением случайных величин $\|\xi_n(\bar{\eta}|\mathbf{v}_0)\|^2$ и $\|P(\bar{\eta}|\mathbf{v}_0) \cdot \xi_n(\bar{\eta}|\mathbf{v}_0)\|^2$ соответственно, где случайный вектор $\bar{\eta} = (\eta_1, \dots, \eta_n)'$ составлен из последовательности случайных величин, образующих стационарный Марковский процесс $(r-1)$ -го прядка с конечным числом состояний $(\eta_t \in \{1, 2, \dots, K\})$ и известными конечномерными вероятностями $P(k_1, \dots, k_r | \mathbf{v}_0)$. В результате процентные точки распределения статистики тестов (19) и (20) можно определить с любой точностью, используя метод Монте-Карло. Заметим, что в вычислении $\xi_n(\bar{\eta}|\mathbf{v}_0)$ не участвуют параметры μ_0 , параметры θ_0 используются только при вычислении градиентов $\nabla_{\theta} g_t(\theta_0)$ в (21), а параметры Q_0 влияют только на величины вероятностей $P^{(r)}$, $P^{(r-1)}$, $P^{(r,i)}$, $P^{(r-1,i)}$.

При больших объемах наблюдений n лучше воспользоваться асимптотическими критическими значениями. Для теста (19) их можно вычислять на основе асимптотической нормальности распределения статистики $\xi_n(\bar{s}|\mathbf{v}_0)$ при гипотезе. Используем для этого обобщение теоремы 7.7.9 из [9] на случай последовательности случайных векторов, которое легко получить, применяя теорему 7.7.7 из [9].

Лемма 1. Пусть числовая последовательность $\{a_t : t \geq 1\}$ и последовательность случайных векторов $\{z_t : t \geq 1\}$ удовлетворяют следующим условиям:

1. Существует целое число $m > 0$ такое, что для любых n и $t_1, \dots, t_n$ ($0 < t_1 < \dots < t_n$) совокупности случайных величин $\{z_{t_1}, \dots, z_{t_n}\}$ и $\{z_1, \dots, z_{t_1-m-1}, z_{t_n+m+1}, \dots\}$ взаимно независимы.

2. $Mz_t = 0$, $t = 1, 2, \dots$

3. $M\|z_t\|^{2+\delta} < M$ для некоторого M и $\delta > 0$, $t = 1, 2, \dots$

4. $|a_t| < L$ для всех t и некоторого $L > 0$.

5. Существует предельная матрица $\Sigma = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \sum_{s=1}^T a_t a_s' Mz_t z_s'$.

Тогда случайный вектор $T^{-1/2} \sum_{t=1}^T a_t z_t$ сходится по распределению к $N(0, \Sigma)$.

Теорема 2. Пусть выполняется гипотеза (3) и следующие условия:

1. $\|\nabla_{\theta} g_t(\theta)\| < L$ для всех t и некоторого $L > 0$.

2. Существует предельная матрица $V = \lim_{n \rightarrow \infty} V_n$, где $V_n = M\xi_n(\bar{s} | v)\xi_n'(\bar{s} | v)$.

3. $P^{(r)}(k_1, \dots, k_r) > 0$ при всех $k_1, \dots, k_r$.

Тогда случайный вектор $\xi_n(\bar{s} | v)$ сходится по распределению к $N(0, V)$, а случайная величина $\zeta_n^2 = \xi_n'(\bar{s} | v)V_n^{-1}\xi_n(\bar{s} | v)$ сходится по распределению к χ_q^2 , где $q = \dim v$.

Доказательство. Если в лемме 1 положить $a_t z_t = \sum_{i=1}^r W_{t+r-1,i}(\bar{s} | v_0)$, $m = 2r$, то доказательство сводится к проверке равномерной (по t) ограниченности элементов вектора (21), а также проверке свойства $M\xi_n(\bar{s} | v) = 0$. Последнее сразу следует из того, что $\sum_{k=1}^K B_1(k)p_k = 0$, $\sum_{k=1}^K B_l(k)p_k = 0$, если применить эти равенства к вычислению $MR_{l,i}^{(r,i)}(v)$, $l \in \{1, 2\}$, в первом и втором блоке вектора (21). Для третьего блока вопрос сводится к проверке равенства $\sum_{k_1, \dots, k_r} [G]_{j_r(k_1, \dots, k_r)}' = 0$ и $\sum_{k_1, \dots, k_{r-1}} [F_{r-1} G]_{j_{r-1}(k_1, \dots, k_{r-1})}' = 0$. Мы проверяли эти условия численно в большом количестве частных случаев при $r \in \{2, \dots, 5\}$, $K \in \{2, 3, 4\}$, при выполнении условий центральной симметрии k -мерных вероятностей и $k \in \{2, \dots, r\}$ без них.

Теорема доказана.

Таким образом, для ограниченных сходящихся планов в асимптотическом тесте можно использовать статистику $\zeta_n^2 = \xi_n'(\bar{s} | v_0)V_n^{-1}\xi_n(\bar{s} | v_0)$ с критическими значениями распределения хи-квадрат. Для элементов матрицы V_n можно получить аналитические выражения, однако мы не приводим их в силу громоздкости. Отметим только, что матрица V_n зависит от параметров θ и Q , но не от μ , а для линейной модели регрессии остается зависимость только от Q .

Заключение

В данной работе получены точные и асимптотические знаковые тесты для проверки простой гипотезы (3) $H_0: v = v_0$ о параметрах модели многоквантильной регрессии (1) с Марковскими ошибками (1) порядка $r-1$. Рассмотрены наиболее интересные случаи одно-, двух- и трехквантильной регрессии.

В теореме 1 показано, что несмотря на непараметрическую постановку задачи, можно получить выражения для градиента правдоподобия знаковых признаков $\nabla P(\bar{s} | v)$, вид которого не зависит от способа параметризации распределения $P(\bar{s} | v)$. Этот факт позволяет получить тесты (19) и (20), основанные на векторной статистике $\xi_n(\bar{s} | v_0)$, см. (21). Эти тесты являются локально наиболее мощными против линейных односторонних одномерных альтернатив при условии равномерного r -мерного распределения процесса знаковых признаков.

В работе рассмотрены вопрос о вычислении критических значений, обеспечивающих заданный уровень значимости с любой точностью при конечных выборках, а также критические значения, основанные на асимптотическом распределении тестовой статистики (теорема 2) для ограниченных сходящихся планов. В результате рекомендуется тест, интерпретируемый как тест, построенный по принципу максимального отношения правдоподобия, с критической областью вида

$$\zeta_n^2 = \xi_n'(\bar{s} | v_0)V_n^{-1}\xi_n(\bar{s} | v_0) > \text{const},$$

где $V_n = V_n(\theta, Q) = M\xi_n(\bar{s} | v)\xi_n'(\bar{s} | v)$, для которого асимптотическое распределение тестовой статистики является распределением хи-квадрат и не зависит от гипотезы.

Полученный тест представляет интерес как основа для оценивания параметров v по принципу максимального асимптотического p -значения (см. основополагающую работу [8]), т.е.

$$v_n = \arg \min_{v_0} \zeta_n^2(v_0).$$

Заметим, что на практике, особенно для линейных моделей, представляет интерес проверка гипотез вида $H_0: [\theta]_j = [\theta_0]_j$, когда остальные параметры являются мешающими. Такие гипотезы можно рассматривать как линейные и применять для их проверки двухэтапные процедуры. Этот подход подробно описан в [1, 7] применительно к знаковым процедурам.

ЛИТЕРАТУРА

1. Болдин М.В., Симонова Г.И., Тюрин Ю.Н. Знаковый статистический анализ линейных моделей. М.: Наука. Физматлит, 1997. 285 с.
2. Koenker R., Bassett G. Regression quantiles // *Econometrica*. 1978. V. 46. P. 349–359.
3. Chen L.A., Tran L.T., Lin L.C. Symmetric regression quantile and its application to robust estimation for the nonlinear regression model // *J. Statist. Plann. And Infer.* 2004. V. 126. P. 423–440.
4. Coudin E., Dufour J. Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form // *Econometrics Journal*. 2009. V. 12. P. 19–49.
5. Tarassenko P.F., Tarima S.S., Zhuravlev A.V., Singh S. On sign-based regression quantiles // *J. of Statist. Comput. and Simul.* 2015. V. 85, No. 7. P. 1420–1441.
6. Dufour J., Kiviet J. Exact inference methods for first-order autoregressive distributed lag models // *Econometrica*. 1998. V. 82. P. 79–104.
7. Тарасенко П.Ф. Индикаторный статистический анализ. Томск: Изд-во Том. ун-та, 2005. 350 с.
8. Hodges J.L.Jr., Lehmann E.L. Estimates of location based on rank tests // *Ann. Math. Statist.* 1963. V. 34, No. 2. P. 528–611.
9. Андерсон Т. Статистический анализ временных рядов. М.: МИР, 1976. 756 с.

Сиренко Михаил Андреевич. E-mail: sma@mail.tsu.ru

Тарасенко Петр Феликсович, канд. физ.-мат. наук, доцент, E-mail: ptara@mail.tsu.ru

Томский государственный университет

Поступила в редакцию 28 апреля 2015 г.

Sirenko Mikhail A., Tarassenko Petr F. (Tomsk State University, Russian Federation).

Finite-sample and asymptotic sign-based tests for parameters of non-linear quantile regression with Markov noise.

Keywords: sign-based statistical procedures; non-linear regression; quantile regression; Markov process; finite-sample test.

DOI 10.17223/19988605/32/7

Sign-based statistical procedures are known to be more robust to outliers than least squares and ability to control precise significance level for finite samples when testing simple hypothesis. In this paper, the sign-based approach is extended to the case of non-linear model with dependent noise. The model of multi-quantile regression is considering, which allows to test hypotheses both on the parameters of the regression $y_t = g_t(\theta) + \varepsilon_t$, $t = \overline{1, n}$, and the parameters of the noise ε_t , which forms a stationary Markov process of order $(r-1)$. According to that, the signs of observations are calculated with respect to a set of quantiles (at specified levels) of one-dimensional distribution of ε_t .

The quantiles depend on the unknown parameter μ . For example, in a symmetric two-quantile regression, the quantiles $(-\mu)$ and μ correspond to levels of p and $1-p$, where p is given, and μ is unknown. For the three-quantile regression, the quantiles $(-\mu)$, 0 , μ correspond to levels of p , $1/2$ and $1-p$. In both cases, the parameter μ is a scale parameter of noise.

According to the sign-based approach, the residuals are substituting with indicators of their belonging to the interquantile intervals $\bar{s} = (s_1, \dots, s_n)$, where s_t takes a finite number of values. The unknown parameters in this scheme are $\nu = (\theta', \mu', Q')'$, where the vector Q contains linearly independent r -dimensional joint probabilities of the states of generated by process of indicators $\{s_t\}$. Since the problem is considering in a nonparametric setting, then each fixed value of parameters μ and Q corresponds to a class of finite-dimensional distributions of the initial process ε_t . However, we can show that for any continuous parameterization of finite-dimensional distributions, all the derivatives of the likelihood of indicators $P(\bar{s} | \nu)$ are expressed in the same way. In the problem of testing a simple hypothesis $H_0: \nu = \nu_0$ it gives the opportunity to build a test based on the principle of maximal likelihood ratio.

In this paper, we consider the problem of calculating the critical values to provide the desired significance level with any accuracy for finite samples, as well as the critical values based on the asymptotic distribution of the test statistic.

The obtained tests can be used as a basis for estimating the parameters ν by the principle of maximal p -values, as well as for the development of tests for linear hypothesis.

REFERENCES

1. Boldin, M.V., Simonova, G.I. & Tyurin, Yu.N. (1997) *Znakovyy statisticheskiy analiz lineynykh modeley* [Sign-based Methods in Linear Statistical Models]. Moscow: Nauka. Fizmatlit.
2. Koenker, R. & Bassett, G. (1978) Regression quantiles. *Econometrica*. 46. pp. 349-359.
3. Chen, L.A., Tran, L.T. & Lin, L.C. (2004) Symmetric regression quantile and its application to robust estimation for the nonlinear regression model. *Journal of Statistical Planning and Inference*. 126. pp. 423-440. DOI: 10.1016/j.jspi.2003.09.014
4. Coudin, E. & Dufour, J. (2009) Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *Econometrics Journal*. 12. pp. 19-49. DOI: 10.1111/j.1368-423X.2009.00285.x
5. Tarassenko, P.F., Tarima, S.S., Zhuravlev, A.V. & Singh, S. (2015) On sign-based regression quantiles. *Journal of Statistical Computation and Simulation*. 85 (7). pp. 1420-1441. DOI: 10.1080/00949655.2013.875176
6. Dufour, J. & Kiviet, J. (1998) Exact inference methods for first-order autoregressive distributed lag models. *Econometrica*. 82. pp. 79-104. DOI: 10.2307/2998541
7. Tarasenko, P.F. (2005) *Indikatornyy statisticheskiy analiz* [Indicator-based statistical analysis]. Tomsk: Tomsk State University.
8. Hodges, J.L.Jr. & Lehmann, E.L. (1963) Estimates of location based on rank tests. *Ann. Math. Statist.* 34 (2). pp. 528-611.
9. Anderson, T.W. (1971) *The statistical analysis of time series*. John Wiley & Sons, Inc.

М.А. Тараник, Г.Д. Копаница**АНАЛИЗ ПРОЦЕССА КОНТРОЛЯ КАЧЕСТВА ОКАЗАНИЯ
МЕДИЦИНСКОЙ ПОМОЩИ В РАМКАХ ПРОГРАММЫ
ОБЯЗАТЕЛЬНОГО МЕДИЦИНСКОГО СТРАХОВАНИЯ**

Рассматривается задача выявления проблем и выработки эффективных решений для медицинских организаций на этапе контроля отчетной медицинской документации с целью прогнозирования получения денежных средств по программе обязательного медицинского страхования. Представлена формальная модель процесса «контроль объемов, сроков, качества и условий предоставления медицинской помощи медицинскими организациями» с применением средств аппарата системного анализа. Декомпозированные подпроцессы представлены с применением диаграмм последовательности. На основании выявленных проблем сделан вывод о необходимости разработки интеллектуальной информационной системы, способной проводить оценку историй болезни на предмет получения денежных средств за оказанные медицинские услуги. В качестве метода логического вывода предложено использовать нечеткую логику и анализ прецедентов.

