

Информатика и её применения

Том 7 Выпуск 2 Год 2013

Тематический выпуск

*Вероятностно-статистические методы
и задачи информатики и информационных технологий*

СОДЕРЖАНИЕ

Предисловие <i>И. А. Соколов, В. Ю. Королев</i>	3
Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях <i>И. Н. Сеницын</i>	4
Методы оценки эффективности и директивных сроков выполнения ресурсоемких вычислительных заданий <i>И. К. Купалов-Ярополк, Ю. Е. Малашенко, И. А. Назарова, А. Ф. Ронжин</i>	17
Об оценивании эффективной пропускной способности системы с регенеративным входным процессом <i>А. В. Бородина, Е. В. Морозов</i>	26
Стационарное распределение времени ожидания в системе обслуживания с отрицательными заявками, бункером для вытесненных заявок, различными интенсивностями обслуживания при дисциплине FIRST-FIFO-FIFO <i>Р. В. Разумчик</i>	34
Центральная предельная теорема для функции обобщенной кросс-валидации при пороговой обработке вейвлет-коэффициентов <i>О. В. Шестаков</i>	40
Статистическая проверка неисполняемости фрагментов кода последовательной программы <i>В. Ю. Королев, Р. Л. Смелянский, Т. Р. Смелянский, А. В. Шалимов</i>	50
Байесовская рекуррентная модель роста надежности: равномерное распределение параметров <i>А. А. Кудрявцев, И. А. Соколов, С. Я. Шоргин</i>	55
Идентификация зависимостей признаков стохастической природы на основе регрессии Деминга <i>В. С. Тимофеев, В. Ю. Щеколдин, А. Ю. Тимофеева</i>	60

Информатика и её применения

Том 7 Выпуск 2 Год 2013

Тематический выпуск

*Вероятностно-статистические методы
и задачи информатики и информационных технологий*

СОДЕРЖАНИЕ

Асимптотическая нормальность оценки параметров многомерной логистической регрессии А. Ю. Хапланов	69
Асимптотические разложения для функций распределения статистик, построенных по выборкам случайного объема В. Е. Бенинг, Н. К. Галиева, В. Ю. Королев	75
О сходимости случайных блужданий, порожденных обобщенными процессами Кокса, к процессам Леви В. Ю. Королев, Л. М. Закс, А. И. Зейфман	84
Статистические механизмы формирования ассоциативных портретов предметных областей на основе естественно-языковых текстов больших объемов для систем извлечения знаний М. М. Шарнин, Н. В. Сомин, И. П. Кузнецов, Ю. И. Морозова, И. В. Галина, Е. Б. Козеренко	92
Information technologies for creating the database of equivalent verbal forms in the Russian-French multivariant parallel corpus S. Loiseau, D. V. Sitchinava, A. A. Zalizniak, and I. M. Zatsman	100
Abstracts	110
Об авторах	115
About Authors	117

Технический редактор *Л. Кокушкина*

Художественный редактор *М. Седакова*

Сдано в набор 03.04.13. Подписано в печать 14.06.13. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 14,2. Уч.-изд. л. 11,5. Тираж 100 экз.

Заказ № 3280.

Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43

torus@torus-press.ru; <http://www.torus-press.ru>

Отпечатано в Академиздатцентре «Наука» РАН с готовых файлов
Москва 121099, Шубинский пер., д. 6.

Предисловие

Вниманию читателей журнала «Информатика и её применения» предлагается традиционный тематический выпуск «Вероятностно-статистические методы и задачи информатики и информационных технологий». Предыдущие тематические выпуски журнала по данному направлению выходили ежегодно в 2008–2012 гг.

Статьи, собранные в данном журнале, посвящены разработке и совершенствованию вероятностно-статистических и смежных с ними методов, ориентированных на применение к решению конкретных задач информатики и информационных технологий, а также — в ряде случаев — и других прикладных задач. Проблематика, охватываемая публикуемыми работами, в значительной степени развивается в рамках научного сотрудничества между Институтом проблем информатики Российской академии наук (ИПИ РАН) и факультетом вычислительной математики и кибернетики Московского государственного университета (МГУ) им. М. В. Ломоносова в ходе работ над совместными научными проектами. Многие из авторов статей, включенных в данный номер журнала, являются активными участниками традиционного международного семинара по проблемам устойчивости стохастических моделей, руководимого В. М. Золотаревым и В. Ю. Королевым; регулярные сессии этого семинара проводятся под эгидой МГУ и ИПИ РАН. В 2012 г. указанный семинар проводился в сентябре в г. Светлогорске Калининградской области РФ, в 2013 г. — в апреле в г. Москве. Среди статей, включённых в настоящий выпуск, некоторые являются развитием докладов, представленных на сессиях этого семинара; это — работы А. В. Бородиной и Е. В. Морозова; Р. В. Разумчика; В. Е. Бенинга, Н. К. Галиевой и В. Ю. Королева; В. Ю. Королева, Л. М. Закс и А. И. Зейфмана.

Наряду с представителями ИПИ РАН и МГУ им. М. В. Ломоносова (факультет вычислительной математики и кибернетики и механико-математический факультет) в число авторов данного выпуска журнала входят ученые и специалисты из Вычислительного центра им. А. А. Дородницына Российской академии наук (РАН), Института точной механики и вычислительной техники им. С. А. Лебедева РАН, Института русского языка им. В. В. Виноградова РАН, Института языкознания РАН, Института прикладных математических исследований Карельского научного центра РАН, Новосибирского государственного технического университета, Вологодского государственного педагогического университета, отдела моделирования и математической статистики Альфа-банка, а также Казахстанского филиала МГУ им. М. В. Ломоносова и Университета Париж-13 (Франция).

Тематика статей данного выпуска включает вопросы математического моделирования и анализа реальных процессов и задач, в том числе моделирование распределений в нелинейных стохастических системах; построение и исследование моделей некоторых специальных систем и сетей передачи информации; задачи теоретической и прикладной математической статистики (включая развитие некоторых статистических методов анализа больших объемов текстов и создания параллельных поливариантных языковых корпусов).

Редакционная коллегия журнала выражает надежду, что данный тематический выпуск будет интересен специалистам в области теории вероятностей и математической статистики и их применения к решению задач информатики и информационных технологий.

Заместитель главного редактора журнала «Информатика и её применения»,
директор Института проблем информатики РАН, академик

И. А. Соколов

Редактор-составитель тематического выпуска,
профессор кафедры математической статистики
факультета вычислительной математики и кибернетики
МГУ им. М. В. Ломоносова,
ведущий научный сотрудник Института проблем информатики РАН,
доктор физико-математических наук

В. Ю. Королев

ПАРАМЕТРИЧЕСКОЕ СТАТИСТИЧЕСКОЕ И АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ В НЕЛИНЕЙНЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ НА МНОГООБРАЗИЯХ*

И. Н. Сеницын¹

Аннотация: Рассматриваются дискретные методы статистического и аналитического моделирования в нелинейных системах на многообразиях описываемых дифференциальными стохастическими уравнениями Ито с винеровскими и пуассоновскими шумами. Предполагается, что в качестве параметров распределений взяты коэффициенты совместных ортогональных плотностей различных порядков. Особое внимание уделено нелинейной корреляционной теории статистического и аналитического моделирования.

Ключевые слова: аналитическое моделирование; метод нормальной аппроксимации; метод ортогональных разложений; метод статистической линеаризации; нелинейная корреляционная теория; нелинейная стохастическая система Ито на многообразии; параметризация одно- и многомерных распределений; статистическое моделирование

1 Введение

Известные методы статистического и аналитического моделирования процессов в стохастических системах (СтС), описываемых дифференциальными стохастическими уравнениями Ито с винеровскими и пуассоновскими шумами, основанные на параметризации распределений, подробно изложены в [1–3].

Обобщение результатов [1–3] на случай многоканальных круговых и сферических СтС выполнено в [4–12].

Статья посвящена развитию методов параметрического статистического и аналитического моделирования в СтС Ито на многообразиях.

В разд. 2 рассмотрены уравнения СтС на многообразиях (МСтС).

В разд. 3 и приложении представлены приближенные методы статистического моделирования (МСМ) различной точности.

Раздел 4 посвящен как методам аналитического моделирования (МAM), основанным на ортогональных разложениях, так и совместным МСМ и МAM.

Нелинейная корреляционная теория МСМ и МAM развита в разд. 5.

2 Уравнения непрерывных стохастических систем на многообразиях

Как известно [1, 2], для дифференциальных СтС в конечномерных пространствах используется дифференциальное стохастическое уравнение Ито вида

$$dY = a(Y, \Theta, t)dt + b(Y, \Theta, t)dW_0 + \int_{R_0^q} c(Y, \Theta, t, v) dP^0(t, dv). \quad (1)$$

Здесь Y — p -мерный вектор состояния, $Y \in \Delta^y$ (Δ^y — многообразие состояний); Θ — вектор случайных параметров размерности p_θ ; $a = a(y, \theta, t)$ и $b = b(y, \theta, t)$ — известные $(p \times 1)$ -мерная и $(p \times r)$ -мерная функции вектора Y и времени t ; $W_0 = W_0(t)$ — r -мерный винеровский случайный процесс интенсивности $\nu_0 = \nu_0(t)$; $c(y, \theta, t, v)$ — $(p \times 1)$ -мерная функция y, t и вспомогательного $(q \times 1)$ -мерного параметра v ; $\int_{\Delta_t} dP^0(t, A)$ — центрированная пуассоновская мера:

$$\int_{\Delta_t} dP^0(t, A) = \int_{\Delta_t} dP(t, A) - \int_{\Delta_t} \nu_P(t, A) dt,$$

где $\int_{\Delta_t} dP(t, A)$ — число скачков пуассоновского процесса в интервале времени Δ ; $\nu_P(t, A)$ — ин-

* Работа выполнена при финансовой поддержке программы «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект 1.7).

¹Институт проблем информатики Российской академии наук, sinitsin@dol.ru

тенсивность пуассоновского процесса $P(t, A)$; A — некоторое борелевское множество пространства R_0^q с выколотым началом координат. Интеграл (1) в общем случае распространяется на все пространство R_0^q с выколотым началом координат. Начальное значение Y_0 вектора Y представляет собой случайную величину, не зависящую от приращений винеровского процесса $W_0(t)$ и пуассоновского процесса $P(t, A)$ на интервалах времени $\Delta_t = (t_1, t_2]$, следующих за t_0 , $t_0 \leq t_1 \leq t_2$, для любого множества A .

В случае, когда подынтегральная функция $c(y, \theta, t, v)$ в уравнении (1) допускает представление

$$c(y, \theta, t, v) = b(y, \theta, t)c'(v), \quad (2)$$

уравнение (1) приводится к виду:

$$\dot{Y} = a(Y, \Theta, t) + b(Y, \Theta, t)V, \quad (3)$$

если принять

$$W(t) = W_0(t) + \int_{R_0^q} c'(u)P^0(t, dv).$$

В некоторых случаях вводят расширенный вектор состояния $\bar{Y} = [Y^T \Theta^T]^T$ размерности $\bar{p} = p + p^\theta$. Тогда уравнения (1) и (3) совместно с уравнением формирующего фильтра для Θ

$$d\Theta = a^\theta(\Theta, t)dt + b^\theta(\Theta, t)dW_0 + \int_{R_0^q} c^\theta(\Theta, t, v)dP^0(t, dv)$$

примут вид:

$$d\bar{Y} = \bar{a}(\bar{Y}, t)dt + \bar{b}(\bar{Y}, t)dW_0 + \int_{R_0^q} \bar{c}(\bar{Y}, t, v)dP_0(t, dv), \quad (4)$$

где

$$\bar{a}(\bar{Y}, t) = \begin{bmatrix} a(\bar{Y}, \Theta, t) \\ a^\theta(\Theta, t) \end{bmatrix}; \quad \bar{b}(\bar{Y}, t) = \begin{bmatrix} b(\bar{Y}, \Theta, t) \\ b^\theta(\Theta, t) \end{bmatrix};$$

$$\bar{c}'(\bar{Y}, t, v) = \begin{bmatrix} c(\bar{Y}, \Theta, t, v) \\ c^\theta(\Theta, t, v) \end{bmatrix}.$$

Аналогично в условиях (2) имеем:

$$\dot{\bar{Y}} = \bar{a}(\bar{Y}, t) + \bar{b}(\bar{Y}, t)V. \quad (5)$$

Замечание 2.1. В дальнейшем будем пользоваться уравнениями (4) и (5), опуская черту над переменными количествами.

Для вычисления вероятностей событий, связанных со случайными функциями, в прикладных задачах достаточно знания многомерных распределений.

Поэтому центральной задачей теории непрерывных СтС является задача вероятностного анализа одно- и многомерных распределений процессов, удовлетворяющих дифференциальным стохастическим уравнениям Ито вида (4) или (5) при соответствующих начальных условиях.

В теории непрерывных СтС различают два принципиально разных подхода к вычислению распределений. Первый общий подход основан на статистическом моделировании, т.е. на прямом численном решении дифференциальных стохастических уравнений (4) или (5) с последующей статистической обработкой результатов. Второй общий подход основан на теории непрерывных марковских процессов и предполагает аналитическое моделирование, т.е. решение детерминированных уравнений в функциональных пространствах (уравнений Фоккера–Планка–Колмогорова, Феллера–Колмогорова, Пугачева и др.) для одно- и многомерных распределений. В практических задачах часто используют и комбинированные методы. При этом будем предполагать, что существуют одно- и многомерные плотности процессов в МСтС (4) и (5). Достаточные условия их существования можно найти, например, в [13].

Будем полагать, что, во-первых, одно- и многомерные плотности распределений существуют и, во-вторых, плотности можно параметризовать с помощью параметров $\{\Xi_n\}$: вероятностных моментов, квазимоментов, семиинвариантов, коэффициентов ортогонального разложения плотностей, канонических разложений и др. В таких случаях наряду с прямыми методами численного решения уравнений (4) и статистической обработки данных и оценивания $\{\Xi_n\}$ используют следующий подход. Пользуясь обобщенной формулой [1–3, 13], составляют в силу (4) дополнительно дифференциальные стохастические уравнения для параметров Ξ_n :

$$d\Xi_n = a^{\Xi_n}(\Xi_n, t)dt + b^{\Xi_n}(\Xi_n, t)dW_0 + \int_{R_0^q} c^{\Xi_n}(\Xi_n, t, v)dP^0(t, dv).$$

При использовании такого подхода под расширенным вектором состояния \bar{Y} будем рассматривать вектор $\tilde{Y} = [Y^T \Theta^T \Xi_n^T]^T$.

3 Параметрическое статистическое моделирование

Численное интегрирование дифференциальных стохастических уравнений имеет некоторые осо-

бенности [2, 14–16]. Дело в том, что все численные методы интегрирования обыкновенных дифференциальных уравнений, кроме простейшего метода Эйлера, основаны на вычислении приращений искомым функций на каждом шаге путем применения интегральной теоремы о среднем значении. В соответствии с этим правые части уравнений (производные искомым функций) берутся в средних точках интервалов. Различные методы численного интегрирования отличаются один от другого, по существу, только способом приближенного нахождения средних значений правых частей уравнений.

Предположим, что требуется заменить уравнение (1) или (3) разностным уравнением для значений процесса $Y(t)$ в заданном дискретном ряду равноотстоящих точек $\{t_k\}$, $t_k = kh$, где h — интервал между соседними точками t_k , $h = t_{k+1} - t_k$. Принципиально задача решается точно, так как значения марковского процесса $Y(t)$ в точках t_k образуют марковскую случайную последовательность $\{\bar{Y}_k\}$, $\bar{Y}_k = Y(kh)$, а всякая марковская последовательность определяется некоторым стохастическим разностным уравнением. Однако составить это точное разностное уравнение по данному дифференциальному уравнению практически невозможно. Для его составления необходимо прежде всего найти переходное распределение марковского процесса $Y(t)$, определяемого уравнением (1) или (3), а потом по найденному переходному распределению можно составить разностное уравнение для последовательности $\{\bar{Y}_n\}$. Но точное определение переходного распределения процесса $Y(t)$ возможно только в некоторых частных случаях. В общем же случае придется довольствоваться приближенным определением переходного распределения процесса $Y(t)$. В результате по этому переходному распределению можно будет получить только приближенное разностное уравнение. Последнее делает нецелесообразным применение чрезвычайно сложного алгоритма вывода точного разностного уравнения для последовательности $\{\bar{Y}_n\}$. Раз уж все равно приходится довольствоваться лишь приближенным разностным уравнением даже при использовании данного алгоритма, то целесообразно сделать это более простыми способами.

Прежде всего заменим интеграл по переменной v в уравнении (1) соответствующей интегральной суммой. В результате (1) заменится уравнением

$$dY = a(Y, t)dt + b(Y, t)dW_0 + \sum_{i=1}^N c_i(Y, t)dP_i^0, \quad (6)$$

где $c_i(y, t)$ — p -мерные векторные функции, представляющие собой значения функции $c(y, t, v)$ в

некоторых средних точках v_i соответствующих элементов A_i разбиения q -мерного шара достаточно большого радиуса, $v_i \in A_i$ ($i = 1, \dots, N$), а $P_i^0(t)$ — центрированные простые пуассоновские процессы:

$$P_i^0(t) = P^0(\Delta_t, A_i) - \mu(\Delta_t, A_i), \quad i = 1, 2, \dots, N.$$

Интенсивности этих процессов определяются через математическое ожидание $\mu(\Delta_t, A)$ пуассоновской меры $P(\Delta_t, A)$ по формуле:

$$v_i(t) = \frac{d\mu(\Delta_t, t), A_i}{dt}.$$

Простейший способ замены уравнения (6) разностным уравнением состоит в замене всех дифференциалов элементами интегральных сумм:

$$\begin{aligned} Y((n+1)h) - Y(nh) &= a(Y(nh), nh)h + \\ &+ b(Y(nh), nh) [W_0((n+1)h) - W_0(nh)] + \\ &+ \sum_{i=1}^N c_i(Y(nh), nh) [P_i^0((n+1)h) - P_i^0(nh)]. \end{aligned}$$

Положим

$$\begin{aligned} \bar{Y}_n &= Y(nh); \\ \varphi_n(\bar{Y}_n) &= Y(nh) + a(Y(nh), nh); \\ \psi_{1n}(\bar{Y}_n) &= b(Y(nh), nh); \\ \psi_{in}(\bar{Y}_n) &= c_{i-1}(Y(nh), nh); \\ V_{1n} &= W_0((n+1)h) - W_0(nh); \\ V_{in} &= P_{i-1}^0((n+1)h) - P_{i-1}^0(nh), \\ & \quad i = 2, \dots, N+1. \end{aligned}$$

В результате получим разностное стохастическое уравнение:

$$\bar{Y}_{n+1} = \varphi_n(\bar{Y}_n) + \sum_{i=1}^{N+1} \psi_{in}(\bar{Y}_n)V_{in}.$$

Вводя блочную матрицу $p \times (r + N)$

$$\psi_n(\bar{Y}_n) = [\psi_{1n}(\bar{Y}_n) \cdots \psi_{N+1,n}(\bar{Y}_n)]$$

и $(r + N)$ -мерный случайный вектор

$$V_n = [V_{1n}^T \ V_{2n} \ \cdots \ V_{N+1,n}^T]^T,$$

можем записать полученное разностное уравнение коротко в виде:

$$\bar{Y}_{n+1} = \omega_n(\bar{Y}_n, V_n) = \varphi_n(\bar{Y}_n) + \psi_n(\bar{Y}_n)V_n. \quad (7)$$

Так как винеровский и пуассоновский процессы являются процессами с независимыми приращениями, то случайные векторы V_n образуют последовательность независимых случайных векторов $\{V_n\}$,

причем блоки V_{1n} векторов V_n имеют нормальное распределение $\mathcal{N}(0, \bar{G}_n)$, где

$$\bar{G}_n = \int_{nh}^{(n+1)h} \nu_0(\tau) d\tau \cong \nu_0(nh)h;$$

$\nu_0(t)$ — интенсивность винеровского процесса $W_0(t)$; скалярные блоки V_{in} ($i = 2, \dots, N+1$) имеют пуассоновские распределения с параметрами

$$\mu_{in} = \int_{nh}^{(n+1)h} \nu_i(\tau) d\tau \cong \nu_i(nh)h.$$

Эти распределения полностью определяют распределения случайных векторов V_n . Ковариационная матрица G_n вектора V_n представляет собой блочно-диагональную матрицу

$$G_n = \begin{bmatrix} \bar{G}_n & 0 & 0 & \dots & 0 \\ 0 & \mu_{2n} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mu_{N+1,n} \end{bmatrix}.$$

Уравнение (7) определяет \bar{Y}_{n+1} при данном \bar{Y}_n с точностью до h в детерминированном слагаемом $\varphi_n(\bar{Y}_n)$ и с точностью до $h^{1/2}$ в случайном слагаемом $\psi_n(\bar{Y}_n)V_n$.

Изложенный метод замены стохастического дифференциального уравнения разностными, по существу, не отличается от метода Эйлера численного интегрирования обыкновенных дифференциальных уравнений.

Более точные уравнения (до h^2 и h^3 , $h^{3/2}$ и $h^{5/2}$ соответственно в детерминированном и случайном слагаемом), лежащие в основе МСМ, приведены в приложении.

Таким образом, с точностью до h в детерминированном слагаемом φ_n и $h^{1/2}$ в случайном слагаемом ψ_n в основе совместного параметрического МСМ лежат разностные стохастические уравнения (7). Пользуясь уточненными уравнениями (П.26) приложения, получим соответствующие уравнения с точностью h^2 и $h^{3/2}$.

4 Совместное параметрическое аналитическое и статистическое моделирование

Как известно [1–3], в задачах корреляционного аналитического моделирования процессов в СтС с аддитивными шумами широкое распространение получил метод статистической линейаризации (МСЛ). Для СтС с параметрическими шумами развит метод нормальной аппроксимации (МНА).

Обобщением МНА распределений являются различные приближенные методы, основанные на параметризации распределений. Аппроксимация одномерной характеристической функции $g_1(\lambda; t)$ и соответствующей плотности $f_1(y, t)$ известными функциями $g_1^*(\lambda; \Xi)$, $f_1^*(y; \Xi)$, зависящими от конечномерного векторного параметра Ξ , сводит задачу приближенного определения одномерного распределения к выводу из уравнения для характеристических функций обыкновенных дифференциальных уравнений, определяющих Ξ как функцию времени. Это относится и к остальным многомерным распределениям. При аппроксимации многомерных распределений целесообразно выбирать последовательности функций $\{g_n^*(\lambda_1, \dots, \lambda_n; \Xi_n)\}$ и $\{f_n^*(y_1, \dots, y_n; \Xi_n)\}$, каждая пара которых находилась бы в такой зависимости от векторного параметра Ξ_n , чтобы при любом n множество параметров, образующих вектор Ξ_n , включало в качестве подмножества множество параметров, образующих вектор Ξ_{n-1} . Тогда при аппроксимации n -мерного распределения придется определять только те координаты вектора Ξ_n , которые не были определены ранее при аппроксимации функций $g_1, f_1, \dots, g_{n-1}, f_{n-1}$.

В зависимости от того, что представляют собой параметры, от которых зависят функции $f_n^*(y_1, \dots, y_n; \Xi_n)$ и $g_n^*(\lambda_1, \dots, \lambda_n; \Xi_n)$, аппроксимирующие неизвестные многомерные плотности $f_n(y_1, \dots, y_n; t_1, \dots, t_n)$ и характеристические функции $g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)$, используются различные приближенные методы решения уравнений, определяющих многомерные распределения вектора состояния системы Y_t , в частности методы моментов, семиинвариантов, ортогональных разложений и др.

Уравнения МАМ для МСТС (7), если использовать метод ортогональных разложений (МОР) [2, 3], имеют следующий вид:

$$f_n^*(y_1, \dots, y_n; \Xi_n) = w_n(y_1, \dots, y_n; \bar{m}_n, \bar{K}_n) \left\{ 1 + \sum_{\rho=3}^N \sum_{|\nu_1| + \dots + |\nu_n| = \rho} c_{\nu_1, \dots, \nu_n}^{l_1, \dots, l_n} p_{\nu_1, \dots, \nu_n}(y_1, \dots, y_n) \right\} \quad (n = 1, 2, \dots); \quad (8)$$

$$c_{\nu}^{l+1} = Mq_{\nu}(\omega_l(Y_l, V_l)) = \left[q_{\nu} \left(\frac{\partial}{i\partial\lambda} \right) M \exp \{ i\lambda^T \omega_l(Y_l, V_l) \} \right]_{\lambda=0}; \quad (9)$$

$$c_{\nu_1, \dots, \nu_n}^{l_1, \dots, l_n+1} = Mq_{\nu_1, \dots, \nu_n}(Y_{l_1}, \dots, Y_{l_{n-1}}, \omega_{l_n}(Y_{l_n}, V_{l_n})). \quad (10)$$

Здесь введены обозначения:

$$\bar{m}_n = [m_{l_1}^T \dots m_{l_n}^T]^T; \quad \bar{K}_n = \begin{bmatrix} K_{l_1} & K_{l_1 l_2} & \dots & K_{l_1 l_n} \\ K_{l_1 l_2}^T & K_{l_2} & \dots & K_{l_2 l_n} \\ \dots & \dots & \dots & \dots \\ K_{l_1 l_n}^T & K_{l_2 l_n}^T & \dots & K_{l_n} \end{bmatrix}; \quad (11)$$

$$c_{\nu_1, \dots, \nu_n}^{l_1, \dots, l_n} = \left[q_{\nu_1, \dots, \nu_n} \left(\frac{\partial}{i\partial\lambda_1} \dots \frac{\partial}{i\partial\lambda_n} \right) g_{l_1, \dots, l_n}(\lambda_1, \dots, \lambda_n) \right]_{\lambda_1 = \dots = \lambda_n = 0};$$

$\{p_{\nu_1, \dots, \nu_n}, q_{\nu_1, \dots, \nu_n}\}$ — биортогональные полиномы, причем

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} w(y_1, \dots, y_n; \bar{m}, \bar{K}) p_{\nu_1, \dots, \nu_n}(y_1, \dots, y_n) q_{\mu_1, \dots, \mu_n}(y_1, \dots, y_n) dy_1 \dots dy_n = \delta_{\nu_1 \mu_1} \dots \delta_{\nu_n \mu_n},$$

где $\nu_l = [\nu_{l_1} \dots \nu_{l_p}]^T$, $l = 1, 2, \dots$; $\nu_{l_1}, \dots, \nu_{l_n} = 0, 1, \dots, N$; $|\nu_1|, \dots, |\nu_n| = 1, \dots, N - n + 1$; $|\nu_l| = |\nu_{l_1}| + \dots + |\nu_{l_p}| = n, \dots, N$; $w_n(y_1, \dots, y_n; \bar{m}_n, \bar{K}_n)$ — эталонная плотность, как правило, нормальная:

$$w_n(y_1, \dots, y_n; \bar{m}_n, \bar{K}_n) = [(2\pi)^n |\bar{K}_n|]^{-1/2} \exp \left\{ -\frac{1}{2} (u_n - \bar{m}_n)^T \bar{K}_n^{-1} (u_n - \bar{m}_n) \right\},$$

где в дополнение к обозначениям (11) принято $u_n = [y_1^T y_2^T \dots y_n^T]^T$. Отсюда как частный случай вытекают уравнения методов моментов, квазимоментов, семиинвариантов и др., а также параметров структурной параметризации [1–3].

Таким образом, с точностью до h и $h^{1/2}$ соответственно при детерминированном и случайном слагаемом уравнениями совместного параметрического МСМ и МАМ служат уравнения (7)–(10).

5 Корреляционные уравнения статистического и аналитического моделирования

Основываясь на результатах разд. 3 и 4, ограничимся уравнениями с точностью до h в детерминированном слагаемом φ_n и с точностью до $h^{1/2}$ в случайном слагаемом ψ_n и заменим нелинейные функции $\varphi_n, \psi_{1n}, \psi_{in}, \psi_n$ в (7) статистически линеаризованными зависимостями:

$$\begin{aligned} \varphi_n &\approx \hat{\varphi}_n = \\ &= \bar{Y}_n + k_0^a(m_n, K_n)m_n + k_1^a(m_n, K_n)(\bar{Y} - m_n) = \\ &= \alpha_n(m_n, K_n)Y_n + \alpha_{0n}(m_n, K_n); \end{aligned}$$

$$\psi_{1n} \approx \hat{\psi}_{1n} = k_0^b(m_n, K_n)m_n + k_1^b(m_n, K_n)(\bar{Y}_n - m_n);$$

$$\begin{aligned} \psi_{in} &\approx \hat{\psi}_{in} = \\ &= k_0^{ci-1}(m_n, K_n)m_n + k_1^{ci-1}(m_n, K_n)(\bar{Y}_n - m_n); \end{aligned}$$

$$\begin{aligned} \psi_n &\approx \hat{\psi}_n = \\ &= k_0^{\psi n}(m_n, K_n)m_n + k_1^{\psi n}(m_n, K_n)(\bar{Y}_n - m_n), \end{aligned}$$

где $m_n = M\bar{Y}_n = MY(nh)$; $K_n = M\bar{Y}_n\bar{Y}_n^T = K(nh)$; $k_0^a(m_n, K_n)$ и $k_1^a(m_n, K_n)$ — коэффициенты нормальной статистической линеаризации соответствующих нелинейных функций. Тогда, введя обозначение

$$\hat{\psi}_n(m_n, K_n) = \gamma_{0n}(m_n, K_n) + \sum_{\rho=1}^p \gamma_{\rho n} \bar{Y}_{\rho n},$$

представим искомое уравнение МСМ в виде:

$$\begin{aligned} \bar{Y}_{n+1} &= \alpha_n(m_n, K_n)\bar{Y}_n + \alpha_{0n}(m_n, K_n) + \\ &+ \left[\gamma_{0n}(m_n, K_n) + \sum_{\rho=1}^p \gamma_{\rho n}(m_n, K_n)\bar{Y}_{\rho n} \right] V_n, \quad (12) \end{aligned}$$

где $\bar{Y}_{\rho n}$ — компонента \bar{Y}_n с номером ρ ($\rho = 1, \dots, p$).

Замечание 5.1. Уравнение (12) является разностным стохастическим уравнением Ито с параметрическими шумами. Оно подробно изучено в [1–3].

Применяя к (12) теорию дискретных стохастических систем с параметрическими шумами [1–3],

получим следующие детерминированные уравнения для m_n, K_n и $K(j, l)$ вместе с соответствующими начальными условиями:

$$m_{n+1} = \alpha(m_n, K_n)m_n + \alpha_{0n}(m_n, K_n), \quad m_1 = MY_1; \quad (13)$$

$$K_{n+1} = \alpha_n(m_n, K_n)K_n\alpha_n(m_n, K_n)^T + \gamma_{0n}(m_n, K_n)G_n\gamma_{0n}(m_n, K_n)^T + \sum_{j=1}^p [\gamma_{0n}(m_n, K_n)G_n\gamma_{jn}(m_n, K_n)^T + \gamma_{jn}(m_n, K_n)G_n\gamma_{0n}(m_n, K_n)^T] + \sum_{j,l=1}^p \gamma_{jn}(m_n, K_n)G_n\gamma_{ln}(m_n, K_n) [m_{nj}m_{nl} + K_{njl}], \quad K_1 = MY_1Y_1^T; \quad (14)$$

$$K(j, L+1) = K(j, l)\alpha_l(m_l, K_l)^T, \quad K(j, j) = K_j \quad (l > j), \quad K(j, l) = K(l, j)^T \quad (l < j). \quad (15)$$

Таким образом, совокупность стохастического разностного уравнения (12) и детерминированных уравнений (13)–(15) представляет собой систему искоемых нелинейных корреляционных уравнений МСМ и МАМ.

Точность совместного корреляционного статистического и аналитического моделирования, особенно для разрывных нелинейных функций a, b, c в (7), можно повысить, если вместо (13)–(15) воспользоваться дискретной версией непрерывных уравнений МНА [1–3].

6 Заключение

Для нелинейных дифференциальных СтС на многообразиях, понимаемых в смысле Ито, разработаны дискретные методы параметрического статистического и математического моделирования различной точности. Предполагается, что в качестве параметров одно- и многомерных распределений выбраны коэффициенты совместных ортогональных разложений плотностей для различных моментов времени. Выведены уравнения нелинейной корреляционной теории статистического и аналитического моделирования.

Полученные результаты положены в основу разрабатываемого в ИПИ РАН символического инструментального программного обеспечения в среде MATLAB [17].

Приложение

1. Обобщая [2, 3], для вывода более точных разностных уравнений, чем (7), заменим (6) соответствующим интегральным уравнением:

$$\Delta Y_n = \int_{nh}^{(n+1)h} a(Y_\tau, \tau) d\tau + \int_{nh}^{(n+1)h} b(Y_\tau, \tau) dW_0(\tau) + \sum_{i=1}^N \int_{nh}^{(n+1)h} c_i(Y_\tau, \tau) dP_i^0(d\tau). \quad (П.1)$$

С целью приближенного вычисления интегралов определим Y_τ путем линейной интерполяции случайной функции $Y(t)$ на интервале $(nh, (n+1)h)$. Тогда, полагая по-прежнему $\Delta Y_n = Y((n+1)h) - Y(nh)$, будем иметь:

$$a(Y_\tau, \tau) \cong a\left(\bar{Y}_n + \frac{\tau - nh}{h} \Delta Y_n, \tau\right); \quad b(Y_\tau, \tau) \cong b\left(\bar{Y}_n + \frac{\tau - nh}{h} \Delta Y_n, \tau\right); \quad c_i(Y_\tau, \tau) \cong c_i\left(\bar{Y}_n + \frac{\tau - nh}{h} \Delta Y_n, \tau\right), \quad i = 1, \dots, N.$$

Для вычисления правых частей в этих формулах применим обобщенную формулу Ито [2, 3]:

$$\varphi(Y + dY, t + dt) = \varphi(Y, t) + \left\{ \varphi_t(Y, t) + \varphi_y(Y, t)^T a(Y, t) + \frac{1}{2} \varphi_{yy}(Y, t) : \sigma(Y, t) + \sum_{i=1}^N [\varphi(Y + c_i(Y, t), t) - \varphi(Y, t) - \varphi_y(Y, t)^T c_i(Y, t)] \nu_i(t) \right\} dt + \varphi_y(Y, t)^T b(X, t) dW_0 + \sum_{i=1}^N [\varphi(Y + c_i(Y, t), t) - \varphi(Y, t)] dP_i^0, \quad (П.2)$$

где $\varphi_t(y, t)$ — частная производная функции $\varphi(y, t)$ (возможно, векторной) по времени t ; $\varphi_y(y, t)$ — матрица, строки которой представляют собой частные производные матрицы-строки $\varphi(y, t)^T$ по компонентам вектора y ;

$\varphi_{yy}(y, t) : \sigma(y, t)$ — вектор, компонентами которого служат следы произведений матриц вторых производных соответствующих компонент векторной функции $\varphi(y, t)$ по компонентам вектора y на матрицу $\sigma(y, t)$:

$$\begin{aligned} [\varphi_{yy}(y, t) : \sigma(y, t)]_k &= \text{tr}[\varphi_{kyy}(y, t)\sigma(y, t)] ; \\ \sigma(y, t) &= b(y, t)\nu_0(t)b(y, t)^T . \end{aligned} \quad (\text{П.3})$$

Формулу (П.2) удобно преобразовать так, чтобы в нее входили непосредственно пуассоновские процессы $P_i(t)$, а не централизованные процессы $P_i^0(t)$. Имея в виду, что

$$P_i^0(t) = P_i(t) - \int_0^t \nu_i(\tau) d\tau , \quad (\text{П.4})$$

можем переписать формулу (П.2) в виде:

$$\begin{aligned} \varphi(Y + dY, t + dt) &= \varphi(Y, t) + \left\{ \varphi_t(Y, t) + \varphi_y(Y, t)^T \left[a(Y, t) - \sum_{i=1}^N c_i(Y, t)\nu_i(t) \right] + \frac{1}{2} \varphi_{yy}(Y, t) : \sigma(Y, t) \right\} dt + \\ &+ \varphi_y(Y, t)^T b(Y, t) dW_0 + \sum_{i=1}^N [\varphi(Y + c_i(Y, t), t) - \varphi(Y, t)] dP_i . \end{aligned} \quad (\text{П.5})$$

Для распространения формул (П.2) и (П.5) на матричные функции φ необходимо видоизменить запись некоторых ее членов, чтобы в рамках алгебры матриц выражения $\varphi_y(y, t)$ имели смысл и для матричной функции φ . Имея в виду, что в случае скалярной или векторной функции φ

$$\varphi_y(y, t)^T u = \sum_{i=1}^n \frac{\partial \varphi(y, t)}{\partial y_i} u_i = u^T \frac{\partial}{\partial y} \varphi(y, t)$$

для любого p -мерного вектора u , можем заменить слагаемые $\varphi_y(Y, t)^T a(Y, t)$, $\varphi_y(Y, t)^T b(Y, t) dW_0$, $\varphi_y(Y, t)^T c_i(Y, t)$ соответственно слагаемыми $a(Y, t)^T (\partial/\partial y)^T \varphi(Y, t)$, $dW_0^T b(Y, t) (\partial/\partial y) \varphi(Y, t)$, $c_i(Y, t)^T (\partial/\partial y) \varphi(Y, t)$. Поэтому формулы (П.2) и (П.5) будут справедливы и для матричных функций φ , так как $u^T (\partial/\partial y)$ представляет собой скалярный дифференциальный оператор для любого p -мерного вектора u и его применение к векторной или матричной функции φ означает его применение ко всем компонентам вектора φ или ко всем элементам матрицы φ . Величина $\varphi_{yy}(Y, t) : \sigma(Y, t)$ в случае матричной функции φ представляет собой матрицу, элементами которой служат следы произведений на матрицу $\sigma(y, t)$ матриц вторых производных соответствующих элементов матрицы $\varphi(y, t)$ по компонентам вектора y на матрицу $\sigma(y, t)$:

$$[\varphi_{yy}(y, t) : \sigma(y, t)]_{kl} = \text{tr}[\varphi_{klyy}(y, t)\sigma(y, t)] .$$

Используя формулу (П.5) и ее модификацию, пригодную для матричной функции φ , а также учитывая (П.3) и то, что согласно (1) и (П.4)

$$\frac{\tau - nh}{h} \Delta Y_n = \frac{\tau - nh}{h} a(\bar{Y}_n, nh)h + \frac{\tau - nh}{h} b(\bar{Y}_n, nh) \Delta W_n + \sum_{i=1}^N \frac{\tau - nh}{h} c_i(\bar{Y}_n, nh) (\Delta P_{in} - \nu_{in}h) , \quad (\text{П.6})$$

где, как и раньше, $\Delta W_n = W_0((n+1)h) - W_0(nh)$, $\Delta P_{in} = P_i((n+1)h) - P_i(nh)$, будем иметь с точностью до малых высшего порядка относительно h следующие равенства:

$$\begin{aligned} a\left(\bar{Y}_n + \frac{\tau - nh}{h} \Delta Y_n, \tau\right) &= a + a_t(\tau - nh) + \frac{\tau - nh}{h} a_y^T \left(a - \sum_{i=1}^N c_i \nu_{in} \right) h + \\ &+ \frac{1}{2} \left(\frac{\tau - nh}{h} \right)^2 (a_{yy} : \sigma)h + \frac{\tau - nh}{h} a_y^T b \Delta W_n + \sum_{i=1}^N \left[a\left(\bar{Y}_n + \frac{\tau - nh}{h} c_i, nh\right) - a \right] \Delta P_{in} ; \end{aligned} \quad (\text{П.7})$$

$$\begin{aligned} b\left(\bar{Y}_n + \frac{\tau - nh}{h} \Delta Y_n, \tau\right) &= b + b_t(\tau - nh) + \frac{\tau - nh}{h} \left(a^T - \sum_{i=1}^N c_i^T \nu_{in} \right) \frac{\partial}{\partial y} b h + \\ &+ \frac{1}{2} \left(\frac{\tau - nh}{h} \right)^2 (b_{yy} : \sigma)h + \frac{\tau - nh}{h} \Delta W_n^T b^T \frac{\partial}{\partial y} b + \sum_{i=1}^N \left[b\left(\bar{Y}_n + \frac{\tau - nh}{h} c_i, nh\right) - b \right] \Delta P_{in} ; \end{aligned} \quad (\text{П.8})$$

$$c_i \left(\bar{Y}_n + \frac{\tau - nh}{h} \Delta Y_n, \tau \right) = c_i + c_{it}(\tau - nh) + \frac{\tau - nh}{h} c_{iy}^T \left(a - \sum_{j=1}^N c_j \nu_{jn} \right) h + \\ + \frac{1}{2} \left(\frac{\tau - nh}{h} \right)^2 (c_{iyy} : \sigma) h + \frac{\tau - nh}{h} c_{iy}^T b \Delta W_n + \sum_{j=1}^N \left[c_i \left(\bar{Y}_n + \frac{\tau - nh}{h} c_j, nh \right) - c_i \right] \Delta P_{in}. \quad (\text{П.9})$$

Здесь аргументы \bar{Y}_n, nh у всех функций, зависящих от них, для краткости опущены.