Ключевые слова: системный анализ; интеллектуальные информационные системы; обязательное медицинское страхование; нечеткая логика; анализ прецедентов.

Применение информационных средств анализа бизнес-процессов во многом способствует их объективному представлению и пониманию. Такой анализ является незаменимым при проектировании автоматизированных программных средств, которые позволили бы повысить эффективность функционирования исследуемого процесса. В настоящем исследовании данный подход применен к системе медицинского страхования. Особенностью российской политики в области социального страхования является система контрагентов, состоящая из территориальных фондов обязательного медицинского страхования (ОМС), основного контролирующего органа – федерального фонда (ФФ) ОМС, медицинских организаций (МО), предоставляющих медицинские услуги населению в рамках программы ОМС, а также страховых медицинских организаций (СМО), производящих оплату медицинских услуг МО. В свою очередь МО подразделяются на фондодержателей (МО-фондодержатели), которые имеют прикрепленных пациентов, а также исполнителей (МО-исполнители). МО-исполнители оказывают внешние медицинские услуги пациентам по направлению (форма 057/У-04) от фондодержателя. Для получения денежных средств за оказанные медицинские услуги по программе ОМС как МО-фондодержатель, так и МО-исполнитель обязаны ежемесячно предоставлять СМО необходимую отчетную документацию – реестр счетов, содержащий персонализированные сведения о пролеченных больных и об оказанной им медицинской помощи. По окончании необходимых проверок СМО производит оплату за оказанные медицинские услуги из средств бюджета фонда ОМС. Однако существуют также и факторы, влияющие на размер денежных выплат. Среди таких факторов наиболее важную роль играет перечень оснований для отказа в оплате медицинской помощи по результатам контроля объемов, сроков, качества и условий предоставления медицинской помощи по ОМС [1]. Таким образом, для МО, которые являются участниками программы ОМС, вопрос получения денежных средств за предоставленные медицинские услуги имеет особый приоритет. В настоящей статье представлен анализ процесса контроля объемов, сроков, качества и условий предоставления медицинской помощи, проводимый СМО, на примере реестра счетов МО-исполнителя.

При решении задач анализа бизнес-процессов наиболее подходящим инструментарием является аппарат системного анализа [2, 3], представляющий целую систему методов исследования. Так, в настоящей работе применяется методология IDEF0 [4] – одна из самых известных и широкоиспользуемых методологий моделирования [2]. Применяя данную методологию, были построены модели исследуемо-

го бизнес-процесса в виде иерархических диаграмм с определенными входными и выходными параметрами, а также сущности управления и исполнения процесса. Полученные на этапе декомпозиции подпроцессы были представлены с использованием UML-диаграмм, а именно диаграмм деятельности (activity diagram) [5]. Такое представление позволяет более детально показать последовательность выполняемых в подпроцессе действий, а также учесть логические компоненты.

Для более детального описания процесса экспертизы качества медицинской помощи и выявления в этом процессе персонифицированных и нечетких экспертных знаний были выбраны 10 случаев лечения для независимой оценки двумя экспертами. Были выявлены и задокументированы разногласия между экспертами. Также был произведен расчет каппы Коэна [6] для оценки уровня разногласий.

1. Основные диаграммы анализа процесса контроля качества оказания медицинской помощи в рамках программы ОМС

Результатом анализа исследуемого бизнес-процесса является его формальное описание с применением диаграмм деятельности, а также диаграмм IDEF0 в терминах естественного языка. Использование диаграмм деятельности позволяет декомпозировать подпроцессы, выявленные на этапе построения моделей в нотации IDEF0, тем самым повысить точность их описания. Начальным пунктом исследования процесса «контроля объема, сроков качества и условий предоставления медицинской помощи медицинскими организациями» [7] стало построение уровня A0 диаграммы IDEF0 (рис. 1), что позволило определить входные, выходные параметры процесса, а также обозначить его исполнителей и элементы, отвечающие за управление.

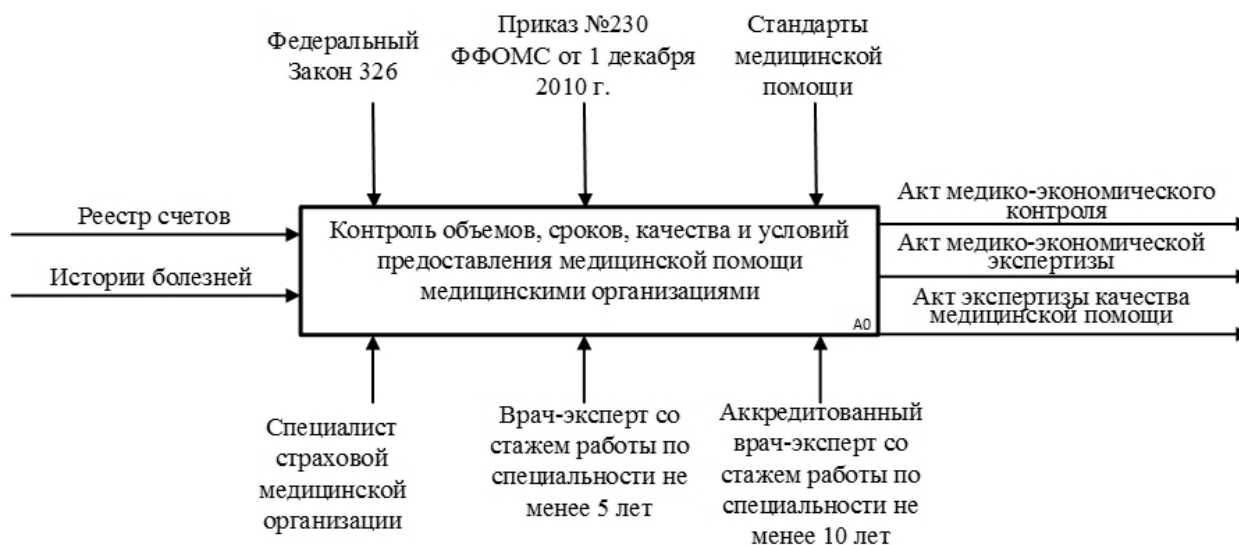


Рис. 1. Уровень A0 процесса контроля объемов, сроков, качества и условий предоставления медицинской помощи медицинскими организациями

Входными параметрами процесса являются реестр счетов, формируемый МО-исполнителем для получения денежных средств за указанные медицинские услуги, а также часть историй болезней по случаям оказания медицинской помощи, коррелирующих с данными поданного реестра. В качестве выходных параметров определены акты медико-экономического контроля, медико-экономической экспертизы, а также экспертизы качества медицинской помощи. Акты формируются по окончании проведения каждого из этапов исследуемого процесса. Управлением являются такие нормативные документы, как Федеральный закон № 326 «Об обязательном медицинском страховании» [7], приказ № 230 ФФОМС от 01 декабря 2010 г., а также федеральные и региональные стандарты медицинской помощи. Реализацию процесса обеспечивают следующие лица: специалист страховой медицинской организации, врач-эксперт со стажем работы по специальности не менее 5 лет и аккредитованный врач-эксперт со стажем работы по специальности не менее 10 лет.

Проводя декомпозицию уровня А0 исследуемого процесса посредством перехода и построения уровня А1, мы получим следующую диаграмму, представленную на рис. 2.

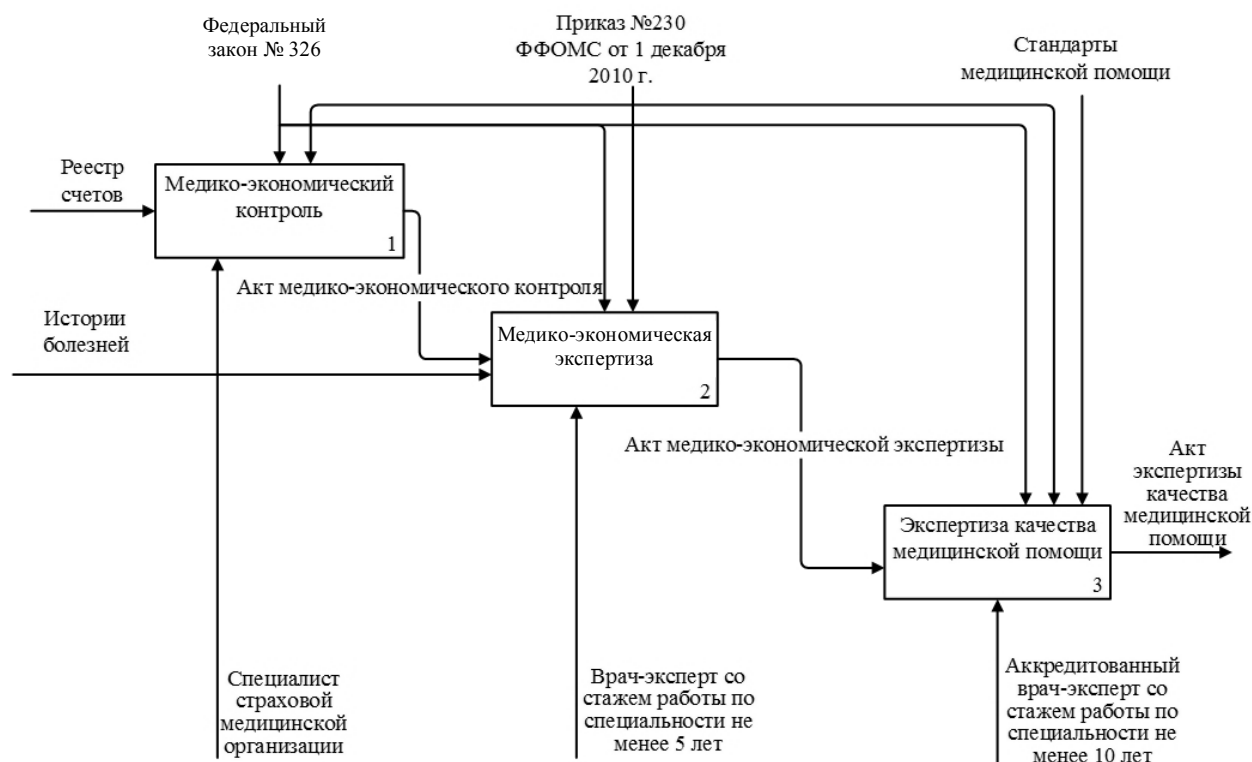


Рис. 2. Уровень А1 процесса контроля объемов, сроков, качества и условий предоставления медицинской помощи медицинскими организациями

Основной исследуемый процесс состоит из трех подпроцессов: медико-экономический контроль, медико-экономическая экспертиза, экспертиза качества медицинской помощи. Все подпроцессы являются последовательными, а выходные данные – входными для последующего процесса.

При реализации процесса «контроля объема, сроков, качества и условий предоставления медицинской помощи медицинскими организациями» сперва выполняется этап медико-экономического контроля (МЭК), состоящий из четырех итераций, представленных на рис. 3.

Данный этап осуществляется при помощи автоматизированного программного обеспечения (ПО) под контролем специалиста страховой медицинской организации в соответствии с [8, 9]. Алгоритм ПО проводит проверку сформированного МО отчетного реестра счетов по следующим позициям:

- Проверка на обязательность заполнения полей.
- Проверка на соответствие справочникам.
- Проверка на соответствие с эталонными таблицами.
- Проверка с помощью составных SQL запросов.

Наличие несоответствий является предпосылкой обоснованного отказа в стопроцентной выплате денежных средств МО по текущему реестру. В результате формируется акт медико-экономического контроля, в котором отражены найденные нарушения и замечания, а сам реестр возвращается в МО. Этап МЭК проводится по текущему реестру до тех пор, пока МО не предоставит на проверку реестр без нарушений. Для дальнейшего этапа медико-экономической экспертизы (МЭЭ) отбирается часть историй болезни, с которыми работает эксперт страховой организации. Диаграмма деятельности данного этапа представлена на рис. 4.

На данном этапе эксперт осуществляет проверку отобранных у МО историй болезни на наличие дефектов оформления медицинской документации. Среди таких нарушений наиболее часто отсутствие листа назначений, температурного листа, флюорографии. Отсутствие данных документов при плановой госпитализации пациента является нарушением.