2. Для повышения точности вычислений, в частности ковариационной матрицы нормально распределенного случайного вектора $\Delta W_n = W_0((n+1)h) - W_0(nh)$ и параметров пуассоновских распределений случайных величин $\Delta P_{in} = P_i((n+1)h) - P_i(nh)$, можно взять значения интенсивностей $\nu_0(t)$ процесса $W_0(t)$ в формуле (П.3) и $\nu_i(t)$ пуассоновских потоков, порождающих процессы $P_i(t)$, в средней точке $nh + h/2$ интервала $(nh, (n+1)h)$:

$$\left. \begin{aligned} \sigma(\bar{Y}_n, nh) &= b(\bar{Y}_n, nh) \nu_0 \left(nh + \frac{h}{2} \right) b(\bar{Y}_n, nh)^T; \\ \nu_{in} &= \nu_i \left(nh + \frac{h}{2} \right), \quad i = 1, \dots, N. \end{aligned} \right\} \quad (\text{П.10})$$

Для облегчения дальнейших вычислений найдем приращения функций a, b, c_i в суммах уравнения (П.6) путем линейной интерполяции на малом интервале $(nh, (n+1)h)$:

$$\left. \begin{aligned} a \left(\bar{Y}_n + \frac{\tau - nh}{h} c_i, nh \right) - a &\cong \frac{\tau - nh}{h} \Delta_i a_n; \\ b \left(\bar{Y}_n + \frac{\tau - nh}{h} c_i, nh \right) - b &\cong \frac{\tau - nh}{h} \Delta_i b_n; \\ c_i \left(\bar{Y}_n + \frac{\tau - nh}{h} c_j, nh \right) - c_i &\cong \frac{\tau - nh}{h} \Delta_j c_{in}; \end{aligned} \right\} \quad (\text{П.11})$$

$$\Delta_i a_n = a(\bar{Y}_n + c_i, nh) - a(\bar{Y}_n, nh); \quad \Delta_i b_n = b(\bar{Y}_n + c_i, nh) - b(\bar{Y}_n, nh); \quad \Delta_j c_{in} = c_i(\bar{Y}_n + c_j, nh) - c_i(\bar{Y}_n, nh).$$

Используя формулы (П.7)–(П.11), найдем приближенные выражения интегралов в (П.1):

$$\int_{nh}^{(n+1)h} a(Y_\tau, \tau) d\tau = ah + \frac{1}{2} \left[a_t + a_y^T \left(a - \sum_{i=1}^N c_i \nu_{in} \right) + \frac{1}{3} a_{yy} : \sigma \right] h^2 + \frac{1}{2} \left(a_y^T b \Delta W_n + \sum_{i=1}^N \Delta_i a_n \Delta P_{in} \right) h;$$

$$\int_{nh}^{(n+1)h} b(Y_\tau, \tau) dW_0(\tau) = b \Delta W_n + \left\{ \left[b_t + \left(a^T - \sum_{i=1}^N c_i^T \nu_{in} \right) \frac{\partial}{\partial y} b \right] h + \right. \\ \left. + \Delta W_n^T b^T \frac{\partial}{\partial y} b + \sum_{i=1}^N \Delta_i b_n \Delta P_{in} \right\} \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dW_0(\tau) + \frac{1}{2} (b_{yy} : \sigma) \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dW_0(\tau),$$

$$\int_{nh}^{(n+1)h} c_i(Y_\tau, \tau) dP_i^0(\tau) = c_i \Delta P_{in} + \left\{ \left[c_{it} + c_{iy}^T \left(a - \sum_{j=1}^N c_j \nu_{jn} \right) \right] h + \right. \\ \left. + c_{iy}^T b \Delta W_n + \sum_{j=1}^N \Delta_j c_{in} \Delta P_{jn} \right\} \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dP_i^0(\tau) + \frac{1}{2} (c_{iyy} : \sigma) \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dP_i^0(\tau);$$

$$\int_{nh}^{(n+1)h} c_i(Y_\tau, \tau) \nu_i(\tau) d\tau = \\ = \left\{ c_i h + \frac{1}{2} \left[c_{it} + c_{iy}^T \left(a - \sum_{j=1}^N c_j \nu_{jn} \right) + \frac{1}{3} c_{iyy} : \sigma \right] h^2 + \frac{1}{2} \left(c_{iy}^T b \Delta W_n + \sum_{j=1}^N \Delta_j c_{in} \Delta P_{jn} \right) h \right\} \nu_{in}.$$

Далее, подставив полученные приближенные выражения интегралов в (П.1) и приняв во внимание, что $\Delta Y_n = Y((n+1)h) - Y_i(nh) = \bar{Y}_{n+1} - \bar{Y}_n$, придем к разностному уравнению:

$$\bar{Y}_{n+1} = \varphi_n(\bar{Y}_n) + \sum_{i=1}^{N+1} \left[\psi_{in}(\bar{Y}_n) V_{in} + \psi'_{in}(\bar{Y}_n, V_{in}^{(1)}) V'_{in} + \psi''_{in}(\bar{Y}_n) V''_{in} \right]. \quad (\text{П.12})$$

Здесь введены следующие обозначения:

$$\begin{aligned} \varphi_n(\bar{Y}_n) = & \bar{Y}_n + \left[a(\bar{Y}_n, nh) - \sum_{j=1}^N c_j(\bar{Y}_n, nh) \nu_{jn} \right] h + \\ & + \frac{1}{2} \left\{ a_t(\bar{Y}_n, nh) - \sum_{j=1}^N c_{jt}(\bar{Y}_n, nh) \nu_{jn} + \left[a_x(\bar{Y}_n, nh)^T - \sum_{j=1}^N c_{jy}(\bar{Y}_n, hn)^T \nu_{jn} \right] \left[a(\bar{Y}_n, nh) - \sum_{j=1}^N c_j(\bar{Y}_n, hn) \nu_{jn} \right] + \right. \\ & \left. + \frac{1}{3} \left[a_{yy}(\bar{Y}_n, nh) - \sum_{j=1}^N c_{jyy}(\bar{Y}_n, nh) \nu_{jn} \right] : \sigma(\bar{Y}_n, nh) \right\} h^2; \quad (\text{П.13}) \end{aligned}$$

$$\psi_{1n}(\bar{Y}_n) = b(\bar{Y}_n, nh) + \frac{1}{2} \left[a_y(\bar{Y}_n, nh)^T - \sum_{j=1}^N c_j(\bar{Y}_n, hn)^T \nu_{jn} \right] b(\bar{Y}_n, nh); \quad (\text{П.14})$$

$$\psi_{in}(\bar{Y}_n) = c_{i-1}(\bar{Y}_n, nh) + \left[\Delta_{i-1} a_n - \sum_{j=1}^N \Delta_{i-1} c_{jn} \nu_{jn} \right] h; \quad (\text{П.15})$$

$$\begin{aligned} \psi'_{1n}(\bar{Y}_n, V_n^{(1)}) = & \left\{ b(\bar{Y}_n, nh) + \left[a(\bar{Y}_n, nh)^T - \sum_{j=1}^N c_j(\bar{Y}_n, hn)^T \nu_{jn} \right] \frac{\partial}{\partial y} b(\bar{Y}_n, nh) + \right. \\ & \left. + V_{1n}^T b(\bar{Y}_n, nh)^T \frac{\partial}{\partial y} b(\bar{Y}_n, nh) \right\} h + \sum_{j=1}^N \Delta_j b_n V_{j+1,n}; \quad (\text{П.16}) \end{aligned}$$

$$\begin{aligned} \psi'_{in}(\bar{Y}_n, V_n^{(1)}) = & \left\{ c_{i-1,t}(\bar{Y}_n, nh) + c_{i-1,y}(\bar{Y}_n, nh)^T \left[a(\bar{Y}_n, nh) - \sum_{j=1}^N c_j(\bar{Y}_n, hn) \nu_{jn} \right] \right\} h + \\ & + c_{i-1,y}(\bar{Y}_n, nh)^T b(\bar{Y}_n, nh) V_{1,n} + \sum_{j=1}^N \Delta_j c_{j-1,n} V_{j+1,n}; \quad (\text{П.17}) \end{aligned}$$

$$\psi''_{1n}(\bar{Y}_n) = \frac{1}{2} [b_{yy}(\bar{Y}_n, nh) : \sigma(\bar{Y}_n, nh)]; \quad (\text{П.18})$$

$$\psi''_{in}(\bar{Y}_n) = \frac{1}{2} [c_{i-1,yy}(\bar{Y}_n, nh) : \sigma(\bar{Y}_n, nh)]; \quad (\text{П.19})$$

$$V_{1n} = \Delta W_n = W_0((n+1)h) - W_0(nh); \quad (\text{П.20})$$

$$V_{in} = \Delta P_{i-1,n} = P_{i-1}((n+1)h) - P_{i-1}(nh), \quad i = 2, \dots, N; \quad (\text{П.21})$$

$$V'_{1n} = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dW_0(\tau), \quad V''_{1n} = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dW_0(\tau); \quad (\text{П.22})$$

$$V'_{in} = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dP_{i-1}^0(\tau); \quad (\text{П.23})$$

$$V''_{in} = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dP_{i-1}^0(\tau), \quad i = 2, \dots, N+1. \quad (\text{П.24})$$

Обозначим $V_n^{(1)} = [V_{1n}^T \ V_{2n} \ \dots \ V_{n+1}]^T$ и введем блочную матрицу

$$\psi_n(\bar{Y}_n, V_n^{(1)}) = [\psi_{1n}(\bar{Y}_n) \ \psi'_{1n}(\bar{Y}_n, V_n^{(1)}) \ \psi''_{1n}(\bar{Y}_n) \ \dots \ \psi_{N+1,n}(\bar{Y}_n) \ \psi'_{N+1,n}(\bar{Y}_n, V_n^{(1)}) \ \psi''_{N+1,n}(\bar{Y}_n)]$$

и блочный случайный вектор

$$V_n = \left[V_{1n}^T \ V_{1n}'^T \ V_{1n}''^T \ V_{2n}^T \ V_{2n}'^T \ V_{2n}''^T \ \cdots \ V_{N+1,n} \ V_{N+1,n}' \ V_{N+1,n}'' \right]^T. \quad (\text{П.25})$$

Тогда можем коротко записать уравнение (П.12) в виде:

$$\bar{Y}_{n+1} = \varphi_n(\bar{Y}_n) + \psi_n(\bar{Y}_n, V_n^{(1)})V_n. \quad (\text{П.26})$$

Входящие в уравнения (П.26) величины определены формулами (П.13)–(П.25).

3. Найдем распределение случайного вектора V_n . Ясно, что математические ожидания случайных величин V_{1n}' , V_{1n}'' равны нулю и что случайный вектор $[V_{1n}^T \ V_{1n}'^T \ V_{1n}''^T]$ имеет нормальное распределение, а скалярные случайные величины $V_{2n}, \dots, V_{N+1,n}$ — пуассоновские распределения с параметрами

$$\mu_{in} = \int_{nh}^{(n+1)h} \nu_{i-1}(\tau) d\tau = \nu_{i-1} \left(nh + \frac{1}{2} h \right) h, \quad i = 2, \dots, N+1.$$

Очевидно также, что тройки случайных величин $V_{1n}, V_{1n}', V_{1n}'', \dots, V_{N+1,n}, V_{N+1,n}', V_{N+1,n}''$ независимы в силу независимости процессов $W_0(t), P_1(t), \dots, P_N(t)$ и что при различных n величины V_n независимы. Однако при любых данных i, n величины $V_{in}, V_{in}', V_{in}''$ зависимы.

Для полного определения распределения случайного вектора V_n в (П.26) достаточно найти ковариационную матрицу нормально распределенного случайного вектора $[V_{in}^T \ V_{in}'^T \ V_{in}''^T]$. Пользуясь известными формулами ковариационных и взаимных ковариационных матриц стохастических интегралов, находим блоки ковариационной матрицы K_{1n} случайного вектора $[V_{in}^T \ V_{in}'^T \ V_{in}''^T]$:

$$\begin{aligned} K_{1n,11} &= MV_{1n}V_{1n}^T = \int_{nh}^{(n+1)h} \nu_0(\tau) d\tau \cong \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,12} &= MV_{1n}V_{1n}'^T = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} \nu_0(\tau) d\tau \cong \frac{1}{2} \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,13} &= MV_{1n}V_{1n}''^T = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 \nu_0(\tau) d\tau \cong \frac{1}{3} \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,21} &= K_{1n,12}, \quad K_{1n,31} = K_{1n,13}; \\ K_{1n,23} &= MV_{1n}V_{1n}''^T = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^3 \nu_0(\tau) d\tau \cong \frac{1}{4} \nu_0 \left(nh + \frac{h}{2} \right) h; \\ K_{1n,31} &= K_{1n,13}; \quad K_{1n,32} = K_{1n,23}; \\ K_{1n,33} &= MV_{1n}''V_{1n}''^T = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^4 \nu_0(\tau) d\tau \cong \frac{1}{5} \nu_0 \left(nh + \frac{h}{2} \right) h. \end{aligned}$$

Практически целесообразно аппроксимировать стохастические интегралы от неслучайных функций в (П.23), определяющие величины V_{in}', V_{in}'' ($i = 2, \dots, N+1$), с помощью аналога интегральной теоремы о среднем для стохастических интегралов:

$$V_{in}' = \int_{nh}^{(n+1)h} \frac{\tau - nh}{h} dP_{i-1}^0(\tau) \cong \left[\frac{\int_{nh}^{(n+1)h} ((\tau - nh)/h) \nu_{i-1}(\tau) d\tau}{\int_{nh}^{(n+1)h} \nu_{i-1}(\tau) d\tau} \right] \Delta P_{i-1,n} \cong \frac{1}{2} \Delta P_{i-1,n} = \frac{1}{2} V_{in};$$

$$V_{in}'' = \int_{nh}^{(n+1)h} \left(\frac{\tau - nh}{h} \right)^2 dP_{i-1}^0(\tau) \cong \left[\frac{\int_{nh}^{(n+1)h} ((\tau - nh)/h)^2 \nu_{i-1}(\tau) d\tau}{\int_{nh}^{(n+1)h} \nu_{i-1}(\tau) d\tau} \right] \Delta P_{i-1,n} \cong \frac{1}{3} \Delta P_{i-1,n} = \frac{1}{3} V_{in}, \quad i = 2, \dots, N + 1.$$

4. При статистическом моделировании системы с помощью уравнения (П.26) также не представляет трудностей моделировать случайные величины, распределенные по нормальному и пуассоновским законам. Легко видеть, что правая часть разностного уравнения (П.26) определена с точностью до h^2 в детерминированном (при данном \bar{Y}_n) слагаемом $\varphi_n(\bar{Y}_n)$ и с точностью до $h^{3/2}$ в случайном слагаемом $\psi_n(\bar{Y}_n, V_n^{(1)})V_n$.

При выводе уравнения (П.26) были допущены две небольшие неточности. Во-первых, при замене \bar{Y}_τ величиной $\bar{Y}_n + (\tau - nh)\Delta Y_n/h$ случайные функции $b(Y_\tau, \tau)$ и $c_i(Y_\tau, \tau)$, независимые от $dW_0(\tau)$ и $dP_i^0(\tau)$ в силу конструкции интеграла Ито, были заменены неслучайными функциями, зависящими от случайного параметра ΔY_n , который зависит от значений $dW_0(\tau)$ и $dP_i^0(\tau)$ в интервале $(nh, (n + 1)h)$. Во-вторых, если $c_i(y, t) \neq 0$ хотя бы при одном i , реализации случайного процесса $Y(t)$ имеют разрывы первого рода в случайных точках, несмотря на его среднеквадратическую непрерывность. Поэтому линейную интерполяцию данного процесса, строго говоря, проводить нельзя. Первую из этих неточностей можно устранить двумя способами. Первый состоит в замене интерполяции процесса $Y(t)$ экстраполяцией, что равноценно замене ΔY_n в получаемом выражении для Y_τ величиной ΔY_{n-1} . Однако это приведет к появлению в правой части разностного уравнения величин \bar{Y}_{n-1} и $V_{n-1}^{(1)}$, т. е. к замене уравнения первого порядка разностным уравнением второго порядка. Второй способ состоит в отказе от интерполяции процесса $Y(t)$ на интервале $(nh, (n + 1)h)$ и непосредственном выражении приращений функций $a(Y_\tau, \tau)$, $b(Y_\tau, \tau)$, $c_i(Y_\tau, \tau)$ на малом интервале $(nh, (n + 1)h)$ по обобщенной формуле Ито с заменой в ней дифференциалов приращениями. При этом способе устраняется и вторая допущенная неточность. Но полученное таким путем разностное уравнение будет более сложным. В него войдут случайные величины, представляющие собой двойные интегралы по компонентам винеровского процесса $W(t) = W_0(t)$ и по пуассоновским процессам:

$$\int_{nh}^{(n+1)h} \int_{nh}^{\tau} dW_j(\sigma) dW_j(\tau); \quad \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_i(\sigma) dP_j^0(\tau); \quad \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_i^0(\sigma) dW_j(\tau); \quad \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_j^0(\sigma) dW_i(\tau).$$

Распределения этих случайных величин найти чрезвычайно сложно, и только первые два из них легко вычисляются при $j = i$:

$$\int_{nh}^{(n+1)h} \int_{nh}^{\tau} dW_i(\sigma) dW_i(\tau) = \frac{[\Delta W_{in}]^2 - \nu_{ii}(nh + h/2)h}{2}; \quad \int_{nh}^{(n+1)h} \int_{nh}^{\tau} dP_i^0(\sigma) dP_i^0(\tau) = \frac{[\Delta P_{in}]^2 - \Delta P_{in}}{2}.$$

Что касается второй неточности, то она не может существенно повлиять на результат, так как вероятность появления скачка пуассоновского процесса в достаточно малом интервале $(nh, (n + 1)h)$ ничтожно мала.

Точность аппроксимации дифференциального стохастического уравнения разностным можно повышать и дальше. В частности, в одном из способов сначала достаточно выразить $a(Y_\tau, \tau)$, $b(Y_\tau, \tau)$, $c_i(Y_\tau, \tau)$ на интервале $(nh, (n + 1)h)$ интегральной формулой Ито, соответствующей дифференциальной формуле:

$$a(Y_\tau, \tau) = a(\bar{Y}_n, nh) + \int_{nh}^{\tau} \left(a_t(Y_s, s) + a_x(Y_s, s)^T \left[a(Y_s, s) - \sum_{i=1}^N c_i(Y_s, \nu_i(s)) + \frac{1}{2} a_{yy}(Y_s, s) : \sigma(Y_s, s) \right] \right) ds + \int_{nh}^{\tau} a_y(Y_s, s)^T b(Y_s, s) dW_0(s) + \sum_{i=1}^N \int_{nh}^{\tau} [a(Y_s + c_i(Y_s, s), s) - a(Y_s, s)] dP_i^0(s). \quad (\text{П.27})$$

При этом формулы для $b(Y_\tau, \tau)$ и $c_i(Y_\tau, \tau)$ аналогичны. Затем применим к полученным интегралам тот же прием, который использовался для вычисления интегралов в (П.1). В результате получим правую часть разностного уравнения с точностью до h^3 в детерминированном (при данном \bar{Y}_n) слагаемом и $h^{5/2}$ в случайном слагаемом. Процесс уточнения разностного уравнения, соответствующего данному стохастическому дифференциальному уравнению, можно продолжать неограниченно. Однако каждое новое уточнение требует существования производных функций a , b , c_i все более высоких порядков.

Для уточнения разностного уравнения можно применить следующий способ: выразить подынтегральные функции в (П.27) и аналогичных формулах для $b(Y_\tau, \tau)$ и $c_i(Y_\tau, \tau)$ их выражениями по обобщенной формуле Ито, заменив в ней дифференциалы приращениями. При этом в разностное уравнение войдут тройные интегралы по компонентам винеровского процесса $W_0(t)$ и пуассоновским процессам $P_i(t)$. Для дальнейшего уточнения аппроксимации стохастического дифференциального уравнения разностным в этом случае следует подынтегральные функции в (П.27) и соответствующих формулах для $b(Y_\tau, \tau)$ и $c_i(Y_\tau, \tau)$, в свою очередь, представить интегральной формулой Ито, а затем уже применять дифференциальную формулу Ито с заменой дифференциалов приращениями. Данный процесс можно продолжать неограниченно, и в итоге он приведет к представлению процесса $Y(t)$ на интервале $(nh, (n+1)h)$ стохастическими аналогами формулы Тейлора. При этом в разностное уравнение войдут кратные стохастические интегралы по компонентам винеровского процесса $W_0(t)$ и по пуассоновским процессам $P_i(t)$. Нахождение распределения этих интегралов представляет практически непреодолимые трудности. И лишь интегралы любой кратности по одной и той же компоненте винеровского процесса $W_0(t)$ или по одному и тому же пуассоновскому процессу $P_i(t)$ вычисляются очень просто.

Чтобы избежать вычисления производных функций $a(Y_\tau, \tau)$, $b(Y_\tau, \tau)$, $c_i(Y_\tau, \tau)$ при применении двух изложенных способов аппроксимации дифференциального стохастического уравнения разностным, можно рекомендовать заменить их отношениями конечных приращений, например на интервале $(nh, (n+1)h)$ по t и на интервалах $(\bar{Y}_{nk}, \bar{Y}_{nk} + a_k(\bar{Y}_n, nh)h)$ по компонентам вектора y .

5. Полученные разностные уравнения можно использовать как при теоретических исследованиях, так и для численного интегрирования дифференциальных стохастических уравнений. При этом нужно знать распределение всех случайных величин, входящих в разностные уравнения. В данном случае разностные уравнения будут представлять собой так называемую сильную аппроксимацию стохастических дифференциальных уравнений [1, 14–16]. При численном интегрировании такая аппроксимация нужна, когда требуется получать реализации процесса $Y(t)$. Однако часто нет нужды в получении реализаций процесса, а достаточно иметь лишь оценки моментов или математических ожиданий каких-либо функций от значения случайного процесса $Y(t)$ в определенный момент. В таких случаях можно отказаться от использования точных распределений входящих в разностные уравнения случайных величин и заменить их какими-нибудь более простыми распределениями с теми же моментными характеристиками. Например, нормально распределенную скалярную величину с нулевым математическим ожиданием и дисперсией D можно заменить дискретной случайной величиной, принимающей два значения $\pm\sqrt{D}$ с вероятностями 1/2. При замене случайных величин на величины с более простыми распределениями разностное уравнение будет представлять собой слабую аппроксимацию дифференциального стохастического уравнения [1, 14–16].

Литература

1. Пугачев В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. [Англ. пер. Stochastic differential systems. Analysis and filtering. — Chichester, New York: John Wiley, 1987.]
2. Пугачев В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. [Англ. пер. Stochastic systems. Theory and applications. — Singapore: World Scientific, 2001.]
3. Сеницын И. Н. Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009.
4. Сеницын И. Н. Стохастические информационные технологии для исследования нелинейных круговых стохастических систем // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 78–89.
5. Sinityn I. N., Belousov V. V., Konashenkova T. D. Software tools for circular stochastic systems analysis // 29th Seminar (International) on Stability Problems for Stochastic Models: Abstracts. — Svetlogorsk, Russia, 2011. P. 86–87.
6. Сеницын И. Н. Математическое обеспечение для анализа нелинейных многоканальных круговых стохастических систем, основанное на параметризации распределений // Информатика и её применения. — М.: ТОРУС ПРЕСС, 2012. Т. 6. Вып. 1. С. 12–18.
7. Сеницын И. Н., Корепанов Э. Р., Белоусов В. В., Конашенкова Т. Д. Развитие математического обеспечения для анализа нелинейных многоканальных круговых стохастических систем // Системы и средства информатики, 2012. Вып. 22. № 1. С. 29–40.
8. Sinityn I. N., Belousov V. V., Konashenkova T. D. Software tools for spherical stochastic systems analysis and filtering // Прикладные задачи теории вероятности и математической статистики, связанные с моделированием информационных систем (АРТР+MS'2012): Сб. тезисов Международного семинара по проблемам устойчивости стохастических моделей (ISSPSM-2012) и VI Международного рабочего семинара. — М.: ИПИ РАН, 2012. С. 91–93.
9. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р., Белоусов В. В., Сергеев И. В., Базилашвили Д. А. Опыт моделирования эрдитарных стохастических систем // Кибернетика и высокие технологии XXI века: Сб. докл. XIII Международного науч.-технич. конф. — Воронеж: Саквее, 2012. Т. 2. С. 346–357.
10. Сеницын И. Н. Развитие методов аналитического моделирования распределений с инвариантной мерой в стохастических системах // Современные проблемы прикладной математики, информатики, автоматизации

- ции, управления: Мат-лы Междунар. семинара. — Севастополь: СевНТУ, 2012. С. 24–35.
11. Сеницын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с автокоррелированными шумами // Информатика и её применения, 2012. Т. 6. Вып. 4. С. 4–8.
 12. Сеницын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с разрывными характеристиками // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 3–11.
 13. Ватанабэ С., Икэда Н. Стохастические дифференциальные уравнения и диффузионные процессы. — М.: Наука, 1986.
 14. Kloeden P., Platen E. Numerical solution of stochastic differential equations. — Berlin – Heidelberg – New York: Springer, 1992.
 15. Артемьев С. С. Численные методы решения задачи Коши для систем обыкновенных и стохастических дифференциальных уравнений. — Новосибирск: ВЦ СО РАН, 1993.
 16. Кузнецов Д. Ф. Численное интегрирование стохастических дифференциальных уравнений. — СПб.: СПбГУ, 2001.
 17. Sinitsyn I. N., Sinitsyn V. I., Korepanov E. R., Belousov V. V. Symbolic software tools for distributions parametrization in stochastic systems // Международный семинар по проблемам устойчивости стохастических моделей (ISSPSM-2013), 2013. С. 91–93.

МЕТОДЫ ОЦЕНКИ ЭФФЕКТИВНОСТИ И ДИРЕКТИВНЫХ СРОКОВ ВЫПОЛНЕНИЯ РЕСУРСОЕМКИХ ВЫЧИСЛИТЕЛЬНЫХ ЗАДАНИЙ

И. К. Купалов-Ярополк¹, Ю. Е. Малашенко², И. А. Назарова³, А. Ф. Ронжин⁴

Аннотация: Рассматривается проблема эффективного использования гетерогенной вычислительной системы при параллельной обработке разнородных заданий. В случае нарушения сроков завершения работ затраченное процессорное время относится к производственным потерям. Планирование и оптимизация управления осуществляются на основе гарантированных оценок, построенных для наихудшего случая.

Ключевые слова: параллельные вычисления; многопроцессорные системы; оптимизация; принцип гарантированного результата

1 Введение

В настоящей работе предлагается модель для анализа эффективности и оперативного планирования процедуры обработки ресурсоемких вычислительных заданий, которые подробно описаны и определены в [1] как *citu*-задания: *computationally intensive task under uncertainty*. Решение каждой из *citu*-задач состоит в просмотре большого массива исходных данных и выделении из него одного уникального фрагмента с наперед заданными свойствами. В ходе поиска реализуется один и тот же переборный алгоритм для различных начальных данных, разбитых на неделимые, содержательно значимые фрагменты. Если при выполнении задания такой фрагмент найден, то говорят, что задача решена, и обработка задания прекращается. Если же просмотрены все предъявленные данные и найти уникальный фрагмент не удалось, то задание считается выполненным, но задача не имеет решения.

Процесс просмотра происходит в условиях неопределенности, связанной как с длительностью поиска, так и с потенциальной возможностью получить решение. Все *citu*-задания выполняются в режиме реального времени и допускают распараллеливание по данным [2].

В современной практике при выполнении *citu*-работ используются гетерогенные высокопроизводительные специализированные вычислительные системы (СВС). Обычно СВС имеет несколько управляющих узлов, на каждом из которых раз-

ворачивается ядро системы управления, и ряд вычислительных узлов, непосредственно обрабатывающих задания. Специализированные вычислительные системы оснащаются специализированными устройствами, позволяющими значительно повысить скорость исполнения отдельных вычислительных процедур по сравнению со стандартными реализациями. В качестве элементной базы таких ускорителей, согласно [3], могут использоваться заказные сверхбольшие интегральные схемы (СБИС) (ASIC — *application-specific integrated circuits*), базовые матричные кристаллы (БМК) (eASIC), системы-на-кристалле (SoC — *systems on chip*), программируемые логические интегральные схемы (ПЛИС) (FPGA — *field-programmable gate array*), графические ускорители GPGPU (*general-purpose graphics processing units*).

Различные типы устройств выполняют одно и то же задание с разной производительностью. Кроме того, некоторые ускорители могут предназначаться для работы только с определенными типами алгоритмов и подходят для решения ограниченного класса задач. Для специализированных устройств в качестве ресурсной единицы (единичного вычислительного модуля) в СВС рассматривается отдельно каждая микросхема (ПЛИС, кристалл) как наименьший возможный объект, имеющий собственное множество состояний и допускающий независимое управление.

При планировании вычислительных работ в СВС возникает научно-техническая проблема, со-

¹Институт точной механики и вычислительной техники им. С. А. Лебедева Российской академии наук, kupyar@rambler.ru

²Вычислительный центр им. А. А. Дородницына Российской академии наук, malash09@ccas.ru

³Вычислительный центр им. А. А. Дородницына Российской академии наук, irina-nazar@yandex.ru

⁴Вычислительный центр им. А. А. Дородницына Российской академии наук, raf-zao-zt@yandex.ru

стоящая в отыскании способов эффективного распределения разнотипных ресурсов между разнородными работами, выполняющимися в заданных временных рамках.

В научной литературе понятие эффективности трактуется весьма широко. В данной работе под эффективностью будем понимать отношение реально произведенных в СВС специализированных вычислительных операций к максимально возможному их числу, взятое на определенном интервале времени.

Рассмотрим специализированную систему, состоящую из большого числа вычислительных модулей различных типов. Пусть ее диспетчеру взаимосвязанные пользователи в случайные моменты времени предоставляют для решения разнородные *situ*-задачи. Все задания вместе и каждое в отдельности должны быть завершены как можно быстрее, но не позднее заранее определенных сроков. В противном случае полученное решение может потерять актуальность и не будет представлять интереса, а вычислительные затраты будут отнесены к производственным потерям (издержкам).

Массив исходных данных для каждой из *situ*-задач состоит из одинаковых по размеру отдельных неделимых содержательно значимых фрагментов и становится известен после предварительного анализа задачи в системе. Выполнение отдельной *situ*-работы прекращается, когда обнаружен фрагмент, удовлетворяющий наперед заданному критерию, — уникальный фрагмент, или если просмотрен весь массив, но ничего найти не удалось. Считается, что архитектура вычислительной системы позволяет просматривать фрагменты каждого *situ*-задания в произвольном порядке и каждая часть данных может обрабатываться независимо, в том числе одновременно с другими.

Администратор или оператор-планировщик в режиме реального времени должен:

- (1) эффективно использовать разнотипные вычислительные модули при обработке разнородных вычислительных заданий;
- (2) завершать каждое конкретное задание до наступления назначенного срока и тем самым минимизировать потери.

При этом обработка происходит в условиях неопределенности, связанной со случайным характером формирования текущего набора *situ*-заданий, длительностью процесса обработки каждого из них и с потенциальной возможностью получить решение.

В литературе при проектировании и анализе современных систем реального времени в основном рассматриваются «жестко» заданные директивные сроки окончания, нарушение которых может иметь

фатальные последствия. В ходе изучения таких систем вначале делается предположение, что директивный срок окончания может быть превышен, а затем предлагаются меры, позволяющие этого избежать [4]. Настоящая работа посвящена изучению функционирования гетерогенной вычислительной системы с более мягкими временными ограничениями. В рассматриваемой модели считается, что при поступлении *situ*-задачи в систему определяется возможное время ее завершения с учетом объективных показателей загруженности последней и имеющихся заданий. Если предварительная оценка — срок, к которому *situ*-задание может быть завершено, — устраивает пользователя, то оно принимается для обработки. При этом заданию назначается директивный срок окончания (ДСО), совпадающий с прогнозируемым. Таким образом, ДСО — это установленный диспетчером и согласованный с пользователем момент календарного времени, до наступления которого *situ*-задача должна быть решена. Дальнейшее планирование осуществляется исходя из предположения, что события будут развиваться по наихудшему сценарию. В условиях объективной неопределенности ДСО используется как ограничение при формировании диспетчерских правил (политики) совместного выполнения всех заданий, находящихся в СВС.

Заметим, что понятие ДСО, введенное выше, несколько отличается от классического *deadline* [5], поскольку устанавливается диспетчером, а не пользователем системы, хотя и по договоренности с последним. Однако принятое допущение больше соответствует реальной ситуации. Действительно, в этом случае *situ*-работа, которая не может быть выполнена по объективным причинам, будет сразу отозвана, что позволит решить другие задачи в срок. У администратора появляется возможность более гибко подходить к формированию пакета текущих заданий и минимизировать возможные производственные потери.

Данная работа продолжает исследования, начатые в [6–8], однако отличается от них методами изучения и конечной целью. В [7, 8] с помощью имитационного моделирования анализировались временные показатели длительности обработки заданий. В настоящей статье показано, как, опираясь на методы оптимизации [9] и принцип гарантированного результата [10], можно оценить и минимизировать максимальную величину возможного превышения установленных сроков завершения для любого набора исходных *situ*-заданий и повысить эффективность использования СВС.

Перейдем к описанию математической модели (М-модели), которая может быть использована

при оперативном планировании работ и реализации различных диспетчерских политик.

2 Описание модели

При описании М-модели верхний индекс любой переменной будет всегда относиться к типу вычислительного модуля, а нижний (один или два) в зависимости от рассматриваемого случая — отвечать номеру и виду задачи соответственно. Для того чтобы избежать путаницы, второй нижний идентификатор, соответствующий виду задачи, помещен в скобки. Кроме того, в зависимости от контекста через t обозначим либо текущий календарный момент времени, либо контрольную точку принятия решения. Таким образом, t не является переменной в традиционном смысле, это обозначение определенных моментов времени в модельном описании процесса. Далее будем говорить об абстрактной гетерогенной высокопроизводительной СВС, в которой в условиях неопределенности выполняются разнородные *situ*-работы.

В рамках М-модели считается, что указанная СВС состоит из центрального управляющего устройства (ЦУП-устройства) и набора независимых исполняющих единичных вычислительных модулей (ЕВ-модулей) различных конструктивных типов. Введем обозначения:

Н — ЦУП-устройство, в котором происходит анализ поступающих заданий, определяются ДСО и порядок их выполнения, а также осуществляется контроль процесса обработки;

Е — набор (множество) ЕВ-модулей нескольких типов.

Пусть M — общее число различных типов ЕВ-модулей, из которых состоит рассматриваемая СВС; e^m — отдельный ЕВ-модуль (устройство) m -го типа; E^m — множество ЕВ-модулей m -го типа e^m , $m = \overline{1, M}$, следовательно, множество **Е** всех ЕВ-модулей в СВС можно записать в виде объединения

$$\mathbf{E} = E^1 \cup E^2 \cup \dots \cup E^M.$$

Под специализированной элементарной вычислительной операцией (СЭВ-операция) будем понимать просмотр отдельного неделимого содержательно значимого фрагмента данных определенного вида и проверку его уникальности. Считается, что любой ЕВ-модуль может обрабатывать задания по крайней мере одного типа. Производительность конкретного ЕВ-модуля при выполнении *situ*-заданий разных видов не одинакова и может меняться (перегрев оборудования, изменение тактовой частоты или работа с другим программным обеспечением).

Кроме того, будем считать, что ЕВ-модули с течением времени могут выходить из строя.

Пусть $p_k^m(t)$ — производительность ЕВ-модуля m -го типа при обработке задания k -го вида, т.е. модуль e^m может выполнять в единицу времени $p_k^m(t)$ СЭВ-операций k -го вида, начиная с момента t ; $R^m(t)$ — общее число работоспособных ЕВ-модулей m -го типа на момент времени t .

Будем считать, что СВС может одновременно обрабатывать как одно, так и несколько заданий на всех действующих в данный момент ЕВ-модулях. Отдельное *situ*-задание может быть разделено на подзадания, каждое из которых, в свою очередь, выполняется самостоятельно или в составе некоторого набора (пакета). Для решения каждой *situ*-задачи достаточно найти хотя бы один уникальный фрагмент.

Для формального описания заданий будут использоваться следующие обозначения:

z_n — задание (задача) с собственным идентификационным номером n ;

K — общее число видов заданий, которые могут обрабатываться в данной СВС;

$z_{n(k)}$ — задание (задача) с собственным идентификационным номером n , для которой явно указан ее вид k , $k = \overline{1, K}$. Данное обозначение введено для удобства описания процесса обработки разнородных заданий;

t_n^0 — календарный момент поступления задания z_n в СВС; $T_n(t) = t - t_n^0$ — длительность промежутка времени, в течение которого задание z_n находится в СВС при условии, что на момент t оно еще не выполнено до конца, т.е. $T_n(t)$ — число единиц времени, которое прошло от поступления z_n в СВС до текущего момента t ; t_n^+ — момент завершения задания z_n и/или его удаления из СВС; $T_n^+ = t_n^+ - t_n^0$ — длительность промежутка времени, в течение которого задание z_n обрабатывалось СВС, т.е. T_n^+ — длительность пребывания z_n в системе.

Для множества $\mathcal{Z}(t)$ заданий z_n , находящихся в СВС в момент t , введем обозначения:

$\mathcal{N}(t)$ — множество номеров заданий z_n из $\mathcal{Z}(t)$;

$N(t)$ — общее число заданий z_n из $\mathcal{Z}(t)$. Таким образом $|\mathcal{Z}(t)| = |\mathcal{N}(t)| = N(t)$;

$\mathcal{Z}_k(t)$ — множество заданий k -го вида;

$\mathcal{N}_k(t)$ — множество номеров (индексов) заданий k -го вида;

$N_k(t)$ — общее число заданий k -го вида.

Тогда

$$\mathcal{Z}(t) = \bigcup_{k=1}^K \mathcal{Z}_k(t); \quad \mathcal{N}(t) = \bigcup_{k=1}^K \mathcal{N}_k(t);$$

$$N(t) = \sum_{k=1}^K N_k(t).$$

В М-модели предполагается, что для каждого задания z_n в момент поступления t_n^0 становится известна нормативная величина \mathbf{Z}_n — общее число СЭВ-операций, которые будет необходимо выполнить для данного задания в случае, когда задача z_n не имеет решения, т. е. в исходной содержательной постановке \mathbf{Z}_n — общее число неделимых фрагментов данных задания z_n , которые необходимо будет обработать, если среди них нет уникального.

В СВС диспетчеризация процесса выполнения situ-заданий осуществляется с помощью программного комплекса планирования и анализа (ПК-план). Для каждой поступившей situ-задачи в ПК-плане определяются необходимые вычислительные затраты и время решения в наихудшем случае с учетом общей загруженности СВС. После этого по согласованию с пользователем назначается ДСО, т. е. календарный момент времени, до наступления которого z_n должно быть завершено. Обозначим через d_n ДСО для z_n .

В СВС вся исходная и текущая информация о situ-заданиях, вновь поступивших или находящихся в обработке, заносится и далее хранится в так называемой базе данных заданий (БД-заданий). Программный комплекс планирования и анализа постоянно просматривает и анализирует БД-заданий. Динамика обработки (просмотров) массивов данных в ПК-плане описывается системой неравенств и конечно-разностных уравнений в дискретном времени. Шаг по времени называется плановым периодом или операционным окном.

В контрольной точке принятия решения t по команде с ЦУП-устройства ПК-план анализирует данные о выполнении всех заданий, находящихся в СВС. Заметим, что здесь t — начальный момент очередного планового периода. Далее диспетчер (специалист и/или алгоритмическая процедура) выбирает число единиц календарного времени $\Delta(t)$ — длительность планового периода. В ПК-плане формируется пакет текущих работ (ТР-пакет), состоящий из подмассивов (поднаборов) неделимых фрагментов данных подзаданий z_n , $n \in \mathcal{N}(t)$, которые будут обрабатываться в СВС в течение $\Delta(t)$. Центральное управляющее устройство помещает ТР-пакет в специальный буфер текущих работ (ТР-буфер) и начиная с момента t СВС выполняет его на всех работоспособных ЕВ-модулях.

3 Параметры планирования и управления

Обозначим через $z_{n(k)}^-(t)$ число СЭВ-операций k -го вида, которые были завершены для задания z_n от его поступления в СВС до момента времени t , или в рамках М-модели — число неделимых фрагментов, которые уже были просмотрены для задания z_n к моменту времени t ; соответственно $z_{n(k)}^+(t)$ — число фрагментов, которые для z_n остались необработанными. Тогда

$$z_{n(k)}^+(t) = \mathbf{Z}_n - z_{n(k)}^-(t), \quad n \in \mathcal{N}(t).$$

Как уже указывалось ранее, подзадания ТР-пакета представляют собой различающиеся по объему поднаборы фрагментов данных для заданий z_n , $n \in \mathcal{N}(t)$. Размер подзадания для $z_{n(k)}$ обозначим через $w_{n(k)}(\Delta(t))$. Таким образом, $w_{n(k)}(\Delta(t))$ — число неделимых фрагментов в поднаборе, взятом из массива начальных данных задания $z_{n(k)}$ в момент t для просмотра в течение $\Delta(t)$, или число СЭВ-операций k -го вида, которые планируется выполнить в операционном окне $\Delta(t)$ для $z_{n(k)}$. Размеры отобранных поднаборов данных (подзаданий) $w_{n(k)}(\Delta(t))$ являются управлениями, или управляющими параметрами, которые определяются в ПК-плане с учетом текущей диспетчерской политики соблюдения ДСО.

Поскольку в момент t для z_n нельзя взять в обработку больше данных, чем осталось, управления $w_{n(k)}(\Delta(t))$ на интервале $\Delta(t)$ должны удовлетворять ограничениям

$$0 \leq w_n(\Delta(t)) \leq z_n^+(t), \quad n \in \mathcal{N}(t). \quad (1)$$

Обозначим через $W_k(\Delta(t))$ суммарное число фрагментов k -го вида, которые планируется включить в ТР-пакет. Тогда $W(\Delta(t))$ — общий объем ТР-пакета для просмотра в операционном окне $\Delta(t)$:

$$W(\Delta(t)) = \sum_{k=1}^K W_k(\Delta(t)) =$$

$$= \sum_{k=1}^K \sum_{n \in \mathcal{N}_k(t)} w_{n(k)}(\Delta(t)) = \sum_{n \in \mathcal{N}(t)} w_n(\Delta(t)). \quad (2)$$

Введем переменную $r_k^m(t)$ — общее число ЕВ-модулей m -го типа, которые начинают обрабатывать подзадания k -го вида в момент времени t . $r_k^m(t)$ также являются управлениями и определяются исходя из текущих установок диспетчерских правил.

Предположим, что ни один из поднаборов $w_{n(k)}(\Delta(t))$ не содержит уникального фрагмента, но все поднаборы планируется выполнить за период $\Delta(t)$. Тогда в момент t для просмотра всех данных для подзаданий k -го вида необходимо выдлить $r_k^m(t)$ ЕВ-модулей m -го типа,

$$W_k(\Delta(t)) = \sum_{m=1}^M p_k^m(t) r_k^m(t) \Delta(t), \quad k = \overline{1, K}, \quad (3)$$

которые за период $\Delta(t)$ должны проделать $W_k(\Delta(t))$ СЭВ-операций k -го вида. Поскольку число ЕВ-модулей, распределяемых для выполнения заданий, не может превышать общего числа работоспособных, то значения $r_k^m(t)$ должны удовлетворять следующим ограничениям:

$$\left. \begin{aligned} \sum_{k=1}^K r_k^m(t) &\leq R^m(t), \quad m = \overline{1, M}; \\ r_k^m(t) &\geq 0, \quad k = \overline{1, K}, \quad m = \overline{1, M}. \end{aligned} \right\} \quad (4)$$

4 Интегральная производительность специализированной вычислительной системы

При диспетчеризации СВС, где на разнотипном оборудовании с различной скоростью выполняются разнородные СЭВ-операции, возникает проблема эффективного использования имеющихся ресурсов.

Обозначим через $\rho(t)$ интегральную производительность СВС. Под $\rho(t)$ будем понимать суммарную производительность СВС на интервале $\Delta(t)$, т. е. общее число СЭВ-операций, которые выполняются на всех действующих ЕВ-модулях, начиная с момента t на отрезке планирования $\Delta(t)$:

$$\rho(t) = \sum_{m=1}^M \sum_{k=1}^K p_k^m(t) r_k^m(t) \Delta(t).$$

В качестве максимальной интегральной производительности СВС $P^*(t)$ на интервале $\Delta(t)$ выберем скалярное решение следующей задачи линейного программирования: найти $P^*(t) = \max_{\rho, r} \rho(t)$ при ограничениях на производительность СВС

$$\left. \begin{aligned} \rho(t) &= \sum_{m=1}^M \sum_{k=1}^K p_k^m(t) r_k^m(t) \Delta(t); \\ \sum_{k=1}^K r_k^m(t) &= R^m(t), \quad m = \overline{1, M}; \\ r_k^m(t) &\geq 0, \quad k = \overline{1, K}, \quad m = \overline{1, M}. \end{aligned} \right\} \quad (5)$$

Значение $P^*(t)$ показывает, какое максимальное число СЭВ-операций суммарно для всех видов заданий может быть выполнено СВС в операционное окно $\Delta(t)$ начиная с момента t . Пусть на интервале $\Delta(t)$ предполагается произвести $W(\Delta(t))$ СЭВ-операций всех видов (см. (2)). Тогда отношение

$$\gamma(t) = \frac{W(\Delta(t))}{P^*(t)}$$

назовем плановой эффективностью использования СВС на интервале $\Delta(t)$ для данного набора заданий $\mathcal{Z}(t)$. Максимальное значение $W(\Delta(t))$ и, соответственно, $\gamma(t)$ существенно зависит от состава и объема заданий, находящихся на обработке в момент времени t (см. (1) и (2)). Сравнивая множества ограничений (1)–(4) и (5), можно утверждать, что для любого набора заданий $\mathcal{Z}(t)$ выполняются неравенства

$$P^*(t) \geq W(\Delta(t)), \quad 0 \leq \gamma(t) \leq 1.$$

Кроме того, при диспетчеризации следует учитывать другие ограничения и требования пользователей, которым должны подчиняться управления $r_k^m(t), w_{n(k)}(\Delta(t))$.

5 Определение директивного срока окончания вновь поступившего задания

Для управления гетерогенными СВС в реальном времени необходимы и разрабатываются интерактивные сценарии диалога диспетчер–пользователь в декларативной, частично вербальной, частично технологической, постановке: задание–требования–возможности–сроки. В рамках М-модели в ПК-плане для каждого вновь поступившего задания вычисляется гарантированная оценка времени завершения. Фактически решается задача быстрого выполнения всех заданий в наилучшем случае.

Пусть в момент t в СВС, которая обрабатывает пакет заданий $\mathcal{Z}(t)$, поступает новая задача \hat{z} , например, \hat{k} -го вида. В ПК-плане этой задаче присваивается порядковый номер $\hat{n}_{(\hat{k})}$. Кроме того, становится известна величина $Z_{\hat{n}_{(\hat{k})}}$ — общее число фрагментов данных, которые необходимо будет просмотреть при условии, что среди них нет уникального.

Для каждого $z_{n(k)}, n \in \mathcal{N}_k(t), k = \overline{1, K}$, и каждого m определим переменную $\nabla_{n(k)}^m(t)$ — число

единиц времени (единичных временных интервалов), в течение которых задание $z_{n(k)}$ планируется выполнять на всех работоспособных ЕВ-модулях m -го типа, $m = \overline{1, M}$, после момента t . При определении ДСО для вновь поступивших заданий рассматривается наихудший случай: ни одна из вновь полученных и ни одна из находящихся в обработке задач не имеет решения. Другими словами, для окончания всех имеющихся заданий необходимо будет просмотреть все имеющиеся фрагменты данных. Следовательно, для заданий, обработка которых уже начата, будет необходимо просмотреть

$$\mathbf{Z}_{n(k)} - z_{n(k)}^-(t) = \sum_{m=1}^M p_k^m(t) R^m(t) \nabla_{n(k)}^m(t),$$

$$n \in \mathcal{N}_k(t), k = \overline{1, K},$$

фрагментов, а для вновь поступившего задания —

$$\mathbf{Z}_{\hat{n}(\hat{k})} = \sum_{m=1}^M p_{\hat{k}}^m(t) R^m(t) \nabla_{\hat{n}(\hat{k})}^m(t),$$

$$\nabla_{n(k)}^m(t) \geq 0, \nabla_{\hat{n}(\hat{k})}^m(t) \geq 0, m = \overline{1, M}, n \in \mathcal{N}(t).$$

Введем переменную $\nabla^m(t)$:

$$\nabla^m(t) = \sum_{k=1}^K \sum_{n \in \mathcal{N}_k(t)} \nabla_{n(k)}^m(t) + \nabla_{\hat{n}(\hat{k})}^m(t) =$$

$$= \sum_{n \in \mathcal{N}(t)} \nabla_{n(k)}^m(t) + \nabla_{\hat{n}(\hat{k})}^m(t), m = \overline{1, M}.$$

Значение $\nabla^m(t)$ показывает, сколько единиц времени после момента t потребуется ЕВ-модулям m -го типа для завершения всех задач, имеющихся в СВС, при условии, что ни одна из них не имеет решения и необходимо просмотреть все данные всех заданий из $\mathcal{Z}(t)$.

В ПК-плане в качестве оценки ДСО для вновь поступившего задания $z_{\hat{n}(\hat{k})}$ используется решение задачи *быстродействия*:

в момент t для заданных $\mathbf{Z}_n, \mathbf{Z}_{\hat{n}(\hat{k})}, z_{n(k)}^-(t), \mathcal{N}(t), R^m(t), d_j$ найти

$$\delta^* = \min_{\delta, \nabla} \delta$$

при условиях:

- (1) ДСО для вновь поступившего задания будет не меньше, чем у находящихся в обработке:

$$\delta \geq d_j, j \in \mathcal{N}(t);$$

- (2) ДСО не меньше гарантированного времени завершения всех заданий на всех ЕВ-модулях:

$$\delta \geq t + \sum_{n \in \mathcal{N}(t)} \nabla_{n(k)}^m(t) + \nabla_{\hat{n}(\hat{k})}^m(t), m = \overline{1, M};$$

- (3) для выполнения всех заданий необходимо будет просмотреть все имеющиеся в момент t данные

$$\mathbf{Z}_{n(k)} - z_{n(k)}^-(t) = \sum_{m=1}^M p_k^m(t) R^m(t) \nabla_{n(k)}^m(t),$$

$$\nabla_{n(k)}^m(t) \geq 0, m = \overline{1, M}, n \in \mathcal{N}_k(t); k = \overline{1, K};$$

$$\mathbf{Z}_{\hat{n}(\hat{k})} = \sum_{m=1}^M p_{\hat{k}}^m(t) R^m(t) \nabla_{\hat{n}(\hat{k})}^m(t),$$

$$\nabla_{\hat{n}(\hat{k})}^m(t) \geq 0, m = \overline{1, M}.$$

Полученное δ^* является гарантированной оценкой ДСО для вновь поступившей задачи. Администратор сообщает значение δ^* пользователю, который предоставил данную задачу. Если пользователя не устраивают прогнозируемые сроки завершения задания, то он отказывается от обработки и снимает задачу со счета. В противном случае задание $z_{\hat{n}(\hat{k})}$ помещается в общий пакет $\mathcal{Z}(t)$ с условием, что его обработка должна быть завершена до момента $d_{\hat{n}(\hat{k})} = \delta^*$.

6 Гарантированные оценки потерь процессорного времени

Рассмотрим проблему диспетчеризации *ситу*-заданий с фиксированными ДСО. В рамках М-модели предполагается, что в случае превышения ДСО произведенные вычислительные затраты записываются в производственные потери, которые администратору предписывается свести к минимуму. На первый взгляд описанный выше способ назначения ДСО для всех заданий, принятых на обработку, гарантирует их окончание в срок. Однако в реальности в СВС могут происходить сбои, изменение приоритетов в обслуживании заявок и др., что неизбежно приведет к нарушению ДСО. В М-модели в момент t анализируются достаточные условия завершения заданий в срок и вычисляются гарантированные оценки возможных превышений ДСО в наихудшем случае.

В соответствии с принятыми ранее обозначениями d_j — ДСО задачи z_j . Каждому $d_j, j \in \mathcal{N}(t)$, ставится в соответствие множество (список) номеров $\mathcal{N}(t, d_j)$ всех заданий z_i из $\mathcal{Z}(t)$, таких что каждое z_i требуется завершить не позднее d_j , т. е.

$$\mathcal{N}(t, d_j) = \{i \mid d_i \leq d_j, i \in \mathcal{N}(t)\}, j \in \mathcal{N}(t).$$

Для получения гарантированных оценок рассмотрим наихудший случай: пусть ни одна из задач z_n ,

$n \in \mathcal{N}(t)$, не имеет решения. Для фиксированного момента времени $t + \Delta(t)$ введем дополнительный параметр управления $\nabla_{n(k)}^m(t + \Delta(t))$ — число единиц времени (единичных временных интервалов), в течение которых задание z_n может обрабатываться на всех работоспособных ЕВ-модулях m -го типа, $m = \overline{1, M}$, начиная с $t + \Delta(t)$.

В рассматриваемом наихудшем случае задача z_n не имеет решения, поэтому для завершения задания необходимо будет перебрать все данные

$$z_n^+(t + \Delta(t)) = Z_{n(k)} - z_{n(k)}^-(t) - w_{n(k)}(\Delta(t)), \quad (6)$$

оставшиеся необработанными для z_n к моменту $t + \Delta(t)$. Здесь управление $w_{n(k)}(\Delta(t))$ — число фрагментов данных для z_n , которые в момент t планируется просмотреть за время $\Delta(t)$.

Таким образом, для оставшихся фрагментов в момент $t + \Delta(t)$ должно выполняться равенство:

$$z_n^+(t + \Delta(t)) = \sum_{m=1}^M p_k^m(t) R^m(t) \nabla_{n(k)}^m(t + \Delta(t)), \quad n \in \mathcal{N}_k(t), k = \overline{1, K}, \quad (7)$$

и условие неотрицательности:

$$\nabla_{n(k)}^m(t + \Delta(t)) \geq 0, \quad m = \overline{1, M}, \quad n \in \mathcal{N}(t).$$

В момент $t + \Delta(t)$ рассмотрим дополнительную переменную

$$\begin{aligned} \nabla^m(t + \Delta(t), d_j) &= \\ &= \sum_{k=1}^K \sum_{n \in \mathcal{N}(t, d_j) \cap \mathcal{N}_k(t)} \nabla_{n(k)}^m(t + \Delta(t)) = \\ &= \sum_{n \in \mathcal{N}(t, d_j)} \nabla_{n(k)}^m(t + \Delta(t)), \quad m = \overline{1, M}, \quad j \in \mathcal{N}(t). \end{aligned} \quad (8)$$

Значение $\nabla^m(t + \Delta(t), d_j)$ показывает, сколько единиц времени после момента $t + \Delta(t)$ потребуется ЕВ-модулям m -го типа для выполнения всех заданий z_n , $n \in \mathcal{N}(t, d_j)$, которые должны быть завершены до наступления d_j . В случае, когда ни одна из задач не имеет решения, величина $\nabla^m(t + \Delta(t), d_j)$ — точная верхняя оценка временных затрат на обработку всех z_n , $n \in \mathcal{N}(t, d_j)$, с помощью работоспособных ЕВ-модулей m -го типа, $m = \overline{1, M}$. В момент t достаточные условия соблюдения всех ДСО в наихудшем случае можно записать в виде:

$$t + \Delta(t) + \nabla^m(t + \Delta(t), d_j) \leq d_j, \quad j \in \mathcal{N}(t), \quad m = \overline{1, M}.$$

Чтобы получить гарантированные оценки относительных величин возможного превышения ДСО для z_n , введем переменные:

$$\begin{aligned} \omega^m(t + \Delta(t), d_j) &= \\ &= \frac{t + \Delta(t) + \nabla^m(t + \Delta(t), d_j) - t_j^0}{d_j - t_j^0}, \\ & \quad j \in \mathcal{N}(t), \quad m = \overline{1, M}, \end{aligned} \quad (9)$$

где t_j^0 — время поступления z_j в СВС. Конкретное значение $\omega^m(t + \Delta(t), d_j)$ характеризует возможность соблюдения ДСО следующим образом:

- (1) если существуют управления $w_n(\Delta(t))$, $\nabla_{n(k)}^m(t + \Delta(t))$ и значение $\nabla^m(t + \Delta(t), d_j)$, удовлетворяющие (6)–(9), при которых $\omega^m(t + \Delta(t), d_j) \leq 1$ для всех $j \in \mathcal{N}(t)$, $m = \overline{1, M}$, то все задания можно гарантированно завершить до наступления их ДСО даже в наихудшем случае;
- (2) если же при любых управлениях $w_n(\Delta(t))$, $\nabla_{n(k)}^m(t + \Delta(t))$ и значениях $\nabla^m(t + \Delta(t), d_j)$ в (6)–(9) хотя бы для одного сочетания m и j величина $\omega^m(t + \Delta(t), d_j) > 1$, то может реализоваться ситуация, при которой хотя бы одно задание в наихудшем случае завершится после назначенного (предписанного) срока.

На самом деле максимальная величина $\omega^m(t + \Delta(t), d_j)$ является своего рода индикатором, который указывает, располагает ли СВС на момент t необходимой мощностью для того, чтобы обработать все имеющиеся разнородные задания до назначенных ДСО в наихудшем случае. Для проверки достаточных условий выполнения ДСО в момент t и получения гарантированных оценок максимально возможного их превышения в рамках М-модели введем параметр управления $\omega(t)$:

$$\omega(t) \leq 1 - \omega^m(t + \Delta(t), d_j), \quad m = \overline{1, M}, \quad j \in \mathcal{N}(t).$$

Величина $\omega(t)$ характеризует максимальное относительное значение возможного превышения ДСО.

7 Формирование пакета текущих работ

Рассмотрим возможный способ формирования ТР-пакета, при котором в процессе выполнения ситу-заданий требуется добиться высокой эффективности использования СВС и минимизировать производственные потери, в данном случае за счет соблюдения ДСО. В рамках М-модели для анализа эффективности и получения гарантированных оценок превышения ДСО в момент времени t решается следующая задача оптимизации:

для заданных $\mathbf{Z}_n, z_{n(k)}^-(t), N(t), R^m(t), d_j, \mathcal{N}(t), P^*(\Delta(t))$ найти

$$\Phi^* = \max_{\gamma, \omega, \nabla, w, r, W} (c_1 \gamma(t) + c_2 \omega(t) + c_3 \omega^\Sigma(t)) \quad (10)$$

при ограничениях:

(1) на эффективность использования разнотипных ЕВ-модулей в СВС

$$0 \leq \gamma(t) P^*(\Delta(t)) = \sum_{k=1}^K \sum_{m=1}^M p_k^m(t) r_k^m(t) \Delta(t); \quad (11)$$

(2) на суммарную величину отклонения от заданных ДСО

$$\left. \begin{aligned} \omega^\Sigma(t) &\leq \sum_{m=1}^M \sum_{j \in \mathcal{N}(t)} \frac{1 - \omega^m(t + \Delta(t), d_j)}{d_j}; \\ \omega^\Sigma(t) &\leq 0; \end{aligned} \right\} \quad (12)$$

(3) на максимальную величину отклонения от заданных ДСО

$$\left. \begin{aligned} \omega(t) &\leq 1 - \omega^m(t + \Delta(t), d_j), \\ &j \in \mathcal{N}(t), m = \overline{1, M}; \\ \omega(t) &\leq 0; \end{aligned} \right\} \quad (13)$$

(4) на балансовые соотношения выполнения заданий после $t + \Delta(t)$

$$\left. \begin{aligned} \mathbf{Z}_n - z_n^-(t) - w_n(\Delta(t)) &= \\ &= \sum_{m=1}^M p_k^m(t) R^m(t) \nabla_{n(k)}^m(t + \Delta(t)), \\ &n \in \mathcal{N}_k(t), k = \overline{1, K}; \\ \nabla_{n(k)}^m(t + \Delta(t)) &\geq 0, n \in \mathcal{N}(t), m = \overline{1, M}; \end{aligned} \right\} \quad (14)$$

$$\begin{aligned} \nabla^m(t + \Delta(t), d_j) &= \sum_{n \in \mathcal{N}(t, d_j)} \nabla_{n(k)}^m(t + \Delta(t)), \\ &m = \overline{1, M}, j \in \mathcal{N}(t); \end{aligned} \quad (15)$$

$$\begin{aligned} \omega^m(t + \Delta(t), d_j) &= \\ &= \frac{t + \Delta(t) + \nabla^m(t + \Delta(t), d_j) - t_j^0}{d_j - t_j^0}, \\ &j \in \mathcal{N}(t), m = \overline{1, M}; \end{aligned} \quad (16)$$

(5) на управления в операционном окне $\Delta(t)$:

(а) суммарное число СЭВ-операций k -го вида, планируемых для ТР-пакета в момент t :

$$\left. \begin{aligned} \sum_{n \in \mathcal{N}_k(t)} w_n(\Delta(t)) &= W_k(\Delta(t)), k = \overline{1, K}; \\ 0 \leq w_n(\Delta(t)) &\leq \mathbf{Z}_n - z_n^-(t), n \in \mathcal{N}(t); \end{aligned} \right\} \quad (17)$$

(б) на число ЕВ-модулей, выделяемых для обработки фрагментов для подзаданий k -го вида в момент t :

$$\left. \begin{aligned} W_k(\Delta(t)) &\leq \sum_{m=1}^M p_k^m(t) r_k^m(t) \Delta(t), k = \overline{1, K}; \\ r_k^m(t) &\geq 0, k = \overline{1, K}, m = \overline{1, M}; \\ \sum_{k=1}^K r_k^m(t) &\leq R^m(t), m = \overline{1, M}. \end{aligned} \right\} \quad (18)$$

Задача (10) может быть решена стандартными методами линейного программирования [11]. Поскольку переменные $\gamma(t)$ и $\omega(t)$ являются относительными, безразмерными и изменяются в ограниченном диапазоне, то с помощью выбора значений c_1, c_2, c_3 можно учитывать текущие приоритеты при выполнении заданий. Например, если в момент времени t во главу угла ставится эффективность и при этом необходимо обработать как можно больше заданий, следует выбрать $c_1 \gg c_2 > c_3$. Если в момент t в множестве $\mathcal{Z}(t)$ много заданий, которые необходимо завершить до наступления их ДСО, то следует выбрать $c_1 \ll c_2 \simeq c_3$ и вновь решить (10)–(18).

Заключение

Решение задачи (10)–(18) в условиях неопределенности помогает оперативно планировать выполнение разнородных ресурсоемких заданий в гетерогенной вычислительной системе. Очевидно, что при управлении СВС можно рассмотреть и другие диспетчерские политики. Однако описанная выше расчетная схема концептуально проста и позволяет оценивать производственные потери и добиваться эффективного использования СВС. Полученные на практике результаты показали, что предлагаемый подход является хорошей основой для диалога администратора и пользователей.

Литература

1. Козлов М. В., Малащенко Ю. Е., Назарова И. А. и др. Анализ режимов управления вычислительным комплексом в условиях неопределенности. — М.: ВЦ РАН, 2011.
2. Sourcebook of parallel computing. — San Francisco: Morgan Kaufmann Pubs., 2003.

3. *Каляев И. А., Левин И. И.* Реконфигурируемые мультитоннельные вычислительные структуры для решения потоковых задач обработки информации и управления // Суперкомпьютерные технологии: разработка, программирование, применение: Тр. Междунар. науч.-практич. конф. — Ростов-на-Дону: ЮФУ, 2010. Т. 1. С. 100–102.
4. *Sha L., Abdelzaher T., Arzen K.-E., et al.* Real time scheduling theory: A historical perspective // *Real-Time Systems*, 2004. Vol. 28. No. 2–3. P. 101–155.
5. *Stankovic J. A., Spuri M., Ramamritham K., et al.* Deadline scheduling for real-time systems, EDF and related algorithms. — Boston: Kluwer, 1998.
6. *Коновалов М. Г., Малашенко Ю. Е., Назарова И. А.* Управление заданиями в гетерогенных вычислительных системах // Изв. РАН. ТиСУ, 2011. № 2. С. 72–90.
7. *Голосов П. Е., Козлов М. В., Малашенко Ю. Е. и др.* Анализ управления специализированными вычислительными заданиями в условиях неопределенности // Изв. РАН. ТиСУ, 2012. № 1. С. 50–66.
8. *Малашенко Ю. Е., Назарова И. А.* Модель управления разнородными вычислительными заданиями на основе гарантированных оценок времени выполнения // Изв. РАН. ТиСУ, 2012. № 4. С. 29–38.
9. *Сухарев А. Г., Тимохов А. В., Федоров В. В.* Курс методов оптимизации. — М.: Наука, 1986.
10. *Гермейер Ю. Б.* Введение в теорию исследования операций. — М.: Наука, 1971.
11. *Данциг Дж. Б.* Линейное программирование, его применения и обобщения. — М.: Прогресс, 1966.

ОБ ОЦЕНИВАНИИ ЭФФЕКТИВНОЙ ПРОПУСКНОЙ СПОСОБНОСТИ СИСТЕМЫ С РЕГЕНЕРАТИВНЫМ ВХОДНЫМ ПРОЦЕССОМ*

А. В. Бородина¹, Е. В. Морозов²

Аннотация: Рассматривается понятие эффективной пропускной способности (ЭПС) коммуникационного узла, которая гарантирует, что вероятность потери или превышения стационарной нагрузкой некоторого уровня ограничена заданной (малой) величиной. Показано, как вычисляется ЭПС в жидкостной системе обслуживания в случае входного процесса с независимыми приращениями. Далее рассматривается жидкостная система с регенеративным входным процессом. Для вычисления ЭПС ключевым является нахождение предельной логарифмической экспоненциальной функции моментов входного процесса. С использованием эвристических соображений получена аппроксимация этого предела, которая выражена в терминах моментов длины цикла регенерации и величины работы, поступающей в систему в течение цикла. Результаты численного моделирования ряда систем с регенеративным входным процессом показывают вполне удовлетворительную точность оценивания вероятности потери в случае, когда в системе используется оценка ЭПС, получаемая на основе найденной аппроксимации.

Ключевые слова: система с постоянной скоростью обслуживания; эффективная пропускная способность; регенеративный процесс; регенеративная оценка; стационарный процесс нагрузки; вероятность потери

1 Введение

В современных коммуникационных системах одним из важнейших показателей качества обслуживания (QoS) является вероятность превышения стационарным процессом нагрузки W (незавершенной работы) некоторого (большого) уровня b . Для системы с конечным буфером b указанная вероятность является вероятностью потери.

Интерес к системам с регенеративным входным процессом обусловлен тем, что такие процессы сохраняют свойство регенерации при прохождении через узлы коммуникационной сети [1, 2]. В то же время пуассоновский процесс и даже общий процесс восстановления не обладают таким свойством. (Исключением являются пуассоновские потоки в стационарной сети Джексона без циклов, состоящей из узлов вида $M/M/1$.) Различные вопросы, связанные с вычислением ЭПС в системе с регенеративным входным процессом, рассматривались в работах [3–7], в которых также предложено оценивать ЭПС на основе *регенеративной* оценки, опирающейся на группировку данных по циклам регенерации. Отметим также близкую по тематике предшествующую работу [8]. Важной работой

в области исследования ЭПС является обзорная статья [9].

Если уровень b задан, а мощность обслуживающего устройства C (величина работы, которую прибор может сделать за единицу времени) можно изменять, то естественная задача QoS состоит в выборе такого значения C , которое гарантирует, что стационарная нагрузка не превысит уровня b с заданной (малой) вероятностью Γ , т. е.

$$P(W > b) \leq \Gamma. \quad (1)$$

Минимальная величина мощности C , удовлетворяющая этому условию, и называется *эффективной пропускной способностью* системы.

Покажем, как решается задача вычисления ЭПС для системы с одним обслуживающим устройством и неограниченным буфером для ожидающих заявок. Удобно считать, что обслуженная работа поступает в систему и покидает ее в (целочисленные) моменты t и что величина $W(t)$ равна незавершенной работе в момент $t - 1$ с учетом работы, поступившей в момент $t - 1$, и за вычетом работы, покинувшей систему в момент $t - 1$, $t = 0, 1, \dots$ [10].

* Работа выполнена при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

¹ Институт прикладных математических исследований Карельского научного центра Российской академии наук; Петрозаводский государственный университет, borodina@krc.karelia.ru

² Институт прикладных математических исследований Карельского научного центра Российской академии наук; Петрозаводский государственный университет, emorozov@karelia.ru

Предполагается, что объем уходящей работы в каждый момент (дискретного) времени равен C , что согласуется с «объемом жидкости», вытекшей из системы в течение интервала времени длины 1. Пусть v_i — величина работы, поступившей в систему в момент i , и тогда величина $V(t) := \sum_{i=0}^{t-1} v_i$ есть суммарная работа, поступившая в систему в интервале $[0, t - 1]$. Обозначим через $W(t)$ незавершенную работу по обслуживанию заявок, находящихся в системе в момент времени $t = 0, 1, \dots$, полагая $W(0) = 0$. Предположим, что $\{v_i, i \geq 0\}$ — независимые, одинаково распределенные случайные величины (н.о.р.с.в.), причем, вообще говоря, $P(v_i = 0) > 0$. Очевидно, имеет место такая рекурсия Линдли (в дискретном времени):

$$W(t + 1) = [W(t) + v_t - C]^+, \quad t \geq 0, \quad (2)$$

откуда следует, что процесс $\{W(t)\}$ образует марковскую цепь (с общим пространством состояний). Обозначим $X_i = v_i - C$ и введем случайное блуждание

$$\begin{aligned} Z(t) &:= \sum_{i=0}^{t-1} X_i = \sum_{i=0}^{t-1} (v_i - C) = \\ &= V(t) - Ct, \quad t \geq 1, \end{aligned} \quad (3)$$

где положено $Z(0) = 0$. Обозначим типичный шаг этого блуждания через $X = v - C$. (Здесь и далее типичный элемент последовательности н.о.р.с.в. обозначается без соответствующего индекса.) Так как буфер для ожидания не ограничен, то предполагается выполненным следующее условие отрицательного сноса у случайного блуждания Z :

$$EX = Ev - C := \lambda - C < 0. \quad (4)$$

Заметим, что (4) является условием стационарности процесса (незавершенной) нагрузки $\{W(t), t \geq 0\}$. Действительно, опираясь на рекурсию (2), определим приращение этого процесса $\Delta(t) = W(t + 1) - W(t)$ между моментами $t - 1$ и t . Предположим, что марковская цепь

$$W(t) \xrightarrow{d} \infty, \quad t \rightarrow \infty, \quad (5)$$

где \xrightarrow{d} означает сходимость по вероятности. Поскольку $\Delta(t) \leq v_t$ и $Ev = \lambda < \infty$, то из (2) легко следует, что $E\Delta(t) \rightarrow \lambda - C < 0, t \rightarrow \infty$. Этот результат, как легко проверить, противоречит сходимости (5) и поясняет термин «отрицательный снос». Для широкого класса цепей Маркова последний результат, в свою очередь, влечет существование стационарного процесса $W(t) \Rightarrow W$ [11]. (Знак \Rightarrow обозначает сходимость по распределению.)

Данная статья является продолжением работы [7], в которой основное внимание было уделено сравнению оценки по методу группового среднего (batch mean) с регенеративной оценкой. Здесь представлена более подробная мотивировка эвристических соображений, позволяющих получить искомую аппроксимацию ЭПС в жидкостной системе с регенеративным входным процессом. Кроме того, представлены результаты численных экспериментов, которые подтверждают, что полученная аппроксимация действительно может быть эффективно использована для оценивания ЭПС в рассматриваемых системах обслуживания. В разд. 2 показано как вычисляется ЭПС, удовлетворяющая условию (1), в случае н.о.р.с.в. $\{v_i\}$.

2 Вероятность большого уклонения и эффективная пропускная способность

Рассмотрим асимптотику вероятности большого уклонения стационарной нагрузки для рассмотренной выше системы с н.о.р. $\{v_i\}$ и с заданной скоростью обслуживания C . Напомним известный результат [11]:

$$\begin{aligned} W(n) &= \sup_{0 \leq t < n} [V(t) - Ct]^+ = \\ &= \sup_{0 \leq t < n} Z(t), \quad n = 1, \dots, \end{aligned} \quad (6)$$

связывающий величину незавершенной работы с максимумом случайного блуждания (3). (Поскольку $Z(0) = 0$, то $\sup[\cdot] = \sup[\cdot]^+$ в соотношении (6), что согласуется с (2) и гарантирует $\inf_n W(n) \geq 0$.) Отметим, что марковская цепь $\{W(t)\}$ с дискретным временем регенерирует в последовательные моменты опустошения системы, т. е. в моменты

$$\beta_{n+1} = \min\{k > \beta_n : W(k) = 0\}, \quad n \geq 0, \quad \beta_0 := 0.$$

Длина (типичного) цикла регенерации процесса $\alpha := \beta_1$ есть непериодическая с.в., поскольку ввиду (4) выполнено условие:

$$\begin{aligned} P(\alpha = 1) &= P(W(t + 1) = 0 | W(t) = 0) = \\ &= P(v < C) > 0. \end{aligned} \quad (7)$$

Как хорошо известно, из (4) и (7) следует, что $E\alpha < \infty$, т. е. процесс $\{v_i\}$ является положительно возвратным и, кроме того, стационарный процесс нагрузки W существует и является пределом по распределению:

$$W(t) \Rightarrow W =_{\text{st}} \sup_{n \geq 0} Z(n),$$

где $=_{st}$ обозначает стохастическое равенство. Предположим, что в некоторой положительной окрестности $(0, \theta_0)$ параметра θ выполнено условие:

$$Ee^{\theta v} < \infty, \quad (8)$$

и приведем ряд необходимых далее понятий и результатов теории больших уклонений [12]. Рассмотрим (нормированную) *логарифмическую производящую функцию моментов* случайного блуждания $Z(n)$

$$\Lambda_n(\theta) := \frac{1}{n} \ln Ee^{\theta Z(n)}, \quad n \geq 1,$$

которая ввиду независимости слагаемых $\{X_i\}$, формирующих случайное блуждание $Z(n)$, может быть записана как

$$\Lambda_n(\theta) := \Lambda(\theta) = \ln Ee^{\theta v} - \theta C, \quad n \geq 1.$$

Поскольку функция $\psi(\theta) := \ln Ee^{\theta v}$ выпуклая, $\psi(0) = 0$ и ввиду условия (4) $\psi'(0) = Ev < C$, то существует единственный корень $\theta^* > 0$ уравнения

$$\ln Ee^{\theta^* v} = \theta^* C. \quad (9)$$

Будем предполагать, что с.в. v такова, что $\theta^* \in (0, \theta_0)$. Из работы [13] следует, что при выполнении условий (4), (8) стационарный процесс W удовлетворяет следующему асимптотическому соотношению [12, 13]:

$$\lim_{x \rightarrow \infty} \frac{1}{x} \ln P(W > x) = -\theta^*. \quad (10)$$

Это влечет такую экспоненциальную аппроксимацию вероятности большого уклонения стационарного процесса нагрузки: $P(W > x) = e^{-\theta^* x + o(x)}$, $x \rightarrow \infty$. Ввиду (9) ЭПС равна

$$C = \frac{\ln Ee^{\theta^* v}}{\theta^*} \quad (11)$$

и удовлетворяет условию (4) (см. лемму 9.1.5 в работе [14]).

Отметим, что в случае, когда с.в. $\{v_i\}$ являются зависимыми, асимптотика (10) также может быть доказана в некоторых случаях [12–14]. В частности, это верно, если последовательность $\{v_i\}$ — стационарная с перемешиванием [10].

Ключевыми для справедливости асимптотики (10) являются свойства логарифмической производящей функции моментов *входного процесса*

$$\Lambda_V^{(n)}(\theta) := \frac{1}{n} \ln Ee^{\theta V(n)}.$$

В первую очередь требуется существование *предела Гартнера–Эллиса* [13, 14]

$$\lim_{n \rightarrow \infty} \Lambda_V^{(n)}(\theta) = \Lambda_V(\theta), \quad n \rightarrow \infty,$$

а также условие его конечности либо для всех $\theta \in (0, \theta_0)$ [12, 13], либо для всех $\theta > 0$ [14].

Когда скорость C задана, параметр θ^* , найденный из условия (9), дает скорость (экспоненциального) убывания хвоста вероятности $P(W > b)$. Рассмотрим теперь обратную задачу: скорость C (т. е. ЭПС) должна быть выбрана так, чтобы обеспечить требование (1). ЭПС должна обеспечивать условие стационарности (4) и поэтому можно опираться на экспоненциальную асимптотику (10). Это дает такое (приближенное) уравнение:

$$P(W > b) = \Gamma = e^{-\theta^* b}, \quad (12)$$

решение которого имеет вид

$$\theta^* = -\frac{\ln \Gamma}{b} > 0. \quad (13)$$

Таким образом, соотношения (11), (13) позволяют (приближенно) определить ЭПС как

$$C = \frac{\Lambda_V(\theta^*)}{\theta^*} = \frac{\Lambda_V(-\ln \Gamma/b)b}{-\ln \Gamma}.$$

В случае н.о.р. $\{v_i\}$ получаем, см. (9),

$$C = -\frac{b}{\ln \Gamma} \ln Ee^{-v \ln \Gamma/b}.$$

Это соотношение позволяет в ряде случаев получить явное значение C .

Приведенный анализ верен также и для системы с конечным буфером (большого) размера b , т. е. для системы с потерями, поскольку вероятность потери $P(W > b)$ в такой системе ведет себя асимптотически (при $b \rightarrow \infty$) так же, как и в системе с неограниченным буфером [12].

3 Вычисление эффективной пропускной способности для регенеративного входного процесса

Как было сказано выше, интерес к системам с регенеративным входным потоком обусловлен сохранением свойства регенерации при прохождении потоков между узлами коммуникационной сети. Предлагаемый ниже подход в значительной мере опирается на эвристические соображения и в

первую очередь мотивирован тем, что широко используемая оценка по методу группового среднего (batch mean) не учитывает зависимости между данными входного процесса. Как отмечено, например, в работах [7, 15], используемое в этой оценке разбиение данных на блоки фиксированной длины, как правило, приводит к недооценке доли потерь в системах с (большим) конечным буфером. Поэтому использование такой оценки для вычисления ЭПС в случае входного процесса с зависимыми данными может привести к нарушению требуемых гарантий QoS, что неприемлемо в системах высокой надежности.

Рассмотрим вновь дискретную рекурсию Линдли (2), и пусть теперь входная последовательность $\{v_n\}$ является регенерирующей с моментами регенерации $\{\beta_k\}$ и периодами регенерации $\alpha_k = \beta_{k+1} - \beta_k$. Таким образом, значения v_i внутри каждого цикла регенерации могут быть зависимыми, но значения v_i и v_j , принадлежащие разным циклам, являются независимыми. Поэтому суммарная работа, поступающая на циклах регенерации входного процесса,

$$V_k := \sum_{i=\beta_k}^{\beta_{k+1}-1} v_i, \quad k \geq 0, \quad \beta_0 = 0, \quad (14)$$

образует последовательность н.о.р. *блоков* (с типичным элементом V). Будем предполагать, что для некоторого $\theta_0 > 0$ при всех $\theta \in (0, \theta_0)$ выполнено условие конечности экспоненциальных моментов величины блока

$$\ln E e^{\theta V} < \infty. \quad (15)$$

Заметим, что поскольку величина блока V и длина цикла α связаны очевидным образом

$$V =_{\text{st}} \sum_{i=0}^{\alpha-1} v_i, \quad (16)$$

то их моментные свойства также тесно связаны. Скажем, в специальном случае, когда длина цикла α не зависит от последовательности н.о.р.с.в. $\{v_n\}$, получаем $E e^{\theta V} = E[E e^{\theta v}]^\alpha$. В дальнейшем потребуется лишь положительная возвратность последовательности $\{v_n\}$, т.е. условие $E\alpha < \infty$. Заметим, что если длина цикла α является моментом остановки относительно последовательности н.о.р.с.в. $\{v_n\}$, то по неравенству Иенсена и тождеству Вальда $E e^{\theta V} \geq \theta E V = \theta E \alpha E v$ и положительная возвратность следует из условия (15). В этом случае положительную возвратность можно также получить из разложения функции $\phi(\theta) := E e^{\theta V} = 1 + \theta E V + o(\theta)$ в ряд Тейлора в окрестности $\theta = 0$.

Обозначим через $k(n) := \max(k \geq 0 : \beta_k \leq n)$ число циклов регенерации среди заявок с номерами $0, 1, \dots, n$, так что $k(0) = 0$. Заметим, что $k(i) = 0, i \leq \beta_1 - 1$ и $k(\beta_i) = i, i \geq 1$. Напомним обозначение:

$$V(n) = \sum_{i=0}^{n-1} v_i, \quad V(0) = 0.$$

В теории регенерации процесс $\{V(n)\}$ называется процессом накопления. Заметим, что

$$\begin{aligned} \ln E e^{\theta \sum_{i=0}^{k(n)-1} V_i} &\leq \ln E e^{\theta V(n)} \leq \\ &\leq \ln E e^{\theta \sum_{i=0}^{k(n)} V_i}, \quad n \geq 0 \quad \left(\sum_{\emptyset} = 0 \right). \end{aligned}$$

Далее основное допущение состоит в том, что *блоки* $V_i, i \leq k(n)$, и *величина* $k(n)$ *независимы при больших* n . (Вообще говоря, эти величины зависимы, поскольку $k(n)$ зависит от длин циклов, полученных к моменту n .) Обозначим $a := E e^{\theta V}$ и, используя свойство условного математического ожидания, получим (при больших n)

$$\frac{1}{n} \ln E a^{k(n)} \leq \frac{1}{n} \ln E e^{\theta V(n)} \leq \frac{1}{n} \ln E a^{k(n)+1}. \quad (17)$$

Следующее допущение состоит в том, что *нижняя и верхняя границы в неравенстве (17) асимптотически близки при больших* n . Приведем некоторые соображения в пользу этого предположения. Рассмотрим величину работы, поступающей в систему с момента n до конца текущего цикла регенерации, т.е.

$$U(n) = \sum_{i=0}^{k(n)} V_i - V(n) = \sum_{i=n}^{\beta_{k(n)}-1} v_i.$$

Как показано в [11], при условии (15) для некоторого $\varepsilon > 0$

$$E e^{\varepsilon U(n)} \rightarrow D, \quad n \rightarrow \infty,$$

где постоянная $D < \infty$. Отметим также, что с.в. $Z(n) := V(n) - \sum_{i=0}^{k(n)-1} V_i$, равная величине работы, которая уже поступила в систему на текущем в момент n цикле регенерации, асимптотически (при $n \rightarrow \infty$) распределена так же, как и величина $U(n)$, и поэтому имеет (в пределе) такие же моментные свойства. (С учетом сделанных замечаний можно легко доказать сближение нижней и верхней границы в (17) в предположении *независимости* величин $V(n), U(n)$ и $Z(n)$.) Далее, по неравенству Иенсена

$$\frac{1}{n} \ln E a^{k(n)+1} \geq \frac{E(k(n)+1)}{n} \ln a,$$

а по элементарной теореме восстановления $E(k(n) + 1)/n \rightarrow 1/E\alpha$. Приведенные выше соображения позволяют заключить, что при больших n аппроксимация

$$\Lambda_V(\theta) = \frac{\ln Ee^{\theta V}}{E\alpha} \quad (18)$$

может дать удовлетворительное для практических целей значение функции $\Lambda_V(\theta)$. Это, в свою очередь, является основанием для аппроксимации искомого ЭПС с помощью формулы

$$C = \frac{\ln Ee^{\theta^* V}}{\theta^* E\alpha}. \quad (19)$$

Подчеркнем, что при получении приведенной выше аппроксимации ЭПС используется приближение (12), а также *предположение*, что в данной системе верна экспоненциальная асимптотика (10). Кроме того, хотя это не оговаривалось ранее, в основе асимптотики (10) лежит также предположение о том, что входная последовательность $\{v_n\}$ является *стационарной*. (Конструкция стационарного регенерирующего процесса описана в [16].) Наконец, отметим, что результат (18) формально можно получить из (17) заменой случайного числа циклов $k(n)$ его математическим ожиданием и переходом к пределу при $n \rightarrow \infty$.

Как и в работе [7], ниже рассмотрены два следующих варианта получения регенеративного входного процесса.

В первом варианте рассматривается двухузловая сеть, в которой на вход узла 1 в соответствии с процессом восстановления поступают заявки с н.о.р. временами обслуживания $\{S_n^{(1)}\}$ и с коэффициентом загрузки $\rho < 1$. В этом случае выходной процесс из узла 1 является положительно возвратным регенерирующим процессом со средней длиной цикла $E\alpha < \infty$. Будем считать длину цикла α равной числу заявок, поступивших в узел 1 на цикле регенерации. (Это обычный подход при рассмотрении процессов, вложенных в моменты прихода заявок. Разумеется, можно рассмотреть длину цикла и в непрерывном времени.) Моментные свойства длины цикла α можно получать исходя из свойств с.в. $S^{(1)}$. В этой связи оказывается полезным следующий результат. Пусть ϕ есть некоторая измеримая функция и $\rho < 1$. Тогда (см. [17, 18])

$$E\phi(S^{(1)}) < \infty \text{ влечет } E\phi(\alpha) < \infty. \quad (20)$$

В частности, если $Ee^{\theta S^{(1)}} < \infty$, то $Ee^{\theta\alpha} < \infty$. (Для получения условий конечности моментов длины цикла регенерации узла 1 в непрерывном времени соответствующее условие надо наложить и на длину интервала входного потока в узел 1, см. [17].

Однако эта конструкция не используется в данной работе.) Далее н.о.р. длины циклов регенерации $\{\alpha_n\}$ узла 1 используются в качестве длин циклов входного процесса в *жидкостной узел 2 с искомой скоростью C*, где процесс загрузки описывается рекурсией (2). Иными словами, α равно длине цикла регенерации (числу интервалов единичной длины) процесса нагрузки $\{v_i\}$, поступающего в узел 2 и описываемого рекурсией (2). Мотивировку такой модели, где оба узла связаны не напрямую, можно найти в работе [7].

Во втором варианте входной регенеративный процесс можно считать заданным и наложить требуемые моментные условия на длины циклов и на объем поступающей на цикле работы.

Точность полученной аппроксимации (18) для обоих случаев иллюстрируется в следующем разделе с помощью имитационного моделирования ряда систем с регенеративным входным процессом.

Замечание 1. С учетом разложения функции $\phi(\theta) := \Lambda_V(\theta)$ в ряд Тейлора в окрестности $\theta = 0$ соотношение (19) можно записать как

$$C = \frac{1}{E\alpha} \left(EV + \frac{E(V^2)\theta^*}{2} + o(\theta^*) \right), \quad \theta^* \rightarrow 0,$$

что в ряде случаев может быть использовано для вычисления C на основе лишь первых двух моментов с.в. V и α , см. (16).

Замечание 2. Используя неравенство Иенсена, из теории регенерации получим

$$\frac{1}{n} \ln Ee^{\theta V(n)} \geq \theta \frac{EV(n)}{n} \rightarrow \theta \frac{EV}{E\alpha}, \quad n \rightarrow \infty. \quad (21)$$

При этом величина $EV/E\alpha$ является интенсивностью входного регенерирующего процесса, и если существует стационарный предел $v_n \Rightarrow v$, то $Ev = EV/E\alpha$ [11]. Если использовать нижнюю границу (21) вместо предела $\Lambda_V(\theta)$, то уравнение (9) дает $EV/E\alpha = C$, что влечет нестационарность системы с регенеративным входным процессом [19].

Замечание 3. В работах [20–22] приводится (в разных формах) сильный принцип инвариантности, позволяющий аппроксимировать с вероятностью 1 процесс восстановления $\{k(n)\}$, а также исходный процесс накопления $\{V(n)\}$, с помощью винеровского процесса и некоторой (случайной) функции *уклонения* $f(n) = o(n)$. Однако применение этих результатов для получения асимптотики $\Lambda_V^{(n)}(\theta)$ опирается на некоторые допущения о независимости и приводит в ряде случаев к значению ЭПС, при котором вероятность потери превышает величину Γ .