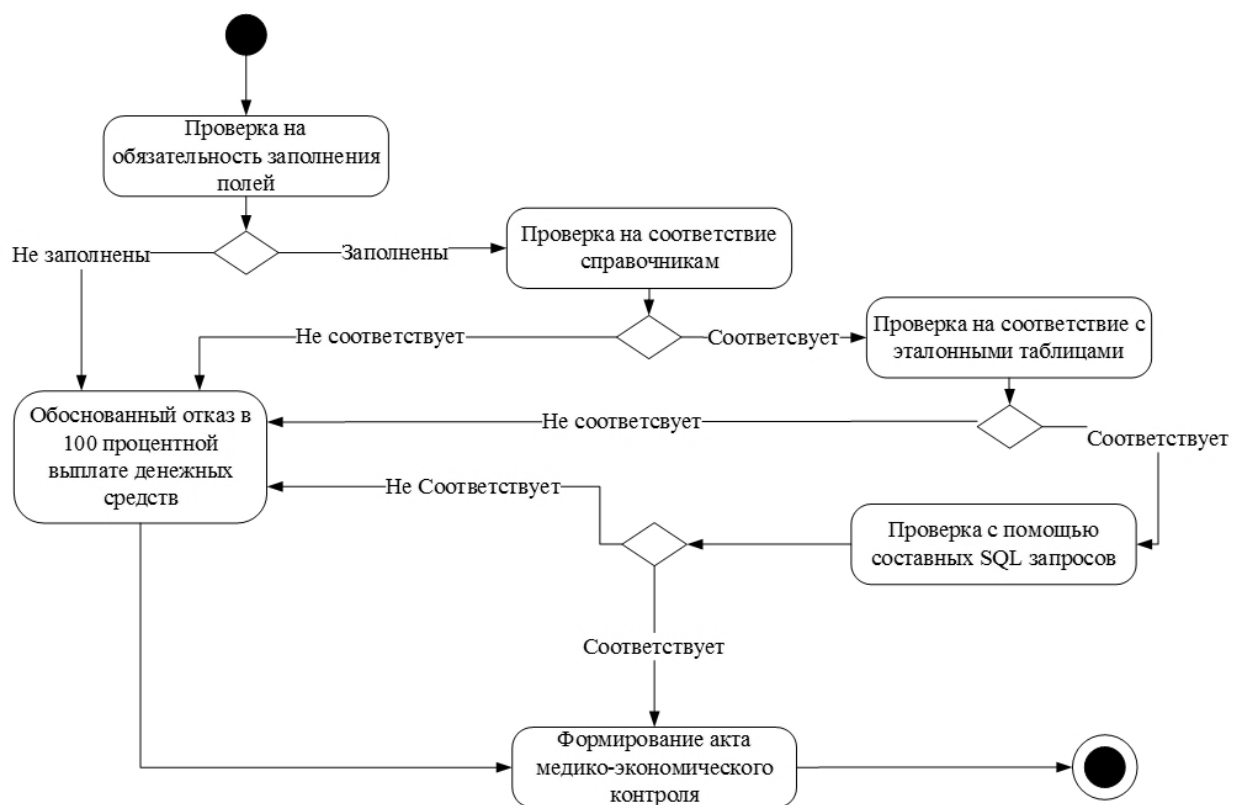


Рис. 3. Диаграмма деятельности этапа проведения медико-экономического контроля

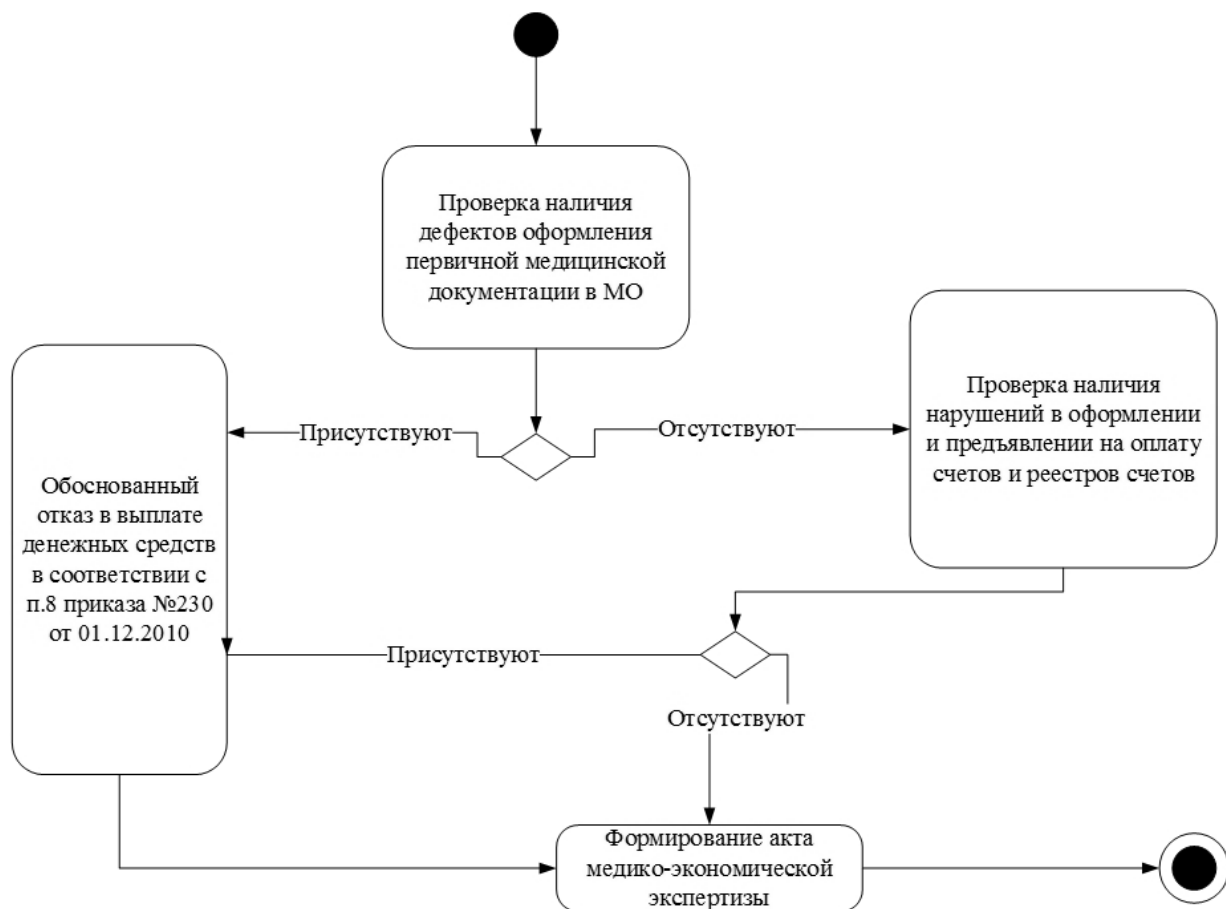


Рис. 4. Диаграмма деятельности этапа проведения медико-экономической экспертизы

Также эксперт осуществляет проверку на наличие нарушений в оформлении и предъявлении на оплату счетов и реестров счетов. При выявлении нарушений следует обоснованный отказ в выплате денежных средств на сумму, установленную в п. 8 приказа ФФ ОМС № 230 от 01.12.2010 г. По окончании этапа МЭЭ составляется соответствующий акт, а истории болезни передаются аккредитованному врачу-эксперту со стажем работы по специальности не менее 10 лет для проведения экспертизы качества медицинской помощи (рис. 5).

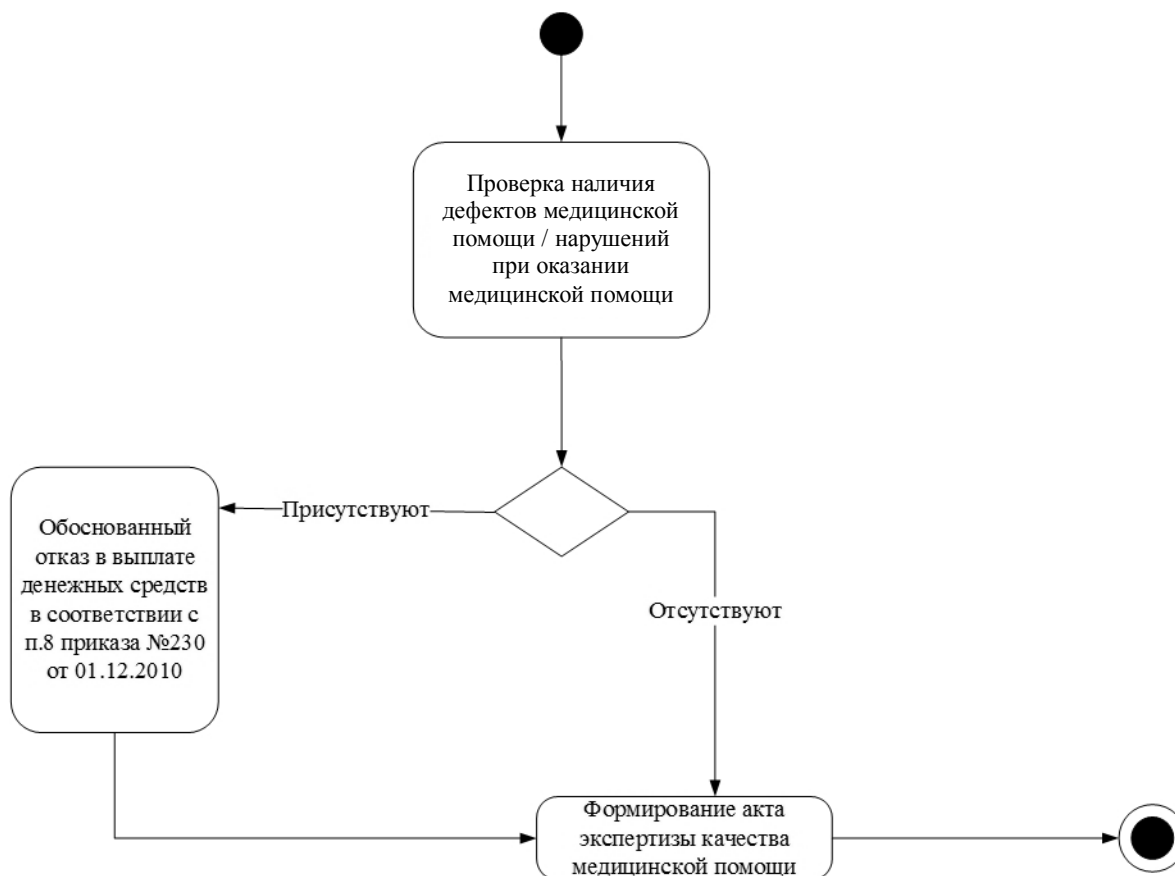


Рис. 5. Диаграмма деятельности этапа проведения экспертизы качества медицинской помощи

Т а б л и ц а 1

Анализ этапов проверки исследуемого процесса

Этап проверки	Условие проверки	Исполнитель
Проверка на обязательность заполнения полей	Заполнены / не заполнены	Программа (Специалист СМО)
Проверка на соответствие справочникам	Соответствует / не соответствует	Программа (Специалист СМО)
Проверка на соответствие эталонным таблицам	Соответствует / не соответствует	Программа (Специалист СМО)
Проверка с помощью составных SQL запросов	Соответствует / не соответствует	Программа (Специалист СМО)
Проверка на наличие дефектов оформления первичной документации	Присутствуют / отсутствуют	Врач-эксперт СМО
Проверка на наличие нарушений в оформлении и представлении на оплату счетов и реестров счетов	Присутствуют / отсутствуют	Врач-эксперт СМО
Проверка на наличие дефектов медицинской помощи / нарушений при оказании медицинской помощи	Присутствуют / отсутствуют	Аккредитованный врач-эксперт

Врач-эксперт на данном этапе осуществляет проверку историй болезни на наличие дефектов медицинской помощи, а также нарушений при оказании медицинской помощи, руководствуясь федеральными и региональными стандартами медицинской помощи, а также личным опытом. Так, например, отсутствие листа назначений и температурного листа при плановой госпитализации пациента является невыполнением необходимых лечебно-диагностических мероприятий, не повлиявших на здоровье.

Также, основываясь на личном опыте врача-эксперта, даются комментарии и пояснения по выявленным нарушениям. Зафиксированные нарушения отражаются в акте экспертизы качества медицинской помощи, а сумма выплаты МО уменьшается в соответствии с п. 8 Приказа № 230 ФФ ОМС от 01.12.2010 г.

Актуальность прогнозирования определяет актуальность разработки информационной системы, способной реализовать алгоритм оценки истории болезни МО-исполнителя на предмет получения денежных средств за оказанные медицинские услуги в рамках программы ОМС, принимая за основу процесс контроля объемов, сроков, качества и условий предоставления медицинской помощи. Проводя анализ исполнителей (см. табл. 1) на каждом этапе исследуемого процесса, можно сделать вывод о влиянии человеческого фактора на трех основных итерациях.

2. Применение индекса каппы Коэна для расчёта коэффициента согласованности между двумя врачами-экспертами СМО

Наличие человеческого фактора при проведении контроля, а также сложности формального описания исследуемой предметной области является предпосылкой возникновения нечеткости и неопределенности. Для более точного определения наличия неопределенности воспользуемся индексом каппы Коэна [6], который обозначает меру согласованности двух оценщиков, классифицирующих n объектов по s категориям. Индекс рассчитывается по следующей формуле:

$$k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}, \quad (1)$$

где $\text{Pr}(a) = \frac{\sum_{i=1}^s n_{ii}}{n}$ – наблюдаемая согласованность между экспертами А и Б, классифицирующими n объектов по s категориям; $\text{Pr}(e) = \frac{\sum_{i=1}^s n_i m_i}{n^2}$ – ожидаемая вероятность случайной согласованности. Если $k > 0,75$, то согласованность считается высокой, если $0,4 < k \leq 0,75$ – то хорошей, иначе – плохой. В качестве примера приведем вариант экспертизы двух врачей-экспертов десяти историй болезни. Каждый из экспертов провел экспертизу качества медицинской помощи следующих клинических ситуаций, представленных в табл. 2.

Т а б л и ц а 2

Описание клинических ситуаций и действий медицинского персонала

Диагноз / клиническая ситуация	Действия медицинского персонала
Атеросклероз артерий нижних конечностей. Хроническая артериальная недостаточность нижних конечностей	Было проведено исследование пациента доплерографией, а выполнена остеопарация
Патология нижних конечностей	Пациент был направлен на медико-социальную экспертизу
Плановая операция холецистэктомии	Пациент был госпитализирован по направлению из поликлиники без необходимого набора анализов. В условиях стационара был проведен полный необходимый перечень исследований. Плановая операция про- шла успешно, пациент выписан
Тромбофлебит	Был произведен забор анализа крови. Пациент был направлен на дальнейшие консультации специалистов
Кровотечение желудочно-кишечного тракта	Пациенту назначено консервативное лечение без проведения эзофагогастродуоденоскопии (ЭГДС)
Острый калькулезный холецистит	Поступившему экстренному пациенту не было проведено срочное опе- ративное вмешательство. Пациент был помещен в стационар
Киста передней поверхности шеи	Пациенты были обследованы в условиях стационара, а также успешно прооперированы и выписаны
Дефект передней грудной стенки	
Постлучевой фиброз левой молочной железы	
Нейропатия правого плечевого сплетения	

При проведении экспертизы специалистам необходимо сделать заключение по каждой клинической ситуации на предмет назначения уменьшения суммы выплаты при обнаружении нарушений либо неприменения экономических санкций к МО и дать рекомендации на выплату полной денежной суммы. Базис расхождений мнений экспертов по клиническим ситуациям заключается в том, что специалисты руководствуются не только стандартами оказания медицинской помощи, но и личным опытом. Разъяснения экспертов А и Б по клиническим ситуациям представлены в табл. 3.