4 Результаты численного моделирования

В этом разделе рассматривается процесс незавершенной работы на интервалах единичной длины (слотах), удовлетворяющий рекурсии (2), с положительно возвратной входной последовательностью $\{v_i\}$, моментами регенерации $\{\beta_k\}$ и н.о.р. длинами циклов $\{\alpha_k\}$. Ниже исследуется точность оценивания скорости обслуживания C (искомой ЭПС) на основе полученной в предыдущем разделе аппроксимации

$$C = \frac{\Lambda_V(\theta^*)}{\theta^*},$$

где

$$\Lambda_V(\theta^*) = \frac{1}{E\alpha} \ln Ee^{\theta^*V} := \Lambda_{\text{REG}}(\theta^*),$$

а параметр $\theta^* = -\ln \Gamma/b$. Поскольку $\beta_k/k \rightarrow E\alpha$, то выборочная оценка

$$\hat{\Lambda}_{\text{REG}}^{(k)}(\theta^*) := \frac{\ln(1/k) \sum_{i=1}^k e^{\theta^*V_i}}{(1/k) \sum_{i=1}^k \alpha_i} = \frac{k}{\beta_k} \ln \frac{1}{k} \sum_{i=1}^k e^{\theta^*V_i}$$

величины $\Lambda_{\text{REG}}(\theta^*)$, полученная по k регенеративным блокам входного потока, является сильно состоятельной: $\hat{\Lambda}_{\text{REG}}^{(k)}(\theta^*) \rightarrow \Lambda_{\text{REG}}(\theta^*)$, $k \rightarrow \infty$ с вероятностью 1. Поэтому ниже соответствующая регенеративная оценка ЭПС вычисляется по формуле

$$\hat{C}(k) := \frac{\hat{\Lambda}_{\text{REG}}^{(k)}(\theta^*)}{\theta^*} \quad (22)$$

для (достаточно большого) числа k циклов регенерации, полученных в процессе имитационного моделирования.

Эксперимент 1. Рассмотрена двухузловая сеть, где узел 1 есть система вида $M/M/1$ с показательным временем обслуживания S с интенсивностью $\mu = 1$ и с интенсивностью пуассоновского входного процесса $\lambda := \rho < 1$. Поскольку время обслуживания имеет конечные экспоненциальные моменты,

$Ee^{\theta S} < \infty$ для всех $\theta < 1$, то ввиду (20) получаем также $Ee^{\alpha\theta} < \infty$ при $\theta < 1$. Подчеркнем, что значения параметра $\theta = \theta^*$ в табл. 1 выбраны с учетом этого условия. (В этой связи отметим, что в случае н.о.р. $\{v_i\}$ с легким хвостом распределение хвоста суммы V в соотношении (16) может иметь тяжелый хвост, если индекс суммирования α имеет тяжелый хвост [23].) Как было упомянуто выше, длины циклов регенерации узла 1 используются как длины циклов регенерации в дискретном времени входного потока в узел 2, на основе которых получают регенеративные блоки (14). В узле 2 надо определить скорость C , гарантирующую неперевышение заданного уровня потерь Γ . В данной модели коэффициент загрузки ρ узла 1 существенно влияет на моментные свойства длины цикла α и, кроме того, значения с.в. $\{v_i\}$ на одном цикле являются зависимыми. Точнее говоря, рассматриваются независимые с.в. $\{\eta_k\}$, имеющие распределение Вейбулла (с легким хвостом):

$$F_\eta(x) = 1 - e^{-\gamma x^c}, \quad \gamma > 0, \quad c \geq 1.$$

Тогда величина работы, поступающей в узел 2 в момент j (любого) текущего цикла регенерации входного процесса, задается соотношением:

$$v_j = \frac{\sum_{k=1}^j \eta_k}{j}, \quad 1 \leq j \leq \alpha, \quad (23)$$

где α есть (типичная) длина цикла. Далее по формуле (22) вычисляется значение оценки $\hat{C}(k)$, которая используется в качестве скорости обслуживания C в узле 2. Затем (в предположении достижения стационарного режима в узле 2) с помощью выборочного среднего оценивается вероятность $P(W > b)$ в узле 2. Заметим, что стандартный способ получения стационарного значения W при имитационном моделировании состоит в пропуске начального, так называемого *переходного*, периода. Кроме того, в рассматриваемой системе пропуск переходного периода служит для получения стационарного режима входного процесса.

В табл. 1 представлены результаты оценивания ЭПС, а также вероятности потери при следующих

Таблица 1 Регенеративная оценка ЭПС для двухузловой сети

#	Γ	θ^*	$\hat{C}(k)$	$\hat{\Gamma}$	Δ/Γ
1	10^{-3}	0,230259	0,264602	$8,15 \cdot 10^{-4}$	0,15
2	10^{-4}	0,307011	0,290134	$2,05 \cdot 10^{-5}$	0,75
3	10^{-5}	0,383764	0,348517	$1,84 \cdot 10^{-6}$	0,816
4	10^{-6}	0,460517	0,527721	$2,97 \cdot 10^{-8}$	0,97
5	10^{-7}	0,53727	0,661887	$0,45 \cdot 10^{-8}$	0,955
6	10^{-8}	0,614023	0,986111	$8,67 \cdot 10^{-10}$	0,913

Таблица 2 Регенеративная оценка ЭПС в случае ограничения объема работы на цикле

#	Γ	θ^*	d	$\hat{\alpha}$	$\hat{C}(k)$	$Var\hat{C}(k)$	$\hat{\Gamma}$	Δ/Γ
1	10^{-4}	0,153506	50	89,1	0,560441	$5,23 \cdot 10^{-6}$	$0,3433 \cdot 10^{-5}$	0,6567
2	10^{-5}	0,191882	50	89,2	0,560947	$7,73 \cdot 10^{-6}$	$0,4153 \cdot 10^{-5}$	0,5847
3	10^{-6}	0,230259	70	124,9	0,561252	$2,64 \cdot 10^{-6}$	$0,8698 \cdot 10^{-6}$	0,1302
4	10^{-7}	0,268635	70	124,8	0,562472	$4,23 \cdot 10^{-6}$	$0,8871 \cdot 10^{-7}$	0,1129
5	10^{-8}	0,307011	70	124,5	0,563537	$6,98 \cdot 10^{-6}$	$0,2116 \cdot 10^{-8}$	0,7884

условиях: параметры распределения Вейбулла $\gamma = 3$, $c = 4$; интенсивность трафика $\rho = 0,4$ в узле 1; число циклов $k = 40\,000$, уровень превышения (буфер) $b = 30$. При этом разность $\Delta := \Gamma - \hat{\Gamma} > 0$, т. е. регенеративная оценка ЭПС (22) обеспечивает требуемый уровень надежности Γ .

Эксперимент 2. Рассматривается система с постоянной искомой скоростью обслуживания C , а цикл регенерации входного процесса завершается, когда величина работы, поступающей в систему на цикле, достигает заданного объема d . Как и выше, в качестве скорости C использована ее оценка $\hat{C}(k)$. В табл. 2 представлены результаты моделирования при $b = 60$, а зависимость между с.в. $\{v_i\}$ внутри цикла регенерации определяется соотношением (23).

Подчеркнем, что во всех случаях $\Delta > 0$, т. е. используемая оценка $\hat{C}(k)$ обеспечивает требуемую гарантию Γ . В этой связи важно отметить, что оценивание ЭПС по методу группового среднего в ряде случаев приводит к значению $\Delta < 0$, т. е. к нарушению гарантии Γ (см. [7]). В то же время величина Δ/Γ показывает, что при использовании оценки $\hat{C}(k)$ уровень Γ обеспечивается с определенным запасом, и это связано с высокой чувствительностью доли потерь к изменению скорости обслуживания. С другой стороны, оценка ЭПС имеет очень небольшую дисперсию и мало чувствительна к изменению величины Γ . Это особенно хорошо заметно в табл. 2. Например, изменение величины Γ с 10^{-4} до 10^{-5} влечет изменение $\hat{C}(k)$ с 0,5604 до 0,5609. Это несомненно вызвано тем, что нагрузка на цикле ограничена постоянной величиной d .

Отметим, что для вычисления оценки $\hat{\Gamma}$ вероятности превышения стационарным процессом нагрузки (высокого) уровня $b = 60$ использовались ресурсы кластера КарНЦ РАН, а также метод ускоренного регенеративного имитационного моделирования, разработанный для оценивания вероятностей редких событий [15].

5 Заключение

В статье обсуждается понятие ЭПС жидкостной системы обслуживания в дискретном времени.

Рассматривается входной процесс с независимыми приращениями, но основное внимание уделено аппроксимации и оцениванию ЭПС в случае входного регенеративного процесса. Представлены эвристические соображения, на основе которых получена аппроксимация ЭПС. Результаты численных экспериментов показывают, что оценка (22), полученная на основе аппроксимации (19), гарантирует уровень надежности Γ . Таким образом, аппроксимация (19) может считаться эффективной альтернативой оценке ЭПС по методу группового среднего для системы с регенеративным входным процессом. Дальнейшие исследования в данном направлении должны включать как теоретическое обоснование предложенной аппроксимации, так и расширение численных экспериментов по проверке ее точности.

Авторы благодарят М. А. Лифшица за ценные замечания.

Литература

1. Морозов Е. Критерий стационарности одного класса непассоновских сетей обслуживания // Изв. АН СССР. Техн. кибернетика, 1988. № 1. С. 129–133.
2. Morozov E. Stability of Jackson type network output // Queueing Syst., 2002. Vol. 40. P. 383–406.
3. Vorobieva I., Morozov E., Pagano M., Procissi G. A new regenerative estimator for effective bandwidth prediction // AMICT 2007 Proceedings. — Petrozavodsk, 2008. P. 175–187.
4. Morozov E., Dyudenko I., Pagano M. Regenerative estimator of the overflow probability in a tandem network // 7th Workshop (International) on Rare Event Simulation Proceedings. — Rennes, France, 2008. P. 283–287.
5. Dyudenko I., Morozov E., Pagano M. Regenerative estimator for effective bandwidth // Mathematical methods for analysis and optimization of information telecommunication networks: Proceedings of the International Conference. — Minsk: Belarusian State University, 2009. P. 58–60.
6. Dyudenko I., Morozov E., Pagano M., Sandmann W. Comparative study of effective bandwidth estimators: Batch

- means and regenerative cycles // 6th St. Petersburg Workshop on Simulation Proceedings. — St-Petersburg, 2009. Vol. II. P. 1003–1007.
7. *Бородина А. В., Морозов Е. В.* Сравнение двух оценок эффективной пропускной способности системы обслуживания // Тр. Карельского научного центра РАН, 2012. № 5. С. 8–17.
 8. *Crosby S., Leslie I., Huggard M., Lewis J. T., McGurk B., Russel R.* Predicting bandwidth requirements of ATM and Ethernet traffic // IEE UK Teletraffic Symposium Proceedings. — Glasgow, U.K., 1996.
 9. *Kelly F.* Notes on effective bandwidths // Stochastic networks: Theory and applications / Eds. F.P. Kelly, S. Zachary, I. B. Ziedins. — Roy. Stat. Soc. Lecture Notes ser., 4. — Oxford University Press, 1996. P. 141–168.
 10. *Lewis J. T., Russell R.* An introduction to large deviation for teletraffic engineers, 1997. <https://engineering.purdue.edu/ece647/notes.html>.
 11. *Asmussen S.* Applied probability and queues. — 2nd ed. — NY: Springer, 2003.
 12. *Ganesh A., O'Connell N., Wischik D.* Big queues. — Berlin: Springer-Verlag, 2004.
 13. *Glynn P. W., Whitt W.* Logarithmic asymptotics for steady-state tail probabilities in a single-server queue // JAP, 1994. Vol. 31. P. 131–156.
 14. *Cheng-Shang Chang.* Performance guarantees in communication networks. — Springer, 2000.
 15. *Бородина А., Дюденко И., Морозов Е.* Ускоренное оценивание вероятности переполнения регенеративных систем обслуживания // ОПИПМ, 2009. Т. 16. № 4. С. 577–593.
 16. *Thorisson H.* Coupling, stationarity, and regeneration. — NY: Springer, 2000.
 17. *Thorisson H.* The queue $GI/GI/k$: Finite moments of the cycle variables and uniform rates of convergence // Commun. Stat.-Stochastic Models, 1985. Vol. 192. P. 221–238.
 18. *Wolff R. W.* Stochastic modeling and the theory of queues. — Prentice-Hall, 1989.
 19. *Morozov E.* Weak regeneration in modeling of queueing processes // Queueing Syst., 2004. Vol. 46. No. 3–4. P. 295–315.
 20. *Csorgo M., Horvath L., Steinebach J.* Invariance principles for renewal processes // Ann. Prob., 1987. Vol. 15. No. 4. P. 1441–1460.
 21. *Damerdji H.* Strong consistency of the variance estimator in steady-state simulation output analysis // Math. Oper. Res., 1994. Vol. 19. No. 2. P. 494–512.
 22. *Sharma V.* Reliable estimation via simulation // Queueing Syst., 1995. Vol. 19. P. 169–192.
 23. *Robert C. Y., Segers J.* Tails of random sums of a heavy-tailed number of light-tailed terms // Insurance: Mathematics and Economics, 2008. Vol. 43. P. 85–92.

СТАЦИОНАРНОЕ РАСПРЕДЕЛЕНИЕ ВРЕМЕНИ ОЖИДАНИЯ В СИСТЕМЕ ОБСЛУЖИВАНИЯ С ОТРИЦАТЕЛЬНЫМИ ЗАЯВКАМИ, БУНКЕРОМ ДЛЯ ВЫТЕСНЕННЫХ ЗАЯВОК, РАЗЛИЧНЫМИ ИНТЕНСИВНОСТЯМИ ОБСЛУЖИВАНИЯ ПРИ ДИСЦИПЛИНЕ FIRST–FIFO–FIFO*

Р. В. Разумчик¹

Аннотация: Рассматривается система обслуживания с одним обслуживающим прибором и пуассоновскими потоками обычных и отрицательных заявок. Обычная заявка, поступающая в систему, занимает одно место в очереди в накопителе неограниченной емкости. Отрицательная заявка при поступлении выбивает заявку, находящуюся на первом месте в очереди в накопителе, перемещает ее в другой накопитель неограниченной емкости (бункер) и сама покидает систему. Если при поступлении отрицательной заявки накопитель пуст, она покидает систему, не оказывая на нее никакого воздействия. После окончания обслуживания очередной заявки на прибор поступает заявка, занимающая первое место в очереди в накопителе или, если накопитель пуст, заявка с первого места в очереди в бункере. Длительности обслуживания заявок из накопителя и из бункера имеют экспоненциальные распределения с различными параметрами. Найдено стационарное распределение времени ожидания (обычной) заявкой начала обслуживания в терминах преобразования Лапласа–Стилтьеса (ПЛС).

Ключевые слова: система массового обслуживания; отрицательные заявки; бункер; различные интенсивности обслуживания; время ожидания

1 Введение и описание системы

Системы и сети с отрицательными заявками по-прежнему являются актуальным предметом исследований, что обусловлено возможностью их применения для моделирования различных аспектов телекоммуникационных систем (таких как перебой в работе, потеря информации), различных стадий процессов планирования, управления и контроля движения материальных, информационных и финансовых ресурсов. Некоторое представление о последних результатах можно найти, например, в [1–7]. Настоящая работа посвящена нахождению стационарного распределения времени ожидания начала обслуживания в системе массового обслуживания (СМО) с одним обслуживающим прибором и специальным видом отрицательных заявок, которые, в отличие от классического, не «убивают» заявки, находящиеся в системе, а перемещают их в другую очередь, откуда те обслуживаются с относительным приоритетом.

Рассматривается однолинейная СМО, в которую поступает пуассоновский поток заявок интенсивности λ . Далее заявки этого потока будем

называть положительными заявками. Для положительных заявок есть накопитель неограниченной емкости. Дополнительно в систему поступает еще один пуассоновский поток отрицательных заявок интенсивности λ^- . Отрицательная заявка, поступающая в систему, вытесняет одну заявку из очереди в накопителе и перемещает ее в накопитель для вытесненных заявок (далее — бункер), который также имеет неограниченную емкость. При этом сама отрицательная заявка уходит из системы. Если в момент поступления отрицательной заявки в очереди в накопителе нет заявок, а на приборе обслуживается заявка, то отрицательная заявка, не прерывая обслуживания на приборе, покидает систему, не оказывая на нее никакого воздействия. То же самое происходит и в случае, когда в момент поступления отрицательной заявки накопитель и обслуживающий прибор пусты.

Выбор заявок на обслуживание производится следующим образом. После окончания обслуживания очередной заявки на прибор становится заявка из накопителя. Если же накопитель пуст, на прибор поступает заявка из бункера. Обслуживание заявок

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты 11-07-00112, 13-07-00223).

¹Институт проблем информатики Российской академии наук, grazumchik@iee.org

не прерывается новыми как положительными, так и отрицательными заявками.

Длительности обслуживания заявок из накопителя имеют экспоненциальное распределение с параметром μ_1 , а из бункера — экспоненциальное распределение с параметром μ_2 .

Будем считать, что поступающая в систему отрицательная заявка «убивает» первую заявку в очереди в накопителе, а в момент окончания обслуживания заявки на приборе на обслуживание выбирается первая заявка из очереди в накопителе или, если накопитель пуст, первая заявка из очереди в бункере. Такую дисциплину, по аналогии с [5], будем обозначать как FIRST–FIFO–FIFO.

Предполагается, что существует стационарный режим функционирования системы. Необходимое и достаточное условие для этого приведено в [8].

Основной результат работы состоит в нахождении в терминах ПЛС стационарного распределения времени ожидания начала обслуживания поступившей в систему (положительной) заявки.

В следующем разделе приводятся некоторые вспомогательные результаты, которые можно найти в [5, 8]. В разд. 3 внимание уделяется нахождению ПЛС времени ожидания начала обслуживания заявкой, поступающей на прибор из накопителя, и заявкой, поступающей на прибор из бункера.

2 Вспомогательные результаты

В работе [8] получены следующие стационарные характеристики, связанные с числом заявок в рассматриваемой системе:

- стационарная вероятность p_0 того, что система находится в состоянии простоя;
- стационарные вероятности $\{p_{i,j,k}, i \geq 0, j \geq 0, k = 0, 1\}$ того, что в накопителе находится i заявок, в бункере ожидают j заявок, вытесненных из накопителя, и на приборе обслуживается заявка либо из накопителя (при $k = 0$), либо из бункера (при $k = 1$), в терминах двойных производящих функций

$$P(u, v) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j,0} u^i v^j;$$

$$N(u, v) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j,1} u^i v^j.$$

Рассмотрим период занятости (ПЗ) системы $M/M/1/\infty$ с входящим потоком интенсивности λ и интенсивностью обслуживания b . Обозначим через

$G(x; b)$ функцию распределения (ФР) ПЗ этой системы, а через $\gamma(s; b) = \int_0^{\infty} e^{-sx} dG(x; b)$ — ПЛС ФР $G(x; b)$. Тогда

$$\gamma(s; b) = \frac{s + \lambda + b - \sqrt{(s + \lambda + b)^2 - 4\lambda b}}{2\lambda}.$$

Обозначим через $G_i(x)$ ФР ПЗ СМО $M/M/1/\infty$ с параметрами λ и $\mu_1 + \lambda^-$, открываемого заявкой экспоненциальной длины с параметром μ_i ($i = 1, 2$). Тогда ПЛС ФР $G_i(x)$ имеет вид:

$$\gamma_i(s) = \frac{\mu_i}{\mu_i + s + \lambda [1 - \gamma(s; \mu_1 + \lambda^-)]}.$$

Далее через $H_i(x; b)$ будем обозначать ФР Эрланга с параметром b и i фазами обслуживания (а через $h_i(x; b)$ — соответствующую плотность), т. е.

$$H_i(x; b) = \begin{cases} u(x), & i = 0, \\ \int_0^x \frac{b^i t^{i-1}}{(i-1)!} e^{-bt} dt = 1 - \sum_{k=0}^{i-1} \frac{b^k x^k}{k!} e^{-bt}, & i \geq 1. \end{cases}$$

Напомним, что распределение Эрланга $H_i(x; b)$ имеет ПЛС $(b/(s+b))^i$.

Наконец, обозначим через $D_k(t)$ вероятность того, что все заявки из накопителя будут обслужены до момента времени t при условии, что в начальный момент на приборе начинает обслуживаться заявка из бункера, а в накопителе имеется $k, k \geq 0$, других заявок. В работе [5] найдено выражение для ПЛС $\delta_k(s) = \int_0^{\infty} e^{-st} dD_k(t), k \geq 0$, этой вероятности, которое можно представить в следующем виде:

$$\delta_k(s) = \alpha(s) [\gamma(s + \mu_2; \lambda^-)]^k + \beta(s) [\gamma(s; \mu_1 + \lambda^-)]^k, k \geq 0,$$

где

$$\alpha(s) = \frac{\mu^-}{\lambda + s + \mu_2 - \lambda \gamma(s + \mu_2; \lambda^-)} \left[1 - \mu / \{ [s + \mu_2 + \lambda^- + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)] \gamma(s; \mu_1 + \lambda^-) - \lambda^- \} \times \frac{s + \mu_2 + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)}{s + \mu_1 + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)} \right],$$

$$\beta(s) = \mu_1 \mu_2 / \{ \{ [s + \mu_2 + \lambda^- + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)] \gamma(s; \mu_1 + \lambda^-) - \lambda^- \} \times (s + \mu_1 + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)) \}.$$

После изложения всех необходимых вспомогательных результатов перейдем к следующему разделу.

3 Распределение времени ожидания начала обслуживания

Обозначим через $V_{\text{nak}}(x)$ стационарную вероятность того, что поступившая в систему положительная заявка будет ожидать начала обслуживания в течение времени меньше x и до этого времени не попадет в бункер. Для нахождения этой вероятности необходимо учитывать, какая заявка (из накопителя или бункера) находится на приборе в момент поступления выделенной заявки в систему.

Пусть $T_n(x)$, $n \geq 1$, — вероятность того, что поступившая в систему положительная заявка и заставшая в накопителе и на приборе n других заявок, а на приборе — заявку из накопителя, поступит на прибор до момента времени x . В силу дисциплины обслуживания вероятность для выделенной заявки поступить на прибор равна $\mu_1/(\mu_1 + \lambda^-)$. Кроме того, также в силу дисциплины обслуживания, времена между соседними уходами заявок из накопителя (на прибор или в бункер) независимы и имеют экспоненциальное распределение с параметром $\mu_1 + \lambda^-$. Положим $t_n(x) = T'_n(x)$. Тогда имеет место

$$t_n(x) = \frac{\mu_1}{\mu_1 + \lambda^-} h_n(x; \mu_1 + \lambda^-).$$

Пусть $R_n(x)$, $n \geq 1$, — вероятность того, что поступившая в систему положительная заявка и заставшая в накопителе и на приборе n других заявок, а на приборе — заявку из бункера, поступит на прибор до момента времени x . Для нахождения этой вероятности необходимо учитывать, сколько заявок было вытеснено из накопителя за время x и успеет ли за время x обслужиться та заявка, которую застала на приборе выделенная заявка при поступлении. Положим $r_n(x) = R'_n(x)$. В случае $n = 1$ поступающая заявка застает накопитель пустым, а на приборе обнаруживает заявку из бункера. Поэтому $r_1(x) = \mu_2 e^{-(\mu_2 + \lambda^-)x}$. Рассмотрим случай $n \geq 2$. Выделенная заявка поступит на прибор, минуя бункер, до момента времени x , только если:

- (1) за время x будут «убиты» $n - 1$ заявок, стоящих перед выделенной заявкой в накопителе, а окончание обслуживания заявки на приборе произойдет в промежутке $(x, x + dx)$;
- (2) за время t , $0 < t < x$, будут «убиты» $j = \overline{0, n - 2}$ заявок, стоящих перед выделенной заявкой в накопителе, в момент t произойдет окончание обслуживания заявки на приборе и за оставшееся время будут обслужены или «убиты» $n - j$ заявок. Заметим, что поскольку все из $n - j$ заявок обслуживаются с одинаковой интенсивностью μ_1 , то времена между соседними

уходами заявок из накопителя (на прибор или в бункер) независимы и имеют экспоненциальное распределение с параметром $\mu_1 + \lambda^-$.

Вообще говоря, возможны и другие случаи обслуживания и «убийства» заявок (например, за время x будут «убиты» больше чем $(n - 1)$ заявок в накопителе и за это время заявка на приборе не успеет обслужиться), но во всех этих случаях вероятность поступления заявки на прибор, минуя бункер, равна нулю. В итоге при $n \geq 2$ имеем:

$$r_n(x) = \frac{\mu_2(\lambda^- x)^{n-1}}{(n-1)!} e^{-(\mu_2 + \lambda^-)x} + \frac{\mu_1 \mu_2}{\mu_1 + \lambda^-} \times \sum_{j=0}^{n-2} \int_0^x \frac{(\lambda^- t)^j}{j!} e^{-(\mu_2 + \lambda^-)t} h_{n-j-1}(x-t; \mu_1 + \lambda^-) dt.$$

Учитывая, что, приходя в свободную систему, положительная заявка сразу же попадает на прибор (время ожидания начала ее обслуживания равно нулю), и воспользовавшись формулой полной вероятности, получаем следующее выражение для плотности $v_{\text{nak}}(x) = V'_{\text{nak}}(x)$:

$$v_{\text{nak}}(x) = p_0 \delta(x) + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j,0} t_{i+1}(x) + \sum_{j=0}^{\infty} p_{0,j,1} r_1(x) + \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} p_{i,j,1} r_{i+1}(x),$$

где $\delta(x)$ — дельта-функция Дирака.

Производя традиционные преобразования, можно показать, что в терминах ПЛС $\varphi_{\text{nak}}(s) = \int_0^{\infty} e^{-sx} dV_{\text{nak}}(x)$ вероятность $V_{\text{nak}}(x)$ имеет вид:

$$\varphi_{\text{nak}}(s) = p_0 + \frac{\mu_1}{\mu_1 + \lambda^- + s} \cdot P\left(\frac{\mu_1 + \lambda^-}{\mu_1 + \lambda^- + s}, 1\right) + \frac{\mu_2}{\mu_2 + \lambda^- + s} \cdot N\left(\frac{\lambda^-}{\mu_2 + \lambda^- + s}, 1\right) + \frac{\mu_1 \mu_2}{(\mu_1 + \lambda^-)(\mu_2 + \lambda^- + s) - \lambda^- (\mu_1 + \lambda^- + s)} \times \left(N\left(\frac{\mu_1 + \lambda^-}{\mu_1 + \lambda^- + s}, 1\right) - N\left(\frac{\lambda^-}{\mu_2 + \lambda^- + s}, 1\right) \right).$$

Перейдем к вычислению стационарной вероятности $V_{\text{bun}}(x)$ того, что поступившая в систему положительная заявка будет ожидать начала обслуживания в течение времени меньше x и до этого времени она попадет в бункер.

В силу того, что заявки из бункера обслуживаются в порядке поступления, для нахождения вероятности $V_{\text{bun}}(x)$ необходимо учитывать, какая

заявка (из накопителя или бункера) находится на приборе в момент поступления выделенной заявки в систему, поскольку от этого зависит число заявок, которые будут вытеснены в бункер до момента ухода туда выделенной заявки.

Пусть $D_n(t, i, j)$, $n \geq 1$, $i \geq 0$, $j = \overline{0, n-1}$, — вероятность того, что поступающая в систему положительная заявка, заставшая в накопителе и на приборе n других заявок, а на приборе заявку из накопителя, будет «убита» до момента t , причем после ее ухода в бункер в накопителе останется i заявок и, кроме того, до ее ухода в бункер туда перейдут j из $(n-1)$ заявок, находившихся в накопителе перед выделенной заявкой. Положим $d_n(t, i, j) = D'_n(t, i, j)$. Тогда, обозначая через C_n^m число сочетаний из n элементов по m , запишем:

$$d_n(t, i, j) = \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right) h_n(t, \mu_1 + \lambda^-) \frac{(\lambda t)^i}{i!} \times e^{-\lambda t} C_{n-1}^j \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right)^j \left(\frac{\mu_1}{\mu_1 + \lambda^-} \right)^{n-1-j}.$$

Обозначим через $\tilde{d}_n(t, z, j) = \sum_{i=0}^{\infty} z^i d_n(t, i, j)$ ПФ $d_n(t, i, j)$. Тогда

$$\tilde{d}_n(t, z, j) = \lambda^- \frac{[(\mu_1 + \lambda^-)t]^{n-1}}{(n-1)!} e^{-(\mu_1 + \lambda^- + \lambda(1-z))t} \times C_{n-1}^j \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right)^j \left(\frac{\mu_1}{\mu_1 + \lambda^-} \right)^{n-1-j}.$$

Теперь найдем $D_n(t, j)$, $n \geq 1$, $j = \overline{0, n-1}$, — вероятность того, что поступающая в систему положительная заявка, заставшая в накопителе и на приборе n других заявок, а на приборе заявку из накопителя, будет «убита», до ее ухода в бункер туда перейдут j из $(n-1)$ заявок, находившихся в накопителе перед поступлением выделенной заявки, и заявки из бункера начнут поступать на прибор до момента t . В терминах ПЛС имеем:

$$\begin{aligned} \hat{d}_n(s, j) &= \int_0^{\infty} e^{-st} dD_n(t, j) = \\ &= \gamma_1(s) \int_0^{\infty} \sum_{i=0}^{\infty} e^{-st} [\gamma(s; \mu_1 + \lambda^-)]^i d_n(t, i, j) dt = \\ &= \frac{\lambda^- \gamma_1(s)}{\mu_1 + \lambda^-} C_{n-1}^j \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right)^j \left(\frac{\mu_1}{\mu_1 + \lambda^-} \right)^{n-1-j} \times \\ &\quad \times \left(\frac{\mu_1 + \lambda^-}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))} \right)^n. \end{aligned}$$

Положим

$$\psi_n(s, z) = \sum_{j=0}^{n-1} z^j \hat{d}_n(s, j) = \frac{\lambda^- \gamma_1(s)}{\mu_1 + \lambda^- z} \times \left(\frac{\mu_1 + \lambda^- z}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))} \right)^n.$$

Теперь перейдем к вычислению вероятности $F_n(t, i, j)$, $n \geq 1$, $i \geq 0$, $j = \overline{0, n-1}$, того, что поступающая в систему положительная заявка, заставшая в накопителе и на приборе n других заявок, а на приборе заявку из бункера, будет «убита» до момента t , причем после ее ухода в бункер в накопителе останется i заявок и, кроме того, до ее ухода в бункер туда перейдут j из $n-1$ заявок, находившихся в накопителе перед поступлением выделенной заявки. Положим $f_n(t, i, j) = F'_n(t, i, j)$. Рассмотрим случай, когда $n = 1$. Тогда, так как j может быть равно только 0, имеем:

$$f_1(t, i, 0) = \lambda^- \frac{(\lambda t)^i}{i!} e^{-(\lambda + \mu_2 + \lambda^-)t}.$$

Для случая $n \geq 2$ получаем, что при $j = \overline{0, n-2}$ выражение для $f_n(t, i, j)$ имеет вид:

$$\begin{aligned} f_n(t, i, j) &= \mu_2 \sum_{k=0}^j \int_0^t \frac{(\lambda^- y)^k (\lambda t)^i}{k! i!} e^{-(\lambda + \mu_2 + \lambda^-)y} \times \\ &\quad \times h_{n-1-k}(t-y, \mu_1 + \lambda^-) dy \times \\ &\quad \times C_{n-2-k}^{j-k} \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right)^{j-k+1} \left(\frac{\mu_1}{\mu_1 + \lambda^-} \right)^{n-2-j}, \end{aligned}$$

а при $j = n-1$ —

$$f_n(t, i, n-1) = \lambda^- \frac{(\lambda^- t)^{n-1} (\lambda t)^i}{(n-1)! i!} e^{-(\lambda + \mu_2 + \lambda^-)t}.$$

Обозначим через $\tilde{f}_n(t, z, j) = \sum_{i=0}^{\infty} z^i f_n(t, i, j)$ ПФ $f_n(t, i, j)$. Тогда в случае $n = 1$

$$\tilde{f}_1(t, z, 0) = \lambda^- e^{-(\mu_2 + \lambda^- + \lambda(1-z))t},$$

а в случае $n \geq 2$ при $j = \overline{0, n-2}$ имеем

$$\begin{aligned} \tilde{f}_n(t, z, j) &= \mu_2 \sum_{k=0}^j \int_0^t \frac{(\lambda^- y)^k}{k!} e^{-(\mu_2 + \lambda^- + \lambda(1-z))y} \times \\ &\quad \times h_{n-1-k}(t-y, \mu_1 + \lambda^-) dy \times \\ &\quad \times C_{n-2-k}^{j-k} \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right)^{j-k+1} \left(\frac{\mu_1}{\mu_1 + \lambda^-} \right)^{n-2-j}, \end{aligned}$$

а при $j = n - 1$ —

$$\tilde{f}_n(t, z, n - 1) = \lambda^- \frac{(\lambda^- t)^{n-1}}{(n-1)!} e^{-(\mu_2 + \lambda^- + \lambda(1-z))t}.$$

Пусть теперь $F_n(t, j)$, $n \geq 1$, $j = \overline{0, n-1}$, — вероятность того, что поступающая в систему положительная заявка, заставшая в накопителе и на приборе n других заявок, а на приборе заявку из бункера, будет «убита», до ее ухода в бункер туда перейдут j из $n - 1$ заявок, находившихся в накопителе перед поступлением выделенной заявки, и заявки из бункера начнут поступать на прибор до момента t . В терминах ПЛС для случая $n = 1$ имеем:

$$\begin{aligned} \hat{f}_1(s, 0) &= \int_0^\infty e^{-st} dF_1(t, 0) = \\ &= \int_0^\infty \sum_{i=0}^\infty e^{-st} f_1(t, i, 0) \delta_i(s) dt = \\ &= \frac{\lambda^- \alpha(s)}{s + \lambda + \mu_2 + \lambda^- - \lambda \gamma(s; \mu_2; \lambda^-)} + \\ &\quad + \frac{\lambda^- \beta(s)}{s + \lambda + \mu_2 + \lambda^- - \lambda \gamma(s; \mu_1 + \lambda^-)}. \end{aligned}$$

Для случая $n \geq 2$ при $j = \overline{0, n-2}$ получаем:

$$\begin{aligned} \hat{f}_n(s, j) &= \int_0^\infty e^{-st} dF_n(t, j) = \\ &= \gamma_1(s) \int_0^\infty \sum_{i=0}^\infty e^{-st} [\gamma(s; \mu_1 + \lambda^-)]^i f_n(t, i, j) dt = \\ &= \frac{\mu_2 \lambda^- \gamma_1(s)}{(\mu_1 + \lambda^-)(s + \mu_2 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-)))} \times \\ &\quad \times \left(\frac{\mu_1 + \lambda^-}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))} \right)^{n-1} \times \\ &\quad \times \sum_{k=0}^j \left\{ \lambda^- (s + \mu_1 + \lambda^- + \right. \\ &\quad \left. + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))) \right\} / \left\{ [s + \mu_2 + \lambda^- + \right. \\ &\quad \left. + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))] (\mu_1 + \lambda^-) \right\}^k \times \\ &\quad \times C_{n-2-k}^{j-k} \left(\frac{\lambda^-}{\mu_1 + \lambda^-} \right)^{j-k} \left(\frac{\mu_1}{\mu_1 + \lambda^-} \right)^{n-2-j}, \end{aligned}$$

а при $j = n - 1$ имеем:

$$\begin{aligned} \hat{f}_n(s, n - 1) &= \int_0^\infty e^{-st} dF_n(t, n - 1) = \\ &= \int_0^\infty \sum_{i=0}^\infty e^{-st} f_n(t, i, n - 1) \delta_i(s) dt = \\ &= \alpha(s) \left(\frac{\lambda^-}{s + \mu_2 + \lambda^- + \lambda - \lambda \gamma(s; \mu_2; \lambda^-)} \right)^n + \\ &\quad + \beta(s) \left(\frac{\lambda^-}{s + \mu_2 + \lambda^- + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)} \right)^n. \end{aligned}$$

Для удобства дальнейшей записи введем следующие обозначения:

$$\begin{aligned} p &= \frac{\lambda^-}{\mu_1 + \lambda^-}; \quad q = \frac{\mu_1}{\mu_1 + \lambda^-}; \\ \theta &= \frac{p [s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))]}{[s + \mu_2 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))]} . \end{aligned}$$

Предположим, что рассматриваемая заявка попала в бункер и перед ней в бункере оказалось n других заявок. Пусть τ — момент времени, когда накопитель и прибор окажутся пусты в первый раз. Тогда нетрудно видеть, что время с момента τ , в течение которого рассматриваемая заявка будет ожидать начала обслуживания по причине обслуживания впереди стоящих n заявок и их потомков, имеет ПЛС $(\delta_0(s))^n$.

Теперь можно выписать вид вероятности $V_{\text{bun}}(x)$ в терминах ПЛС $\varphi_{\text{bun}}(s) = \int_0^\infty e^{-sx} dV_{\text{bun}}(x)$:

$$\begin{aligned} \varphi_{\text{bun}}(s) &= \sum_{i=0}^\infty \sum_{j=0}^\infty \sum_{k=0}^i p_{i,j,0} \hat{d}_{i+1}(s, k) (\delta_0(s))^{j+k} + \\ &\quad + \sum_{i=0}^\infty \sum_{j=0}^\infty \sum_{k=0}^i p_{i,j,1} \hat{f}_{i+1}(s, k) (\delta_0(s))^{j+k}. \end{aligned}$$

Данное выражение можно упростить, подставляя найденные выше выражения для величин $\hat{d}_i(s, k)$, $\hat{f}_i(s, k)$, $\delta_0(s)$ и производя несложные, но утомительные преобразования. Окончательное выражение для $\varphi_{\text{bun}}(s)$ имеет вид:

$$\begin{aligned} \varphi_{\text{bun}}(s) &= \frac{\lambda^- \gamma_1(s)}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))} \times \\ &\quad \times P \left(\frac{\lambda^- \delta_0(s) + \mu_1}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))}, \delta_0(s) \right) + \end{aligned}$$

$$\begin{aligned}
 & + \left[\frac{\lambda^- \alpha(s)}{s + \lambda + \mu_2 + \lambda^- - \lambda \gamma(s; \mu_2; \lambda^-)} + \right. \\
 & \quad \left. + \frac{\lambda^- \beta(s)}{s + \lambda + \mu_2 + \lambda^- - \lambda \gamma(s; \mu_1 + \lambda^-)} \right] \times \\
 & \times N \left(\frac{\lambda^- \delta_0(s)}{s + \mu_2 + \lambda^- + \lambda - \lambda \gamma(s; \mu_1 + \lambda^-)}, \delta_0(s) \right) + \\
 & \quad + \mu_2 \lambda^- \gamma_1(s) / \left\{ (\mu_1 + \lambda^-) (s + \mu_2 + \lambda^- + \right. \\
 & \quad \left. + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))(p\delta_0(s) + q - \theta\delta_0(s))) \right\} \times \\
 & \times \left[N \left(\frac{(\mu_1 + \lambda^-)(p\delta_0(s) + q)}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))}, \delta_0(s) \right) - \right. \\
 & \quad \left. - N \left(\frac{\theta(\mu_1 + \lambda^-)\delta_0(s)}{s + \mu_1 + \lambda^- + \lambda(1 - \gamma(s; \mu_1 + \lambda^-))}, \delta_0(s) \right) \right].
 \end{aligned}$$

В итоге получаем, что стационарное распределение времени ожидания произвольной положительной заявкой начала обслуживания имеет ПЛС

$$\varphi(s) = \varphi_{\text{nak}}(s) + \varphi_{\text{bun}}(s).$$

Заметим, что обратить ПЛС $\varphi(s)$ в явном виде едва ли представляется возможным, но путем его дифференцирования можно получить выражения для моментов любого порядка стационарного распределения времени ожидания начала обслуживания заявки.

4 Заключение

В статье в терминах преобразования Лапласа–Стилтьеса найдено распределение времени ожидания заявкой начала обслуживания в СМО с отрицательными заявками и бункером для вытесненных заявок, в которой заявки из накопителя и бункера обслуживаются с различными интенсивностями. Предполагается, что отрицательная заявка при поступлении выбивает заявку, находящуюся на первом месте в очереди в накопителе, а после окончания обслуживания очередной заявки на прибор

поступает заявка, занимающая первое место в очереди в накопителе или, если накопитель пуст, заявка с первого места в очереди в бункере.

Все полученные аналитические результаты были проверены путем сравнения с результатами работы имитационной модели, разработанной с помощью программных средств GPSS (General Purpose Simulation System).

Литература

1. *Jinting W., Yunbo H., Zhangmin D.* A discrete-time on-off source queueing system with negative customers // *Computers Ind. Eng.*, 2011. Vol. 61. No. 4. P. 1226–1232.
2. *Klimenok V., Dudin A.* A BMAP/PH/N queue with negative customers and partial protection of service // *Comm. Statistics Simulation Comput.*, 2012. Vol. 41. Iss. 7. P. 1062–1082.
3. *Rakhee, Sharma G., Priya K.* Analysis of G-queue with unreliable server // *OPSEARCH*, 2012. DOI 10.1007/s12597-012-0117-y. P. 1–12.
4. *Pechinkin A., Razumchik R.* A method for calculating a stationary queue distribution in a queueing system with flows of ordinary and negative claims and a bunker for superseded claims // *J. Comm. Technol. Electronics*, 2012. Vol. 57. No. 8. P. 882–891.
5. *Pechinkin A., Razumchik R.* Stationary waiting time distribution in queueing system with negative customers and bunker for ousted customers under LAST–LIFO–LIFO service discipline // *J. Comm. Technol. Electronics*, 2012. Vol. 57. No. 12. P. 1331–1339.
6. *Hannah Revathy P., Muthu Ganapathi Subramanian A.* Two server (s, S) inventory system with positive service time, positive lead time, retrial customers and negative arrivals // *Int. J. Computer Appl.*, 2013. Vol. 62. No. 10. P. 9–13.
7. *Bojarovich J., Marchenko L.* An open queueing network with temporarily non-active customers and rounds modern probabilistic methods for analysis of telecommunication networks // *Comm. Computer Information Sci.*, 2013. Vol. 356. P. 33–36.
8. *Разумчик Р. В.* Система массового обслуживания с отрицательными заявками, бункером для вытесненных заявок и различными интенсивностями обслуживания // *Информатика и её применения*, 2011. Т. 5. Вып. 3. С. 39–43.

ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ ФУНКЦИИ ОБОБЩЕННОЙ КРОСС-ВАЛИДАЦИИ ПРИ ПОРОГОВОЙ ОБРАБОТКЕ ВЕЙВЛЕТ-КОЭФФИЦИЕНТОВ*

О. В. Шестаков¹

Аннотация: Исследуются асимптотические свойства функции обобщенной кросс-валидации при пороговой обработке коэффициентов вейвлет-разложения функции, удовлетворяющей некоторым условиям гладкости. Рассматривается процедура выбора порога, минимизирующей функцию обобщенной кросс-валидации. Доказывается асимптотическая нормальность функции обобщенной кросс-валидации при таком выборе порога.

Ключевые слова: вейвлеты; пороговая обработка; обобщенная кросс-валидация; адаптивный порог; несмещенная оценка риска; асимптотическая нормальность

1 Введение

Вейвлет-разложение применяется для обработки сигналов и изображений в самых разнообразных областях, включая геофизику, физику плазмы, вычислительную томографию, компьютерную графику и т. д. Одна из основных задач, для решения которой используется вейвлет-разложение, — это подавление шума. При этом наиболее популярным методом является пороговая обработка вейвлет-коэффициентов, которая обнуляет коэффициенты, не превышающие заданного порога. Порог можно выбирать различными способами, исходя из постановки задачи и целей обработки (см., например, [1–4]). Наличие шума неизбежно приводит к погрешностям в оцениваемом сигнале/изображении. Свойства оценки таких погрешностей (риска) исследовались в работах [1–10]. В частности, в работах [7–10] показано, что при выполнении некоторых условий гладкости и выборе «универсального» порога или адаптивного порога, минимизирующего несмещенную оценку риска, оценка риска является состоятельной и асимптотически нормальной. Пороги, используемые в указанных работах, пропорциональны дисперсии шума. Однако во многих практических ситуациях уровень шума неизвестен. В работах [7–10] исследуется ситуация, в которой вместо дисперсии шума подставляется ее оценка. В качестве альтернативы в работах [11, 12] предложено использовать для выбора порога процедуру минимизации функции обобщенной кросс-валидации. В [11] показывается, что порог, выбранный на основе этой проце-

дуры, является в некотором смысле асимптотически оптимальным. В данной работе доказывается асимптотическая нормальность функции обобщенной кросс-валидации при выборе такого порога. Этот факт служит дополнительным обоснованием для выбора порога, минимизирующего указанную функцию.

2 Пороговая обработка вейвлет-коэффициентов и оценка риска

При использовании вейвлет-разложения функция $f \in L^2(\mathbf{R})$, описывающая сигнал, представляется в виде ряда из сдвигов и растяжений некоторой вейвлет-функции ψ :

$$f = \sum_{j,k \in \mathbf{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}, \quad (1)$$

где $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$ (семейство $\{\psi_{j,k}\}_{j,k \in \mathbf{Z}}$ образует ортонормированный базис в $L^2(\mathbf{R})$). Индекс j в (1) называется масштабом, а индекс k — сдвигом. Функция ψ должна удовлетворять определенным требованиям [13], однако ее можно выбрать таким образом, чтобы она обладала некоторыми полезными свойствами, например была дифференцируемой нужное число раз и имела заданное число M нулевых моментов [13], т. е.

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0, \quad k = 0, \dots, M - 1.$$

* Работа выполнена при финансовой поддержке РФФИ (гранты 11-01-00515 и 11-01-12026-офи-м).

¹ Московский государственный университет им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, oshestakov@cs.msu.ru

В дальнейшем будут рассматриваться функции сигнала $f \in L^2(\mathbf{R})$ на конечном отрезке $[a, b]$, равномерно регулярные по Липшицу с некоторым параметром $\gamma > 0$, т.е. такие функции, для которых существует константа $L > 0$ и полином P_γ степени $n = \lfloor \gamma \rfloor$ такой, что для любого $y \in [a, b]$ и любого $x \in \mathbf{R}$

$$|f(x) - P_\gamma(x)| \leq L|x - y|^\gamma.$$

Для таких функций f известно [14], что если вейвлет-функция M раз непрерывно дифференцируема ($M \geq \gamma$), имеет M нулевых моментов и быстро убывает на бесконечности вместе со своими производными, т.е. для всех $0 \leq k \leq M$ и любого $m \in N$ найдется константа C_m такая, что при всех $x \in \mathbf{R}$

$$|\psi^{(k)}(x)| \leq \frac{C_m}{1 + |x|^m},$$

то найдется такая константа $A > 0$, что

$$\langle f, \psi_{j,k} \rangle \leq \frac{A}{2^{j(\gamma+1/2)}}. \quad (2)$$

На практике функции сигнала всегда заданы в дискретных отсчетах на конечном отрезке. Не ограничивая общности, будем считать, что это отрезок $[0, 1]$ и функция f задана в точках i/N ($i = 1, \dots, N$, где $N = 2^J$ для некоторого J): $f_i = f(i/N)$. Дискретное вейвлет-преобразование представляет собой умножение вектора значений функции f (обозначим его через \vec{f}) на ортогональную матрицу W , определяемую вейвлет-функцией ψ : $\vec{f}^W = W\vec{f}$ [14]. При этом если перейти к двойному индексу (j, k) , как в непрерывном случае, то дискретные вейвлет-коэффициенты будут связаны с непрерывными следующим образом: $f_{j,k}^W \approx \sqrt{N} \langle f, \psi_{j,k} \rangle$ (см., например, [2] или [15]). Это приближение тем точнее, чем больше N . Не будем обсуждать методы борьбы с краевыми эффектами, связанными с использованием вейвлет-разложения на конечном отрезке. Познакомиться с этими методами можно, например, в [16]. В дальнейшем для удобства будем нумеровать дискретные вейвлет-коэффициенты так же, как отсчеты функции f , одним индексом i вместо двойного индекса (j, k) .

В реальных наблюдениях всегда присутствует шум. Будем рассматривать следующую модель:

$$Y_i = f_i + z_i, \quad i = 1, \dots, N,$$

где z_i — независимые случайные величины, имеющие нормальное распределение с нулевым средним и дисперсией σ^2 . Тогда в силу ортогональности матрицы W для дискретных вейвлет-коэффициентов применим следующую модель:

$$Y_i^W = f_i^W + z_i^W, \quad i = 1, \dots, N,$$

где z_i^W также независимы и нормально распределены с нулевым средним и дисперсией σ^2 , а f_i^W равны соответствующим непрерывным вейвлет-коэффициентам, умноженным на \sqrt{N} .

Для подавления шума часто используется процедура пороговой обработки вейвлет-коэффициентов. Смысл ее заключается в удалении достаточно маленьких коэффициентов, которые считаются шумом. Если функция сигнала достаточно гладкая, то используется так называемая мягкая пороговая обработка с порогом T . К каждому вейвлет-коэффициенту применяется функция

$$\rho_T(x) = \text{sgn}(x) (|x| - T)_+,$$

т.е. при такой пороговой обработке коэффициенты, которые по модулю меньше порога T , обнуляются, а абсолютные величины остальных коэффициентов уменьшаются на величину порога. Погрешность (или риск) мягкой пороговой обработки определяется следующим образом:

$$R_N(f, T) = \sum_{i=1}^N \mathbb{E} (f_i^W - \rho_T(Y_i^W))^2. \quad (3)$$

В выражении (3) присутствуют неизвестные величины f_i^W , поэтому вычислить значение $R_N(f, T)$ нельзя. Однако его можно оценить. В каждом слагаемом если $|Y_i^W| > T$, то вклад этого слагаемого в риск составляет $\sigma^2 + T^2$, а если $|Y_i^W| \leq T$, то вклад составляет $(f_i^W)^2$. Поскольку $\mathbb{E}(Y_i^W)^2 = \sigma^2 + (f_i^W)^2$, то величину $(f_i^W)^2$ можно оценить разностью $(Y_i^W)^2 - \sigma^2$.

Таким образом, в качестве оценки риска можно использовать следующую величину:

$$\hat{R}_N(f, T) = \sum_{i=1}^N F[(Y_i^W)^2, T], \quad (4)$$

где

$$F[x, T] = (x - \sigma^2) \mathbb{1}(|x| \leq T^2) + (\sigma^2 + T^2) \mathbb{1}(|x| > T^2).$$

Если дисперсия шума известна, то оценка риска оказывается несмещенной [14].

В работах [1, 3] было предложено использовать порог $T_U = \sigma \sqrt{2 \ln N}$. Было показано, что при таком пороге риск близок к минимальному [1]. Этот порог получил название «универсальный». В работе [2] рассматривается метод пороговой обработки с названием SureShrink (от Stein Unbiased Risk Estimate — несмещенная оценка риска Стейна), заключающийся в минимизации оценки риска (4) на множестве $T \in [0, T_U]$ (исследования, проведенные

в работах [2, 4], показывают, что можно не рассматривать $T > T_U$, т. е. порог выбирается следующим образом:

$$\widehat{R}_N(f, T_{SURE}) = \min_{T \in [0, T_U]} \widehat{R}_N(f, T).$$

Этот порог имитирует теоретический «идеальный» порог T_{\min} , минимизирующий риск:

$$R_N(f, T_{\min}) = \min_{T \in [0, T_U]} R_N(f, T). \quad (5)$$

В то время как значение порога T_{\min} найти нельзя, если неизвестны незашумленные значения f_i (можно лишь в некоторых случаях выяснить его асимптотическое поведение), алгоритм поиска порога T_{SURE} очень прост и его описание можно найти в [14] или [17]. Порог T_{SURE} является адаптивным, поскольку использует только наблюдаемые данные и «автоматически адаптируется» к гладкости сигнала.

В работе [10] доказываются следующие утверждения об асимптотической нормальности оценки риска.

Теорема 1. Пусть $f \in L^2(\mathbf{R})$ задана на отрезке $[0, 1]$ и является равномерно регулярной по Липшицу с параметром $\gamma = 1/2 + \alpha$ ($\alpha > 0$). Тогда имеет место сходимость по распределению

$$P \left(\frac{\widehat{R}_N(f, T_{\min}) - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) \Rightarrow \Phi(x),$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Теорема 2. Пусть $f \in L^2(\mathbf{R})$ задана на отрезке $[0, 1]$ и является равномерно регулярной по Липшицу с параметром $\gamma = 1/2 + \alpha$ ($\alpha > 0$). Тогда имеет место сходимость по распределению

$$P \left(\frac{\widehat{R}_N(f, T_{SURE}) - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) \Rightarrow \Phi(x),$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

3 Обобщенная кросс-валидация

Далее для удобства будем обозначать Y_i^W через X_i , а f_i^W — через a_i .

Цель процедуры обобщенной кросс-валидации заключается в минимизации ошибки без использования ненаблюдаемых истинных значений функции сигнала и точного значения дисперсии шума.

Для этого строится следующая функция обобщенной кросс-валидации, которая зависит только от наблюдаемых данных и порога T [11]:

$$\widehat{G}_N(f, T) = \frac{\sum_{i=1}^N (X_i - \rho_T(X_i))^2}{\mu_T^2},$$

где

$$\mu_T = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(|X_i| \leq T).$$

Для того чтобы $\widehat{G}_N(f, T)$ не принимала бесконечного значения, при $\mu_T = 0$ полагают $\widehat{G}_N(f, T) = 0$. Выбор порога, основанного на процедуре обобщенной кросс-валидации, заключается в минимизации функции $\widehat{G}_N(f, T)$ на некотором множестве $T \in [T_0, T_U]$:

$$\widehat{G}_N(f, T_{GCV}) = \min_{T \in [T_0, T_U]} \widehat{G}_N(f, T),$$

где T_0 — достаточно большое, но не зависящее от N число. Выбор T_0 обусловлен тем, что при стремлении T к нулю $\widehat{G}_N(f, T)$ также может стремиться к нулю [6, 11, 12], т. е. к своему абсолютному минимуму. В то же время в [6] показывается, что «разумный» порог должен возрастать с увеличением N (далее будет показано, что величина границы T_0 не имеет большого значения). При этом, как уже отмечалось выше, можно не рассматривать $T > T_U$ (подробнее с методом поиска T_{GCV} и поведением функций $\widehat{G}_N(f, T)$ и $E\widehat{G}_N(f, T)$ при значениях T , близких к нулю, можно познакомиться в [6]). Порог T_{GCV} имитирует теоретический порог T^* :

$$E\widehat{G}_N(f, T^*) = \min_{T \in [T_0, T_U]} E\widehat{G}_N(f, T).$$

В работе [11] показано, что

$$\frac{R_N(f, T^*)}{R_N(f, T_{\min})} \downarrow 1 \text{ при } N \rightarrow \infty.$$

Это утверждение служит некоторым обоснованием для выбора порога T_{GCV} (особенно в случаях, когда дисперсия шума неизвестна). В следующем разделе будет доказана асимптотическая нормальность функции $\widehat{G}_N(f, T_{GCV})$, что является дополнительным доводом для выбора такого порога.

4 Асимптотическая нормальность функции обобщенной кросс-валидации

Сначала докажем вспомогательную теорему об асимптотической нормальности функции

$\widehat{G}_N(f, T_{\min})$ при выборе по критерию (5) «идеального» порога T_{\min} .

Теорема 3. Пусть $f \in L^2(\mathbf{R})$ задана на отрезке $[0, 1]$ и является равномерно регулярной по Липшицу с параметром $\gamma = 1/2 + \alpha$ ($\alpha > 0$). Тогда имеет место сходимость по распределению

$$\mathbb{P} \left(\frac{\widehat{G}_N(f, T_{\min}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) \Rightarrow \Phi(x), \quad (6)$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Доказательство. Для $\widehat{G}_N(f, T_{\min})$ справедливо следующее представление [12]:

$$\begin{aligned} \widehat{G}_N(f, T_{\min}) - N\sigma^2 &= \\ &= \frac{1}{\mu_{T_{\min}}^2} \left[\widehat{R}_N(f, T_{\min}) - N\sigma^2(1 - \mu_{T_{\min}})^2 \right] \end{aligned} \quad (7)$$

(верное не только для T_{\min} , но и для любого $T > 0$), из которого в силу теоремы 1 и следует утверждение (6), поскольку $\mu_{T_{\min}} \xrightarrow{P} 1$ и $\sqrt{N}(1 - \mu_{T_{\min}})^2 \xrightarrow{P} 0$ при $N \rightarrow \infty$. Теорема доказана.

Докажем теперь асимптотическую нормальность $\widehat{G}_N(f, T_{GCV})$ при выборе адаптивного порога T_{GCV} по критерию [6].

Теорема 4. Пусть $f \in L^2(\mathbf{R})$ задана на отрезке $[0, 1]$ и является равномерно регулярной по Липшицу с параметром $\gamma = 1/2 + \alpha$ ($\alpha > 0$). Тогда имеет место сходимость по распределению

$$\mathbb{P} \left(\frac{\widehat{G}_N(f, T_{GCV}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) \Rightarrow \Phi(x), \quad (8)$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Доказательство. Запишем выражение из левой части (8) в виде

$$\begin{aligned} \frac{\widehat{G}_N(f, T_{GCV}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} &= \\ &= \frac{\widehat{G}_N(f, T_{\min}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} + \\ &+ \frac{\widehat{G}_N(f, T_{GCV}) - \widehat{G}_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} = S_1 + S_2. \end{aligned}$$

В силу предыдущей теоремы первое слагаемое сходится по распределению к стандартному нормальному закону. Покажем, что второе слагаемое стремится по вероятности к нулю.

Выберем $T_1 = \sigma\sqrt{\lambda \ln N}$ с некоторым $0 < \lambda < 1$ и $T_2 = \sigma\sqrt{\beta \ln N}$ с $1 < \beta < \min(3/2, 4\gamma/(2\gamma + 1))$. Для некоторого $\varkappa > 0$ (ограничения на которое будут наложены ниже) справедливо

$$\begin{aligned} \mathbb{P}(S_2 > N^{-\varkappa/2}) &\leq \mathbb{P}(T_{GCV} \in [T_0, T_1]) + \\ &+ \mathbb{P}(T_{GCV} \in [T_1, T_2]) + \\ &+ \mathbb{P}(S_2 > N^{-\varkappa/2}, T_{GCV} \in [T_2, T_U]) = P_1 + P_2 + P_3. \end{aligned}$$

Обозначим

$$\widehat{H}_N(f, T) = \widehat{G}_N(f, T_{\min}) - \widehat{G}_N(f, T);$$

$$\begin{aligned} \widehat{H}_N^*(f, T) &= \frac{\sum_{i=1}^N (X_i - \rho_{T_{\min}}(X_i))^2}{(\mathbf{E}\mu_{T_{\min}})^2} - \\ &- \frac{\sum_{i=1}^N (X_i - \rho_T(X_i))^2}{(\mathbf{E}\mu_T)^2}. \end{aligned}$$

Поскольку $\widehat{G}_N(f, T_{GCV}) = \min_{T \in [T_0, T_U]} \widehat{G}_N(f, T)$, для P_1 имеем

$$\begin{aligned} P_1 &= \mathbb{P}(T_{GCV} \in [T_0, T_1]) \leq \\ &\leq \mathbb{P} \left(\sup_{T \in [T_0, T_1]} \widehat{H}_N(f, T) \geq 0 \right) \leq \\ &\leq \mathbb{P} \left(\sup_{T \in [T_0, T_1]} \left| \widehat{H}_N^*(f, T) - \mathbf{E}\widehat{H}_N^*(f, T) \right| \geq \right. \\ &\quad \geq \inf_{T \in [T_0, T_1]} -\mathbf{E}\widehat{H}_N^*(f, T) - \\ &\quad \left. - \sup_{T \in [T_0, T_1]} \left| \widehat{H}_N(f, T) - \widehat{H}_N^*(f, T) \right| \right) \leq \\ &\leq \mathbb{P} \left(\sup_{T \in [T_0, T_1]} \left| \widehat{H}_N^*(f, T) - \mathbf{E}\widehat{H}_N^*(f, T) \right| \geq \right. \\ &\quad \geq \inf_{T \in [T_0, T_1]} -\mathbf{E}\widehat{H}_N^*(f, T) - r'_N \left. \right) + \\ &+ \mathbb{P} \left(\sup_{T \in [T_0, T_1]} \left| \widehat{H}_N(f, T) - \widehat{H}_N^*(f, T) \right| > r'_N \right). \quad (9) \end{aligned}$$

Заметим, что $(X_i - \rho_T(X_i))^2 \leq T^2$. Следовательно,

$$\begin{aligned} \mathbb{P} \left(\sup_{T \in [T_0, T_1]} \left| \widehat{H}_N(f, T) - \widehat{H}_N^*(f, T) \right| > r'_N \right) &\leq \\ &\leq \mathbb{P} \left(T_1 N \sup_{T \in [T_0, T_1]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbf{E}\mu_T)^2} \right| > \frac{r'_N}{2} \right) + \\ &+ \mathbb{P} \left(T_{\min} N \left| \frac{1}{\mu_{T_{\min}}^2} - \frac{1}{(\mathbf{E}\mu_{T_{\min}})^2} \right| > \frac{r'_N}{2} \right). \quad (10) \end{aligned}$$

Для первого слагаемого начиная с некоторого N

$$\begin{aligned} & \mathbb{P} \left(T_1 N \sup_{T \in [T_0, T_1]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E}\mu_T)^2} \right| > \frac{r'_N}{2} \right) \leq \\ & \leq \mathbb{P}(\mu_{T_0} = 0) + \\ & + \mathbb{P} \left(T_1 N \sup_{T \in [T_0, T_1]} |\mu_T - \mathbb{E}\mu_T| > C'_r r'_N \right) \end{aligned} \quad (11)$$

для некоторой положительной константы C'_r . Выберем $r'_N = N^{1-\lambda/2}/(\ln N)^2$. Применяя экспоненциальное неравенство для эмпирических случайных процессов из работы [18], для некоторых положительных констант K_1^* и C_1^* имеем:

$$\begin{aligned} & \mathbb{P} \left(T_1 N \sup_{T \in [T_0, T_1]} |\mu_T - \mathbb{E}\mu_T| > C'_r r'_N \right) \leq \\ & \leq K_1^* \exp \left[-\frac{C_1^* N^{1-\lambda}}{(\ln N)^5} \right]. \end{aligned}$$

Событие $\{\mu_{T_0} = 0\}$ означает, что ни одна величина X_i не оказалась меньше порога T_0 . Следовательно, начиная с некоторого N

$$\mathbb{P}(\mu_{T_0} = 0) \leq \frac{N^{-C_{\mu_{T_0}} N}}{(\ln N)^{C_{\mu_{T_0}} N}}$$

с некоторой константой $C_{\mu_{T_0}} > 0$. Таким образом, начиная с некоторого N для некоторых положительных констант \tilde{K}_1 и \tilde{C}_1 ,

$$\begin{aligned} & \mathbb{P} \left(T_1 N \sup_{T \in [T_0, T_1]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E}\mu_T)^2} \right| > \frac{r'_N}{2} \right) \leq \\ & \leq \tilde{K}_1 \exp \left[-\frac{\tilde{C}_1 N^{1-\lambda}}{(\ln N)^5} \right]. \end{aligned} \quad (12)$$

Второе слагаемое в (10) оценивается аналогично с применением неравенства Бернштейна [19] и учетом того, что при выполнении условий теоремы $2\gamma/(2\gamma+1) > 1/2$ и T_{\min} ведет себя асимптотически как [6]

$$T_{\min} \sim \sigma \sqrt{\frac{2\gamma}{2\gamma+1}} \sqrt{2 \ln N}. \quad (13)$$

Оценим теперь первое слагаемое в (9). В работе [6, 11] показано, что

$$\begin{aligned} & (X_i - \rho_T(X_i))^2 = \sigma^2 + \mathbb{E}(a_i - \rho_T(X_i))^2 - \\ & - 2\sigma^2 [1 - \Phi_\sigma(T - a_i) + \Phi_\sigma(-T - a_i)]; \end{aligned} \quad (14)$$

$$\begin{aligned} & \mathbb{E}(a_i - \rho_T(X_i))^2 = \sigma^2 + T^2 + \\ & + (a_i^2 - T^2 - \sigma^2) [\Phi_\sigma(T - a_i) - \Phi_\sigma(-T - a_i)] - \\ & - \sigma^2 [(T - a_i)\phi_\sigma(T + a_i) + (T + a_i)\phi_\sigma(T - a_i)], \end{aligned} \quad (15)$$

где $\phi_\sigma(y)$ и $\Phi_\sigma(y)$ — плотность и функция распределения нормального закона с нулевым средним и дисперсией σ^2 . Поскольку выполнено (2) и для плотности $\phi(y)$ и функции распределения $\Phi(y)$ стандартного нормального закона справедливо [20]

$$\frac{\phi(y)}{y} > 1 - \Phi(y) > \phi(y) \left(\frac{1}{y} - \frac{1}{y^3} + \frac{3}{y^5} - \frac{15}{y^7} \right)$$

при $y > 0$, используя в (14) и (15) формулу Лагранжа и учитывая (13), можно показать, что если T_0 достаточно велико, то найдется такая константа $C'_h > 0$, что

$$\inf_{T \in [T_0, T_1]} -\mathbb{E}\hat{H}_N^*(f, T) \geq C'_h \frac{N^{1-\lambda/2}}{(\ln N)^{3/2}}. \quad (16)$$

Пусть

$$h^*(x, T) = \frac{(x - \rho_{T_{\min}}(x))^2}{(\mathbb{E}\mu_{T_{\min}})^2} - \frac{(x - \rho_T(x))^2}{(\mathbb{E}\mu_T)^2}.$$

Таким образом,

$$\hat{H}_N^*(f, T) = \sum_{i=1}^N h^*(X_i, T).$$

Обозначим через \mathbf{H} класс функций $h^*(x, T)$, индексированный параметром $T \in [T_0, T_U]$. Далее определим

$$\begin{aligned} & \mathbf{H}(\mathbf{H}, \varepsilon, L^2(\mathbb{P})) = \min\{k : \text{существуют} \\ & h_1, \dots, h_k \in L^2(\mathbb{P}) \text{ такие, что} \\ & \min_{i \leq k} \|h - h_i\|_{L^2(\mathbb{P})} \leq \varepsilon \text{ для всех } h \in \mathbf{H}\}; \end{aligned}$$

$$\begin{aligned} & \mathbf{H}_{\square}(\mathbf{H}, \varepsilon, L^2(\mathbb{P})) = \min\{k : \text{существуют} \\ & h_1^u, h_1^l, \dots, h_k^u, h_k^l \in L^2(\mathbb{P}) \text{ такие, что} \\ & \text{для любой } h \in \mathbf{H} \text{ найдется } i \text{ такое, что} \\ & h_i^l \leq h \leq h_i^u \text{ и } \|h_i^u - h_i^l\|_{L^2(\mathbb{P})} \leq \varepsilon\}. \end{aligned}$$

Функции $H(\mathbf{H}, \varepsilon, L^2(\mathbb{P})) = \ln N(\mathbf{H}, \varepsilon, L^2(\mathbb{P}))$ и $H_{\square}(\mathbf{H}, \varepsilon, L^2(\mathbb{P})) = \ln N_{\square}(\mathbf{H}, \varepsilon, L^2(\mathbb{P}))$ называются энтропией и энтропией с брэкетами пространства \mathbf{H} [21].

Для всех $T \in [T_0, T_U]$ выполнены неравенства $\sup_{x \in \mathbb{R}} |h^*(x, T)| \leq T_U^2/\mathbb{E}(\mu_{T_0})^2$ и $Dh^*(Z, T) \leq T_U^4/(\mathbb{E}\mu_{T_0})^4$ для произвольной случайной величины Z . Можно показать [22], что для некоторых констант K и K_{\square}

$$\sup_{\mathbb{P}} H(\mathbf{H}, \varepsilon, L^2(\mathbb{P})) \leq K \ln \left(\frac{T_U}{\varepsilon} \right); \quad (17)$$

$$\sup_{\mathbb{P}} H_{\square}(\mathbf{H}, \varepsilon, L^2(\mathbb{P})) \leq K_{\square} \ln \left(\frac{T_U}{\varepsilon} \right), \quad (18)$$

где супремум берется по всем вероятностным распределениям P .

Применяя неравенство для вероятности уклонения эмпирического процесса (см., например, [18, 21, 23]) с учетом (16)–(18), можно показать, что для некоторых положительных констант K_1 и C_1 выполнено

$$\begin{aligned} P \left(\sup_{T \in [T_0, T_1]} \left| \widehat{H}_N^*(f, T) - \mathbb{E} \widehat{H}_N^*(f, T) \right| \geq \right. \\ \left. \geq \inf_{T \in [T_0, T_1]} -\mathbb{E} \widehat{H}_N^*(f, T) - r'_N \right) \leq \\ \leq K_1 \exp \left[-\frac{C_1 N^{1-\lambda}}{(\ln N)^5} \right]. \quad (19) \end{aligned}$$

Оценим теперь P_2 . Поступая так же, как при оценке P_1 , имеем:

$$\begin{aligned} P_2 = P(T_{GCV} \in [T_1, T_2]) \leq \\ \leq P \left(\sup_{T \in [T_1, T_2]} \widehat{H}_N(f, T) \geq 0 \right) \leq \\ \leq P \left(\sup_{T \in [T_1, T_2]} \left| \widehat{H}_N^*(f, T) - \mathbb{E} \widehat{H}_N^*(f, T) \right| \geq \right. \\ \left. \geq \inf_{T \in [T_1, T_2]} -\mathbb{E} \widehat{H}_N^*(f, T) - r''_N \right) + \\ + P \left(\sup_{T \in [T_1, T_2]} \left| \widehat{H}_N(f, T) - \widehat{H}_N^*(f, T) \right| > r''_N \right). \quad (20) \end{aligned}$$

Далее

$$\begin{aligned} P \left(\sup_{T \in [T_1, T_2]} \left| \widehat{H}_N(f, T) - \widehat{H}_N^*(f, T) \right| > r''_N \right) \leq \\ \leq P \left(T_2 N \sup_{T \in [T_1, T_2]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E} \mu_T)^2} \right| > \frac{r''_N}{2} \right) + \\ + P \left(T_{\min} N \left| \frac{1}{\mu_{T_{\min}}^2} - \frac{1}{(\mathbb{E} \mu_{T_{\min}})^2} \right| > \frac{r''_N}{2} \right). \quad (21) \end{aligned}$$

Для первого слагаемого начиная с некоторого N

$$\begin{aligned} P \left(T_2 N \sup_{T \in [T_1, T_2]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E} \mu_T)^2} \right| > \frac{r''_N}{2} \right) \leq \\ \leq P(\mu_{T_1} = 0) + \\ + P \left(T_2 N \sup_{T \in [T_1, T_2]} |\mu_T - \mathbb{E} \mu_T| > C_r'' r''_N \right) \quad (22) \end{aligned}$$

для некоторой положительной константы C_r'' . Далее

$$\begin{aligned} d_N = \sup_{T \in [T_1, T_2]} \frac{1}{N} \sum_{i=1}^N D \mathbb{1}(|X_i| \leq T) \leq \\ \leq C_d \sup_{T \in [T_1, T_2]} \frac{\phi(T)}{T} \leq C_d \frac{N^{-\lambda/2}}{(\ln N)^{1/2}}, \quad (23) \end{aligned}$$

где $\phi(y)$ — плотность стандартного нормального закона. Выберем $r''_N = N^{-\beta/2}/(\ln N)^2$. Применяя экспоненциальное неравенство для эмпирических случайных процессов из работы [18] с учетом (23), для некоторых положительных констант K_2^* и C_2^* имеем:

$$\begin{aligned} P \left(T_2 N \sup_{T \in [T_1, T_2]} |\mu_T - \mathbb{E} \mu_T| > C_r'' r''_N \right) \leq \\ \leq K_2^* \exp \left[-\frac{C_2^* N^{1-\beta+\lambda/2}}{(\ln N)^{9/2}} \right]. \end{aligned}$$

Первое слагаемое в (22) оценивается аналогично первому слагаемому в (11):

$$P(\mu_{T_1} = 0) \leq \frac{N^{-C_{\mu_{T_1}} N}}{(\ln N)^{C_{\mu_{T_1}} N}}$$

с некоторой константой $C_{\mu_{T_1}} > 0$. Таким образом, начиная с некоторого N для некоторых положительных констант \tilde{K}_2 и \tilde{C}_2 ,

$$\begin{aligned} P \left(T_2 N \sup_{T \in [T_1, T_2]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E} \mu_T)^2} \right| > \frac{r''_N}{2} \right) \leq \\ \leq \tilde{K}_2 \exp \left[-\frac{\tilde{C}_2 N^{1-\beta+\lambda/2}}{(\ln N)^{9/2}} \right]. \quad (24) \end{aligned}$$

Второе слагаемое в (21) оценивается аналогично второму слагаемому в (10).

Для оценки первого слагаемого в (20), так же как для (16), можно показать, что найдется такая константа $C_h'' > 0$, что

$$\inf_{T \in [T_1, T_2]} -\mathbb{E} \widehat{H}_N^*(f, T) \geq C_h'' \frac{N^{1-\beta/2}}{(\ln N)^{3/2}}. \quad (25)$$

Кроме того,

$$\begin{aligned} v_N = \sup_{T \in [T_1, T_2]} \frac{1}{N} \sum_{i=1}^N D h^*(X_i, T) \leq \sup_{T \in [T_1, T_2]} \frac{2}{N} \times \\ \times \sum_{i=1}^N \left[\left\{ \mathbb{E} [(T^2 - X_i^2)^2 \mathbb{1}(T < |X_i| \leq T_{\min})] + \right. \right. \\ \left. \left. + (T^2 - T_{\min}^2)^2 \mathbb{E} \mathbb{1}(|X_i| > T_{\min}) \right\} / (\mathbb{E} \mu_{T_{\min}})^4 + \right. \\ \left. + T_{\min}^4 \frac{(\mathbb{E} \mu_{T_{\min}} + \mathbb{E} \mu_T)^2 (\mathbb{E} \mu_{T_{\min}} - \mathbb{E} \mu_T)^2}{(\mathbb{E} \mu_{T_{\min}} \mathbb{E} \mu_T)^4} \right] \leq \\ \leq \tilde{C}_v T_{\min}^4 \sup_{T \in [T_1, T_2]} \frac{\phi(T)}{T} \leq C_v N^{-\lambda/2} (\ln N)^{3/2} \quad (26) \end{aligned}$$

для некоторых положительных констант \tilde{C}_v и C_v .

Применяя неравенство для вероятности уклонения эмпирического процесса (см., например, [18, 21, 23]) с учетом (17), (18), (25) и (26), можно показать, что для некоторых положительных констант K_2 и C_2 выполнено

$$\begin{aligned} & \mathbb{P} \left(\sup_{T \in [T_1, T_2]} \left| \widehat{H}_N^*(f, T) - \mathbb{E} \widehat{H}_N^*(f, T) \right| \geq \right. \\ & \quad \left. \geq \inf_{T \in [T_1, T_2]} -\mathbb{E} \widehat{H}_N^*(f, T) - r_N'' \right) \leq \\ & \quad \leq K_2 \exp \left[-\frac{C_2 N^{1-\beta+\lambda/2}}{(\ln N)^{9/2}} \right]. \quad (27) \end{aligned}$$

Можно выбрать $0 < \lambda < 1$ и $1 < \beta < \min(3/2, 4\gamma/(2\gamma + 1))$ такими, что $1 - \beta + \lambda/2 > 0$. Тогда правая часть (27) будет стремиться к нулю.

Наконец, оценим P_3 . Обозначим

$$h(x, T) = \frac{(x - \rho_{T_{\min}}(x))^2}{\mu_{T_{\min}}^2} - \frac{(x - \rho_T(x))^2}{\mu_T^2}.$$

Имеем

$$\begin{aligned} P_3 & \leq \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - \mathbb{E} h^*(X_i, T)] \right| + \right. \\ & \quad \left. + \sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - h(X_i, T)] \right| + \right. \\ & \quad \left. + \sup_{T \in [T_2, T_U]} \sum_{i=1}^N \mathbb{E} h^*(X_i, T) \geq \sigma^2 \sqrt{2} N^{1/2-\varkappa/2} \right) \leq \\ & \leq \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - \mathbb{E} h^*(X_i, T)] \right| \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa/2} \right) + \\ & \quad + \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - h(X_i, T)] \right| \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa/2} \right) + \\ & \quad + \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \sum_{i=1}^N \mathbb{E} h^*(X_i, T) \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa/2} \right). \quad (28) \end{aligned}$$

Поскольку найдется такая константа $C_E > 0$, что

$$\sup_{T \in [T_2, T_U]} \sum_{i=1}^N \mathbb{E} h^*(X_i, T) \leq C_E \frac{N^{1-\beta/2}}{(\ln N)^{1/2}},$$

начиная с некоторого N третья вероятность в (28) равна нулю, если $\varkappa < \beta - 1$. Для второго слагаемого найдется такая положительная константа C_r^* , что

$$\begin{aligned} & \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - h(X_i, T)] \right| \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa/2} \right) \leq \\ & \leq \mathbb{P} \left(T_U N \sup_{T \in [T_2, T_U]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E} \mu_T)^2} \right| > \right. \\ & \quad \left. > C_r^* N^{-1/2-\varkappa/2} \right) + \\ & \quad + \mathbb{P} \left(T_{\min} N \left| \frac{1}{\mu_{T_{\min}}^2} - \frac{1}{(\mathbb{E} \mu_{T_{\min}})^2} \right| > C_r^* N^{-1/2-\varkappa/2} \right). \end{aligned}$$

Поступая так же, как при оценке вероятностей в (21), имеем:

$$\begin{aligned} & \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - h(X_i, T)] \right| \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa/2} \right) \leq \tilde{K}_3 \exp \left[-\frac{\tilde{C}_3 N^{\beta/2-\varkappa}}{(\ln N)^{1/2}} \right] \quad (29) \end{aligned}$$

с некоторыми положительными константами \tilde{K}_3 и \tilde{C}_3 . Применяя к первому слагаемому в (28) неравенство для вероятности уклонения эмпирического процесса с учетом того, что для некоторой константы $C_{\tilde{v}}$

$$\begin{aligned} \tilde{v}_N & = \sup_{T \in [T_2, T_U]} \frac{1}{N} \sum_{i=1}^N D h^*(X_i, T) \leq \\ & \leq C_{\tilde{v}} N^{-\beta/2} (\ln N)^{3/2}, \end{aligned}$$

имеем

$$\begin{aligned} & \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - \mathbb{E} h^*(X_i, T)] \right| \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa/2} \right) \leq \\ & \leq \mathbb{P} \left(\sup_{T \in [T_2, T_U]} \left| \sum_{i=1}^N [h^*(X_i, T) - \mathbb{E} h^*(X_i, T)] \right| \geq \right. \\ & \quad \left. \geq \frac{\sigma^2 \sqrt{2}}{3} N^{1/2-\varkappa'/2} \right) \leq K_3 \exp \left[-\frac{C_3 N^{\beta/2-\varkappa'}}{(\ln N)^{3/2}} \right] \quad (30) \end{aligned}$$

для некоторых констант $K_3 > 0$ и $C_3 > 0$ и $\beta - 1 < \varkappa' < \beta/2$ (поскольку $\varkappa < \beta - 1$).

Объединяя (10), (12), (19), (21), (24) и (27)–(30), получаем, что S_2 стремится к нулю по вероятности. Теорема доказана.

Замечание 1. Из доказательства теоремы видно, что вероятность попадания порога T_{GCV} в множество $[T_0, T_2]$ мала. Поэтому в практических ситуациях можно искать минимум функции $\widehat{G}_N(f, T)$ сразу на множестве $[T_2, T_U]$. Следовательно, величина T_0 не имеет большого значения.

Следствие. При выполнении условий теоремы 4 справедливо

$$\mathbb{P} \left(\frac{\widehat{R}_N(f, T_{GCV}) - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) \Rightarrow \Phi(x). \quad (31)$$

Доказательство. Утверждение (31) следует из представления (7) и теоремы 4.

Таким образом, доказанные утверждения обосновывают разумность выбора порога T_{GCV} , поскольку риск в этом случае оказывается близким к минимальному.

5 Оценки скорости сходимости к нормальному закону

В процессе доказательства теоремы 4 используются методы, позволяющие оценить скорость сходимости распределения оценки риска к нормальному закону.

Теорема 5. Пусть выполнены условия теоремы 4, тогда существуют такие константы \tilde{C}_0 и \tilde{C}_1 , что

$$\sup_{x \in \mathbf{R}} \left| \mathbb{P} \left(\frac{\widehat{G}_N(f, T_{GCV}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) - \Phi(x) \right| \leq \frac{\tilde{C}_0 (\ln N)^{1+1/(2(\alpha+1))}}{N^{1/2-1/(2(\alpha+1))}}, \quad (32)$$

если $0 < \alpha \leq 1$, и

$$\sup_{x \in \mathbf{R}} \left| \mathbb{P} \left(\frac{\widehat{G}_N(f, T_{GCV}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) - \Phi(x) \right| \leq \frac{\tilde{C}_1 (\ln N)^{-1/2}}{N^{1/4}}, \quad (33)$$

если $\alpha > 1$.

Доказательство. Как и в предыдущей теореме, обозначим

$$\frac{\widehat{G}_N(f, T_{GCV}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} = S_1 + S_2,$$

где

$$S_1 = \frac{\widehat{G}_N(f, T_{\min}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}};$$

$$S_2 = \frac{\widehat{G}_N(f, T_{GCV}) - \widehat{G}_N(f, T_{\min})}{\sigma^2 \sqrt{2N}}.$$

Имеем

$$\sup_{x \in \mathbf{R}} \left| \mathbb{P} \left(\frac{\widehat{G}_N(f, T_{GCV}) - N\sigma^2 - R_N(f, T_{\min})}{\sigma^2 \sqrt{2N}} < x \right) - \Phi(x) \right| = \sup_{x \in \mathbf{R}} |\mathbb{P}(S_1 + S_2 < x) - \Phi(x)| \leq$$

$$\leq \sup_{x \in \mathbf{R}} |\mathbb{P}(S_1 < x) - \Phi(x)| + \frac{\varepsilon_N}{\sqrt{2\pi}} + \mathbb{P}(S_2 > \varepsilon_N). \quad (34)$$

Поступая так же, как в работе [9], и используя представление (7), можно показать, что для некоторой константы \tilde{C}'_0 справедливо

$$\sup_{x \in \mathbf{R}} |\mathbb{P}(S_1 < x) - \Phi(x)| \leq \frac{\tilde{C}'_0 (\ln N)^{3/2+1/(4(\alpha+1))}}{N^{1/2-1/(4(\alpha+1))}}. \quad (35)$$

Для оценки третьего слагаемого в (34) нужно более точно оценить вероятности P_1 , P_2 и P_3 в теореме 4. Возьмем $r'_N = C_r^* N^{1/2} \ln N$ и $T_1 = \sigma \sqrt{\ln N} - d_1 \sigma \ln \ln N / \sqrt{\ln N}$. Константы $C_r^* > 0$ и $d_1 > 0$ можно выбрать таким способом, чтобы в (12) было справедливо

$$\mathbb{P} \left(T_1 N \sup_{T \in [T_0, T_1]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E}\mu_T)^2} \right| > \frac{r'_N}{2} \right) \leq \frac{C_{r'}}{N^{1/2}}, \quad (36)$$

а в (19) было справедливо

$$\mathbb{P} \left(\sup_{T \in [T_0, T_1]} \left| \widehat{H}_N^*(f, T) - \mathbb{E}\widehat{H}_N^*(f, T) \right| \geq \inf_{T \in [T_0, T_1]} -\mathbb{E}\widehat{H}_N^*(f, T) - r'_N \right) \leq \frac{C_{d_1}}{N^{1/2}} \quad (37)$$

с некоторыми константами $C_{r'}$ и C_{d_1} . Далее возьмем $r''_N = C_r^{**} N^{1-\beta/2} (\ln N)^{3/4}$ и $T_2 = \sigma \sqrt{\beta \ln N} - d_2 \sigma \ln \ln N / \sqrt{\ln N}$, где $\beta = \min(3/2, 4\gamma/(2\gamma+1))$. Константы $C_r^{**} > 0$ и $d_2 > 0$ можно выбрать таким способом, чтобы в (24) было справедливо

$$\mathbb{P} \left(T_2 N \sup_{T \in [T_1, T_2]} \left| \frac{1}{\mu_T^2} - \frac{1}{(\mathbb{E}\mu_T)^2} \right| > \frac{r''_N}{2} \right) \leq \frac{C_{r''}}{N^{1/2}}, \quad (38)$$

а в (27) было справедливо

$$P \left(\sup_{T \in [T_1, T_2]} \left| \widehat{H}_N^*(f, T) - E \widehat{H}_N^*(f, T) \right| \geq \right. \\ \left. \geq \inf_{T \in [T_1, T_2]} -E \widehat{H}_N^*(f, T) - r_N'' \right) \leq \frac{C_{d_2}}{N^{1/2}}. \quad (39)$$

с некоторыми положительными константами $C_{r''}$ и C_{d_2} . Наконец, для оценки P_3 при таком выборе β вместо $N^{-\alpha/2}$ возьмем

$$\varepsilon_N = \begin{cases} C_\varepsilon N^{(1-\beta)/2} (\ln N)^{-1/2}, & \text{если } \frac{4\gamma}{2\gamma+1} > \frac{3}{2}, \\ C_\varepsilon N^{(1-\beta)/2} (\ln N)^{1+1/(2(\alpha+1))}, & \text{если } \frac{4\gamma}{2\gamma+1} \leq \frac{3}{2}, \end{cases}$$

с некоторой положительной константой C_ε . Найдется такая константа $\widetilde{C}_E > 0$, что

$$\sup_{T \in [T_2, T_U]} \sum_{i=1}^N E h^*(X_i, T) \leq \widetilde{C}_E N^{1-\beta/2} (\ln N)^{-1/2},$$

если $4\gamma/(2\gamma+1) > 3/2$, и

$$\sup_{T \in [T_2, T_U]} \sum_{i=1}^N E h^*(X_i, T) \leq \widetilde{C}_E N^{1-\beta/2} (\ln N)^{1+1/(2(\alpha+1))},$$

если $4\gamma/(2\gamma+1) \leq 3/2$. Следовательно, рассуждая как при выводе соотношений (28)–(30), можно показать, что при $C_\varepsilon \sigma^2 \sqrt{2}/3 > \widetilde{C}_E$ найдется такая константа C_{P_3} , что

$$P_3 \leq \frac{C_{P_3}}{N^{1/2}}. \quad (40)$$

Объединяя (34)–(40) и учитывая выбор $r'_N, r''_N, \varepsilon_N$ и β , получаем (32) и (33). Теорема доказана.

Замечание 2. В следствии из теоремы 4 можно получить такие же оценки скорости сходимости к нормальному закону, быть может, с другими константами.