Разъяснения экспертов по клиническим ситуациям

№ ситуации	Разъяснения экспертов
1	Эксперт А: в истории болезни (ИБ) отсутствуют мероприятия непрерывной консервативной терапии (сосудистой терапии), наличие которой необходимо при данном диагнозе. Вывод: невыполнение необходимых лечебно-диагностических мероприятий (ЛДМ), не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: отсутствуют мероприятия непрерывной консервативной терапии, наличие которой необходимо при данном диагнозе. Результаты доплерографии носят формальный характер. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
2	Эксперт А: в ИБ отсутствует заключение ангиохирурга. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: при диагнозе патологии нижних конечностей с действиями медицинских специалистов согласен. Вывод: согласие эксперта. Результат: выплата полной денежной суммы по данному случаю
3	Эксперт А: несмотря на нарушения плановой госпитализации, тактика медицинской помощи верна. Пациент находился в отдаленном районе, действия медицинского персонала оправданы. Вывод: согласие эксперта. Результат: выплата полной денежной суммы по данному случаю Эксперт Б: отсутствие проведенных в поликлинике исследований означает нарушение порядка плановой госпитализации пациента. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
4	Эксперт А: при диагнозе тромбозов необходимо ультразвуковое исследование (УЗИ), которое не было оказано. Причиной является отсутствие прибора УЗИ и специалиста. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье, которые были обоснованы отсутствием оборудования. Результат: выплата полной денежной суммы по данному случаю
	Эксперт Б: в ИБ отсутствуют результаты УЗИ. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
5	Эксперт А: в ИБ нет результатов ЭГДС, что не позволяет определить верную тактику лечения. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: отсутствуют результаты ЭГДС. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
6	Эксперт А: помещенному в стационар пациенту была оказана консервативная помощь, позволившая купировать процесс. Признаки острого деструктивного холецистита отсутствуют. Вывод: согласие эксперта. Результат: выплата полной денежной суммы по данному случаю
	Эксперт Б: в проведении экстренного оперативного вмешательства не было необходимости. В результате консервативной терапии наблюдается уменьшение в размере желчного пузыря, отсутствие слоистости стенки, а также приведение лейкоцитов в норму. Вывод: согласие эксперта. Результат: выплата полной денежной суммы по данному случаю
7	Эксперт А: отсутствует обязательная флюорография. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: нет флюорографии, что для плановой госпитализации является обязательным. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
8	Эксперт А: отсутствие листа назначений, также температурного листа. Вывод: дефект оформления медицинской документации, препятствующий экспертизе. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: нет листа назначений, а также температурного листа. Вывод: дефект оформления медицинской документации, препятствующий экспертизе. Результат: уменьшение суммы оплаты по данному случаю
9	Эксперт А: отсутствие обязательной флюорографии. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: нет обязательной флюорографии. Вывод: невыполнение необходимых ЛДМ, не повлиявших на здоровье. Результат: уменьшение суммы оплаты по данному случаю
10	Эксперт А: отсутствие листа назначений, а также температурного листа. Вывод: дефект оформления медицинской документации, препятствующий экспертизе. Результат: уменьшение суммы оплаты по данному случаю
	Эксперт Б: нет листа назначений, а также температурного листа. Вывод: дефект оформления медицинской документации, препятствующий экспертизе. Результат: уменьшение суммы оплаты по данному случаю

Таким образом, мнение обоих экспертов совпало по 7 историям болезни, по которым необходимо применить финансовые санкции в виде уменьшения суммы оплаты. Среди клинических случаев, по которым можно производить выплату в полном объеме, оба эксперта сошлись лишь в одном случае. Расхождения экспертов возникли в случаях 2–4. Поместим полученные результаты в табл. 4, в которой обозначим «+» уменьшение суммы выплаты, т.е. наличие санкций, а «–» – их отсутствие, т.е. выплату в полном объеме.

Т а б л и ц а 4

Общие результаты оценок экспертов

		Эксперт Б	
		+	–
Эксперт А	+	6	1
	–	2	1

В этом примере параметр согласованности $Pr(a) = 0,7$. Эксперт А определил 7 клинических случаев, по которым необходимо уменьшить конечную сумму выплаты, а по 3 случаям из 10 – осуществить полную выплату. Таким образом, его решения распределились в процентном соотношении 70 и 30%. В 8 одобренных случаях, для которых необходимы экономические санкции, определенные экспертом Б, всего 2 случая заслуживают выплаты полной стоимости денежных средств. Итог эксперта Б: 80 и 20%. Ожидаемая вероятность случайной согласованности $Pr(e) = 0,62$, коэффициент каппы Козна $k = 0,21$.

Данный показатель обосновывает низкую согласованность между двумя экспертами в вышеприведенном примере и дополнительно подтверждает наличие неопределенности.

Следовательно, возникает необходимость использования соответствующих математических аппаратов в алгоритме проектируемой системы. Эффективная реализация достоверного логического вывода в таких условиях может быть реализована с применением нечеткой логики [10], которая активно применяется в создании средств искусственного интеллекта для решения клинических задач [11–16]. Данная технология позволит не только достаточно точно формализовать как медицинские знания экспертов, так и входные данные для работы системы, но и определить соответствующие вероятностные коэффициенты логического вывода. Однако, помимо реализации эффективного логического вывода в условиях неопределенности, необходимо также учитывать ранее накопленный опыт. Одной из таких технологий является case-based reasoning (cbr) [17–21]. Использование данной технологии в совокупности с нечеткой логикой может позволить более точно осуществлять прогнозирование историй болезни на предмет выплаты денежных средств. Таким образом, появляется необходимость формирования базы прецедентов результирующих параметров системы при соответствующих входных данных. Автоматизация пополнения такой базы обеспечит реализацию процесса самообучаемости системы. Более того, анализ базы прецедентов может быть использован для динамического формирования коэффициентов, применяемых на этапе нечеткого логического вывода, что позволит сделать систему более адаптированной и повысить ее эффективность.

Заключение

Представленное в настоящей статье исследование процесса «контроля объема, сроков качества и условий предоставления медицинской помощи медицинскими организациями» с применением методов аппарата системного анализа позволяет сделать выводы о необходимости реализации интеллектуальной информационной системы (ИИС) оценки историй болезни на предмет получения денежных средств за оказанные медицинские услуги в рамках программы ОМС. Необходимость принятия решений в условиях неопределенности, обоснованная наличием влияния человеческого фактора, обуславливает использование нечеткой логики в качестве метода логического вывода будущей системы. Наряду с применением эффективного логического вывода использование опыта на основе прецедентов позволит эффективно повысить качество вывода для принятия управленческих решений. Реализация такой ИИС, содержащей в своей основе комбинированный метод, позволит производить оценку вероятности получения денежных средств до формирования реестра счетов.

ЛИТЕРАТУРА

1. Приказ ФФОМС № 230 «Об утверждении Порядка организации и проведения контроля объемов, сроков, качества и условий предоставления медицинской помощи по обязательному медицинскому страхованию» от 01 декабря 2010 г.
2. Силич В.А., Силич М.П. Теория систем и системный анализ : учеб пособие. Томск : Томский политехнический университет, 2010. 281 с.
3. Перегудов Ф.И., Тарасенко Ф.П. Основы системного анализа : учеб. 2-е изд., доп. Томск : Изд-во НТЛ, 1997. 396 с.
4. Методология IDEF0. Стандарт. Русская версия. М. : Метатехнология, 1993. 107 с.
5. Буч Г., Рамбо Д., Джекобсон А. Язык UML. Руководство пользователя / пер. с англ. М. : ДМК, 2000. 432 с.
6. Cohen's kappa. URL: http://en.wikipedia.org/wiki/Cohen%27s_kappa (дата обращения: 15.05.2015).
7. Федеральный закон № 326 «Об обязательном медицинском страховании в Российской Федерации» от 29 ноября 2010 г.
8. Федеральный закон № 323 «Об основах охраны здоровья граждан в Российской Федерации» от 21 ноября 2011 г.
9. Приложение № 5 к Регламенту проведения предварительного медико-экономического контроля реестров счетов на оплату медицинской помощи, представляемых в электронном виде, утвержденному приказом ТФОМС Томской области от 13.11.2013 № 265.
10. Zadeh L.A. Fuzzy Sets // Information and Control. 1965. No. 8. P. 338–353.
11. Тараник М., Копаница Г. Анализ задач и методов построения интеллектуальных медицинских систем // Врач и информационные технологии. 2014. № 3. С. 6–12.
12. Anooj P. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules // Journal of King Saud University – Computer and Information Science. 2012. No. 24. P. 27–40.
13. Samuel O., Omisore M., Ojokoh B. A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever // Expert Systems with Applications. 2013. No. 40. P. 4164–4171.
14. Castanho M., Hernandez F. Re A., Rautenberg S., Billis A. Fuzzy expert system for predicting pathological stage of prostate cancer // Expert Systems with Applications. 2013. No. 40. P. 466–470.
15. Pal D., Mandana K., Pal S. [etс.]. Fuzzy expert system approach for coronary artery disease screening using clinical parameters // Knowledge-Based Systems. 2012. No. 36. P. 162–174.
16. Castillo O., Melin P., Ramirez E., Soria J. Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system // Expert Systems with Applications. 2012. No. 39. P. 2947–2955.
17. Haghighi P., Burstein F., Zaslavsky A., Arbon P. Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings // Decision Support Systems. 2013. No. 54. P. 1192–1204.
18. Bequm S., Barua S., Ahmed M.U. Physiological sensor signals classification for healthcare using sensor data fusion and case-based reasoning // Sensors. 2014. No. 14. P. 11770–11785.
19. Gonzalez C., Lopez D.M., Blobel B. Case-based reasoning in intelligent Health Decision Support Systems // pHHealth 2013: Studies in Health Technology and Informatics. No. 189. P. 44–49.
20. Sharaf-El-Deen D.A., Moawad I.F., Khalifa M.E. A new hybrid case-based reasoning approach for medical diagnosis system // Journal of Medical Systems. 2014. No. 38 (9). P. 8–19.
21. El-Sappagh SH, El-Masri S, Elmogy M. [etс.]. An ontological case base engineering methodology for diabetes management // Journal of Medical Systems. 2014. No. 38(8). P. 66–80.

Тараник Максим Алексеевич. E-mail: taranik@tpu.ru

Копаница Георгий Дмитриевич, канд. техн. наук. E-mail: georgy.kopanitsa@gmail.com

Томский политехнический университет

Поступила в редакцию 3 апреля 2015 г.

Taranik Maksim A., Kopanitsa Georgy D. (Tomsk Polytechnic University, Russian Federation).

Analysis of the process of control quality of medical service in the scope of compulsory health insurance program.

Keywords: system analysis; intellectual information systems; medical insurance; fuzzy logic; case-based reasoning.

DOI 10.17223/19988605/32/8

The relevance of the presented paper deals with the necessity of determining problems and effective solutions for medical organizations on the stage of medical documentation reports control purposely for forecasting the average of financial resources that can be obtained in the scope of compulsory health insurance program.

The aim of the study. For the purpose of further issues definition to present formal model of the analyzed process using a set of system analysis methods.

The methods. System analysis methods, especially IDEF0 diagrams and activity diagrams; for estimation of medical expert's agreement Cohen's kappa was used.

The results. Based on the specification documents and expert's experience the spread description on the process «the control of volume, duration, quality and conditions of medical service assignment by medical organizations» provided by medical insurance organization is presented. Input and output parameters, elements of process management were determined. As a result of decomposition, subprocesses were presented within activity diagrams.

Conclusions. The obtained results allows to conclude that there is a set of problems which appear when medical organizations send reports for getting financial resources for clinical service realization in the scope of compulsory health insurance program. On the grounds of determined problems, we can conclude that it is necessary to develop an intellectual information system for estimating clinical records concerning getting financial resources for clinical service. In respect that human factor influences on the main stages of the analyzed process, we propose to use fuzzy logic as an inference engine. The self-learning function of the system will provide case-based reasoning.

REFERENCES

1. Federal Compulsory Medical Insurance Fund. (2010) *The order of the Federal Compulsory Medical Insurance Fund (FCMIF) № 230 (2010) "On approval of the organization and control of the volume, timing, quality and conditions of providing medical care for compulsory health insurance", dated December 1, 2010.*
2. Silich, V.A. & Silich, M.P. (2010) *Teoriya sistem i sistemnyy analiz* [System theory and system analysis]. Tomsk: Tomsk Polytechnic University.
3. Peregudov, F.I. & Tarasenko, F.P. (1997) *Osnovy sistemnogo analiza* [Introduction to system analysis]. 2nd ed. Tomsk: NTL.
4. Metodologiya IDEF0. (1993) Metodologiya IDEF0. Standart. Russkaya versiya [IDEF0 Methodology. Standard. Russian Version]. Moscow: Metatekhnologiya.
5. Buch, G., Rambo, D. & Dzhekbson, A. (2000) *Yazyk UML* [UML]. Translated from English. Moscow: DMK.
6. Wikipedia.org. *Cohen's kappa*. [Online] Available from: http://en.wikipedia.org/wiki/Cohen%27s_kappa. (Accessed: 15th May 2015).
7. Russian Federation. (2010) The Federal Law №326 "On compulsory medical insurance in the Russian Federation" dated 29th November 2010. (In Russian).
8. Russian Federation. (2011) The Federal Law №323 "On the basis of the health of citizens in the Russian Federation" dated 21st November 2011. (In Russian).
9. Tomsk Compulsory Medical Insurance Fund. (2013) Appendix №5 to the Regulations # 265 of the pre-medical and economic control registers invoices for medical care to be submitted in electronic form, approved by the Order of Territorial Compulsory Medical Insurance Fund of Tomsk region from 13th November 2013. (In Russian).
10. Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*. 8. pp. 338-353. DOI: 10.1016/S0019-9958(65)90241-X
11. Taranik, M. & Kopanitsa, G. (2014) Analiz zadach i metodov postroeniya intellektual'nykh meditsinskikh sistem []. *Vrach i informatsionnye tekhnologii – Information technologies for the Physician*. 3. pp. 6-12.
12. Anooj, P. (2012) Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University – Computer and Information Science*. 24. pp. 27-40.
13. Samuel, O., Omisore, M. & Ojokoh, B. (2013) A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever. *Expert Systems with Applications*. 40. pp. 4164-4171. DOI: 10.1016/j.eswa.2013.01.030
14. Castanho, M., Hernandez, F., Re, A., Rautenberg, S. & Billis, A. (2013) Fuzzy expert system for predicting pathological stage of prostate cancer. *Expert Systems with Applications*. 40. pp. 466-470. DOI: 10.1016/j.eswa.2012.07.046
15. Pal, D., Mandana, K., Pal, S. et al. (2012) Fuzzy expert system approach for coronary artery disease screening using clinical parameters. *Knowledge-Based Systems*. 36. pp. 162-174. DOI: 10.1016/j.knosys.2012.06.013
16. Castillo, O., Melin, P., Ramirez, E. & Soria, J. (2012) Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system. *Expert Systems with Applications*. 39. pp. 2947-2955.
17. Haghighi, P., Burstein, F., Zaslavsky, A. & Arbon, P. (2013) Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings. *Decision Support Systems*. 54. pp. 1192-1204. DOI: 10.1016/j.dss.2012.11.013
18. Bequm, S., Barua, S. & Ahmed, M.U. (2014) Physiological sensor signals classification for healthcare using sensor data fusion and case-based reasoning. *Sensors*. 14. pp. 11770-11785. DOI: 10.3390/s140711770
19. Gonzalez, C., Lopez, D.M. & Blobel, B. (2013) Case-based reasoning in intelligent Health Decision Support Systems. *Health 2013: Studies in Health Technology and Informatics*. 189. pp. 44-49.
20. Sharaf-El-Deen, D.A., Moawad, I.F. & Khalifa, M.E. (2014) A new hybrid case-based reasoning approach for medical diagnosis system. *Journal of Medical Systems*. 38 (9). pp. 8-19. DOI: 10.1007/s10916-014-0009-1
21. El-Sappagh, SH, El-Masri, S, Elmogy, M. et al. (2014) An ontological case base engineering methodology for diabetes management. *Journal of Medical Systems*. 38(8). pp. 66-80. DOI: 10.1007/s10916-014-0067-4

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

УДК 004.415.2

DOI 10.17223/19988605/32/9

М.Н. Головчинер, В.Ю. Юшин**АВТОМАТИЗИРОВАННАЯ СИСТЕМА УЧЕТА И КОНТРОЛЯ
ПОСТАНОВЛЕНИЙ УЧЕНОГО СОВЕТА ВУЗА**

Рассматриваются автоматизированная система учета и контроля документооборота, связанного с постановлениями Ученого совета вуза, на примере Томского государственного университета (ТГУ); проблемы документооборота, связанного с постановлениями Ученого совета ТГУ. Проанализированы основные задачи автоматизации, решаемые данной системой, и набор предоставляемых возможностей. Также рассмотрена обобщенная структурно-функциональная схема и приведен стек технологий реализации.

Ключевые слова: электронный документооборот; система контроля; Ученый совет; стек технологий.

Системы электронного документооборота (СЭД) формируют новое поколение систем автоматизации предприятий. Среди основных функций средств электронного документооборота выделяют [1]:

- согласование и утверждение документов;
- контроль исполнения документов и поручений;
- электронный архив документов и т.д.

Постоянная работа с документами ведётся и в Томском государственном университете, и весомую часть в этой работе занимают заседания Ученого совета (УС), которые проходят с определенной периодичностью. Результатом является список постановлений, взятых на контроль. Секретарем Ученого совета создается текстовый документ, содержащий перечень этих постановлений. Подразделения университета должны выполнить поставленные УС задачи и представить соответствующий отчет. С течением времени происходит накопление банка постановлений УС (ПУС) и отчетов подразделений, что усложняет работу по отслеживанию исполнения решений. Кроме того, как секретарю, так и подразделениям требуется выполнять одинаковую монотонную работу по заполнению вручную и просмотру десятков документов, очень похожих друг на друга.

Очевиден вывод о недостатках организации процесса работы с ПУС. Было принято решение о создании специализированной системы электронного документооборота.

Данная система выполняет следующие функции автоматизации:

- 1) разовую регистрацию ПУС, т.е. исключение необходимости многократного копирования и распечатки постановлений;
- 2) хранение всех постановлений в единой базе, что позволяет избежать потери бумаг или возникновения дубликатов;
- 3) доступ к системе с любого компьютера, имеющего подключение к Интернету, что значительно повышает оперативность;
- 4) возможность прикрепления к рассылаемым подразделениям задачам вспомогательных информационных файлов, что избавляет от необходимости передачи кип бумаг или самостоятельного поиска подразделениями соответствующих документов и форм;
- 5) внутрисистемный обмен сообщениями для упрощения коммуникации, необходимой для решения возникающих вопросов;
- 6) возможность архивирования как постановлений УС, так и задач, поставленных перед подразделениями;
- 7) простую систему контроля исполнения, позволяющую безотлагательно и эффективно взаимодействовать подразделениям;

8) своевременное напоминание о невыполненных задачах, срок сдачи которых уже прошел, что уменьшает возможный ущерб от подобных ситуаций;

9) прозрачность деятельности, относящейся к исполнению решений УС, и как результат – возможность качественного анализа этой деятельности для поиска узких мест.

Анализ отечественных СЭД [2, 3] показывает, что наряду с высокой стоимостью они не решают в комплексе всех перечисленных задач. Поэтому, по мнению авторов, предлагаемая в статье система представляет определенную практическую ценность.

2. Концептуальная модель системы

В системе определены четыре типа пользователей:

- Секретарь;
- Контролирующий;
- Исполнитель;
- Администратор.

Ниже перечислены возможности, предоставляемые системой различным типам пользователей.

Для Секретаря Ученого совета:

- унифицированные механизмы накопления и хранения всей информации о решениях УС;
- надежное хранение и гарантия, что нужная информация никогда не потеряется;
- единые средства контроля исполнения решений УС, интегрированные со средствами календарного планирования. Отсутствие необходимости отслеживать своевременность исполнения задания, система предоставляет всю необходимую информацию;

- централизованные механизмы поиска информации;
- возможность отслеживать процессы выполнения решений УС в подразделениях ТГУ.

Для Контролирующего (как правило, проректора):

- подтверждение выполнения задачи;
- добавление файлов к задачам;
- в случае неудовлетворительных результатов снятие статуса выполнения.

Для Исполнителя:

- общие средства доступа;
- унифицированный интерфейс, нет необходимости в дополнительном обучении представителей подразделений;
- полная поддержка интернет-доступа, пользователь может работать с любого компьютера в сети Интернет;

Для Администратора:

- единая инфраструктура управления;
- не требуется времени на установку и обновление клиентских приложений;
- веб-доступ ко всем функциям системы;
- единая система разграничения прав доступа.

Основной набор требований к системе представлен на диаграмме прецедентов [4] (рис. 1), а аспекты взаимодействия различных категорий пользователей с системой – на рис. 2.

Термин «задача» в системе обозначает объем работ, который должно выполнить подразделение по реализации решения УС.

3. Архитектура системы

Архитектура системы изображена на рис. 3.

«Контроль доступа». Данный модуль определяет набор полномочий для конкретного пользователя и предоставляет функционал для авторизации в системе.

«Базовый функционал» – модуль, содержащий всю основную логику работы системы, позволяет осуществлять всевозможные манипуляции с данными подразделений, задач и отчетов. Набор возможностей зависит от типа зарегистрированного пользователя.