Замечание 3. В работе [2] отмечается, что порог T_{SURE} в случае, когда больших вейвлет-коэффициентов слишком мало, несколько недооценивает порог T_{min} . Как видно из доказательства теорем 4 и 5, если $\alpha \leq 1$, т. е. регулярность функции сигнала относительно невелика, то вероятность события $\{T_{GCV} < T_{min}\}$ стремится к нулю и порог T_{GCV} может, скорее, переоценивать T_{min} , что справедливо и для порога T_{SURE} . Если же $\alpha > 1$, то $T_{min} > \sqrt{(3/2) \ln N}$, но можно лишь утверждать, что вероятность события $\{T_{GCV} < \sqrt{(3/2) \ln N}\}$ стремится к нулю, и в этом случае T_{GCV} может недооценивать T_{min} так же, как порог T_{SURE} . Таким образом, доказанные теоремы дают количественную характеристику регулярности, влияющую на свойства порога T_{GCV} .

Литература

1. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // *Biometrika*, 1994. Vol. 81. No. 3. P. 425–455.
2. Donoho D., Johnstone I. M. Adapting to unknown smoothness via wavelet shrinkage // *J. Amer. Stat. Assoc.*, 1995. Vol. 90. P. 1200–1224.
3. Donoho D. L., Johnstone I. M., Kerkycharian G., Picard D. Wavelet shrinkage: Asymptopia? // *J. R. Statist. Soc. Ser. B.*, 1995. Vol. 57. No. 2. P. 301–369.
4. Marron J. S., Adak S., Johnstone I. M., Neumann M. H., Patil P. Exact risk analysis of wavelet regression // *J. Comput. Graph. Stat.*, 1998. Vol. 7. P. 278–309.
5. Antoniadis A., Fan J. Regularization of wavelet approximations // *J. Amer. Statist. Assoc.*, 2001. Vol. 96. No. 455. P. 939–967.
6. Jansen M. Noise reduction by wavelet thresholding // *Lecture notes in Statistics*. Vol. 161. — Springer Verlag, 2001.
7. Маркин А. В., Шестаков О. В. О состоятельности оценки риска при пороговой обработке вейвлет-коэффициентов // *Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн.*, 2010. № 1. С. 26–34.
8. Маркин А. В. Предельное распределение оценки риска при пороговой обработке вейвлет-коэффициентов // *Информатика и её применения*, 2009. Т. 3. Вып. 4. С. 57–63.
9. Шестаков О. В. Аппроксимация распределения оценки риска пороговой обработки вейвлет-коэффициентов нормальным распределением при использовании выборочной дисперсии // *Информатика и её применения*, 2010. Т. 4. Вып. 4. С. 73–81.
10. Шестаков О. В. Асимптотическая нормальность оценки риска пороговой обработки вейвлет-коэффициентов при выборе адаптивного порога // *Докл. РАН*, 2012. Т. 445. № 5. С. 513–515.
11. Jansen M., Malfait M., Bultheel A. Generalized cross validation for wavelet thresholding // *Signal Processing*, 1997. Vol. 56. No. 1. P. 33–44.
12. Jansen M. Minimum risk methods in the estimation of unknown sparsity. Technical Report, 2010. U.L.B.
13. Добеши И. Десять лекций по вейвлетам. — Ижевск: НИЦ Регулярная и хаотическая динамика, 2001.
14. Mallat S. A wavelet tour of signal processing. — N.Y.: Academic Press, 1999.
15. Abramovich F., Silverman B. W. Wavelet decomposition approaches to statistical inverse problems // *Biometrika*, 1998. Vol. 85. No. 1. P. 115–129.
16. Boggess A., Narkowich F. A first course in wavelets with Fourier analysis. — Upper Saddle River: Prentice Hall, 2001.

17. *Захарова Т. В., Шестаков О. В.* Вейвлет-анализ и его приложения: Учебное пособие. — М.: МАКС Пресс, 2009.
18. *Alexander K.* Probability inequalities for empirical processes and a law of the iterated logarithm // *Ann. Probab.*, 1984. Vol. 12. No. 4. P. 1041–1067.
19. *Bennett G.* Probability inequalities for sums of independent random variables // *J. Amer. Statist. Assoc.*, 1962. Vol. 57. P. 33–45.
20. *Феллер В.* Введение в теорию вероятностей и ее приложения. — М.: Мир, 1984.
21. *Vaart A. W., Wellner J. A.* Weak convergence and empirical processes. — N.Y.: Springer Verlag, 1996.
22. *Колмогоров А. Н., Тихомиров В. М.* ε -энтропия и ε -емкость множеств в функциональных пространствах // *УМН*, 1959. Т. 14. № 2(86). С. 3–86.
23. *Shen X., Wong W. H.* Convergence rate of sieve estimates // *Ann. Statist.*, 1994. Vol. 22. No. 2. P. 580–615.

СТАТИСТИЧЕСКАЯ ПРОВЕРКА НЕИСПОЛНЯЕМОСТИ ФРАГМЕНТОВ КОДА ПОСЛЕДОВАТЕЛЬНОЙ ПРОГРАММЫ*

В. Ю. Королев¹, Р. Л. Смелянский², Т. Р. Смелянский³, А. В. Шалимов⁴

Аннотация: Рассматривается задача статистической проверки неисполняемости фрагментов кода последовательной программы. Рассматриваются методы, основанные на минимизации как априорных, так и апостериорных вероятностей ошибок.

Ключевые слова: проверка статистических гипотез; геометрическое распределение; вероятность ошибки первого рода; вероятность ошибки второго рода; лемма Неймана–Пирсона; апостериорная вероятность ошибки

1 Введение

В данной статье рассматривается задача статистической проверки неисполняемости фрагментов кода последовательной программы. Такая задача возникает в некоторых приложениях, связанных с необходимостью оптимизации размера программы. Эта проблема весьма актуальна при создании программного обеспечения миниатюрных устройств, в частности таких, как кардиостимуляторы или кардиомониторы, а также при создании бортового программного обеспечения специального назначения, когда небольшой физический объем носителя программы вынуждает оптимизировать объем самой программы. При этом, естественно, потери функциональности программы должны быть незначительными.

В данной статье используется строгое определение частоты выполнения фрагмента кода последовательной программы, предложенное и проанализированное в работе [1]. Под фрагментом кода понимается линейный участок программы, то есть такой участок кода, где не нарушается естественный порядок выполнения.

Рассматриваемая задача имеет тесную формальную связь с областью профилировки программ [2, 3], где возникает задача определения частоты выполнения инструкций программы. Существующие решения в этой области сводятся к применению различных эмпирических данных или

к предположению, что все входные данные программы равномерно распределены на области значений, что в общем случае не всегда верно. Данная статья продолжает исследования методов оценивания частоты выполнения фрагментов кода последовательных программ, начатые в работах [1, 4].

2 Описание задачи

Пусть $\Pi(x_1, \dots, x_p) = \{V, E\}$ — исходная последовательная программа с p входными параметрами x_1, \dots, x_p . Программа представлена в виде графа потока управления с вершинами $V = \{b_j\}$, где b_j — линейные участки программы, $j = 1, \dots, m$, и дугами $E = \{(b_{j_1}, b_{j_2})\}$, где (b_{j_1}, b_{j_2}) — переход управления от j_1 -го линейного участка к j_2 -му линейному участку.

Конкретное значение параметра x_i будем обозначать \hat{x}_i , $i = 1, \dots, p$. Множество $\mathcal{X} = \{(\hat{x}_1, \dots, \hat{x}_p)\}$ всех допустимых входных параметров программы будем считать метрическим пространством, что позволит формально рассмотреть борелевскую σ -алгебру его подмножеств и определить на ней распределение вероятностей P , которое позволяет интерпретировать вектор $(\hat{x}_1, \dots, \hat{x}_p)$ как значение случайного элемента, принимающего значения в \mathcal{X} и имеющего известное распределение P . Заметим, что, вообще говоря, в силу дискретного характера представления информации в компьютере для

*Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м, 12-07-00109а).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, victoryukorolev@yandex.ru

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, smel@sc.msu.ru

³Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, smelyanskiy.t@bk.ru

⁴Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, ashlimov@lvk.cs.msu.su

каждого входного параметра x_1, \dots, x_p множество допустимых значений является конечным, так что упомянутое выше множество \mathcal{X} конечно, а распределение входных значений дискретно. Однако для удобства будем рассматривать более общую модель, поскольку для многих реальных программ мощность $|\mathcal{X}|$ множества \mathcal{X} может иметь порядок 10^{100} [5], что, например, превосходит число атомов водорода (наиболее распространенного химического элемента) в видимой части Вселенной, которое по современным оценкам составляет примерно 10^{48} [6]. Поэтому методы теории вероятностей и математической статистики, ориентированные на конечные модели, в рассматриваемом случае неэффективны.

Приведем простой пример, свидетельствующий в пользу рассмотрения в данной задаче общих (в том числе непрерывных) вероятностных моделей. Рассмотрим входной параметр, отвечающий за показания альтиметра (пилотажно-навигационного прибора, указывающего высоту полета). Погрешности показаний альтиметра, которые, безусловно, имеются и обусловлены суммарным воздействием большого числа случайных факторов, могут считаться случайными и, в силу центральной предельной теоремы теории вероятностей, нормально распределенными с дисперсией, обусловленной точностью прибора.

Рассматриваемая постановка задачи восходит к работам [7, 8], в которых впервые дана ее общая строгая формализация. Позднее с развитием техники профилировки программ появились частные случаи этой постановки [2, 3]. Будем придерживаться уточненной формализации стохастического функционирования программы, предложенной в [9] (также см., например, [10]).

Будем считать, что программа $\Pi(x_1, \dots, x_p)$ не заикливается на допустимых наборах входных данных, т.е. каждый линейный участок выполняется конечное число раз. Введем обозначение $\Pi_j(\hat{x}_1, \dots, \hat{x}_p)$ для числа выполнений j -го линейного участка на векторе входных значений $\hat{x}_1, \dots, \hat{x}_p$.

Итак, пусть $X = (X_1, \dots, X_p)$ — случайный элемент со значениями в \mathcal{X} , интерпретируемый как случайный вектор входных значений программы. Пусть $Y_j = \Pi_j(X_1, \dots, X_p)$ — соответствующее число выполнений j -го линейного участка программы. Очевидно, что Y_j — это дискретная случайная величина, принимающая значения в $\mathbb{N} \cup \{0\}$ и имеющая собственное распределение (т.е. принимающая бесконечное значение с нулевой вероятностью). Математическое ожидание, соответствующее распределению P , будет обозначаться E .

Возможно несколько подходов к вычислению вероятности активизации (выполнения) линейно-

го участка b_j . В работах [7, 8] предложены методы построения функции распределения числа выполнений линейных участков программы по распределению входных параметров в аналитическом виде. Фактически используются формулы свертки и при реализации требуются значительные вычислительные затраты. В работе [1] предложен подход, позволяющий не вычислять точное значение частоты выполнения, так как это сопряжено с большими вычислительными затратами, сравнимыми с прогонами программы на всех входных данных, а оценивать значение вероятности выполнения с заранее заданной точностью. Для оценки вероятности выполнения линейного участка, которая заключается в приближенном вычислении математического ожидания бернуллиевой случайной величины, в [1] предложено использовать метод статистических испытаний (метод Монте-Карло). Идея этого метода заключается в проведении многократных опытов с исследуемой случайной величиной, и на основе полученных значений делается вывод о различных характеристиках этой случайной величины.

3 Метод, основанный на лемме Неймана–Пирсона

Проведем n независимых испытаний, в результате каждого из которых наблюдается значение случайной величины Y_j . На практике это означает запуски программы $\Pi(x_1, \dots, x_p)$ на значениях $\hat{x}_1, \dots, \hat{x}_p$, сгенерированных независимым образом в соответствии с распределением P на множестве \mathcal{X} входных параметров [1]. При этом на каждом испытании используется новый экземпляр программы $\Pi(x_1, \dots, x_p)$. Получим выборку из n значений $Y_j^{(1)}, \dots, Y_j^{(n)}$ — числа выполнений j -го линейного участка при запусках программы на элементах из \mathcal{X} , отвечающих распределению P . При этом можно считать, что моменты любых порядков случайной величины Y_j существуют [1].

Обозначим $\chi_i = \mathbb{I}(Y_j^{(i)} > 0)$, где $\mathbb{I}(A)$ — индикатор события A :

$$\mathbb{I}(A) = \begin{cases} 1, & \text{событие } A \text{ произошло;} \\ 0, & \text{событие } A \text{ не произошло.} \end{cases}$$

Описанный выше способ организации испытаний позволяет считать случайные величины χ_1, \dots, χ_n независимыми и одинаково распределенными, причем

$$P(\chi_i = 1) = p = 1 - P(\chi_i = 0), \quad i = 1, \dots, n.$$

Очевидно, что при этом число p имеет смысл вероятности того, что рассматриваемый линейный

участок программы будет задействован при случайном выборе значений входных параметров в соответствии с данным распределением вероятностей на множестве \mathcal{X} . Положим $q = 1 - p$. Ясно, что q — это вероятность неисполнения рассматриваемого линейного участка (фрагмента) программы. Будем считать, что если $q = 1$, то участок неактивен, неисполняем (и может быть удален из программы), а если $q = q_0 < 1$, то фрагмент активен. Предположим, что на вход подано n стандартных независимых наборов значений входных параметров. Положим

$$\nu = \sum_{i=1}^n \chi_i.$$

При этом если рассматриваемый участок программы активен (исполняем), то

$$P(\nu = k) \equiv p_0(k) = C_n^k q_0^{n-k} (1 - q_0)^k, \quad k = 0, \dots, n, \quad (1)$$

а если рассматриваемый участок программы неактивен (неисполняем), то

$$P(\nu = k) \equiv p_1(k) = \begin{cases} 1, & k = 0, \\ 0, & k = 1, \dots, n. \end{cases} \quad (2)$$

Таким образом, способ проверки исполняемости фрагмента программы может быть описан в терминах задачи проверки статистических гипотез. Имеется случайная величина ν , распределение которой при гипотезе $H_0 : q = q_0 < 1$ имеет вид (1), а при гипотезе $H_1 : q = 1$ — вид (2). По наблюдаемому значению величины ν требуется сделать вывод о том, какое распределение — (1) или (2) — случайная величина ν имеет в действительности. В данной работе рассматриваются два подхода к решению этой задачи: традиционный подход, основанный на стандартной постановке задачи проверки статистических гипотез, решение которой дается леммой Неймана–Пирсона [11], и подход, основанный на рассмотрении апостериорных вероятностей ошибок [11–13].

Сначала рассмотрим первый подход. При проверке неисполняемости линейного участка программы следует формально допустить возможность ошибок двух типов: (1) исполняемый участок может быть признан неактивным и (2) неисполняемый участок может быть признан активным. Вероятности этих ошибок соответственно обозначим α и β . Ошибка первого типа значительно более нежелательна, чем ошибка второго типа, поскольку она может привести к значительной потере функциональности и надежности программы после удаления признанного неисполняемым активного участ-

ка. Зафиксируем некоторое число $a > 0$ и потребуем, чтобы метод контроля неисполняемости удовлетворял условию $\alpha \leq a$. При этом желательно (например, по экономическим соображениям), чтобы вероятность β была бы минимально возможной. Другими словами, методика контроля неисполняемости должна быть такой, чтобы средняя доля неисполняемых фрагментов, признанных активными, была минимальной при условии, что средняя доля исполняемых фрагментов, признанных неисполняемыми, не превосходит a .

Как известно, правило проверки статистических гипотез, удовлетворяющее указанному критерию оптимальности, имеет следующий вид (см., например, [9, с. 78]). Пусть $d(\nu) = (d_0(\nu), d_1(\nu))$ — вектор-функция такая, что

$$d_0(\nu) \geq 0, \quad d_1(\nu) \geq 0; \quad d_0(\nu) + d_1(\nu) \equiv 1.$$

При заданном значении ν условимся с вероятностью $d_0(\nu)$ принимать гипотезу H_0 и с вероятностью $d_1(\nu)$ принимать гипотезу H_1 .

Положим

$$L = L(\nu) = \frac{p_0(\nu)}{p_1(\nu)}, \quad G_i(\ell) = P(L < \ell | H_i), \quad i = 0, 1.$$

Согласно фундаментальной лемме Неймана–Пирсона оптимальное правило имеет вид:

$$d(\nu) = d(L) = (d_0(L), d_1(L)) = \begin{cases} (0, 1), & L < \ell_1, \\ (r_1, 1 - r_1), & L = \ell_1, \\ (1, 0), & L > \ell_1, \end{cases} \quad (3)$$

где

$$\ell_1 = \sup \{ \ell : G_0(\ell - 0) \leq a \};$$

$$r_1 = \frac{G_0(\ell_1) - a}{G_0(\ell_1) - G_0(\ell_1 - 0)}.$$

В рассматриваемом случае

$$L = L(\nu) = \begin{cases} q_0^n, & \nu = 0; \\ \infty, & \nu > 0; \end{cases}$$

$$G_0(\ell) = \begin{cases} 0, & \ell < q_0^n; \\ q_0^n, & q_0^n \leq \ell < \infty; \\ 1, & \ell = \infty; \end{cases}$$

$$G_1(\ell) = \begin{cases} 0, & \ell < q_0^n; \\ 1, & \ell \geq q_0^n. \end{cases}$$

Возможны два случая: (А) $q_0^n \leq a$ и (В) $q_0^n > a$. В случае (А) $\ell_1 = \ell_1^A \equiv \infty$ и потому согласно (3) заведомо активные участки ($\nu > 0$) с ненулевой вероятностью будут признаваться неисполняемыми,

что, очевидно, нежелательно. Поэтому с помощью выбора достаточно большого n надо добиваться, чтобы было выполнено неравенство $q_0^n > a$, т.е. чтобы имел место случай (В). В случае (В) имеем

$$\begin{aligned} \ell_1 &= \ell_1^B = q_0^n; \\ \beta &= 1 - G_1(\ell_1^B) + r_1 [G_1(\ell_1^B) - G_1(\ell_1^B - 0)] = \\ &= 1 - \frac{a}{q_0^n}. \end{aligned}$$

При этом для фиксированного n вероятность β не может быть уменьшена. Однако с помощью надлежащего выбора n можно уменьшить β до величины, близкой к нулю. А именно: положим

$$n = n_1 = \max \left\{ m : 1 - \frac{a}{q_0^m} \geq 0 \right\} = \lceil \log_{q_0} a \rceil.$$

При таком выборе n , во-первых, имеет место ситуация (В), во-вторых, $q_0^n \approx a$, т.е. $\beta \approx 0$, и, в-третьих, $r_1 = 0$, т.е. если $\nu > 0$, то участок сразу признается активным.

При этом испытания можно прекратить после появления первого значения $\chi_i = 1$. Таким образом, число N испытаний, необходимых для проверки неисполнимости фрагмента программы по правилу, удовлетворяющему указанному критерию оптимальности, случайно, причем

$$\begin{aligned} P(N = k | H_0) &= q_0^{k-1} (1 - q_0), \quad k = 1, \dots, n_1 - 1; \\ P(N = n_1 | H_0) &= q_0^{n_1-1}. \end{aligned}$$

Несложно убедиться, что

$$E(N | H_0) = \frac{1 - q_0^{n_1}}{1 - q_0} \approx \frac{1 - a}{1 - q_0}.$$

При этом величина $E(N | H_0)$ является минимально возможной среди всех последовательных критериев с $\alpha = a$ и $\beta = 0$, поскольку предложенный метод является урезанным последовательным критерием отношения вероятностей, оптимальным в смысле [12].

В табл. 1 для $q_0 = 0,99$ приведены значения n_1 и $E(N | H_0)$ в зависимости от a .

Для $q_0 = 0,999$, $q_0 = 0,9999$ и т.д. хорошие аппроксимации для n_1 и $E(N | H_0)$ получаются из табл. 1 с помощью умножения соответствующих значений на 10, 100 и т.д.

Таблица 1 Значения n_1 и $E(N | H_0)$ в зависимости от a (метод, основанный на лемме Неймана–Пирсона)

a	n_1	$E(N H_0)$
0,05	300	95
0,01	460	99
0,005	530	99,5
0,001	920	99,99

4 Метод, основанный на оптимизации апостериорных вероятностей ошибок

Теперь рассмотрим подход, основанный на апостериорных вероятностях ошибок. Как уже отмечалось, величина α имеет смысл средней доли фрагментов, признанных неисполняемыми, среди исполняемых участков. С практической точки зрения намного важнее уметь управлять средней долей активных участков среди систем, признанных неисполняемыми (апостериорной вероятностью ошибки первого рода), которую будем обозначать γ . Правило различения двух простых гипотез, минимизирующее β при условии $\gamma \leq c$, где $c > 0$ — наперед заданное число, построено в работах [13–15]. Это правило имеет вид (3), где ℓ_1 и r_1 заменены соответственно величинами

$$\begin{aligned} \ell_2 &= \sup \left\{ \ell : \frac{G_0(\ell - 0)}{G_1(\ell - 0)} \leq \frac{c(1 - w)}{w(1 - c)} \right\}; \\ r_2 &= \left[1 + \frac{w(1 - c)G_0(\ell_2 - 0) - c(1 - w)G_1(\ell_2 - 0)}{c(1 - w)G_1(\ell_2) - w(1 - c)G_0(\ell_2)} \right]^{-1}. \end{aligned}$$

Здесь $w = P(H_0)$ — априорная вероятность справедливости гипотезы H_0 — имеет смысл средней (ожидаемой) доли активных участков среди всех фрагментов программы (ребер графа $\{V, E\}$).

Возможны два случая:

- (C) $q_0^n \geq c(1 - w)/(w(1 - c))$;
- (D) $q_0^n < c(1 - w)/(w(1 - c))$.

Рассмотрим ситуацию (C). Здесь $\ell_2 = \ell_2^C \equiv q_0^n$, $\alpha = (1 - r_2)q_0^n$, $\beta = r_2$ и γ не зависит от r_2 . Поэтому из соображений минимальности β следует положить $r_2 = 0$. При этом в силу условия $q_0^n \geq c(1 - w)/(w(1 - c))$ имеет место неравенство $\gamma \geq c$. Так как γ убывает с ростом n , то следует положить

$$\begin{aligned} n = n_2 &= \max \left\{ n : q_0^n \geq \frac{c(1 - w)}{w(1 - c)} \right\} = \\ &= \left\lceil \log_{q_0} \frac{c(1 - w)}{w(1 - c)} \right\rceil. \end{aligned}$$

При таком выборе n имеет место приближенное равенство $\gamma \approx c$ и для среднего числа испытаний $E(N | H_0)$ справедлива аппроксимация $E(N | H_0) \approx (w - c)/(w(1 - c)(1 - q_0))$.

В ситуации (D) критерий имеет те же недостатки, что и в ситуации (A). Поэтому с помощью

Таблица 2 Значения n_2 и $E(N|H_0)$ в зависимости от c и w (метод, основанный на оптимизации апостериорных вероятностей ошибок)

w	$a = 0,05$		$a = 0,01$		$a = 0,005$		$a = 0,001$	
	n_2	$E(N H_0)$	n_2	$E(N H_0)$	n_2	$E(N H_0)$	n_2	$E(N H_0)$
0,05	—	—	164	80	233	91	394	99
0,1	73	52	238	91	308	96	468	99
0,2	155	78	319	96	388	98	549	99
0,3	208	87	372	98	442	99	602	100
0,4	252	92	416	99	486	99	646	100
0,5	292	94	457	99	526	100	687	100
0,6	333	96	497	100	566	100	727	100
0,7	377	97	541	100	610	100	771	100
0,8	430	98	595	100	664	100	825	100
0,9	511	99	675	100	745	100	905	100
0,95	585	99	749	100	819	100	979	100

надлежащего выбора n и c надо добиваться, чтобы имела место ситуация (С).

В табл. 2 для $q_0 = 0,99$ приведены значения n_2 и $E(N|H_0)$ в зависимости от c и w .

Для $q_0 = 0,999$, $q_0 = 0,9999$ и т.д. хорошие аппроксимации для n_2 и $E(N|H_0)$ получаются из табл. 2 с помощью умножения соответствующих значений на 10, 100 и т.д.

Литература

1. Шалимов А. В. Метод оценки частоты выполнения фрагментов кода последовательной программы // Моделирование и анализ информационных систем, 2010. Т. 17. Вып. 2. С. 122–132.
2. Wu Youfeng, Larus J. R. Static branch frequency and program profile analysis // 27th Annual Symposium (International) on Microarchitecture Proceedings, 1994. P. 1–11.
3. Ball T., Larus J. R. Optimally profiling and tracing programs // ACM Transactions on Programming Languages and Systems (TOPLAS), 1994. P. 1319–1360.
4. Королев В. Ю., Смелянский Р. Л., Смелянский Т. Р., Шалимов А. В. Об оценивании частоты выполнения фрагментов кода последовательной программы // Программирование, 2013 (в печати).
5. Пальчун Б. П. Метод испытаний программ на надежность // Функциональная устойчивость специального математического обеспечения автоматизированных систем. — М., 1989. С. 111–117.
6. Парнов Е. И. На перекрестке бесконечностей. — М.: Атомиздат, 1967. 464 с.
7. Smelianski R. L., Alanko T. On the calculation of control transition probabilities in a program // Inform. Process. Letters, 1984. No. 3.
8. Смелянский Р. Л., Гурьев Д. Е., Бахмутов А. Г. Об одной математической модели для расчета динамических характеристик программы // Программирование, 1986. Вып. 6.
9. Королев В. Ю. Предельные распределения для случайных последовательностей со случайными индексами и некоторые их приложения: Дисс. . . . докт. физ.-мат. наук. — М.: МГУ, 1994. 265 с.
10. Королев В. Ю., Соколов И. А. Основы математической теории надежности модифицируемых систем. — М.: ИПИ РАН, 2006. 108 с.
11. Леман Э. Проверка статистических гипотез. — М.: Наука, 1979.
12. Вальд А. Последовательный анализ. — М.: Наука, 1960.
13. Королев В. Ю. Наиболее мощные критерии проверки простой гипотезы против простой альтернативы с апостериорным уровнем значимости // Проблемы устойчивости стохастических моделей: Труды семинара. — М.: ВНИИСИ, 1985. С. 87–91.
14. Королев В. Ю. Различение двух простых гипотез с неопределенными решениями // IV Международная Вильнюсская конференция по теории вероятностей и математической статистике: Тезисы докладов. Т. 2. — Вильнюс: Институт математики и кибернетики АН Литовской ССР, 1985. С. 100–103.
15. Акбулатов Н. А., Белокуров Д. В., Королев В. Ю. Задачи проверки надежности БИС ЭВМ // Разработка и применение в народном хозяйстве ЕС ЭВМ: Тезисы докладов Всесоюзной школы-семинара (Кишинев, 1985). — М., 1985. С. 91–94.

БАЙЕСОВСКАЯ РЕКУРРЕНТНАЯ МОДЕЛЬ РОСТА НАДЕЖНОСТИ: РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ ПАРАМЕТРОВ*

А. А. Кудрявцев¹, И. А. Соколов², С. Я. Шоргин³

Аннотация: Работа посвящена обоснованию целесообразности байесовского подхода при решении задач, связанных с определением надежности сложных модифицируемых систем. В качестве иллюстрации приводится среднее значение надежности системы, в которой показатели «дефективности» и «эффективности» средства, исправляющего недостатки системы, имеют равномерное распределение.

Ключевые слова: модифицируемые информационные системы; теория надежности; байесовский подход

1 Введение

Задача прогнозирования надежности сложных модифицируемых информационных систем была сформулирована в [1], а в дальнейшем более подробно рассмотрена в [2]. Изложение во введении строится, в основном, на [2].

Любой впервые созданный более или менее сложный агрегат, предназначенный для переработки или передачи информационных потоков, например новая программная система для компьютера, новая информационно-вычислительная сеть или новая административно-информационная система, как правило, не обладает требуемой надежностью. Для единства терминологии впредь будет говорить о сложных информационных системах. Такие системы подвергаются изменениям (модификациям) в ходе:

- разработки;
- испытаний и опытной эксплуатации;
- штатного функционирования.

Модификации информационных систем на этапе разработки имеют своей целью как уточнение задач, для решения которых предназначена система, так и оптимальную адаптацию систем к решению этих задач.

Модификации информационных систем в ходе испытаний и опытной эксплуатации имеют своей целью обнаружение и устранение имеющихся дефектов, препятствующих правильному функционированию.

Модификации сложных информационных систем в ходе штатного функционирования как результат усовершенствования их отдельных подсистем

или блоков, например замены отдельных морально устаревших узлов более современными, имеют своей целью повышение эффективности функционирования систем.

Целью таких модификаций является увеличение надежности информационных систем. В связи с этим возникает необходимость формализации понятия надежности модифицируемых информационных систем и разработки методов и алгоритмов оценивания и прогнозирования различных надежностных характеристик.

Рассматриваемый круг вопросов имеет две отличительные особенности.

1. Из-за огромной сложности современных информационных систем практически невозможно реализовать детерминированный подход к тестированию системы на всех возможных вариантах ее функционирования. Поэтому, по видимому, единственным возможным подходом к исследованию таких задач является вероятностно-статистический.
2. После каждой модификации свойства информационной системы изменяются, и потому данные, используемые для вероятностно-статистического анализа ее надежности, например длины интервалов времени между отказами системы (или между модификациями), вообще говоря, не могут интерпретироваться как одинаково распределенные случайные величины или, другими словами, не образуют однородных выборок, традиционно изучаемых в классической теории надежности. Однако методы

* Работа выполнена при поддержке РФФИ (проекты 11-01-00515-а, 11-07-00112-а, 12-07-00109-а).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, nubigena@hotmail.com

² Институт проблем информатики Российской академии наук, isokolov@ipiran.ru

³ Институт проблем информатики Российской академии наук, sshorgin@ipiran.ru

теории вероятностей и математической статистики позволяют найти приемлемое решение и в таких задачах.

Первые работы, описывающие математические модели роста надежности модифицируемых систем, появились в середине 1950-х гг., и сейчас их число огромно. Достаточно сказать, что по данным, приведенным в докладе [3], к концу 1980-х гг. было опубликовано более четырехсот научных работ только по надежности программного обеспечения. В то же время необходимо отметить, что универсального метода решения задачи прогнозирования надежности модифицируемых систем нет. Более того, до последнего времени не было и унифицированного подхода к созданию математических моделей. Возможно, именно поэтому модели роста надежности модифицируемых систем обходились стороной в канонических учебниках и монографиях по математической теории роста надежности. Такой подход был предложен в книге [1]. Именно на нем и основано изложение в книге [2].

В указанной книге приводятся общие соображения, служащие основой для построения математических моделей, описывающих изменение надежности модифицируемых информационных систем, а затем обсуждаются конкретные аналитические модели роста (изменения) надежности модифицируемых информационных систем.

К числу таких моделей относятся рекуррентные модели роста надежности. Они могут использоваться в случае, когда удобно иметь дело непосредственно с параметром, интерпретируемым как надежность системы. Рассмотрим произвольную систему, на вход которой подаются некоторые сигналы (например, команды оператора или внешние воздействия). Реакция системы на поданные сигналы может быть либо правильной (корректной), либо неправильной (некорректной). В каждый момент времени t надежность системы можно характеризовать параметром $p(t)$ — вероятностью того, что на сигнал, поданный на вход системы в момент t , система отреагирует правильно. По смыслу такая характеристика надежности ближе всего к традиционно используемому коэффициенту готовности. В случайные моменты времени $0 = Y_0 \leq Y_1 \leq Y_2 \leq \dots$ система подвергается (мгновенной) модификации, в результате чего изменяется параметр $p(t)$. Следует обратить внимание на то обстоятельство, что ниже рассматривается непрерывное время, без привязки напрямую процесса модифицирования системы к процессу ее тестирования. Предположим, что траектории процесса $p(t)$ непрерывны справа и кусочно-постоянны, так что $p(t) = p(Y_j)$ при $Y_j \leq t < Y_{j+1}$.

Задача прогнозирования поведения процесса $p(t)$ чрезвычайно важна. Описанная выше очень общая схема может быть переформулирована в терминах, традиционных для столь разных областей знания, как медицина, программирование или менеджмент. Например, в программировании параметр $p(t)$ можно рассматривать как надежность программного обеспечения, в котором по ходу отладки в моменты $0 = Y_0 \leq Y_1 \leq Y_2 \leq \dots$ вносятся изменения для исправления замеченных ошибок. Оценивание $p(t)$ и прогнозирование поведения этого параметра здесь важно как для оценивания надежности всего комплекса, составной частью которого является программное обеспечение, так и для прогнозирования продолжительности отладки. Более подробно об этом см. в книгах [1, 4]. В медицине параметр $1 - p(t)$ (называемый индексом летальности) характеризует вероятность летального исхода в момент времени t для пациента, организм которого в моменты $0 = Y_0 \leq Y_1 \leq Y_2 \leq \dots$ подвергается какому-либо медицинскому вмешательству (операции, инъекции, приему лекарств и т. п.). Здесь прогнозирование $p(t)$ чрезвычайно важно с точки зрения принятия решений о стратегии лечения. Наконец, в менеджменте параметр $p(t)$ может характеризовать надежность и дееспособность коллектива, организации или предприятия, структура которых в моменты времени $0 = Y_0 \leq Y_1 \leq Y_2 \leq \dots$ претерпевает изменения. Будем предполагать, что в результате каждой модификации системы параметр $p(t)$ изменяется случайным образом.

Обозначим $p_j = p(Y_j)$. Рассмотрим поведение p_j в зависимости от изменения j . Другими словами, будем изучать изменение надежности системы в зависимости от номера модификации.

В книге [2] рассматривается, в частности, следующая рекуррентная модель роста надежности. Пусть $\{(\theta_j, \eta_j)\}$, $j \geq 1$, — последовательность независимых одинаково распределенных двумерных случайных векторов таких, что $0 < \eta_1 < 1$; $0 < \theta_1 < 1$ почти наверное. Отметим, что независимость и совпадение распределений θ_j и η_j внутри каждой пары не предполагается.

Задав начальную надежность p_0 , рассмотрим модель, определяемую рекуррентным соотношением

$$p_{j+1} = \eta_{j+1}p_j + \theta_{j+1}(1 - p_j). \quad (1)$$

Эта модель названа дискретной экспоненциальной моделью. В такой модели случайные величины η_j описывают возможное уменьшение надежности из-за некачественных модификаций, в ходе которых вместо исправления существующих дефектов в систему могут быть внесены новые, в то время как величины θ_j описывают повышение надежности за

счет исправления дефектов. Частные случаи модели (1) с двухточечными распределениями случайных величин η_j и θ_j рассматривались в книгах [5, 6]. В свою очередь, эти частные случаи представляют собой переформулировку в терминах теории надежности одной модели обучаемости, рассмотренной в книге [7].

Обозначим $\lambda = 1 - E\theta_j$, $\mu = E\eta_j$. В [2] доказано, что при условии $\lambda + \mu \neq 1$

$$p = \lim_{j \rightarrow \infty} E p_j = \frac{\mu}{\lambda + \mu}.$$

2 Постановка задачи

Изучение предельного значения средней величины $E p_j$ представляет значительный интерес, поскольку эта величина характеризует асимптотическое значение надежности системы в рамках некоторой рекуррентной модели, задаваемой набором $\{(\theta_j, \eta_j)\}$. Из результатов [2] следует, что это асимптотическое значение зависит только от средних значений величин $\{(\theta_j, \eta_j)\}$, $j \geq 1$.

Каков «физический смысл» указанных случайных величин? Очевидно, что величины $\{\eta_j\}$ являются количественной характеристикой недостатков в работе средств (устройств, людей и т.п.), отвечающих за исправление дефектов, а величины $\{\theta_j\}$ — количественной характеристикой успешности работы таких средств (устройств, людей и т.п.). Если говорят об одной модифицируемой системе или объекте, то можно считать (в рамках рассматриваемой модели), что речь идет об одной и той же группе средств и т.п. (будем условно говорить — об одной ремонтной бригаде (РБ)), которая занимается исправлением дефектов данного сложного объекта в течение достаточно длительного времени. В этом случае естественно ставить задачу вычисления средних параметров λ и μ , характеризующих данную РБ, и на основе этих параметров вычислять величину p .

Однако можно представить себе более сложную ситуацию, при которой рассматривается целый набор однотипных сложных модифицируемых объектов (МО), каждый из которых обслуживается собственной РБ. Исследователю хотелось бы определить усредненное значение p по всем МО. Для решения этой задачи целесообразно перейти к так называемой байесовской постановке.

Будем считать, что рассматривается целая группа однотипных МО и группа им соответствующих однотипных РБ. Пусть $m = 1, 2, \dots$ — номера этих объектов. Для каждого МО (вместе с его РБ) существует собственный набор $\{(\theta_j^m, \eta_j^m)\}$ ($j \geq 1$, $m \geq 1$) независимых одинаково распределенных при

каждом фиксированном j двумерных случайных векторов таких, что $0 < \eta_1^m < 1$; $0 < \theta_1^m < 1$ почти наверное. Но средние значения величин θ_j^m , η_j^m , $j \geq 1$, $m \geq 1$, не предполагаются известными; более того, они не предполагаются даже одинаковыми. Вводится предположение, что величины $\lambda = 1 - E\theta_j^m$, $\mu = E\eta_j^m$ сами по себе являются случайными, т.е. на вероятностном пространстве, в которое в качестве элементарных событий входят все рассматриваемые в рамках данной постановки МО вместе с их РБ, заданы случайные величины λ и μ (которые будем полагать независимыми), имеющие смысл $\lambda = 1 - E\theta_j^m$, $\mu = E\eta_j^m$, где m — случайный номер МО. Принимаемые исследователем за основу распределения величин λ и μ будем называть априорными.

При этом подлежащие вычислению характеристики такой «рандомизированной» группы МО, естественно, являются рандомизацией аналогичных характеристик «отдельно взятой» МО с учетом априорного распределения параметров λ и μ , взятого исследователем за основу. Наиболее естественной и удобной для изучения характеристикой является усредненное по всем МО значение предельной вероятности надежности, т.е.

$$p_{\text{сред}} = E p = E \frac{\mu}{\lambda + \mu},$$

где усреднение ведется по совместному распределению случайных величин (λ, μ) .

Данный подход близок к байесовскому анализу систем массового обслуживания, которому посвящены статьи [8–12], но имеет следующее существенное отличие. В рамках байесовской постановки в моделях массового обслуживания возникали задачи рандомизации «обычных» характеристик систем массового обслуживания с учетом весьма разнообразных априорных распределений входных параметров; скажем, принималось предположение о показательном, эрланговском, равномерном и многих других распределениях одного или нескольких параметров входящих потоков и/или обслуживания. В рассматриваемой сейчас ситуации величины η_1^m и θ_1^m удовлетворяют ограничениям $0 < \eta_1^m < 1$, $0 < \theta_1^m < 1$. Значит, и средние значения λ и μ величин $1 - E\theta_j$ и $E\eta_j$ соответственно также находятся на отрезке $[0, 1]$. Поэтому в качестве априорных распределений параметров λ и μ следует выбирать только распределения, сосредоточенные на $[0, 1]$.

В настоящей заметке будут рассмотрены независимые случайные параметры λ и μ , распределенные равномерно на некоторых (вообще говоря, разных) отрезках, являющихся подмножествами отрезка $[0, 1]$.

3 Предварительные выкладки

Итак, пусть случайная величина λ имеет равномерное распределение на отрезке $[a_\lambda, b_\lambda]$, а случайная величина μ — равномерное распределение на отрезке $[a_\mu, b_\mu]$, причем $0 \leq a_\lambda \leq b_\lambda \leq 1, 0 \leq a_\mu \leq b_\mu \leq 1$.

Как уже отмечалось, надежность системы p имеет формально такой же вид, как и коэффициент готовности в системе массового обслуживания $M|M|1|0$ с интенсивностью входящего потока λ и интенсивностью обслуживания μ . Очевидно, что при байесовском подходе вычисление вероятностных характеристик p удобно производить, базирываясь на известном распределении величины $\rho = \lambda/\mu$, которая для моделей массового обслуживания имеет смысл коэффициента загрузки системы. При этом $p = 1/(1 + \rho)$.

В работе [9] были получены формулы для плотности и функции распределения коэффициента загрузки ρ . Приведем их здесь для удобства дальнейшего изложения для случая $a_\lambda/a_\mu \leq b_\lambda/b_\mu$.

Обозначим $K = ((b_\lambda - a_\lambda)(b_\mu - a_\mu))^{-1}$. Тогда (см. [9])

$$P(\rho < x) = \begin{cases} 0 & \text{при } x < \frac{a_\lambda}{b_\mu}; \\ K \frac{(b_\mu x - a_\lambda)^2}{2x} & \text{при } \frac{a_\lambda}{b_\mu} \leq x \leq \frac{a_\lambda}{a_\mu}; \\ K \left(\frac{a_\mu + b_\mu}{2} x - a_\lambda \right) (b_\mu - a_\mu) & \text{при } \frac{a_\lambda}{a_\mu} \leq x \leq \frac{b_\lambda}{b_\mu}; \\ 1 - K \frac{(b_\lambda - a_\mu x)^2}{2x} & \text{при } \frac{b_\lambda}{b_\mu} \leq x \leq \frac{b_\lambda}{a_\mu}; \\ 1 & \text{при } x > \frac{b_\lambda}{a_\mu}. \end{cases}$$

Выпишем плотность случайной величины ρ :

$$f_\rho(x) = \begin{cases} 0 & \text{при } x < \frac{a_\lambda}{b_\mu} \text{ и } x > \frac{b_\lambda}{a_\mu}; \\ K \left(\frac{b_\mu^2}{2} - \frac{a_\lambda^2}{2x^2} \right) & \text{при } \frac{a_\lambda}{b_\mu} \leq x \leq \frac{a_\lambda}{a_\mu}; \\ K \frac{b_\mu^2 - a_\mu^2}{2} & \text{при } \frac{a_\lambda}{a_\mu} \leq x \leq \frac{b_\lambda}{b_\mu}; \\ K \left(\frac{b_\lambda^2}{2x^2} - \frac{a_\mu^2}{2} \right) & \text{при } \frac{b_\lambda}{b_\mu} \leq x \leq \frac{b_\lambda}{a_\mu}. \end{cases}$$

4 Основные результаты

Основываясь на характеристиках распределения параметра ρ и учитывая, что $p = 1/(1 + \rho)$,

имеем для функции распределения и плотности p следующие соотношения:

$$F_p(x) = 1 - F_\rho \left(\frac{1-x}{x} \right); \quad f_p(x) = \frac{1}{x^2} f_\rho \left(\frac{1-x}{x} \right),$$

а следовательно, для случая $a_\lambda/a_\mu \leq b_\lambda/b_\mu$ имеем:

$$F_p(x) = \begin{cases} 0 & \text{при } x < \frac{a_\mu}{a_\mu + b_\lambda}; \\ 1 - K \frac{((a_\mu + b_\lambda)x - a_\mu)^2}{2x(1-x)} & \text{при } \frac{a_\mu}{a_\mu + b_\lambda} \leq x \leq \frac{b_\mu}{b_\mu + b_\lambda}; \\ 1 - K(b_\mu - a_\mu) \left(\frac{(a_\mu + b_\mu)(1-x) - a_\lambda}{2x} \right) & \text{при } \frac{b_\mu}{b_\mu + b_\lambda} \leq x \leq \frac{a_\mu}{a_\mu + a_\lambda}; \\ 1 - K \frac{((b_\mu + a_\lambda)x - b_\mu)^2}{2x(1-x)} & \text{при } \frac{a_\mu}{a_\mu + a_\lambda} \leq x \leq \frac{b_\mu}{b_\mu + a_\lambda}; \\ 1 & \text{при } x > \frac{b_\mu}{b_\mu + a_\lambda}. \end{cases}$$

Для плотности

$$f_p(x) = \begin{cases} 0 & \text{при } x < \frac{a_\mu}{a_\mu + b_\lambda} \text{ и } x > \frac{b_\mu}{b_\mu + a_\lambda}; \\ K \left(\frac{b_\lambda^2}{2(1-x)^2} - \frac{a_\mu^2}{2x^2} \right) & \text{при } \frac{a_\mu}{a_\mu + b_\lambda} \leq x \leq \frac{b_\mu}{b_\mu + b_\lambda}; \\ K \left(\frac{b_\mu^2}{2x^2} - \frac{a_\lambda^2}{2(1-x)^2} \right) & \text{при } \frac{a_\mu}{a_\mu + a_\lambda} \leq x \leq \frac{b_\mu}{b_\mu + a_\lambda}. \end{cases}$$

Теперь можно легко найти интересующую величину $p_{\text{сред}} = E p$. Имеем:

$$p_{\text{сред}} = \int_{-\infty}^{\infty} x f_p(x) dx = \frac{1}{2} + \frac{K b_\mu^2}{2} \ln \frac{b_\mu + b_\lambda}{b_\mu + a_\lambda} - \frac{K a_\mu^2}{2} \ln \frac{a_\mu + b_\lambda}{a_\mu + a_\lambda} + \frac{K b_\lambda^2}{2} \ln \frac{a_\mu + b_\lambda}{b_\mu + b_\lambda} - \frac{K a_\lambda^2}{2} \ln \frac{a_\mu + a_\lambda}{b_\mu + a_\lambda}. \quad (2)$$

При соотношении параметров $a_\lambda/a_\mu \geq b_\lambda/b_\mu$ аналогичные выкладки приводят к такому же результату.