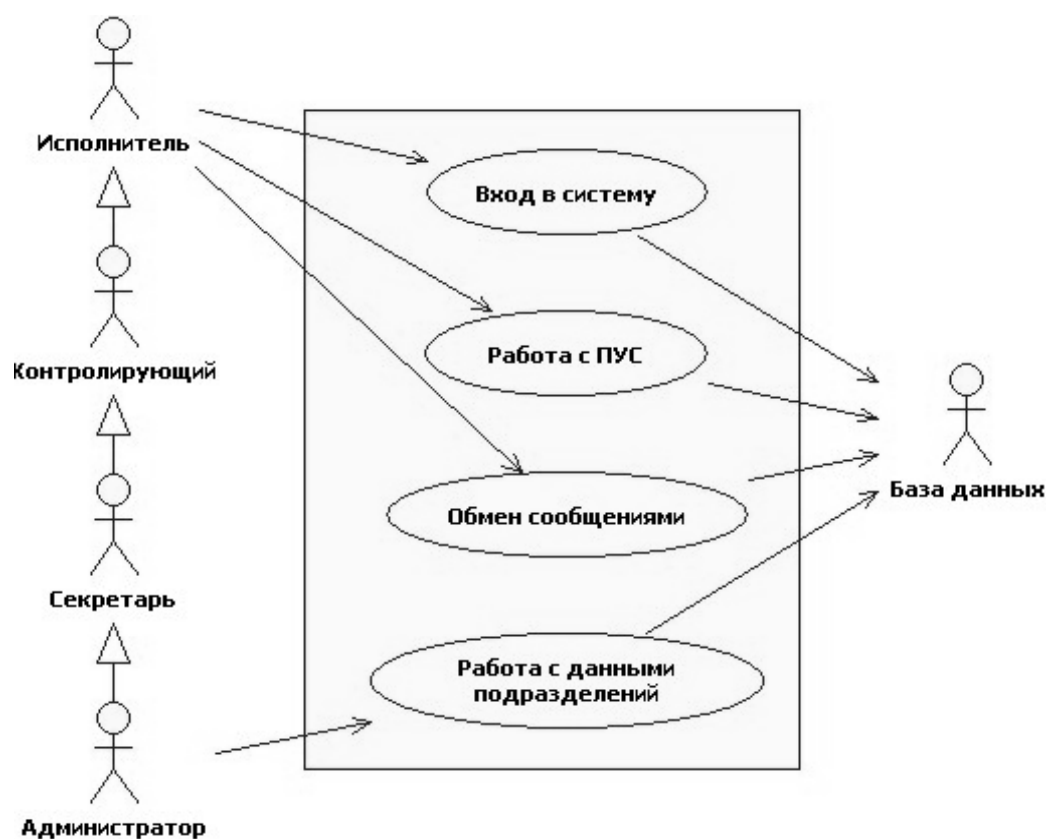


Рис. 1. Общая диаграмма вариантов использования

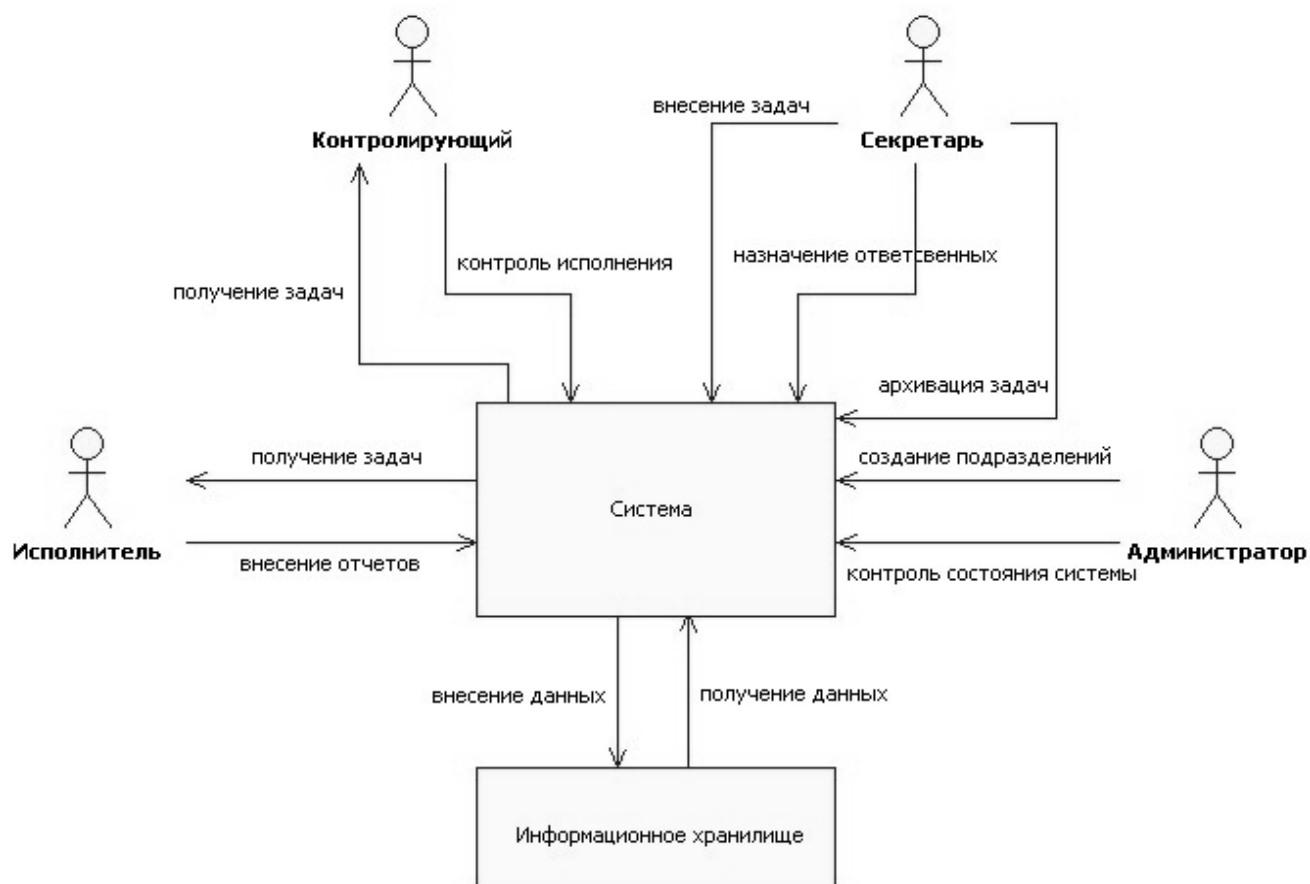


Рис. 2. Схема взаимодействия пользователей с системой

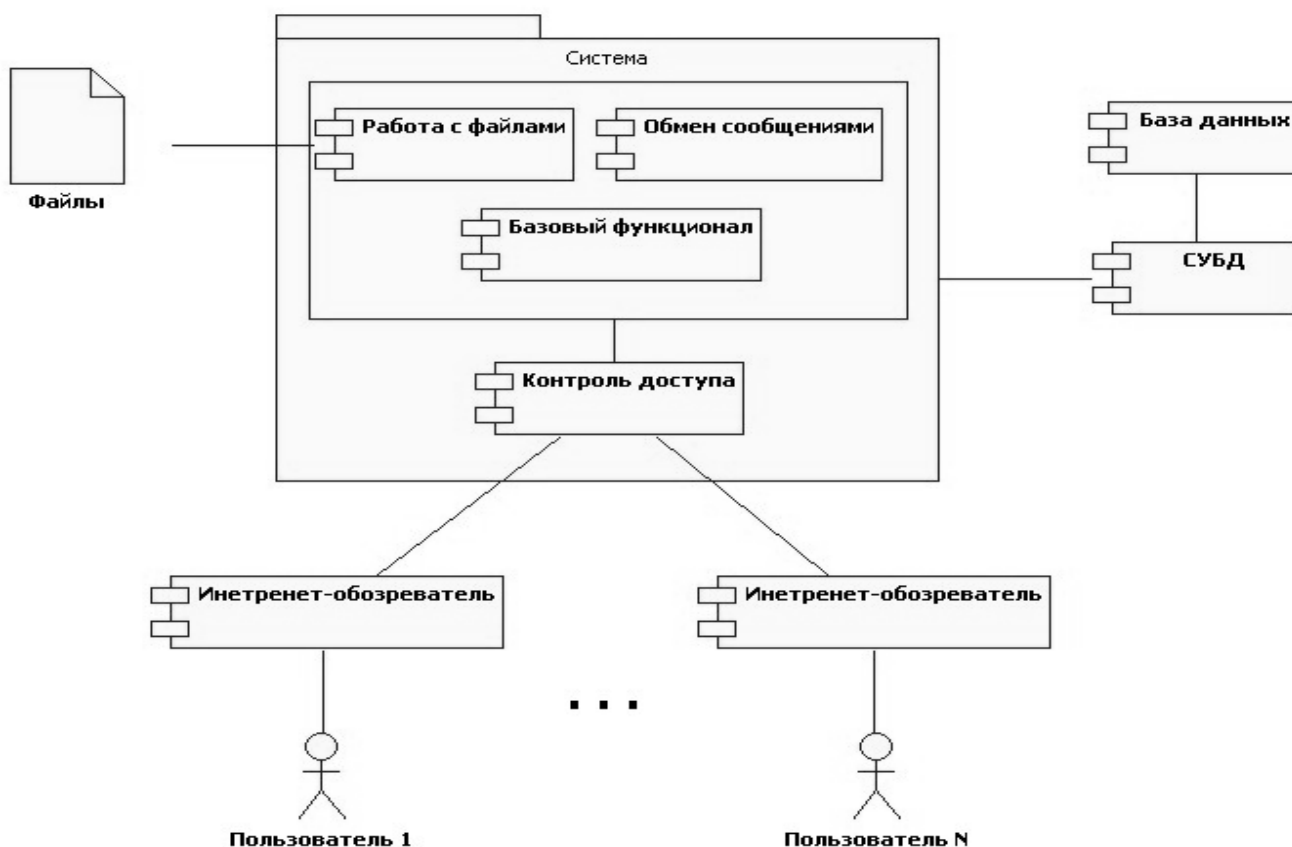


Рис. 3. Обобщенная структурно-функциональная схема

«Работа с файлами» – модуль, позволяющий сохранять прикрепленные к задачам файлы на сервере и осуществлять управление ими.

«Обмен сообщениями» – модуль коммуникации, предоставляющий пользователям возможность обмениваться сообщениями.

«База данных» – хранилище, содержащее всю информацию системы.

«СУБД» (система управления базами данными) – модуль, обеспечивающий полноценное взаимодействие системы (всех ее модулей) с базой данных.

С системой могут одновременно работать множество пользователей, получая к ней доступ с любого компьютера, имеющего подключение к Интернету, используя какой-либо интернет-обозреватель (браузер) (Mozilla Firefox, Opera, Google Chrome, Internet Explorer и т.п.). Для начала работы необходимо просто в адресной строке указать имя сайта, на котором расположена система.

Для реализации системы, удовлетворяющей поставленным требованиям, был выбран язык программирования PHP, а инструментальной основой приложения – Yii фреймворк [5].

Хранение данных предполагается посредством базы данных под управлением СУБД MySQL, поскольку такой способ является наиболее надежным и удобным при работе с информацией структурированного типа.

Для отображения графических интерфейсов используется связка технологий:

- CSS – каскадные таблицы стилей [6];
- HTML – гипертекстовый язык разметки;
- JavaScript – скриптовый язык программирования;
- библиотека JQuery [7].

Наиболее предпочтительным веб-сервером является Apache HTTP-сервер, хотя не исключаются возможности использования и иных инструментов.

Заключение

Разработана и находится в стадии опытной эксплуатации система электронного документооборота, имеющая следующие функциональные возможности:

1. Сведение всех постановлений Ученого совета в единую базу данных и её пополнение после каждого заседания.
2. Формирование подразделениями университета отчетов по выполненным постановлениям.
3. Контроль над общим состоянием системы.
4. Контроль подразделениями, имеющими на то полномочия, за отчетностью других подразделений по определенным задачам.
5. Возможность вынесения выполненных и подтвержденных задач в архив.
6. Использование режима администратора, позволяющего манипулировать как подразделениями, так и данными задач.

ЛИТЕРАТУРА

1. Васюхин О.В., Варзунов А.В. Информационный менеджмент. СПб. : СПбГУ ИТМО, 2010. 119 с.
2. Обзор систем электронного документооборота // iXBT – интернет-издание о компьютерной технике. URL: <http://www.ixbt.com/soft/sed.shtml> (дата обращения: 21.06.15).
3. Пахчанян А., Романов Д. Рынок ПО: Обзор систем электронного документооборота // CNEWS.RU: издание в сфере высоких технологий, 2002. URL: <http://www.cnews.ru/reviews/index.shtml?2002/05/17/140012> (дата обращения: 10.04.2015).
4. Буч Г., Рамбо Д., Якобсон И. Язык UML. Руководство пользователя. 2-е изд. М. : ДМК Пресс, 2006. 496 с.
5. Полное руководство по Yii // Yii Software LLC. 2014. URL: <http://www.yiiframework.com/doc/guide/1.1/ru> (дата обращения: 12.05.2015).
6. Хеник Б. HTML и CSS: путь к совершенству. СПб. : Питер, 2011. 336 с.
7. Ленгсторф Дж. PHP и jQuery для профессионалов. М. : Вильямс, 2011. 352 с.

Головчинер Михаил Наумович, канд. техн. наук, доцент. E-mail: golovchiner@mail.ru

Юшин Владимир Юрьевич. E-mail: krash-5678@yandex.ru

Томский государственный университет

Поступила в редакцию 18 мая 2015 г.

Golovchiner Michael N., Yushin Vladimir Yu. (Tomsk State University, Russian Federation).

Automated system of accounting and control of decisions of the Scientific Council of University.

Keywords: electronic workflow; system of control; scientific council; technology stack.

DOI 10.17223/19988605/32/9

There is a permanent work with documents in Tomsk State University. A sizable part in making this work is brought in by meeting of the Scientific Council (SC), which takes place with certain periodicity. A result is a list of the decisions taken on control. The secretary of the SC creates a text document, containing the list of these decisions. Subdivisions of University must execute task and present a corresponding report. In time, there is the accumulation of bank of decisions and reports of subdivisions, that complicate work on watching execution of decisions. In addition, both a secretary and subdivisions require to execute identical monotonous work on the hand filling and viewing dozens of documents very alike on each other.

A conclusion is obvious about the serious lack of organization efficiency of work with the SC decisions. So, it was decided to create the dedicated system electronic circulation of documents.

This system solves the next problems on automation:

1. one-time registration of the SC regulations, exception of necessity of frequent printing-down and unsealing of decisions;
2. storage of all decisions in a single base that allows to avoid loss papers or origin of doublets;
3. access to the system from any computer with an internet connection, which greatly improves the efficiency;
4. possibility of attaching to the tasks, sent to subdivisions, of auxiliary informative files, that releases from the necessity of transmission of stacks of papers or independent search subdivisions of corresponding documents and forms;
5. intrasystem exchange by reports for simplification of communication needed to address emerging issues;
6. possibility of archiving both decisions of the SC and tasks put before subdivisions;
7. simple monitoring system allowing without delay and effectively to communicate to subdivisions;
8. timely reminder about unfulfilled tasks, term of handing over of which passed already, that diminished a possible damage from similar situations;
9. transparency of all activity related to execution of the SC decisions, and as a result is possibility qualitatively to analyze this activity for the search of bottlenecks.

REFERENCES

1. Vasyukhin, O.V. & Varzunov, A.V. (2010) *Informatsionnyy menedzhment* [Information management]. St. Petersburg: SPbGU ITMO.
2. Ixht.com. (2015) *Obzor sistem elektronnoy dokumentooborota* [Review of document management systems]. Available from: <http://www.ixht.com>soft/sed.shtml>. (Accessed: 21st June 2015).
3. Pakhchanyan, A. & Romanov, D. (2002) *Rynok PO: Obzor sistem elektronnoy dokumentooborota* [Software market: Overview of electronic document management systems]. [Online] Available from: <http://www.cnews.ru/reviews/index.shtml?2002/05/17/140012>. (Accessed: 10th April 2015).
4. Buch, G., Rambo, D. & Yakobson, I. (2006) *Yazyk UML* [UML]. 2nd ed. Moscow: DMK Press.
5. Yii Software LLC. (2014) *Polnoe rukovodstvo po Yii* [The complete guide to Yii]. [Online] Available from: <http://www.yiiframework.com/doc/guide/1.1/ru>. (Accessed: 12th May 2015).
6. Khenik, B. (2011) *HTML i CSS: put' k sovershenstvu* [HTML and CSS: way to perfection]. St. Petersburg: Pite.
7. Langsdorf, J. (2011) *PHP i jQuery dlya professionalov* [PHP and jQuery for the professionals]. Moscow: Vil'yams.

О.А. Змеев, А.М. Политов, Я.М. Чайка

КОНЦЕПЦИЯ УНИФИЦИРОВАННОЙ МОДЕЛИ СЦЕНАРИЯ ВАРИАНТА ИСПОЛЬЗОВАНИЯ ДЛЯ ФИКСАЦИИ ФУНКЦИОНАЛЬНЫХ ТРЕБОВАНИЙ К ПРОГРАММНОМУ ПРОДУКТУ

Разработана унифицированная модель сценария варианта использования, поддерживающая различные его представления. Реализовано упрощение одновременной поддержки нескольких различных способов представления сценария варианта использования, например текстовых и графических. Предложена собственная графическая нотация модели. Построен прототип инструмента для работы с моделью, поддерживающий двустороннее преобразование в текстовую нотацию R.J. Wirfs-Brock.

Ключевые слова: объектно-ориентированный подход; управление требованиями; вариант использования; сценарий варианта использования; спецификация варианта использования; UML.

Предложенная Иваром Якобсоном (Ivar Jacobson) в конце 90-х годов прошлого века концепция определения требований к программным системам в виде вариантов использования [1] на сегодняшний день является основой рабочего потока определения требований в большинстве процессов разработки программного обеспечения, в основе которых лежит объектно-ориентированная парадигма. Согласно стандарту унифицированного языка моделирования (UML) вариант использования – спецификация множества действий, выполняемых системой, имеющего значимый результат для одного или нескольких актеров (пользователей или других систем) [2]. С помощью модели вариантов использования в современных процессах разработки фиксируются две важнейшие составляющие модели требований: структурная (в виде диаграмм вариантов использования) и динамическая (в виде спецификаций сценариев вариантов использования).

К сожалению, ориентированный прежде всего на разработку графических моделей артефактов программных систем UML предлагает для описания сценариев вариантов использования формат графических диаграмм состояний и практически не рассматривает инструменты для работы с текстовыми спецификациями сценариев вариантов использования. Отсутствие принятого, удобного для всех участников процесса разработки и заинтересованных сторон стандарта описания функциональных требований приводит к широкому разнообразию форматов текстовых спецификаций, например, в [4] рассматривается порядок сорока различных способов описания сценариев вариантов использования. Несмотря на большое число предлагаемых форматов, разнообразные представления вариантов использования можно разбить на два принципиально различных вида:

1. Текстовое представление – описывают последовательность действий на естественном языке, с той или иной степенью формализма.
2. Графическое представление – описывают последовательность действий при помощи диаграмм, как правило, имеющих формальную модель.

Необходимо отметить еще тот факт, что различные участники процесса разработки, задействованные в работе с вариантами использования, очень часто предпочитают применять разные формы представления функциональных требований. Заказчики предпочитают текстовые описания на естественном языке, поскольку для применения диаграмм необходимо знание модели и графической нотации, более того, наличие такого текстового представления является одним из достоинств подхода, основанного на применении вариантов использования. С другой стороны, разработчики предпочитают графические представления в виде диаграмм, поскольку такие представления имеют формальную модель, которая может использоваться для дальнейших преобразований в другие необходимые артефакты

процесса разработки. В результате в больших проектах со сложной структурой модели вариантов использования, когда они необходимы в различных форматах, нужно вручную создавать все необходимые представления и впоследствии выполнять синхронизацию изменений. Для больших программных систем как само такое преобразование, так и поддержка различных представлений вариантов использования в актуальном состоянии ведут к достаточно серьезным издержкам.

Для решения данной проблемы в настоящей работе предлагается применять разработанную унифицированную модель вариантов использования, которая поддерживает различные способы представления. В качестве примера поддерживаемых представлений в настоящей работе рассматриваются диаграммы деятельности UML (Activity Diagram) [2], текстовые спецификации, которые разработал А. Cockburn [3] и двухколоночный табличный формат текстовых спецификаций, разработанный R. Wirfs-Brock [5].

1. Представления сценариев варианта использования

Для классификации различных определений вариантов использования можно применять следующие критерии, предлагаемые Alistair Cockburn в [3]:

1. Цель: является ли целью сбор пользовательских историй или описание требований.
2. Содержимое: является ли содержимое варианта использования согласованным или оно может быть противоречивым. Если содержимое согласованно, то является ли оно при этом формальным.
3. Множественность: является ли вариант использования представлением одного или множества сценариев.
4. Структура: образует ли множество вариантов использования формальную (или полуформальную) структуру.

И Якобсон [1], и Коберн [3] определяют вариант использования согласно данной классификации как требование, имеющее согласованное содержимое, представляющее множество сценариев и образующее полуформальную структуру.

Для представления структуры варианта использования Коберн [1] предложил модель «полосатых брюк» (рис. 1). «Ремень брюк» – цель варианта использования, которая «держит» все сценарии. «Полосы» – индивидуальные сценарии, характеризующиеся условием их наступления и результатом. Множество «Полос» разделено на две группы: достигающие цели и не достигающие цели. Основным сценарием в данной модели является наиболее простой, в котором все промежуточные цели достигаются. Все остальные сценарии являются альтернативными. Альтернативные сценарии делятся на восстанавливаемые (если одна или несколько промежуточных целей не были достигнуты, но итоговая цель была достигнута) и неудачные (итоговая цель не была достигнута).

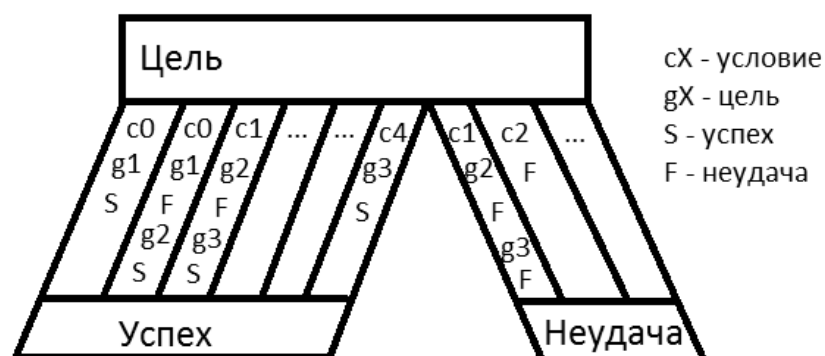


Рис. 1. Модель «полосатых брюк»

Как уже упоминалось выше, в процессе разработки программного обеспечения варианты использования применяются (могут использоваться) для фиксирования функциональных требований к разрабатываемому продукту. Часто при этом их необходимо представлять заказчику (или иным заинтересо-

ванным лицам, не занимающимся напрямую разработкой), который может не знать специализированных нотаций (например, UML). В таком случае единственным способом представления варианта использования для заказчика является текстовое описание на естественном языке, для применения которого не требуется знаний специализированных нотаций.

С другой стороны, на основании вариантов использования происходят дальнейшие этапы разработки программного обеспечения (анализ, проектирование, тестирования). В таком случае для возможности автоматизированной обработки вариантов использования необходима формальная модель, которая, как правило, отсутствует у текстовых представлений, но присутствует у графических.

Таким образом, при необходимости применения различных представлений варианта использования возможны следующие варианты:

1. Вариант использования специфицируется в различных его представлениях вручную. В таком случае спецификатору ВИ необходимо проделывать фактически одну и ту же работу (спецификацию одного и того же ВИ) несколько раз. В случае необходимости внесения изменений в вариант использования такие изменения должны быть отражены во всех его представлениях, что также добавляет дополнительную работу и может являться причиной возникновения ошибок (из-за расхождений между представлениями).

2. Вариант использования специфицируется в одном представлении. После этого, с помощью специализированных инструментов, выполняется его преобразование (в том числе и автоматическое) в необходимые представления. При внесении изменений они вносятся в первичное представление, что ограничивает возможность спецификатора вариантов использования редактировать производные представления (после повторного преобразования будут теряться изменения производных представлений вариантов использования, изменения в производных представлениях вариантов использования не будут отражены в первичном). Такая модель реализован, например, в Visual Paradigm [6].

3. Решение, которое предлагается в настоящей работе: создание унифицированной модели вариантов использования, которая поддерживала бы необходимые форматы представления вариантов использования и не ограничивала редактирование варианта использования одной моделью (рис. 2).

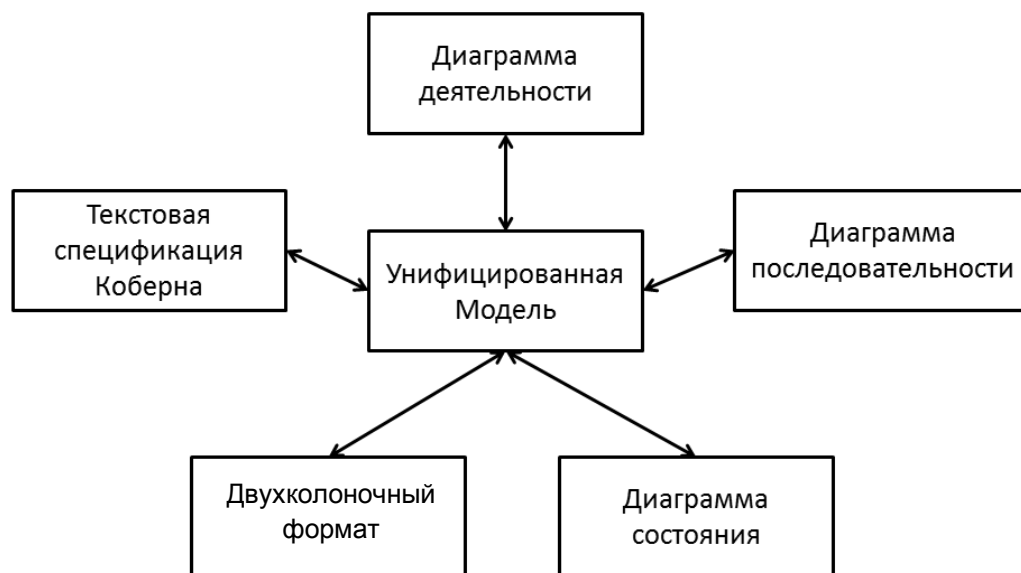


Рис. 2. Переход между представлениями с помощью унифицированной модели

2. Представления сценариев варианта использования

Существует достаточно много способов представления вариантов использования [4]. В качестве примера в данной работе будут рассмотрены три из них.

2.1. Диаграммы деятельности UML

Одним из способов графического представления вариантов использования являются диаграммы деятельности UML. Данный вид диаграмм позволяет определить поведение с помощью последовательного исполнения поведений более низкого уровня. Исполнение следующего действия может начинаться в результате наступления одного из следующих событий: завершение исполнения предыдущего действия (потoki управления), появление необходимых данных (потoki данных), наступление определенного события (потoki управления).

2.2. Текстовая спецификация А. Коберна

Согласно [1] текстовая спецификация вариантов использования состоит из общей информации: названия варианта использования, контекста использования, области действия, уровня цели, основных актеров, предусловия, триггера, минимальных гарантий и успешного постусловия.

Кроме общей информации о варианте использования, описываются его основной сценарий и расширения к нему. Основной сценарий представляет собой последовательность действий, при успешном выполнении которых достигается цель варианта использования. Расширения основного сценария описывают действия при возникновении исключительных ситуаций (действия, не предусмотренные основным сценарием, ошибки, внешние события).

Основной сценарий описывается в виде пронумерованной последовательности шагов:

<номер шага>. <действие>

Расширения привязываются к определенным шагам основного сценария (в том числе ко всему сценарию) и описываются в виде пронумерованной последовательности шагов:

<номер шага + идентификатор расширения>. <условие>: <действие или вложенный вариант использования>

2.3. Двухколоночный формат Р. Вирфс-Брок

В [5] Ребеккой Вирфс-Брок было предложено оформлять сценарии варианта использования в виде диалога между основным актером и системой (не предполагается участие в одном варианте использования более двух актеров). Сценарий записывается в форме таблицы, состоящей из двух колонок: действия основного актера и действия системы.

3. Основа унифицированной модели

В результате анализа различных представлений вариантов использования были выделены следующие особенности:

1. Вариант использования имеет множество сценариев, завершающихся успешно или неудачно.
2. Каждый такой сценарий представляет собой последовательность действий.
3. Каждое действие исполняется определенным актером (например, пользователем, системой).
4. Множество таких сценариев может быть бесконечно (из-за возможности образования циклов) и в результате неприменимо в чистом виде на практике. Для решения данной проблемы выделим основной сценарий варианта использования (сценарий, все действия которого успешно выполняются), а остальное множество сценариев определим через расширения к отдельным действиям основного сценария (в том числе через расширения ко множеству действий и, возможно, ко всему сценарию).

4. Предлагаемая модель сценария

Предлагаемая модель представляет собой расширение к модели UML (версии 2.4.1, стандартизованной ISO) [2]. Модель основана на пакетах `Classes::Kernel` и `CommonBehaviors::BasicBehaviors`. На рис. 3 показана диаграмма пакетов предлагаемой модели.

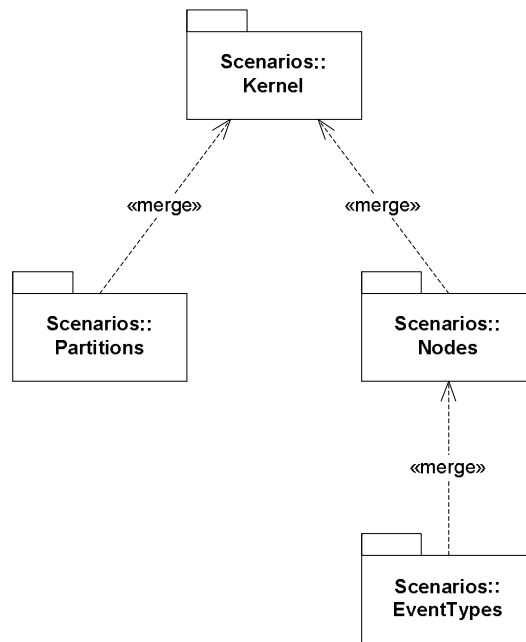


Рис. 3. Диаграмма пакетов предлагаемой модели

Модель состоит из следующих пакетов:

1. Scenarios::Kernel содержит ядро модели, определяющее базовые элементы сценария и потоки управления для перехода между элементами.
 2. Scenarios::Nodes содержит конкретные элементы сценария (например, действия, обработчики событий и т.д.).
 3. Scenarios::Partitions содержит разделы, используемые для группировки элементов сценария.
 4. Scenarios::EventTypes содержит типы событий.
- Рассмотрим каждый пакет более подробно.

4.1. Пакет Scenarios::Kernel

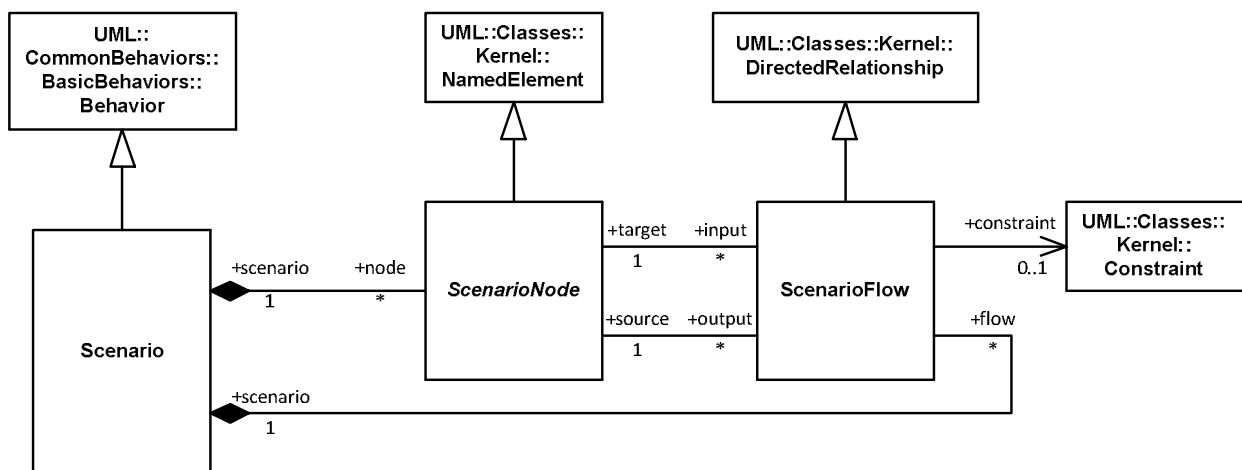


Рис. 4. Диаграмма классов пакета Scenarios::Kernel

Пакет Scenario::Kernel представляет собой ядро унифицированной модели (рис. 4). В нем определяются сценарий варианта использования, элементы сценария и потоки управления, используемые для перехода между элементами сценария.

Описание классов пакета:

1. Scenario представляет собой сценарий варианта использования.

2. ScenarioNode представляет собой элемент сценария варианта использования.
3. ScenarioFlow представляет направленный поток управления, передающий исполнение от одного элемента сценария другому. Поток может иметь ассоциированное с ним условие, в таком случае передача управления выполняется, только если условие выполняется.