Заметим, что полученный результат остается справедливым для случая $a_\lambda = b_\lambda, a_\mu = b_\mu$, соответствующего вырожденному распределению параметров λ и μ . При этом, очевидно, $p_{\text{сред}} = 1/2$.

Частные значения средней надежности

λ	μ									
	[0, 1/4]	[0, 1/2]	[0, 3/4]	[0, 1]	[1/4, 1/2]	[1/4, 3/4]	[1/4, 1]	[1/2, 3/4]	[1/2, 1]	[3/4, 1]
[0, 1/4]	0,50	0,63	0,70	0,74	0,76	0,80	0,83	0,84	0,86	0,88
[0, 1/2]	0,37	0,50	0,58	0,63	0,63	0,68	0,72	0,73	0,76	0,79
[0, 3/4]	0,30	0,42	0,50	0,56	0,55	0,60	0,64	0,66	0,69	0,72
[0, 1]	0,25	0,37	0,44	0,50	0,48	0,54	0,58	0,60	0,63	0,67
[1/4, 1/2]	0,24	0,37	0,45	0,52	0,50	0,56	0,61	0,63	0,66	0,70
[1/4, 3/4]	0,20	0,32	0,40	0,46	0,44	0,50	0,55	0,56	0,60	0,64
[1/4, 1]	0,17	0,28	0,36	0,42	0,39	0,46	0,50	0,51	0,55	0,60
[1/2, 3/4]	0,16	0,27	0,34	0,40	0,37	0,44	0,49	0,50	0,54	0,58
[1/2, 1]	0,14	0,24	0,31	0,37	0,34	0,40	0,45	0,46	0,50	0,54
[3/4, 1]	0,12	0,21	0,28	0,33	0,30	0,36	0,40	0,42	0,46	0,50

Такое же значение величины $p_{\text{сред}}$ получаем в случае, когда оба параметра λ и μ равномерно распределены на отрезке $[0, 1]$. Приведем еще несколько частных значений (с точностью до сотых) средней надежности, основываясь на формуле (2). В таблице строкам соответствуют отрезки, на которых сосредоточено распределение λ , а столбцам — распределение μ .

5 Заключение

Полученные результаты могут применяться, например, для вычисления других моментов и построения доверительных интервалов для характеристики p .

В дальнейшем предполагается рассмотреть более широкий класс априорных распределений величин λ и μ , сосредоточенных на $[0, 1]$, в частности бета-распределение.

Литература

1. Gnedenko B. V., Korolev V. Yu. Random summation: Limit theorems and applications. — Boca Raton, FL: CRC Press, 1996.
2. Королев В. Ю., Соколов И. А. Основы математической теории надежности модифицируемых систем. — М.: ИПИ РАН, 2006.
3. Van Pul M. C. Asymptotic properties of statistical models in software reliability // 2nd Bernoulli Society World Congress: Abstracts of communications. — Uppsala, 1990. P. 43–44.
4. Королев В. Ю. Прикладные задачи теории вероятностей: модели роста надежности модифицируемых систем. — М.: Диалог-МГУ, 1997.
5. Волков Л. И., Шишкевич А. М. Надежность летательных аппаратов. — М.: Высшая школа, 1975.
6. Волков Л. И. Управление эксплуатацией летательных комплексов. — М.: Высшая школа, 1981.
7. Буш Р., Мостеллер Ф. Стохастические модели обучаемости. — М.: ГИФМЛ, 1962.
8. Шоргин С. Я. О байесовских моделях массового обслуживания // II Научная сессия Института проблем информатики РАН: Тезисы докладов. — М.: ИПИ РАН, 2005. С. 120–121.
9. Кудрявцев А. А., Шоргин С. Я. Байесовский подход к анализу систем массового обслуживания и показателей надежности // Информатика и её применения, 2007. Т. 1. Вып. 2. С. 76–82.
10. Кудрявцев А. А., Шоргин С. Я. Байесовские модели массового обслуживания и надежности: экспоненциально-эрланговский случай // Информатика и её применения, 2009. Т. 3. Вып. 1. С. 44–48.
11. Кудрявцев А. А., Шоргин В. С., Шоргин С. Я. Байесовские модели массового обслуживания и надежности: общий эрланговский случай // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 30–34.
12. Кудрявцев А. А., Шоргин С. Я. Байесовские модели массового обслуживания и надежности: характеристики среднего числа заявок в системе $M|M|1|\infty$ // Информатика и её применения, 2010. Т. 4. Вып. 3. С. 16–21.

ИДЕНТИФИКАЦИЯ ЗАВИСИМОСТЕЙ ПРИЗНАКОВ СТОХАСТИЧЕСКОЙ ПРИРОДЫ НА ОСНОВЕ РЕГРЕССИИ ДЕМИНГА

В. С. Тимофеев¹, В. Ю. Щеколдин², А. Ю. Тимофеева³

Аннотация: Рассмотрены подходы к построению регрессионных зависимостей в предположении стохастичности входных и выходных показателей. Предложена оригинальная геометрическая интерпретация функционалов задачи оценивания параметров для частных случаев регрессии Деминга. Доказано утверждение о взаимном расположении прямой, обратной, диагональной и ортогональной регрессий. Получена величина смещения и стандартного отклонения оценок параметров регрессий в зависимости от соотношения весовых коэффициентов в модели Деминга.

Ключевые слова: метод наименьших квадратов; регрессия Деминга; геометрическая интерпретация; эллипс рассеяния

1 Введение

Одной из основных задач прикладного статистического анализа данных является выявление и построение зависимостей. Для ее решения разработано большое количество самых разнообразных методов, которые отличаются друг от друга как применяемым аналитическим аппаратом, так и набором условий и ограничений, при которых они могут быть корректно применены.

Наиболее широко проработана теоретическая база для метода наименьших квадратов (МНК) и различных его модификаций, которые отличаются простотой вычислений и понятной статистической интерпретацией получаемых на их основе результатов. Одним из предположений МНК является то, что значения входных переменных детерминированы (неслучайны), а значения отклика имеют стохастическую природу. Однако на практике при проведении экспериментов не всегда возможно обеспечить выполнение этого условия. Например, даже при использовании современных технических средств не всегда удается обеспечить точные заранее заданные условиями эксперимента значения всех или некоторых входных факторов. Это приводит к тому, что входные факторы также оказываются стохастическими. Следовательно, применение МНК для идентификации рассматриваемых моделей будет приводить к некоторым нежелательным эффектам, таким как смещенность и неэффективность оценок, гетероскедастичность и др.

Возможным выходом из этой ситуации является построение моделей, в которых изначально предполагается стохастический характер всех переменных. Для оценивания таких моделей разработан ряд методов, каждый из которых имеет определенные ограничения по области применения. Так, в методах группировки [1, 2] предполагается, что наблюдаемые значения независимых переменных упорядочены в соответствии с их истинными значениями [3].

Использование метода инструментальных переменных [2] связано с трудностями другого плана: новые переменные, выбираемые в качестве инструментальных, должны быть независимы от ошибок измерений регрессоров и тесно коррелировать с заменяемыми переменными.

Наличие проблем такого рода приводит к серьезному изменению постановки решаемой задачи, а также к необходимости разработки, анализа и применения специальных методов и алгоритмов ее решения.

В представленной работе задача идентификации статистической зависимости решается при помощи регрессии Деминга [4]. Исследуется техника оценивания такой регрессии, рассматриваются различные варианты интерпретации получаемых результатов. Кроме того, акцентируется внимание на особенностях этой постановки задачи и ее взаимосвязях с традиционными методами, такими как МНК.

¹Новосибирский государственный технический университет, netsc@rambler.ru

²Новосибирский государственный технический университет, gaix@ngs.ru

³Новосибирский государственный технический университет, supernasty@mail.ru

2 Постановка задачи

Пусть исследуется зависимость между двумя признаками (x, y) , причем признак x будем рассматривать как входной, а признак y — как выходной. По результатам экспериментов получена выборка $\{(x_i, y_i)\}_{i=1, \dots, N}$ значений этих признаков объемом N . В зависимости от выбираемой схемы эксперимента и характера самих признаков (стохастический или детерминированный) существует несколько вариантов проведения экспериментов [5]:

- схема S_1 : предполагается детерминированность входной переменной при стохастическом характере выходной;
- схема S_2 : предполагается детерминированность выходной переменной при стохастическом характере входной;
- схема S_3 : предполагается стохастический характер обеих переменных.

Для формализации предположений рассмотренных схем будем полагать, что стохастический характер переменных обуславливается наличием неконтролируемых ошибок, т. е. фактически вместо пар (x_i, y_i) фиксируются значения

$$\xi_i = x_i + \varepsilon_i^{(x)}; \quad (1)$$

$$\eta_i = y_i + \varepsilon_i^{(y)}, \quad (2)$$

где $\varepsilon_i^{(x)}$ и $\varepsilon_i^{(y)}$ — статистически независимые случайные величины с нулевыми математическими ожиданиями и постоянными дисперсиями $\sigma_{\varepsilon^{(x)}}^2$ и $\sigma_{\varepsilon^{(y)}}^2$ соответственно. Распределения ошибок во всех экспериментах считаются одинаковыми.

Удобной графической интерпретацией результатов эксперимента является корреляционное по-

ле [6]. Примеры корреляционных полей, соответствующих схемам S_1 , S_2 и S_3 , приведены на рис. 1, а, б и в соответственно.

Отметим, что применение одних и тех же методов для оценивания зависимостей, соответствующих рассмотренным корреляционным полям, может приводить к существенным искажениям получаемых оценок и строящихся на их основе интерпретационных выводов.

3 Контурные эллипсы для анализа корреляционного поля

Корреляционные поля, как правило, используются для визуального (предварительного) анализа собранной статистической информации, по итогам которого выдвигаются различные предположения о наличии, характере и форме изучаемой взаимосвязи. Поскольку визуальный анализ всегда отличается известной степенью субъективности, целесообразно проводить оценивание формы корреляционного поля на основе специальных методов, среди которых одним из наиболее известных считается построение контурных эллипсов (эллипсов рассеяния).

Базовым предположением этого метода является нормальность распределения рассматриваемых случайных величин: $\xi \sim N(\mu_\xi, \sigma_\xi)$ и $\eta \sim N(\mu_\eta, \sigma_\eta)$. Согласно [6], при статистической независимости переменных ξ и η верно соотношение:

$$\left(\frac{\xi - \mu_\xi}{\sigma_\xi}\right)^2 + \left(\frac{\eta - \mu_\eta}{\sigma_\eta}\right)^2 \sim \chi^2(2). \quad (3)$$

В этом случае граница соответствующего корреляционного поля может быть описана некоторым

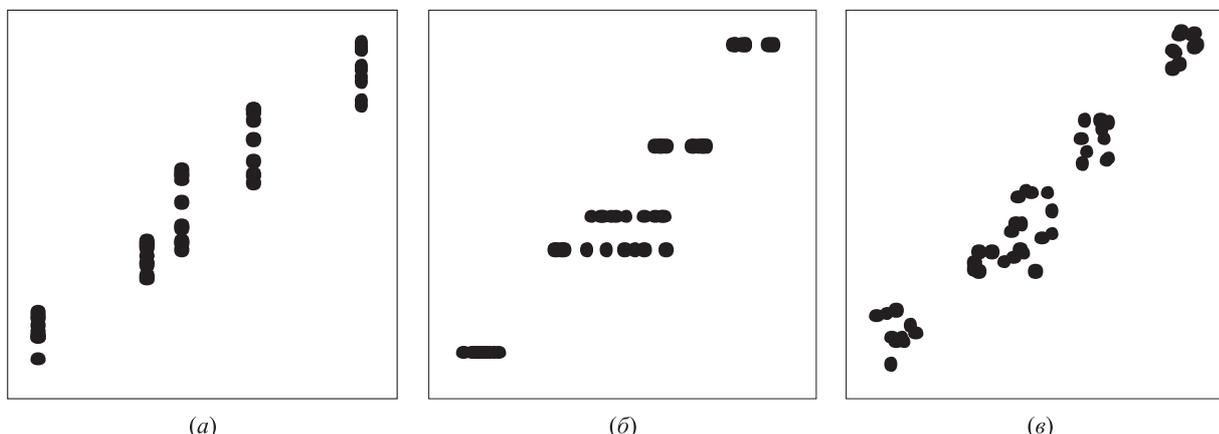


Рис. 1 Корреляционные поля рассматриваемых схем эксперимента

контурным эллипсом C . При этом вероятность попадания произвольной точки $(\xi^{(0)}, \eta^{(0)})$ внутрь эллипса равна

$$P \left\{ \left(\xi^{(0)}, \eta^{(0)} \right) \in C \right\} = P \left\{ \chi^2 < \chi_{кр}^2(1 - \alpha, 2) \right\}.$$

Отметим, что главные оси эллипса C будут параллельны координатным осям вследствие независимости ξ и η .

При работе с реальными данными необходимо строить оценку эллипса рассеяния, определяемую выражением:

$$\left(\frac{\xi - \bar{\xi}}{S_\xi} \right)^2 + \left(\frac{\eta - \bar{\eta}}{S_\eta} \right)^2 = \chi_{кр}^2(1 - \alpha, 2),$$

где $\chi_{кр}^2(1 - \alpha, 2)$ — критическое значение, определяемое по таблицам χ^2 -распределения; α — уровень значимости; $\bar{\xi}$ и $\bar{\eta}$ — выборочные средние; S_ξ и S_η — выборочные среднеквадратические отклонения по ξ и η соответственно.

При наличии линейной зависимости между переменными ξ и η , характеризуемой отличным от нуля коэффициентом корреляции $\rho_{\xi\eta}$, выражение (3) будет принимать вид [7, 8]:

$$\frac{1}{1 - \rho_{\xi\eta}^2} \left(\left(\frac{\xi - \mu_\xi}{\sigma_\xi} \right)^2 - 2\rho_{\xi\eta} \frac{\xi - \mu_\xi}{\sigma_\xi} \frac{\eta - \mu_\eta}{\sigma_\eta} + \left(\frac{\eta - \mu_\eta}{\sigma_\eta} \right)^2 \right) \sim \chi^2(2).$$

Оценка контурного эллипса может быть определена следующим образом:

$$\left(\frac{\xi - \bar{\xi}}{S_\xi} \right)^2 - 2r_{\xi\eta} \frac{\xi - \bar{\xi}}{S_\xi} \frac{\eta - \bar{\eta}}{S_\eta} + \left(\frac{\eta - \bar{\eta}}{S_\eta} \right)^2 = (1 - r_{\xi\eta}^2) \chi_{кр}^2(1 - \alpha, 2), \quad (4)$$

где $r_{\xi\eta}$ — оценка коэффициента корреляции $\rho_{\xi\eta}$. Очевидно, что при статистической независимости признаков эллипс занимает наибольший возможный объем, что соответствует нулевому значению коэффициента корреляции и максимальной неопределенности при прогнозировании.

4 Обобщенная модель регрессии Деминга

Следующий этап анализа состоит в построении и оценивании аналитической формы изучаемой

взаимосвязи. Дальнейшее изложение будет проводиться в рамках наиболее простой линейной зависимости, модель которой имеет вид:

$$y = \alpha + \beta x, \quad (5)$$

где α и β — параметры, подлежащие оцениванию. Заметим, что вследствие (1) и (2) будет рассматриваться зависимость вида $\eta = \alpha + \beta\xi$.

При оценивании параметров уравнения (5) в рамках схемы S_1 традиционно применяется метод наименьших квадратов [2, 5, 7], который может быть использован и для схемы S_2 путем идентификации модели обратной регрессии ($\xi = \alpha' + \beta'\eta$). В случае третьей схемы эксперимента такой подход может приводить к появлению существенно смещенных, несостоятельных и неэффективных оценок [5]. Возможным решением этой проблемы может быть переход к обобщенной модели регрессии Деминга [4, 9].

Регрессия Деминга, согласно [4], определяется как результат решения оптимизационной задачи

$$\sum_{i=1}^N \left[w_x^2 (\xi_i - x_i)^2 + w_y^2 (\eta_i - \alpha - \beta x_i)^2 \right] \rightarrow \min_{\alpha, \beta, x_i}, \quad (6)$$

где w_x и w_y — весовые коэффициенты по переменным x и y соответственно. Далее будем полагать, что они постоянны и не зависят от номера наблюдения. Поскольку, в отличие от стандартной постановки задачи оценивания параметров регрессии, решение задачи (6) зависит от неизвестных истинных значений переменной x , то приравнявая к нулю частную производную функционала в задаче (6) по значениям x_i , получим следующее выражение:

$$x_i = \frac{w_x^2 \xi_i + w_y^2 \beta (\eta_i - \alpha)}{w_x^2 + w_y^2 \beta^2}. \quad (7)$$

Учитывая соотношения (1) и (2), запишем задачу (6) в виде:

$$\sum_{i=1}^N \left[\left(w_x \varepsilon_i^{(x)} \right)^2 + \left(w_y \varepsilon_i^{(y)} \right)^2 \right] \rightarrow \min_{\alpha, \beta, x_i}. \quad (8)$$

Выражения, находящиеся под знаком суммы в (8), с геометрической точки зрения представляют собой квадраты диагоналей прямоугольников со сторонами $|w_x \varepsilon_i^{(x)}|$ и $|w_y \varepsilon_i^{(y)}|$, вершины которых соответствуют точкам (ξ_i, η_i) (см. рис. 1). Обозначая через

$$e_i^{(x)} = -\frac{1}{\beta} (\eta_i - \alpha - \beta \xi_i), \quad e_i^{(y)} = \eta_i - \alpha - \beta \xi_i$$

величины отклонений наблюдаемых значений от линии регрессии вдоль координатных осей и используя (7), выразим взвешенные ошибки в (8) через исходные данные $\{(\xi_i, \eta_i)\}_{i=\overline{1, N}}$:

$$w_x \varepsilon_i^{(x)} = -\frac{w_x w_y^2 \beta^2}{w_x^2 + w_y^2 \beta^2} e_i^{(x)};$$

$$w_y \varepsilon_i^{(y)} = \frac{w_y w_x^2}{w_x^2 + w_y^2 \beta^2} e_i^{(y)}.$$

Тогда задачу минимизации (6) можно упростить, исключив входную переменную x :

$$\frac{w_x^2 w_y^2}{w_x^2 + w_y^2 \beta^2} \sum_{i=1}^N (e_i^{(y)})^2 =$$

$$= \frac{1}{1/w_y^2 + (1/w_x^2)\beta^2} \sum_{i=1}^N (e_i^{(y)})^2 \rightarrow \min_{\alpha, \beta}. \quad (9)$$

Из (9) очевидно, что регрессия Деминга является обобщением прямой и обратной регрессий при $w_x^2 \gg w_y^2$ и $w_y^2 \gg w_x^2$ соответственно. Если предположить $w_x^2 = w_y^2$, то полученная регрессия будет называться ортогональной.

Для того чтобы выразить оценки параметров α , β через теоретические моменты распределения случайных величин ξ , η и оценить смещение оценок, перейдем в задаче (9) от суммы к математическому ожиданию:

$$\frac{1}{(1/w_y^2) + (1/w_x^2)\beta^2} E(\eta - \alpha - \beta\xi)^2 \rightarrow \min_{\alpha, \beta}. \quad (10)$$

Приравняв к нулю частную производную функционала в задаче (10) по параметру α , получаем выражение:

$$\alpha = E\eta - \beta E\xi.$$

Подставив это соотношение в частную производную от функционала в (10) по параметру β , приравняв ее к нулю и произведя некоторые упрощения, получим следующее уравнение:

$$\frac{1}{w_x^2} \beta E((\eta - E\eta) - \beta(\xi - E\xi))^2 +$$

$$+ \left(\frac{1}{w_y^2} + \frac{1}{w_x^2} \beta^2 \right) E(\xi(\eta - E\eta) - \beta\xi(\xi - E\xi)) = 0.$$

После упрощения этого соотношения с учетом справедливости равенства $E(\zeta_1(\zeta_2 - E\zeta_2)) = E((\zeta_1 - E\zeta_1)(\zeta_2 - E\zeta_2))$ для случайных величин ζ_1 и ζ_2 , уравнение примет такой вид:

$$\lambda^2 \sigma_{\xi\eta} \beta^2 - (\lambda^2 \sigma_\eta^2 - \sigma_\xi^2) \beta - \sigma_{\xi\eta} = 0, \quad (11)$$

где $\sigma_{\xi\eta}$ — ковариация признаков ξ и η , $\lambda^2 = w_y^2/w_x^2$. Решая квадратное уравнение относительно β , получаем:

$$\beta = \frac{1}{\sigma_{\xi\eta}} \left(\frac{\lambda^2 \sigma_\eta^2 - \sigma_\xi^2}{2\lambda^2} \pm \sqrt{\left(\frac{\lambda^2 \sigma_\eta^2 - \sigma_\xi^2}{2\lambda^2} \right)^2 + \frac{\sigma_{\xi\eta}^2}{\lambda^2}} \right). \quad (12)$$

Поскольку решение (12) должно иметь тот же знак, что и ковариация $\sigma_{\xi\eta}$, и второе слагаемое в скобках всегда больше первого, то второй множитель в (12) должен быть положительным, что достигается путем выбора знака «+». Это позволяет далее ограничиться рассмотрением только ситуации $\sigma_{\xi\eta} > 0$, в противном случае все последующие рассуждения останутся справедливыми с точностью до знака ковариации.

Проанализируем полученное решение (12). Предполагая, что переменная x имеет дисперсию σ_x^2 , воспользуемся следующими выражениями для моментов случайных величин ξ и η [10]:

$$\sigma_\xi^2 = \sigma_x^2 + \sigma_{\varepsilon^{(x)}}^2; \quad \sigma_\eta^2 = \beta^2 \sigma_x^2 + \sigma_{\varepsilon^{(y)}}^2; \quad \sigma_{\xi\eta} = \beta \sigma_x^2$$

и, подставляя их в (12), получим:

$$\beta_D = \frac{\Psi_0}{2\lambda^2} + \sqrt{\left(\frac{\Psi_0}{2\lambda^2} \right)^2 + \frac{1}{\lambda^2}}, \quad (13)$$

где $\Psi_0 = \lambda^2 \beta - 1/\beta + (\gamma_x^2/(\beta\delta^2))(\lambda^2 - \delta^2)$; $\delta^2 = \sigma_{\varepsilon^{(x)}}^2/\sigma_{\varepsilon^{(y)}}^2$; $\gamma_x^2 = \sigma_{\varepsilon^{(x)}}^2/\sigma_x^2$ — доля дисперсии случайной ошибки $\varepsilon^{(x)}$ в дисперсии входной переменной x , называемая уровнем шума. На практике при нахождении оценки $\hat{\beta}_D$ регрессии Деминга в (13) следует заменить все теоретические моменты их выборочными аналогами.

Из (13) следует, что если $\lambda = \delta$, т.е. веса в (6) обратно пропорциональны стандартным отклонениям ошибок соответствующих переменных, то $E\hat{\beta}_D = \beta$. Но на практике, как правило, дисперсии ошибок исследуемых признаков неизвестны, поэтому следует рассмотреть различные варианты соотношения весов w_x^2 и w_y^2 .

1. Пусть $w_x^2 \gg w_y^2$, т.е. $\lambda \rightarrow 0$. Тогда несложно заметить, что в пределе уравнение (11) становится линейным по β , а угловой коэффициент регрессии Деминга, с учетом определения коэффициента корреляции $\rho_{\xi\eta} = \sigma_{\xi\eta}/(\sigma_\xi\sigma_\eta)$, будет определяться как

$$\beta^{(y)} = \frac{\sigma_{\xi\eta}}{\sigma_\xi^2} = \frac{\sigma_\eta}{\sigma_\xi} \rho_{\xi\eta}, \quad (14)$$

что соответствует случаю прямой регрессии. Из (14) можно получить смещение оценки Деминга

$$\lim_{\lambda \rightarrow 0} E\hat{\beta}^{(y)} = E\left(\frac{\beta\sigma_x^2}{\sigma_x^2 + \sigma_{\varepsilon(x)}^2}\right) = \frac{1}{1 + \gamma_x^2}\beta.$$

В рамках схемы S_1 (т.е. при отсутствии ошибок по x) считается, что $\gamma_x^2 = 0$; следовательно, оценка параметра β будет несмещенной. Если же модель прямой регрессии будет оцениваться при наличии шума по x , оценка параметра β будет занижаться относительно истинного его значения.

2. Пусть $w_x^2 \ll w_y^2$, т.е. $\lambda \rightarrow \infty$. В этом случае при делении уравнения (11) на λ^2 и переходе к пределу имеем:

$$\beta(\sigma_{\xi\eta}\beta - \sigma_\eta^2) = 0.$$

Опуская тривиальное решение $\beta = 0$, получим, что угловой коэффициент регрессии Деминга имеет вид:

$$\beta^{(x)} = \frac{\sigma_\eta^2}{\sigma_{\xi\eta}} = \frac{\sigma_\eta}{\sigma_\xi} \frac{1}{\rho_{\xi\eta}}, \quad (15)$$

что соответствует случаю обратной регрессии. Из (15) смещение оценки Деминга определяется следующим образом:

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} E\hat{\beta}^{(x)} &= \beta + \frac{\gamma_x^2}{\beta\delta^2} = \left(1 + \frac{\gamma_x^2}{\beta^2\delta^2}\right)\beta = \\ &= (1 + \gamma_y^2)\beta. \end{aligned}$$

Если отсутствуют ошибки по y ($\gamma_y^2 = 0$), что соответствует схеме S_2 , оценка параметра β оказывается несмещенной. При наличии таких ошибок эта оценка будет приобретать смещение в сторону завышения.

3. Пусть $w_x^2 = w_y^2$, т.е. $\lambda = 1$. Получаемая регрессия будет ортогональной с угловым коэффициентом, равным

$$\begin{aligned} \beta^\perp &= \frac{\sigma_\eta^2 - \sigma_\xi^2}{2\sigma_{\xi\eta}} + \sqrt{\left(\frac{\sigma_\eta^2 - \sigma_\xi^2}{2\sigma_{\xi\eta}}\right)^2 + 1} = \\ &= \frac{2\rho_{\xi\eta}\Lambda}{1 - \Lambda^2 + \sqrt{(1 - \Lambda^2)^2 + 4\rho_{\xi\eta}^2\Lambda^2}}, \quad (16) \end{aligned}$$

где $\Lambda = \sigma_\eta/\sigma_\xi$. Этот случай соответствует схеме S_3 с одинаковой дисперсией ошибок.

4. Пусть отношение дисперсий ошибок переменных пропорционально отношению дисперсий наблюдаемых переменных ($\lambda^2 = 1/\Lambda^2$ или, для выборочных оценок, $\hat{\lambda}^2 = S_\xi^2/S_\eta^2$), что также соответствует схеме S_3 . В этом случае оценка Деминга принимает вид:

$$\hat{\rho}_D = \text{sign}(r_{\xi\eta}) \frac{S_\eta}{S_\xi}. \quad (17)$$

Оценка (17) получила название оценки диагональной регрессии, введенной Р. Фришем в 1929 г. [11]. Отметим, что без учета знака выражение (17) представляет собой среднее геометрическое оценок прямой и обратной регрессии. Следовательно, ее смещение будет зависеть от уровней шума по x и y одновременно, т.е.

$$E\hat{\beta}_D = \beta \sqrt{\frac{1 + \gamma_y^2}{1 + \gamma_x^2}}.$$

5 Графическая интерпретация регрессии Деминга

Построенная при тех или иных предположениях о соотношении дисперсий ошибок оценка параметра β модели (5) представляет собой тангенс угла наклона линии регрессии относительно положительного направления оси абсцисс. Следовательно, расположение линии регрессии в пространстве признаков будет варьироваться в зависимости от выбора весовых коэффициентов w_x, w_y , что определяет свойства минимизируемого функционала (9). Для построения графической интерпретации регрессии Деминга (рис. 2) функционал (9) может быть записан в следующих эквивалентных формах:

$$\frac{1}{1 + \lambda\beta^2} \sum_{i=1}^N (w_y e_i^{(y)})^2 \rightarrow \min_{\alpha, \beta} \quad (18)$$

или

$$\frac{\lambda\beta^2}{1 + \lambda\beta^2} \sum_{i=1}^N (w_x e_i^{(x)})^2 \rightarrow \min_{\alpha, \beta}. \quad (19)$$

Поскольку МНК предполагает минимизацию квадратов расстояний от точек корреляционного поля до искомой линии регрессии, то можно сопоставить каждой из этих точек квадрат, сторона которого определяется соответствующим расстоянием (см. рис. 2). В случае прямой регрессии это квадрат со стороной $e^{(y)}$, т.е. отрезком AA_y , что при суммировании по всем наблюдениям приводит

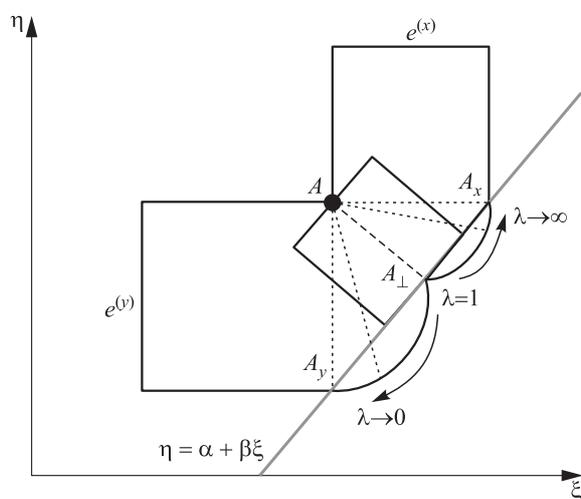


Рис. 2 Графическая интерпретация остатков регрессий Деминга

к задаче (18) при $\lambda = 0$. В случае обратной регрессии сторона квадрата равна $e^{(x)}$, т. е. длине отрезка AA_x , а решаемая задача имеет вид (19) при $\lambda = \infty$. Промежуточное положение между ними занимает ортогональная регрессия, для которой сторона квадрата равна AA_{\perp} , причем оценивание параметров можно осуществлять по любой из форм (18) или (19). Заметим, что длина отрезка AA_{\perp} есть минимальное расстояние от точки до линии регрессии, а условие $\lambda = 1$ означает равенство дисперсий ошибок по обоим переменным. Из рис. 2 также следует, что при увеличении дисперсии ошибок по переменной x рассматриваемый квадрат поворачивается относительно точки A вверх, пока не достигнет предельного положения, соответствующего обратной регрессии. Если же увеличивается дисперсия ошибок по переменной y , то квадрат поворачивается в другую сторону до предельного положения, определяемого линией прямой регрессии.

Следовательно, крайними случаями являются прямая и обратная регрессии, в то время как линии ортогональной и диагональной регрессии всегда расположены между ними. Кроме того, линия ортогональной регрессии совпадает с главной осью контурного эллипса (4) [6, 7]. Авторами доказано утверждение, определяющее расположение линии ортогональной регрессии относительно линий прямой, обратной и диагональной регрессий.

Утверждение. При любом значении коэффициента корреляции $r_{\xi\eta}$ ортогональная регрессия находится на наибольшем расстоянии (по значению угла наклона):

- (1) от прямой регрессии, если соответствующая диагональная регрессия образует угол $\pi/3$ с положительным направлением оси X ;

- (2) от обратной регрессии, если соответствующая диагональная регрессия образует угол $\pi/6$ с положительным направлением оси X .

Доказательство. Вследствие симметрии расположения ортогональной регрессии относительно прямой и обратной (см. выражения (14)–(16) для $\beta^{(y)}$, $\beta^{(x)}$ и β^{\perp}), достаточно ограничиться рассмотрением только прямой регрессии. Кроме того, без потери общности будем считать положительной корреляцию признаков $\rho_{\xi\eta} > 0$. Вычислим угол между ортогональной и прямой регрессией как угол между двумя прямыми [12] с угловыми коэффициентами β^{\perp} и $\beta^{(y)}$:

$$\begin{aligned} \operatorname{tg} \varphi &= \frac{\beta^{\perp} - \beta^{(y)}}{1 + \beta^{\perp}\beta^{(y)}} = \\ &= \frac{\rho_{\xi\eta}\Lambda \left(1 + \Lambda^2 - \sqrt{(1 - \Lambda^2)^2 + 4\rho_{\xi\eta}^2\Lambda^2}\right)}{1 - \Lambda^2 + 2\rho_{\xi\eta}\Lambda + \sqrt{(1 - \Lambda^2)^2 + 4\rho_{\xi\eta}^2\Lambda^2}}. \end{aligned}$$

Для определения максимального расстояния между регрессиями при фиксированном коэффициенте корреляции вычислим производную по Λ и приравняем к нулю числитель полученного выражения:

$$\begin{aligned} 2\rho_{\xi\eta}(1 - \rho_{\xi\eta})(1 + \rho_{\xi\eta})(\Lambda^2 - 3) \times \\ \times \left(\Lambda^2 - 1 - \sqrt{(\Lambda^2 - 1)^2 + (2\rho_{\xi\eta}\Lambda)^2}\right) = 0. \end{aligned}$$

Очевидно, что последний множитель никогда не обращается в нуль. Решение, получаемое при $\rho_{\xi\eta} = 0$, не имеет смысла, поскольку ему соответствует случай отсутствия линейной зависимости между признаками, а случаи $\rho_{\xi\eta} = 1$ и $\rho_{\xi\eta} = -1$ означают функциональную линейную зависимость, когда все рассматриваемые регрессии совпадают. Тогда наибольшее расстояние между регрессиями будет достигаться при выполнении условия $\Lambda^2 = \sigma_{\eta}^2/\sigma_{\xi}^2 = 3$. Из (17) очевидно, что угловой коэффициент диагональной регрессии есть $\Lambda = \sqrt{3}$, что соответствует углу $\pi/3$, образуемому с положительным направлением оси X . Утверждение доказано.

6 Свойства оценки регрессии Деминга

Кроме расположения линий регрессии в пространстве признаков задаваемое соотношение весовых коэффициентов w_x и w_y однозначно определяет и смещение оценок регрессии. Поэтому

определенный интерес представляет исследование влияния выбора значений λ на отклонение оценки Деминга β_D от истинного значения β .

При фиксированных значениях γ_x^2 и β определим зависимость $\Delta\beta = \beta_D/\beta$ от задаваемых величин λ^2 и коэффициента $\kappa = \lambda^2/\delta^2$, показывающего, во сколько раз заданное соотношение весов λ^2 отличается от соотношения дисперсий ошибок δ^2 . Далее будем использовать величину $\Delta\beta$ как относительное смещение оценки Деминга β_D от истинного значения параметра β . Без потери общности считаем $\beta > 0$. Используя (13), выразим $\Delta\beta$ через κ , γ_x^2 и

$$\gamma_y^2 = \frac{\sigma_{\varepsilon(y)}^2}{\sigma_y^2} = \frac{\sigma_{\varepsilon(y)}^2}{\beta^2 \sigma_x^2} = \frac{\sigma_{\varepsilon(x)}^2}{\delta^2} \frac{1}{\beta^2 \sigma_x^2} = \frac{\gamma_x^2}{\beta^2 \delta^2} = \frac{\kappa}{\beta^2 \lambda^2} \gamma_x^2,$$

тогда

$$\Delta\beta = \Psi_1 + \sqrt{\Psi_1^2 + \frac{\gamma_y^2}{\kappa \gamma_x^2}}, \quad (20)$$

где

$$\Psi_1 = \frac{1}{2} + \frac{1}{2} \gamma_y^2 \left(1 - \frac{1}{\kappa} - \frac{1}{\kappa \gamma_x^2} \right).$$

Оказывается, что в (20) $\Delta\beta$ не зависит от истинного значения параметра β . Поэтому можно проследить взаимосвязь между $\Delta\beta$ и κ при фиксированных значениях уровней шума по x и y (рис. 3). Для удобства интерпретации зависимости представлены в двойном логарифмическом масштабе и взяты по модулю. При этом левая часть графика попадает во вторую координатную четверть, зеркально отражаясь относительно оси абсцисс. Тогда очевидно, что при отклонении λ^2 от истинного значения δ^2 в большую сторону ($\ln \kappa > 0$) значение параметра β_D будет завышено по сравнению с истинным, а при $\kappa < 1$ ($\ln \kappa < 0$), наоборот, занижено. При этом

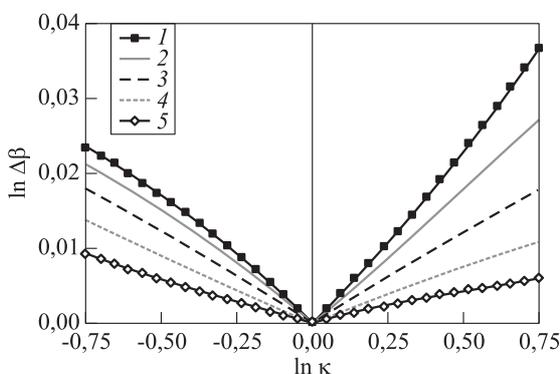


Рис. 3 Зависимость соотношения оценки и истинного значения параметра β от $\ln \kappa$ при $\gamma_x^2 = 0,05$ и заданном уровне шума γ_y^2 : 1 – 0,2; 2 – 0,1; 3 – 0,05; 4 – 0,025; 5 – 0,0125

величина отклонения $\Delta\beta$ при значениях κ и $1/\kappa$ различна и зависит от уровня шума. Отметим, что графики на рис. 3 несимметричны относительно оси ординат, за исключением случая, когда $\gamma_x^2 = \gamma_y^2$. Это позволяет условно разделить всю совокупность функций (20) на два подмножества. Если $\gamma_x^2 < \gamma_y^2$ (линии, проходящие выше графика функции при $\gamma_x^2 = \gamma_y^2$), то ошибка в оценивании коэффициента β будет возрастать быстрее при $\kappa > 1$, чем при $\kappa < 1$, т. е. если оценка отношения дисперсий ошибок λ^2 будет больше их истинного отношения. В другом подмножестве, когда уровень шума по x больше уровня шума по y , ошибка в оценивании коэффициента β будет возрастать быстрее при $\kappa < 1$, чем при $\kappa > 1$.

Такая особенность поведения функций (20) позволяет предложить следующую рекомендацию для выбора величины соотношения весов λ^2 : если известно, что уровень шума по переменной x больше уровня шума по y , то с точки зрения повышения точности оценивания параметра β выгоднее завышать отношение весов λ^2 , чем занижать его в такое же число раз. Аналогичный вывод можно сделать и для случая, когда уровень шума по x меньше уровня шума по y .

Еще одним вариантом графической интерпретации результатов оценивания при использовании регрессии Деминга может стать предлагаемая авторами «воронка Деминга» (рис. 4). Она представляет собой изображение семейства кривых, описывающих зависимость оценок стандартных отклонений $\hat{\sigma}_{\beta_D}$ оценок регрессии Деминга от их смещения $\Delta\beta$ в логарифмическом масштабе при фиксированном значении шума по x ($\gamma_x^2 = 0,05$) и разных уровнях шума по y . Оценки стандартных отклонений $\hat{\sigma}_{\beta_D}$ получены по результатам статистического моделирования с параметрами: $\beta = 1$, $X \sim N(10, 5)$, $N = 100$, $\varepsilon_i^{(x)} \sim N(0, \sigma_{\varepsilon(x)})$, $\varepsilon_i^{(y)} \sim N(0, \sigma_{\varepsilon(y)})$, число повторений – 10 000. При этом значения стандартных отклонений ошибок $\sigma_{\varepsilon(x)}$ и $\sigma_{\varepsilon(y)}$ рассчитывались исходя из фиксированных уровней шума γ_x^2 и γ_y^2 , а также оценки стандартного отклонения переменной X . В процессе моделирования соотношение весовых коэффициентов λ^2 задавалось регулярным образом по правилу:

$$\lambda^2 = \begin{cases} \delta^2 e^{0,3t}, & t = -12 + j, \gamma_y^2 > 0,0125; \\ \delta^2 e^{0,5t}, & t = -30,25 + 1,25j, \gamma_y^2 \leq 0,0125, \end{cases}$$

где $j = \overline{1, 30}$. Кроме того, отдельно рассмотрены частные случаи регрессии Деминга: прямая, обратная и ортогональная, оценки параметра β которых определялись из (14)–(16) путем замены теоретических моментов выборочными.

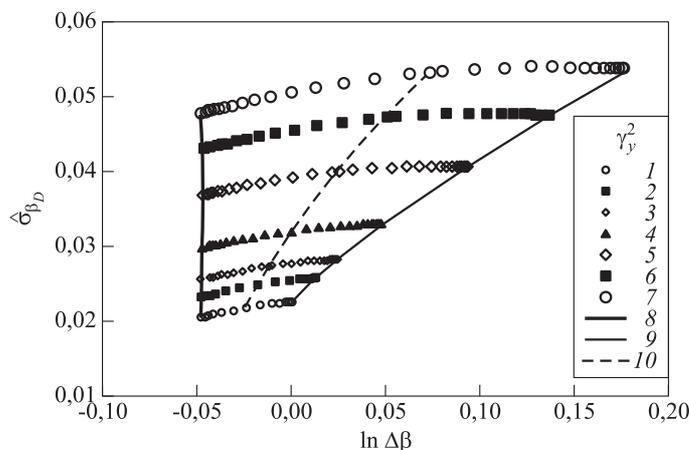


Рис. 4 «Воронка Деминга»: 1 — $\gamma_y^2 = 0,000001$; 2 — 0,0125; 3 — 0,025; 4 — 0,05; 5 — 0,1; 6 — 0,15; 7 — $\gamma_y^2 = 0,2$; 8 — прямая регрессия; 9 — обратная регрессия; 10 — ортогональная регрессия

Как видно из получаемого представления (см. рис. 4), границы воронки соответствуют прямой и обратной регрессиям (левая и правая границы соответственно), а линия, разделяющая воронку на две части — ортогональной. При этом, поскольку в эксперименте предполагается варьирование отношения весов при фиксированном значении γ_x^2 , левая граница воронки представляет собой прямую, параллельную оси ординат, а правая — кривую, которая удаляется от начала координат по мере увеличения веса, соответствующего переменной y . В целом, если оценка параметра β занижена по сравнению с его истинным значением, с увеличением смещения снижается стандартная ошибка оценки, достигая своего минимального значения в случае прямой регрессии. При завышении оценки параметра β ее стандартное отклонение увеличивается с ростом смещения, причем в ряде случаев (при уровне шума, превышающем 5%) достигает максимального значения и несколько снижается.