4.2. Пакет Scenarios::Nodes

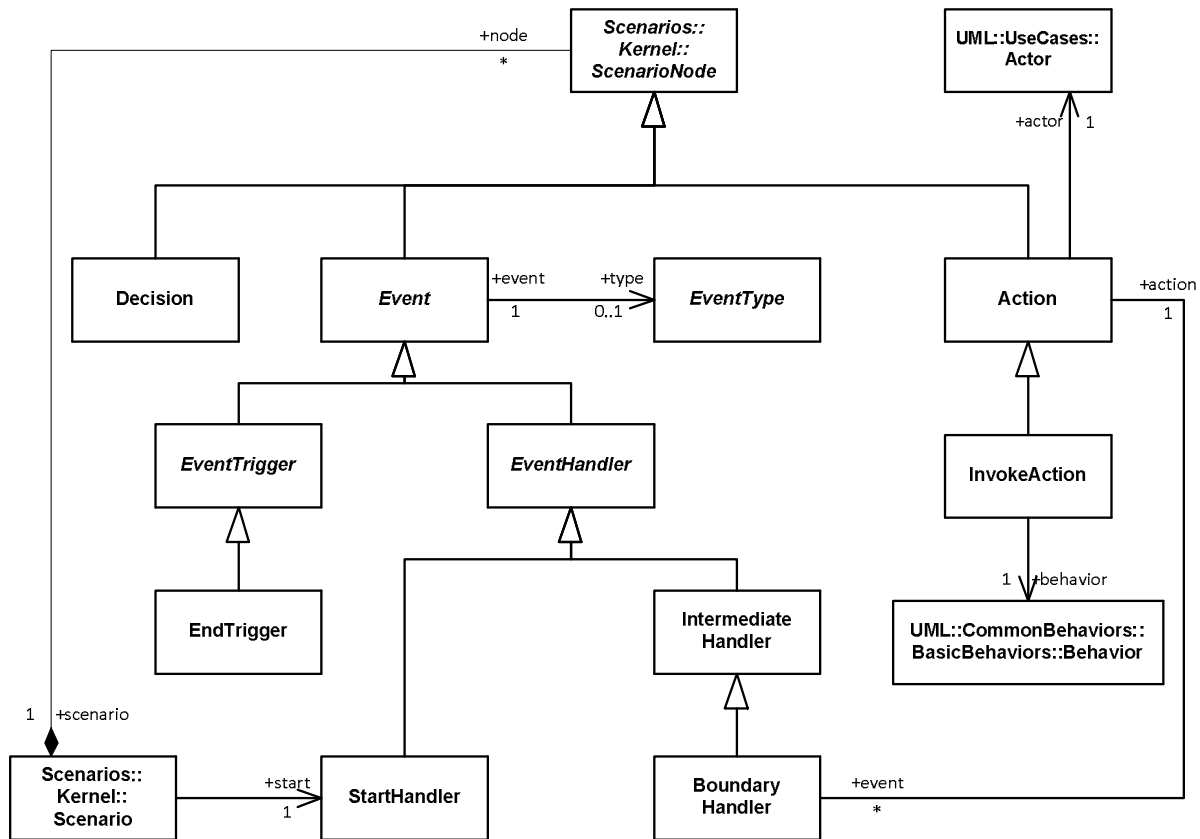


Рис. 5. Диаграмма классов пакета Scenarios::Nodes

Пакет Scenarios::Nodes содержит различные элементы сценария унифицированной модели (рис. 5). В нем определяются действия сценария, триггеры и обработчики событий и т.д.

Описание классов пакета:

1. Scenario – класс сценария расширен для связи с обработчиком начала сценария.
2. Action представляет действие в сценарии.
3. InvokeAction представляет действие сценария, являющееся вызовом другого поведения (в том числе, возможно, и сценария).
4. Decision представляет элемент сценария, который выбирает один из исходящих потоков для продолжения исполнения. Для определения потока, по которому передается управление, используются условия потоков.
5. Event – абстрактный класс, представляющий событие в сценарии: элементы, вызывающие события, и элементы, обрабатывающие события.
6. EventType – абстрактный класс EventType, представляет тип события.
7. EventTrigger – абстрактный класс EventTrigger, представляет триггер события – элемент, при исполнении которого происходит определенное событие.
8. EndTrigger – класс EndTrigger, представляет завершающий сценарий элемент. При исполнении данного элемента сценарий завершается. Если с завершающим элементом ассоциирован тип события, то происходят вызов данного события и передача его в сценарий, вызвавший данный сценарий.
9. EventHandler – класс EventHandler, представляет абстрактный обработчик события.

10. StartHandler – класс StartHandler, представляет обработчик начала сценария – элемент сценария, с которого начинается его исполнение. У сценария может быть только один обработчик начала.
11. IntermediateHandler – класс IntermediateHandler, представляет промежуточный обработчик события.
12. BoundaryHandler – класс BoundaryHandler, представляет промежуточный обработчик события, который обрабатывает события в связанном с ним действии сценария.

Заключение

В результате была разработана унифицированная модель вариантов использования, имеющая собственную графическую нотацию.

На основе предложенной модели разработан прототип инструмента для работы со сценариями варианта использования в предложенной модели. Реализация поддерживает преобразование сценариев из текстовой нотации Ребекки Вирфс-Брок (R. J. Wirfs-Brock) в графическую нотацию предложенной модели и наоборот. В дальнейшем планируется анализ возможности трассировки вариантов использования в предложенной модели в артефакты дальнейших этапов процесса разработки, например модель анализа и модель тестирования.

ЛИТЕРАТУРА

1. *Jacobson I.* Modeling with Use Cases: Formalizing use-case modeling // JOOP. 1995. June. V. 8, No. 3.
2. ISO/IEC 19505-2:2012(E). Information technology – Object Management Group Unified Modeling Language (OMG UML). Part 2: Superstructure. April 2012. International Organization for Standardization, 2012. 740 p.
3. *Cockburn A.* Structuring use cases with goals. Electronic version of printed publication. 1995. URL: <http://alistair.cockburn.us/Structuring+use+cases+with+goals> (access date: 30.11.2011).
4. *Hurlbut R.R.* A Survey of Approaches For Describing and Formalizing Use Cases. Expertech, Wheaton, Illinois, 1997. URL: <http://www.iit.edu/~rhurlbut/xpt-tr-97-03.html> (access: 14.12.2011).
5. *Wirfs-Brock R.* Designing Scenarios: Making the Case for a Use Case Framework // The Smalltalk Report. 1993. November-December. V. 3, No. 3.
6. *Software Design Tools for Agile Teams with UML, BPMN and More.* URL: <http://www.visual-paradigm.com> (access date: 23.03.2014).

Змеев Олег Алексеевич, д-р физ.-мат. наук, профессор. E-mail: ozmeyev@gmail.com

Политов Арсентий Михайлович. E-mail: m.politov@gmail.com

Чайка Яна Михайловна. E-mail: chaykayana@gmail.com

Томский государственный университет

Поступила в редакцию 26 мая 2015 г.

Zmeyev Oleg A., Politov Arseny M., Chayka Yana M. (Tomsk State University, Russian Federation).

Concept of unified use case model for software requirements specification.

Keywords: object-oriented approach; requirements management; use case, use case scenario; use case specification; UML

DOI 10.17223/19988605/32/10

Use case is a specification of actions sequence executed by system for achieving a significant result for any or all actors. The use cases model is used for capturing both structural and behavioral parts of product's functional requirements in current developing processes.

Unified Modelling Language offers several graphical representations of use cases scenarios, such as activity diagrams, rather than textual formats. Nevertheless, there are many different textual formats used for use cases scenarios specifications, e.g. Wirfs-Brock's two-column table style and Cockburn style. Different stakeholders may prefer a different specification format, e.g. customer may prefer textual formats, while developers may prefer graphical formats. In addition, use cases can be used for subsequent development stages (e.g. analysis, design, testing). For use cases to be automatically traced into these stages they must be formally defined. Such definition is usually possible in the graphical representation but not in textual.

When several formats are required, following approaches can be applied:

1. Use case should be created manually in every required formats and every change should be to be synchronized between used formats to keep requirements specifications up to date. Such synchronization is time consuming and error-

prone process due to the human factor. To simplify this process, the development of the unified use cases model that will support different formats of use cases scenarios is proposed.

2. Use case is specified in a single format. Then transformation into required formats is performed by special tools. Any changes in the use case are made in the primary format of use case thus limiting an ability to edit derived formats (subsequent transformation will cause a loss of derived formats edits and these edits will not affect primary format).

3. An approach proposed in this paper is to create the unified use case model that will support different formats of use cases scenarios and not restricting scenario editing to one of them.

There are the following use cases features:

1. Use case has multiple scenarios that can either succeed or fail.

2. Each scenario is described by actions sequence.

3. Each action is executed by a specific actor.

4. A set of these scenarios can be infinite due to loops and cannot be applied. To make scenarios more useful, a main success scenario can be selected and other scenarios can be defined as extensions to specific actions of a main success scenario.

The model proposed in this paper is an extension of UML 2.4.1 (formally published as ISO/IEC 19505-1 and ISO/IEC 19505-2). It is based on the following packages: `Classes::Kernel` and `CommonBehaviors::BasicBehaviors` and consists of the following:

1. `Scenarios::Kernel` package that contains the unified model kernel. In this package use case scenarios, scenario nodes and flows are defined.

2. `Scenarios::Nodes` package that contains specific scenario nodes, such as scenario actions, triggers, handlers, decisions, etc.

3. `Scenarios::Partitions` package that defines partitions used for scenario nodes grouping.

4. `Scenarios::EventTypes` package that define events types, e.g. generic errors, system events, user events, time events, etc.

As a result, the unified use case model that has its own graphical notation is developed. In addition, prototype of a tool that supports editing use cases in the developed model and transformation between Wirfs-Brock's textual format and model's graphical notation is implemented.

REFERENCES

1. Jacobson, I. (1995) Modeling with Use Cases: Formalizing use-case modeling. *JOOP*. 8 (3).
2. ISO/IEC 19505-2:2012(E). (2012) Information technology. *Object Management Group Unified Modeling Language (OMG UML) Part 2: Superstructure*. April 2012. International Organization for Standardization.
3. Cockburn, A. (2011) *Structuring use cases with goals*. [Online] Available from: <http://alistair.cockburn.us/Structuring+use+cases+with+goals>. (Accessed: 30th November 2011).
4. Hurlbut, R.R. (1997) *A Survey of Approaches For Describing and Formalizing Use Cases*. Expertech, Wheaton, Illinois. Available from: <http://www.iit.edu/~rhurlbut/xpt-tr-97-03.html>. (Accessed: 14th December 2011)
5. Wirfs-Brock, R. (1993) *Designing Scenarios: Making the Case for a Use Case Framework*. The Smalltalk Report. 3(3).
6. *Software Design Tools for Agile Teams with UML, BPMN and More*. [Online] Available from: <http://www.visual-paradigm.com>. (Accessed: 23th March 2014).

СВЕДЕНИЯ ОБ АВТОРАХ

БЕРЕЗИН Дмитрий Владимирович – магистрант факультета прикладной математики и кибернетики Томского государственного университета. E-mail: berezin14@mail.ru

ГОЛОВЧИНЕР Михаил Наумович – кандидат технических наук, доцент кафедры программирования Томского государственного университета. E-mail: golovchiner@mail.ru

ДОМБРОВСКИЙ Владимир Валентинович – профессор, доктор технических наук, заведующий кафедрой математических методов и информационных технологий в экономике экономического факультета Томского государственного университета. E-mail: dombrovs@ef.tsu.ru

ЗМЕЕВ Олег Алексеевич – профессор, доктор физико-математических наук, заведующий кафедрой программной инженерии факультета информатики Томского государственного университета. E-mail: ozmeyev@gmail.com

КАЛЯГИН Алексей Андреевич – аспирант факультета прикладной математики и кибернетики Томского государственного университета. E-mail: redall@inbox.ru

КОПАНИЦА Георгий Дмитриевич – кандидат технических наук, инженер-исследователь кафедры информатики и проектирования систем Института кибернетики Томского политехнического университета. E-mail: georgy.kopanitsa@gmail.com

КУСАИНОВ Марат Исламбекович – аспирант факультета прикладной математики и кибернетики Национального исследовательского Томского государственного университета. E-mail: rjrltsk@gmail.com

МЕДВЕДЕВ Геннадий Алексеевич – профессор, доктор физико-математических наук, профессор факультета прикладной математики и информатики Белорусского государственного университета (г. Минск, Беларусь). E-mail: MedvedevGA@cosmostv.by

МОИСЕЕВ Александр Николаевич – доцент, кандидат технических наук, доцент кафедры программной инженерии факультета информатики Томского государственного университета. E-mail: moiseev.tsu@gmail.com

НЕЖЕЛЬСКАЯ Людмила Алексеевна – доцент, кандидат технических наук, доцент кафедры исследования операций факультета прикладной математики Томского государственного университета. E-mail: ludne@mail.ru

ОБЪЕДКО Татьяна Юрьевна – кандидат физико-математических наук, доцент кафедры математических методов и информационных технологий в экономике экономического факультета Томского государственного университета. E-mail: tatyana.obedko@mail.ru

ПОЛИТОВ Арсентий Михайлович – магистрант факультета информатики Томского государственного университета. E-mail: a.m.politov@gmail.com

САМОРОДОВА Мария Владимировна – аспирант кафедры прикладной математики факультета прикладной математики и кибернетики Томского государственного университета. E-mail: samorodova21@gmail.com

СИРЕНКО Михаил Андреевич – аспирант факультета прикладной математики и кибернетики Томского государственного университета. E-mail: sma@mail.tsu.ru

ТАРАНИК Максим Алексеевич – аспирант кафедры оптимизации систем управления Института кибернетики Томского политехнического университета. E-mail: taranik@tpu.ru

ТАРАСЕНКО Петр Феликсович – доцент, кандидат физико-математических наук, декан Международного факультета управления Томского государственного университета. E-mail: ptara@mail.tsu.ru

ЧАЙКА Яна Михайловна – магистрант факультета информатики Томского государственного университета. E-mail: chaykayana@gmail.com

ЮШИН Владимир Юрьевич – магистрант факультета прикладной математики и кибернетики Томского государственного университета. E-mail: krash-5678@yandex.ru

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2015. № 3 (32)

Редактор Н.А. Афанасьева
Оригинал-макет А.И. Лелоюр
Редакторы-переводчики: Г.М. Кошкин; В.Н. Горенинцева
Дизайн обложки Л.Д. Кривцова

Подписано к печати 15.09.2015 г. Формат 60x84¹/₈.

Гарнитура Times.

Печ. л. 12,5; усл. печ. л. 11,6.

Тираж 250 экз. Заказы № 1284, 1284/1.

Журнал отпечатан на полиграфическом оборудовании
Издательского Дома Томского государственного университета

634050, г. Томск, Ленина, 36

Тел. 8+(382-2)-53-15-28

Сайт: <http://publish.tsu.ru>; E-mail: rio.tsu@mail.ru