Все представленные выше исследования проводились в предположении (явном или неявном), что соотношение дисперсии ошибок является известной величиной. В то же время на практике проблема оценивания соотношения дисперсий ошибок остается одной из самых сложных. Традиционный подход предполагает, что дисперсии в точках факторного пространства должны оцениваться по повторным наблюдениям. Однако такие наблюдения зачастую или отсутствуют в выборке, или их доля достаточно мала, что не позволяет с требуемой степенью надежности оценивать соответствующие дисперсии. Выходом из такой ситуации может стать применение методов репликации выборки таких, например, как бутстреп или метод «складного ножа» (jack-knife) [13].

7 Заключение

В работе рассмотрены три основные схемы эксперимента, отличающиеся разной степенью детерминированности входных и выходных переменных модели. Отмечены некоторые статистические свойства контурных эллипсов, применяемых для оценивания конфигурации корреляционных полей.

Исследованы различные варианты регрессии Деминга как одной из наиболее общих регрессионных моделей, включая формы представления оптимизируемых функционалов, свойства получаемых оценок регрессии. Предложена оригинальная геометрическая интерпретация регрессии Деминга, иллюстрирующая особенности изменения элементов оптимизируемого функционала в зависимости от соотношения весовых коэффициентов. Доказано утверждение о взаимном расположении линий прямой, обратной, ортогональной и диагональной регрессий. Исследован характер влияния соотношения уровней шума по обоим переменным на величину смещения и стандартного отклонения оценки регрессии Деминга. Обнаруженные закономерности представлены в виде «воронки Деминга».

Полученные результаты и предложенные способы их интерпретации позволяют уточнить выводы, формируемые в процессе построения и последующего анализа зависимостей признаков стохастической природы.

Литература

1. Wald A. The fitting of straight lines if both variables are subject to error // Ann. Math. Stat., 1940. Vol. 11. No. 3. P. 284–300.

2. Демиденко Е. З. Линейная и нелинейная регрессия. — М.: Финансы и статистика, 1981.
3. Айвазян С. А., Богдановский И. М. Методы статистического исследования парных зависимостей в схемах конъюнктного анализа и их применение // Заводская лаборатория, 1974. Т. 40. № 3. С. 285–295.
4. Deming W. E. Statistical adjustment of data. — N.Y.: Dover Publications, 2011.
5. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Исследование зависимостей. — М.: Финансы и статистика, 1985.
6. Хальд А. Математическая статистика с техническими приложениями. — М.: Изд-во иностранной литературы, 1956.
7. Крамер Г. М. Математические методы статистики. — М.: Мир, 1975.
8. Львовский Е. Н. Статистические методы построения эмпирических формул. — М.: Высшая школа, 1988.
9. Jones T. A. Fitting straight lines when both variables are subject to error. I. Maximum likelihood and least-squares estimation // Math. Geology, 1979. Vol. 11. No. 1. P. 1–25.
10. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973.
11. Frisch R. Correlation and scatter in statistical variables // Nord. Stat. J., 1929. No. 1. P. 36–102.
12. Гельфанд И. М. Лекции по линейной алгебре. — М.: Наука, 1971.
13. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. — М.: Финансы и статистика, 1988.

АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ ОЦЕНКИ ПАРАМЕТРОВ МНОГОМЕРНОЙ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

А. Ю. Хапланов¹

Аннотация: Построена оценка вектора коэффициентов многомерной логистической регрессии, когда размерность этого вектора растет с увеличением объема выборки. Установлена оценка скорости сходимости предложенной оценки, а также ее асимптотическая нормальность.

Ключевые слова: логистическая регрессия; скорость сходимости; асимптотическая нормальность; предикторы высокой размерности

1 Введение

В современной информатике методы машинного обучения (см., например, [1]) широко используются для решения многих типов вероятностных задач. Достаточно упомянуть кластеризацию и понижение размерности наблюдений. В представленной работе исследуется проблема классификации данных. Распространенным подходом к ее решению является применение метода логистической регрессии (см., например, [2, 3]). Этот метод часто используется для распознавания образов, анализа текстов и решения многих других прикладных задач (см., например, [4, 5]). В модели логистической регрессии предполагается, что бинарная случайная величина Y определенным образом зависит от вектора факторов $X = (X_1, \dots, X_p)$, принимающего значения в пространстве \mathbb{R}^p . В статьях [6, 7] изучался случай, когда Y принимает значения в множестве $\{0, 1, \dots, k\}$. В данной работе исследуется обобщение этой модели. Пусть для каждого $n \in \mathbb{N}$ величина $Y = Y(n)$ отображает пространство элементарных исходов Ω в множество $\{0, 1, \dots, k\}$, а $X = X(n)$ — случайный вектор со значениями в \mathbb{R}^{p_n} . Предположим, что существует параметр $\alpha^{0,n} \in \mathbb{R}^{k p_n}$ логистической зависимости $Y(n)$ от $X(n)$ (см., например, [8]). Необходимые определения даются в разд. 2. По независимым наблюдениям векторов $(X_q(n)^T, Y_q(n)^T)$, $q = 1, \dots, n$, имеющих такое же распределение, как $(X(n)^T, Y(n)^T)$, строится оценка $\hat{\alpha}_n$ вектора $\alpha^{0,n}$. Символ T обозначает транспонирование. Основным результатом данной работы является доказательство асимптотической нормальности величины $\hat{\alpha}_n - \alpha^{0,n}$ при $n \rightarrow \infty$. Тем самым обобщаются работы [8, 9], в которых рассматривался случай $k = 1$. Кроме того, устанавливается оценка скорости сходимости $\hat{\alpha}_n$ к $\alpha^{0,n}$ при $n \rightarrow \infty$.

2 Основные результаты

Пусть для каждого $n \in \mathbb{N}$ дан случайный вектор $(X(n)^T, Y(n)^T)$ такой, что $X(n)$ принимает значения в \mathbb{R}^{p_n} , а $Y(n)$ — в множестве $\{0, \dots, k\}$, где k и p_n являются натуральными числами. Для удобства записи введем $K = \{1, \dots, k\}$. В работе исследуется многомерная логистическая зависимость изучаемой переменной от факторов (см., например, [6]). А именно: пусть для $n \in \mathbb{N}$, $j \in K$ и $x \in \mathbb{R}^{p_n}$ справедливы соотношения

$$P(Y(n) = j | X(n) = x) = \frac{\exp\{-(\alpha_j^{0,n})^T x\}}{1 + \sum_{t=1}^k \exp\{-(\alpha_t^{0,n})^T x\}};$$

$$P(Y(n) = 0 | X(n) = x) = \frac{1}{1 + \sum_{t=1}^k \exp\{-(\alpha_t^{0,n})^T x\}},$$

где $\alpha_j^{0,n}$ — неслучайные векторы в пространстве \mathbb{R}^{p_n} . Обозначим $\alpha^{0,n} = \{(\alpha_1^{0,n})^T, \dots, (\alpha_k^{0,n})^T\}^T$. В дальнейшем для любого вектора $\alpha \in \mathbb{R}^{k p_n}$ будет удобно использовать двойную нумерацию компонент, т.е. $\alpha^{(r,l)} = \alpha^{l+(r-1)p_n}$, $r \in K$, $l \in \{1, \dots, p_n\}$. Тогда $(\alpha^{0,n})^{(r,l)} = (\alpha_r^{0,n})^l$. Положим $\alpha^{(r,\cdot)} := \{\alpha^{(r,1)}, \dots, \alpha^{(r,p_n)}\}^T$. Определим функции $\{\pi_j^n(\alpha, x)\}_{j=0}^k$ такие, что для $j \in K$, $\alpha \in \mathbb{R}^{k p_n}$ и $x \in \mathbb{R}^{p_n}$

$$\pi_j^n(\alpha, x) = \frac{\exp\left\{-\left(\alpha^{(j,\cdot)}\right)^T x\right\}}{1 + \sum_{t=1}^k \exp\left\{-\left(\alpha^{(t,\cdot)}\right)^T x\right\}};$$

$$\pi_0^n(\alpha, x) = \frac{1}{1 + \sum_{t=1}^k \exp\left\{-\left(\alpha^{(t,\cdot)}\right)^T x\right\}}.$$

¹Московский государственный университет им. М. В. Ломоносова, Khaplanova@gmail.com

Предполагая, что случайный вектор $X(n)$ имеет плотность распределения $g_n(x)$, запишем функцию правдоподобия для наблюдений $(X_q(n)^T, Y_q(n))^T$, $q = 1, \dots, n$. Имеем:

$$L_n(\alpha) = \prod_{q=1}^n \left(\pi_{Y_q(n)}^n(\alpha, X_q(n)) g_n(X_q(n)) \right) = \prod_{q=1}^n \left\{ \prod_{j=0}^k [\pi_j^n(\alpha, X_q(n))]^{I\{Y_q(n)=j\}} \right\} \prod_{i=1}^n g_n(X_i(n)),$$

где $I\{A\}$ — индикатор множества A . Будем считать, что последний множитель отличен от нуля. Тогда он никак не повлияет на исследование поведения функции $L_n(\alpha)$ по аргументу α . Поэтому функция плотности $g_n(x)$ не учитывается, когда исследуется модель логистической регрессии (см., например, [10]). Введем

$$\tilde{L}_n(\alpha) = \prod_{q=1}^n \left(\pi_{Y_q(n)}^n(\alpha, X_q(n)) \right) = \prod_{q=1}^n \left\{ \prod_{j=0}^k [\pi_j^n(\alpha, X_q(n))]^{I\{Y_q(n)=j\}} \right\}.$$

Положим $\mathcal{L}_n(\alpha) = \ln \tilde{L}_n(\alpha)$. Справедливо равенство:

$$\mathcal{L}_n(\alpha) = \sum_{q=1}^n \left[\sum_{j=0}^k I\{Y_q(n) = j\} \ln \pi_j^n(\alpha, X_q(n)) \right].$$

Рассмотрим функцию:

$$R_n(\alpha) = -\nabla \mathcal{L}_n(\alpha) = \left\{ R_n^{(1,1)}(\alpha), \dots, \dots, R_n^{(1,p_n)}(\alpha), R_n^{(2,1)}(\alpha), \dots, R_n^{(k,p_n)}(\alpha) \right\}^T, \quad \alpha \in \mathbb{R}^{kp_n},$$

где

$$R_n^{(r,l)}(\alpha) = \sum_{q=1}^n [I\{Y_q(n) = r\} - \pi_r^n(\alpha, X_q(n))] X_q^l(n).$$

Пусть $\hat{\alpha}_n$ — вектор с минимальной евклидовой нормой, для которого $R_n(\hat{\alpha}_n) = \bar{0}$, здесь $\bar{0}$ — нулевой вектор в пространстве \mathbb{R}^{kp_n} . Если решение этого уравнения определяется неоднозначно, то в качестве оценки используется корень, минимальный в лексикографическом смысле. В случае когда корней не существует, полагаем $\hat{\alpha}_n = \bar{0}$. Далее будет доказано, что с вероятностью, стремящейся к единице при $n \rightarrow \infty$, у функции $R_n(\alpha)$ существует

как минимум один корень, а построенный вектор $\hat{\alpha}_n$ является асимптотически нормальной оценкой $\alpha^{0,n}$. Для того чтобы сформулировать основные результаты данной работы, понадобятся новые обозначения. Заметим, что для исследуемой модели $\mathbb{E}(I\{Y(n) = r\} | X(n)) = \pi_r^n(\alpha^{0,n}, X(n))$. Значит, $\mathbb{E}R_n(\alpha^{0,n}) = \bar{0}$. Введем матрицу вторых производных функции $-\mathcal{L}_n(\alpha)$:

$$Q_n(\alpha) = \nabla R_n(\alpha)^T = \left(u_{(m,s)}^{(r,l)}(\alpha) \right),$$

где (r, l) и (m, s) обозначают номера строки и столбца элементов матрицы $Q_n(\alpha)$ в двойной нумерации. Функции $u_{(m,s)}^{(r,l)}(\alpha)$ для $r, m \in K, l, s \in \{1, \dots, p_n\}$ и $\alpha \in \mathbb{R}^{kp_n}$ удовлетворяют соотношению

$$u_{(m,s)}^{(r,l)}(\alpha) = \sum_{q=1}^n H_{r,m}^n(\alpha, X_q(n)) X_q^l(n) X_q^s(n),$$

где $H_{r,m}^n(\alpha, x)$ для $n \in \mathbb{N}, \alpha \in \mathbb{R}^{kp_n}$ и $x \in \mathbb{R}^{p_n}$ определяются следующим образом:

$$H_{r,m}^n(\alpha, x) = \begin{cases} -\pi_r^n(\alpha, x) \pi_m^n(\alpha, x), & \text{если } r \neq m, \\ \pi_r^n(\alpha, x) (1 - \pi_r^n(\alpha, x)), & \text{если } r = m. \end{cases} \quad (1)$$

Введем также квадратную матрицу $B_n = \left(b_{(m,s)}^{(r,l)}(\alpha^{0,n}) \right)$ порядка kp_n , для которой $b_{(m,s)}^{(r,l)}(\alpha) = H_{r,m}^n(\alpha, X(n)) X^l(n) X^s(n)$. Положим $G_n(\alpha) = \mathbb{E}Q_n(\alpha)$. Важно отметить, что $G_n(\alpha^{0,n})$ совпадает с ковариационной матрицей вектора $R_n(\alpha^{0,n})$. Для упрощения записи далее пишем G_n вместо $G_n(\alpha^{0,n})$. Легко заметить, что $G_n = nB_n$. Далее понадобятся следующие условия.

1. Существует константа $C > 0$ такая, что для всех $n \in \mathbb{N}$ неравенство $\|X(n)\| \leq C$ выполняется почти наверное.

2. Существует число $c > 0$, для которого неравенство $m_n^T B_n m_n \geq c \|m_n\|^2$ выполняется при всех $n \in \mathbb{N}$ и $m_n \in \mathbb{R}^{kp_n}$, где $\|\cdot\|$ — евклидова норма.

Замечание 1. Из определения матрицы B_n следует, что условие 2 будет иметь место, например, когда вектор $X(n)$ нормально распределен с единичной ковариационной матрицей и существует натуральное число s такое, что $(\alpha^{0,n})^{(r,l)} = 0$ при всех $n \in \mathbb{N}, r \in K$ и $l \geq s$.

Чтобы сформулировать основной результат, введем неслучайную матрицу U_n размера $[k \times kp_n]$:

$$U_n = \begin{pmatrix} \overbrace{\frac{1}{\sqrt{p_n}} \dots \frac{1}{\sqrt{p_n}}}^{p_n} & \overbrace{0 \dots 0}^{p_n} & \dots & \overbrace{0 \dots 0}^{p_n} \\ 0 & \dots & 0 & \overbrace{\frac{1}{\sqrt{p_n}} \dots \frac{1}{\sqrt{p_n}}}^{p_n} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & \overbrace{\frac{1}{\sqrt{p_n}} \dots \frac{1}{\sqrt{p_n}}}^{p_n} \end{pmatrix}.$$

Теорема 1. *Предположим, что выполняются условия 1 и 2. Тогда для любой последовательности $\delta_n > 0$, для которой $\delta_n p_n / \sqrt{n} \rightarrow 0$ и $\sqrt{p_n} / \delta_n \rightarrow 0$ при $n \rightarrow \infty$, имеет место соотношение:*

$$\mathbb{P} \left(\|\alpha^{0,n} - \hat{\alpha}_n\| \geq \frac{\delta_n}{\sqrt{n}} \right) \rightarrow 0, \quad n \rightarrow \infty.$$

Теорема 2. *Пусть выполняются условия 1, 2 и $p_n = o(n^{1/3})$ при $n \rightarrow \infty$. Тогда*

$$U_n G_n^{1/2} (\hat{\alpha}_n - \alpha^{0,n}) \xrightarrow{d} Z, \quad Z \sim N(0, E_k), \quad n \rightarrow \infty,$$

где E_k — единичная матрица порядка k .

Следствие 1. *Если выполняются условия теоремы 2, то при $n \rightarrow \infty$ справедливы соотношения:*

$$\left\| \frac{Q_n(\hat{\alpha}_n)}{n} - \frac{G_n}{n} \right\|_2 \xrightarrow{P} 0;$$

$$U_n Q_n(\hat{\alpha}_n)^{1/2} (\hat{\alpha}_n - \alpha^{0,n}) \xrightarrow{d} Z, \quad Z \sim N(0, E_k),$$

где $\|M\|_2$ — операторная норма матрицы M .

Для доказательства этих теорем понадобится несколько элементарных вспомогательных утверждений.

Лемма 1. *Пусть дано отображение $F : \mathbb{R}^u \rightarrow \mathbb{R}^v$ с непрерывными частными производными. Тогда для $x_1, x_2 \in \mathbb{R}^u$ выполняется тождество:*

$$F(x_1) - F(x_2) = \int_0^1 (\nabla_x (F(x)^T)|_{x=x_2+u(x_1-x_2)}) du \Big|^T (x_1 - x_2),$$

где интеграл от матрицы берется поэлементно.

Если записать определение первообразной функции для $F(x)$, утверждение этой леммы станет очевидным.

Лемма 2 (см., например, [11]). *Пусть дано непрерывное инъективное отображение $\Gamma : \mathbb{R}^p \rightarrow \mathbb{R}^p$. Фиксируем точку $x_0 \in \mathbb{R}^p$. Предположим, что $\inf_{\|x-x_0\|=\delta} \|\Gamma(x) - \Gamma(x_0)\| \geq R$ для некоторых констант $\delta, R > 0$. Тогда для любого $y \in \{u \in \mathbb{R}^p : \|u - \Gamma(x_0)\| \leq R\}$ существует такая точка $x(y)$, что $\Gamma(x(y)) = y$ и $\|x(y) - x_0\| \leq \delta$.*

Лемма 3. *Если выполняются условия теоремы 1, то при $n \rightarrow \infty$ имеет место предельное соотношение*

$$\sup_{\alpha \in N_n(\delta_n)} \|G_n^{-1/2} Q_n(\alpha) G_n^{-1/2} - E_{kp_n}\|_2 \xrightarrow{P} 0, \quad (2)$$

где

$$N_n(\delta_n) = \{\alpha \in \mathbb{R}^{kp_n} : \|G_n^{1/2}(\alpha - \alpha^{0,n})\| \leq \delta_n\};$$

E_{kp_n} — единичная матрица порядка kp_n .

3 Доказательства

Доказательство леммы 3. Оценим норму, фигурирующую в (2). Имеем:

$$\begin{aligned} & \left\| G_n^{-1/2} Q_n(\alpha_n) G_n^{-1/2} - E_n \right\|_2 = \\ & = \left\| G_n^{-1/2} [Q_n(\alpha_n) - G_n(\alpha^{0,n})] G_n^{-1/2} \right\|_2 = \\ & = \left\| G_n^{-1/2} [Q_n(\alpha_n) - Q_n(\alpha^{0,n}) + Q_n(\alpha^{0,n}) - G_n(\alpha^{0,n})] G_n^{-1/2} \right\|_2 \leq \left\| G_n^{-1/2} \sqrt{n} \right\|_2 \times \\ & \quad \times \left\| \frac{Q_n(\alpha_n) - Q_n(\alpha^{0,n})}{n} \right\|_2 \left\| G_n^{-1/2} \sqrt{n} \right\|_2 + \\ & \quad + \left\| G_n^{-1/2} \sqrt{n} \right\|_2 \left\| \frac{Q_n(\alpha^{0,n}) - G_n(\alpha^{0,n})}{n} \right\|_2 \times \\ & \quad \times \left\| G_n^{-1/2} \sqrt{n} \right\|_2. \quad (3) \end{aligned}$$

Величина $\left\| G_n^{-1/2} \sqrt{n} \right\|_2$ ограничена константой, которая по условию 2 не зависит от n . Используя неравенство $\|A\|_2 \leq v \|A\|_\infty$, где A — квадратная матрица порядка v и $\|A\|_\infty = \max_{i,j} \{ |a_{i,j}| \}$, получаем, что

$$\begin{aligned} & \left\| \frac{Q_n(\alpha_n) - Q_n(\alpha^{0,n})}{n} \right\|_2 \leq \\ & \leq kp_n \left\| \frac{Q_n(\alpha_n) - Q_n(\alpha^{0,n})}{n} \right\|_\infty = \\ & = \frac{kp_n}{n} \max_{r,m,l,s} \left\{ \left| \sum_{q=1}^n [H_{r,m}^n(\alpha_n, X_q(n)) - H_{r,m}^n(\alpha^{0,n}, X_q(n))] X_q^l(n) X_q^s(n) \right| \right\} \leq \\ & \leq kp_n \max_{r,m,l,s,q} \left\{ |H_{r,m}^n(\alpha_n, X_q(n)) - H_{r,m}^n(\alpha^{0,n}, X_q(n))| C^2 \right\}, \quad (4) \end{aligned}$$

где C — константа, для которой $\mathbb{P}(\|X(n)\| \leq C) = 1$ при $n \in \mathbb{N}$. Из определения функции $H_{r,m}^n(\alpha, x)$ и равенства (1) следует существование числа $L \in \mathbb{R}_+$ такого, что для $n \in \mathbb{N}$ выполняется неравенство:

$$|H_{r,m}^n(\alpha_n, X_q(n)) - H_{r,m}^n(\alpha^{0,n}, X_q(n))| \leq \|\alpha_n - \alpha^{0,n}\| L. \quad (5)$$

Пользуясь определением множества $N_n(\delta_n)$, оценим правую часть неравенства (5):

$$\|\alpha_n - \alpha^{0,n}\| = \|G_n^{-1/2} G_n^{1/2}(\alpha_n - \alpha^{0,n})\| \leq \|G_n^{-1/2}\|_2 \|G_n^{1/2}(\alpha_n - \alpha^{0,n})\| \leq \|G_n^{-1/2}\|_2 \sqrt{n} \left\| \frac{\delta_n}{\sqrt{n}} \right\|.$$

Цепочка неравенств (4) может быть продолжена следующим образом:

$$kp_n \max_{r,m,l,s,q} \{ |H_{r,m}^n(\alpha_n, X_q(n)) - H_{r,m}^n(\alpha^{0,n}, X_q(n))| C^2 \} \leq \frac{kp_n \delta_n}{\sqrt{n}} C^2 L \|G_n^{-1/2}\|_2. \quad (6)$$

Заметим, что при выполнении условий леммы 3 первое слагаемое в (3) стремится к нулю, когда $n \rightarrow \infty$. Далее

$$\begin{aligned} & \left\| \frac{Q_n(\alpha^{0,n}) - G_n(\alpha^{0,n})}{n} \right\|_2 \leq \left\| \frac{Q_n(\alpha^{0,n}) - G_n(\alpha^{0,n})}{n} \right\|_F = \frac{1}{n} \times \\ & \times \left(\sum_{j_1, j_2, i_1, i_2}^{k, k, p, p} \left\{ \sum_{q=1}^n X_q^{j_1}(n) X_q^{j_2}(n) H_{i_1, i_2}^n(\alpha^{0,n}, X_q(n)) - \mathbb{E} [X_q^{j_1}(n) X_q^{j_2}(n) H_{i_1, i_2}^n(\alpha^{0,n}, X_q(n))] \right\}^2 \right)^{1/2}, \end{aligned}$$

где $\|A\|_F = \sqrt{\sum_{i,j} (a_{i,j})^2}$ является нормой Фробениуса матрицы $A = (a_{i,j})$. В силу неравенства Чебышёва для векторов (см., например, [12]) имеем:

$$\begin{aligned} & \mathbb{P} \left\{ \left\| \frac{Q_n(\alpha^{0,n}) - G_n(\alpha^{0,n})}{n} \right\|_2 \geq \varepsilon \right\} \leq \\ & \leq \mathbb{P} \left(\sum_{j_1, j_2, i_1, i_2} \left\{ \sum_{q=1}^n X_q^{j_1}(n) X_q^{j_2}(n) H_{i_1, i_2}^n(\alpha^{0,n}, X_q(n)) - \mathbb{E} [X_q^{j_1}(n) X_q^{j_2}(n) H_{i_1, i_2}^n(\alpha^{0,n}, X_q(n))] \right\}^2 \geq \varepsilon^2 n^2 \right) \end{aligned}$$

$$\begin{aligned} & - \mathbb{E} [X_q^{j_1}(n) X_q^{j_2}(n) H_{i_1, i_2}^n(\alpha^{0,n}, X_q(n))] \}^2 \geq n^2 \varepsilon^2 \} \leq \\ & \leq \left(\sum_{j_1, j_2, i_1, i_2}^{k, k, p, p} \left\{ \sum_{q=1}^n \mathbb{D} (X_q^{j_1}(n) X_q^{j_2}(n) \times \right. \right. \\ & \left. \left. \times H_{i_1, i_2}^n(\alpha^{0,n}, X_q(n))) \right\} \right) / (\varepsilon^2 n^2) \leq \\ & \leq \frac{k^2 p_n^2 n B}{\varepsilon^2 n^2} = \frac{p_n^2 k^2 B}{n \varepsilon^2}. \quad (7) \end{aligned}$$

Существование такой константы B , для которой $\mathbb{D} (X_q^{j_1} X_q^{j_2} H_{i_1, i_2}(\alpha_n X_q)) \leq B$ при всех $n \in \mathbb{N}$, $i_1, i_2 \in K$, $j_1, j_2 \in \{1, \dots, p_n\}$ и $q \in \{1, \dots, n\}$, следует из ограниченности функций $H_{i_1, i_2}^n(\alpha, x)$ и условия 2. Подставляя (6) и (7) в (3), получаем утверждение леммы.

Доказательство теоремы 1. Напомним, что $N_n(\delta_n) = \{\alpha \in \mathbb{R}^{kp_n} : \|G_n^{1/2}(\alpha - \alpha^{0,n})\| \leq \delta_n\}$. Основная идея доказательства теоремы состоит том, что уравнение $R_n(\alpha) = \bar{0}$ имеет решения $\alpha(n) \in N_n(\delta_n)$ с вероятностью, стремящейся к единице при $n \rightarrow \infty$. Для проверки этого рассмотрим функцию $F_n(\alpha) = G_n^{-1/2} [R_n(\alpha) - R_n(\alpha^{0,n})]$. Легко заметить, что с вероятностью, стремящейся к единице при $n \rightarrow \infty$, эта функция является непрерывным автоморфизмом некоторой окрестности точки $\alpha^{0,n}$ в пространстве \mathbb{R}^{kp_n} . Нетрудно видеть, что $F(\alpha^{0,n}) = \bar{0}$. Рассмотрим цепочку неравенств, в которой будет использоваться неравенство Чебышёва для векторов и тот факт, что G_n — ковариационная матрица вектора $R_n(\alpha^{0,n})$.

$$\begin{aligned} & \mathbb{P} \left(\|G_n^{-1/2} R_n(\alpha^{0,n})\| \leq c_0 \delta_n \right) \geq \\ & \geq 1 - \frac{\mathbb{E} \|G_n^{-1/2} R_n(\alpha^{0,n})\|^2}{(c_0 \delta_n)^2} = 1 - \frac{kp_n}{(c_0 \delta_n)^2}. \end{aligned}$$

Таким образом, $\mathbb{P}(\|G^{-1/2} R_n(\alpha^{0,n})\| \leq c_0 \delta_n) \rightarrow 1$ при $n \rightarrow \infty$. Теперь покажем, что $\mathbb{P} \left(\inf_{\alpha \in \partial N_n(\delta_n)} \|F_n(\alpha)\| \geq c_0 \delta_n \right) \rightarrow 1$, где $\partial N_n(\delta_n) = \{\alpha \in \mathbb{R}^{kp_n} : \|G_n^{1/2}(\alpha - \alpha^{0,n})\| = \delta_n\}$. Применим лемму 1 к функции $R_n(\alpha)$. Тогда

$$\begin{aligned} & R_n(\alpha) - R_n(\alpha^{0,n}) = \\ & = \left[\int_0^1 \nabla_x (R_n(x)^T) |_{x=\alpha^{0,n} + u(\alpha - \alpha^{0,n})} du \right]^T (\alpha - \alpha^{0,n}). \end{aligned}$$

Обозначая через $\tilde{Q}_n(\alpha)$ последний интеграл, получаем $R_n(\alpha) - R_n(\alpha^{0,n}) = \tilde{Q}_n(\alpha)(\alpha - \alpha^{0,n})$. Отсюда

$$\begin{aligned} & \inf_{\alpha \in \partial N_n(\delta_n)} \|G_n^{-1/2}(R_n(\alpha) - R_n(\alpha^{0,n}))\|^2 = \\ & = \inf_{\alpha \in \partial N_n(\delta_n)} \|G_n^{-1/2}\tilde{Q}_n(\alpha)(\alpha - \alpha^{0,n})\|^2 = \\ & = \inf_{\alpha \in \partial N_n(\delta_n)} \left[(\alpha - \alpha^{0,n})^T \tilde{Q}_n(\alpha)^T G_n^{-1} \times \right. \\ & \quad \left. \times \tilde{Q}_n(\alpha)(\alpha - \alpha^{0,n}) \delta_n^2 \right] / \delta_n^2. \end{aligned}$$

В этих переходах применялось равенство $\|m\|^2 = m^T m$, где m — произвольный вектор. Воспользуемся неравенством Коши–Буняковского–Шварца для векторов $(\alpha - \alpha^{0,n})^T \tilde{Q}_n(\alpha) G_n^{-1/2}$ и $G_n^{1/2}(\alpha - \alpha^{0,n})$. В конце следующей цепочки неравенств $(\alpha - \alpha^{0,n})^T G_n^{1/2} / \delta_n$ заменяется произвольным вектором e единичной длины. Имеем

$$\begin{aligned} & \mathbb{P} \left(\inf_{\alpha \in \partial N_n(\delta_n)} \|G_n^{-1/2}\{R_n(\alpha) - R_n(\alpha^{0,n})\}\| \geq c_0 \delta_n \right) \geq \\ & \geq \mathbb{P} \left(\inf_{\alpha \in \partial N_n(\delta_n)} \frac{(\alpha - \alpha^{0,n})^T \tilde{Q}_n(\alpha)^T (\alpha - \alpha^{0,n})}{\delta_n} \geq \right. \\ & \quad \left. \geq c_0 \delta_n \right) \geq \\ & \geq \mathbb{P} \left(\inf_{\substack{\alpha \in \partial N_n(\delta_n), \\ \|e\|=1}} \left[e^T G_n^{-1/2} \tilde{Q}_n(\alpha)^T G_n^{-1/2} e \right] \geq c_0 \right). \end{aligned}$$

Из леммы 3 следует, что

$$\mathbb{P} \left(\inf_{\alpha \in \partial N_n(\delta_n)} \|G_n^{-1/2}\{R_n(\alpha) - R_n(\alpha^{0,n})\}\| \geq \geq c_0 \delta_n \right) \rightarrow 1, \quad n \rightarrow \infty. \quad (8)$$

Значит, для любого $c_0 \in (0, 1)$ при $n \rightarrow \infty$ будет верно

$$\mathbb{P} \left(\inf_{\alpha \in \partial N_n(\delta_n)} \|G_n^{-1/2}\{R_n(\alpha) - R_n(\alpha^{0,n})\}\| \geq \geq \|G_n^{-1/2} R_n(\alpha^{0,n})\| \right) \rightarrow 1.$$

Из леммы 2 и последнего соотношения вытекает, что с вероятностью, стремящейся к единице при $n \rightarrow \infty$, существует такое $\hat{\alpha}_n \in N_n(\delta_n)$, для которого $R_n(\hat{\alpha}_n) = \bar{0}$. Чтобы доказать, что $\hat{\alpha}_n$ является единственным корнем уравнения $R_n(\hat{\alpha}) = \bar{0}$, достаточно повторить вывод соотношения (8), вычисляя нижнюю грань для $\alpha \in \mathbb{R}^{k p_n} \setminus N_n(\delta_n)$ вместо $\alpha \in \partial N_n(\delta_n)$.

Доказательство теоремы 2. По многомерной центральной предельной теореме для центрированных, независимых и одинаково распределенных векторов с конечным вторым моментом нормы

$$\begin{aligned} U_n G_n^{-1/2} R_n(\alpha^{0,n}) &= U_n \left(\frac{G_n}{n} \right)^{-1/2} \frac{R_n(\alpha^{0,n})}{\sqrt{n}} \xrightarrow{d} Z, \\ Z &\sim N(0, E_k), \quad n \rightarrow \infty. \end{aligned}$$

Далее воспользуемся результатом теоремы 1. Тогда

$$\begin{aligned} U_n G_n^{1/2}(\alpha^{0,n} - \hat{\alpha}_n) &= U_n \left[G_n^{-1/2} \tilde{Q}_n(\hat{\alpha}_n) \times \right. \\ & \quad \left. \times G_n^{-1/2} \right]^{-1} G_n^{-1/2} (R_n(\alpha^{0,n}) - R_n(\hat{\alpha}_n)) = \\ &= U_n [E + o_p(1)]^{-1} G_n^{-1/2} (R_n(\alpha^{0,n}) - R_n(\hat{\alpha}_n)) = \\ &= U_n G_n^{-1/2} R_n(\alpha^{0,n}) + U_n G_n^{-1/2} o_p(1), \end{aligned}$$

где $o_p(1)$ обозначает случайные матрицы, операторные нормы которых стремятся к нулю по вероятности с ростом n . Вектор $U_n G_n^{-1/2} R_n(\alpha^{0,n})$ является асимптотически нормальным. Следовательно, $U_n G_n^{1/2}(\alpha^{0,n} - \hat{\alpha}_n) \xrightarrow{d} Z$ при $n \rightarrow \infty$, где $Z \sim N(0, E_k)$. Теорема доказана. \square

Для доказательства следствия 1 достаточно заметить, что $\|(Q_n(\hat{\alpha}_n) - G_n(\alpha^{0,n}))/n\|$ стремится к нулю по вероятности при $n \rightarrow \infty$. Это было получено при доказательстве леммы 3.

В заключение автор хотел бы выразить глубокую благодарность профессору А. В. Булинскому за постановку задачи, помощь в работе и неизменное внимание.

Литература

1. *Murphy K. P.* Machine learning: A probabilistic perspective. — Cambridge, MA: MIT Press, 2012.
2. *Bach F.* Self-concordant analysis for logistic regression // Electron. J. Stat., 2010. Vol. 4. P. 384–414.
3. *Yu H., Huang F., Lin C.* Dual coordinate descent methods for logistic regression and maximum entropy models // Machine Learning, 2011. Vol. 85. Issue 1-2. P. 41–75.
4. *Genkin A., Lewis D. D., Madigan D.* Large-scale bayesian logistic regression for text categorization // Technometrics, 2007. Vol. 49. No. 3. P. 291–304.
5. *Guo W., Kotsia I., Patras I.* Tensor learning for regression // IEEE Trans. Image Proc., 2012. Vol. 21. No. 2. P. 816–827.
6. *Anderson J. A.* Separate sample logistic regression // Biometrika, 1972. Vol. 59. No. 1. P. 19–35.

7. *Hossain S., Ejaz Ahmed S., Howlader H.* Model selection and parameter estimation of a multinomial logistic regression model // *J. Stat. Comput. Simulation*, 2012. P. 1–15.
8. *Wang L.* GEE analysis of clustered binary data with diverging number of covariates // *Ann. Stat.*, 2011. Vol. 39. No. 1. P. 389–417.
9. *Liang H.* Maximum likelihood estimation in logistic regression models with a diverging number of covariates // *Electron. J. Stat.*, 2012. Vol. 6. P. 1838–1846.
10. *Gramacy R., Polson N.* Simulation-based regularized logistic regression // *Bayesian Anal.*, 2012. Vol. 7. No. 3. P. 567–590.
11. *Rudin W.* Principles of mathematical analysis. — New York, Francisco, Toronto, London: McGraw-Hill Book Co., 1964.
12. *Chen X.* A new generalization of Chebyshev inequality for random vectors. <http://arxiv.org/pdf/0707.0805v2.pdf>. 2007.

АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ДЛЯ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ СТАТИСТИК, ПОСТРОЕННЫХ ПО ВЫБОРКАМ СЛУЧАЙНОГО ОБЪЕМА*

В. Е. Бенинг¹, Н. К. Галиева², В. Ю. Королев³

Аннотация: Доказана общая теорема переноса, позволяющая получать асимптотические разложения (а.р.) для функций распределения (ф.р.) статистик, основанных на выборках случайного объема, из а.р. для ф.р. случайного объема выборки и а.р. для ф.р. статистик, построенных по выборкам неслучайного объема.

Ключевые слова: выборка случайного объема; асимптотическое разложение; теорема переноса; смесь вероятностных законов; распределение Лапласа; распределение Стьюдента

1 Введение

В классических задачах математической статистики объем выборки, доступной исследователю, традиционно считается детерминированным и в асимптотических постановках играет роль (как правило, неограниченно возрастающего) *известного* параметра. В то же время на практике часто возникают ситуации, когда размер выборки не является заранее определенным и может рассматриваться как случайный. Такие ситуации, как правило, связаны с тем, что статистические данные накапливаются в течение фиксированного времени. Это имеет место, в частности, в страховании, когда в течение разных отчетных периодов одинаковой длины (скажем, месяцев) происходит разное число страховых событий (страховых выплат и/или заключений страховых контрактов); в медицине, когда число пациентов с тем или иным заболеванием варьируется от года к году; в технике, когда при испытании на надежность (скажем, при определении наработки на отказ) разных партий приборов (изделий), число отказавших приборов в разных партиях оказывается разным. В таких ситуациях заранее не известно число наблюдений, которые будут доступны исследователю, разумно считать случайной величиной (с.в.). Другими словами, в таких ситуациях объем выборки является не (известным) параметром, а сам становится *наблюдением*, т. е. статистикой. В силу указанных обстоятельств вполне естественным

становится изучение асимптотического поведения распределений статистик достаточно общего вида, основанных на выборках случайного объема.

На естественность такого подхода, в частности, обратил внимание Б. В. Гнеденко в работе [1], в которой рассматривались асимптотические свойства распределений выборочных квантилей, построенных по выборкам случайного объема, и было продемонстрировано, что при замене неслучайного объема выборки случайной величиной асимптотические свойства статистик могут радикально измениться. К примеру, если объем выборки является геометрически распределенной с.в., то вместо ожидаемого в соответствии с классической теорией нормального закона в качестве асимптотического распределения выборочной медианы возникает распределение Стьюдента с двумя степенями свободы, хвосты которого столь тяжелы, что у него отсутствуют моменты порядков, больших второго. «Тяжесть» хвостов асимптотических распределений имеет же критически важное значение, в частности, в задачах проверки гипотез.

Простейшей статистикой является сумма наблюдений. Для выборок случайного объема число слагаемых в таких суммах само становится случайным. Асимптотическим свойствам распределений сумм случайного числа с.в. посвящено много работ (см., например, [1–7]). Такого рода суммы находят широкое применение в страховании, экономике, биологии и т. п. [2, 5, 7, 8]. В классической

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м), Министерством образования и науки РФ (госконтракт 16.740.11.0133).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, bening@yandex.ru

² Казахстанский филиал Московского государственного университета им. М. В. Ломоносова, nurgul_u@mail.ru

³ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, vkorolev@cs.msu.ru

статистике суммирование наблюдений, как правило, возникает при определении выборочных средних. При статистическом анализе, основанном на моделях, в которых объем выборки считается неслучайным, асимптотическое поведение статистик типа сумм и статистик типа средних арифметических одинаково — эти статистики после нормирования, обязательной для получения нетривиальных предельных распределений, становятся неразличимыми. Однако, как уже говорилось, в реальной практике очень часто объем выборки сам является статистикой и, как недавно показано, например, в работе [9], асимптотическое поведение статистик типа сумм и статистик типа средних арифметических при их неслучайной нормировке оказывается различным. Заметим, что, конечно же, формально допустима и случайная нормировка, но для построения разумных асимптотических аппроксимаций для распределений статистик (а именно это и является целью асимптотической статистики) она не применима. Именно использованием неслучайной нормировки и объясняется возникновение не «чистого» нормального закона, а (разных!) смешанных нормальных предельных распределений у статистик типа сумм и типа средних арифметических. При этом различие этих предельных законов может дать дополнительную информацию о структуре исходных данных.

Более того, в математической статистике и ее приложениях часто встречаются статистики, которые не являются суммами наблюдений. Примерами могут служить ранговые статистики, U -статистики, линейные комбинации порядковых статистик (L -статистики) и т. п.

В данной работе получены а.р. для ф.р. статистик, построенных по выборкам случайного объема. Эти а.р. непосредственно зависят от а.р. ф.р. случайного объема выборки и а.р. ф.р. статистики, основанной на неслучайной выборке. Подобного рода утверждения принято называть теоремами переноса. Таким образом, в данной работе доказаны теоремы переноса для а.р. статистик, построенных по выборкам случайного объема.

В работе приняты следующие обозначения: \mathbb{R} — множество вещественных чисел; \mathbb{N} — множество натуральных чисел; $\Phi(x)$ и $\varphi(x)$ — соответственно ф.р. и плотность стандартного нормального закона.

В разд. 2 приведен эвристический вывод основного результата, в разд. 3–5 содержатся строгая формулировка основной теоремы, ее доказательство и примеры.

Рассмотрим с.в. N_1, N_2, \dots и X_1, X_2, \dots , заданные на одном и том же вероятностном пространстве (Ω, \mathcal{A}, P) . В статистике с.в. X_1, X_2, \dots, X_n имеют смысл наблюдений, n — неслучайный объем выборки,

а с.в. N_n — случайный объем выборки, зависящий от натурального параметра $n \in \mathbb{N}$. Например, если с.в. N_n имеет геометрическое распределение вида

$$P(N_n = k) = \frac{1}{n} \left(1 - \frac{1}{n}\right)^{k-1}, \quad k \in \mathbb{N},$$

то

$$EN_n = n,$$

т. е. среднее значение случайного объема выборки равно n .

Предположим, что при каждом $n \geq 1$ с.в. N_n принимают только натуральные значения (т. е. $N_n \in \mathbb{N}$) и независимы от последовательности с.в. X_1, X_2, \dots . Всюду далее считаем с.в. X_1, X_2, \dots независимыми и одинаково распределенными.

Обозначим через $T_n = T_n(X_1, \dots, X_{N_n})$ некоторую статистику, т. е. действительную измеримую функцию от наблюдений X_1, \dots, X_n . Для каждого $n \geq 1$ определим с.в. T_{N_n} , полагая

$$T_{N_n}(\omega) \equiv T_{N_n(\omega)}(X_1(\omega), \dots, X_{N_n(\omega)}(\omega)), \quad \omega \in \Omega.$$

Можно сказать, что T_{N_n} — это статистика, построенная на основе статистики T_n по выборке случайного объема N_n .

Сформулируем условие, определяющее а.р. для ф.р. статистики T_n при неслучайном объеме выборки.

Условие 1.1. *Существуют константы $l \in \mathbb{N}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > l/2$, $\gamma \geq 0$, $C_1 > 0$, дифференцируемая ф.р. $F(x)$ и дифференцируемые ограниченные функции $f_j(x)$, $j = 1, \dots, l$, такие что*

$$\sup_x \left| P(\sigma n^\gamma (T_n - \mu) < x) - F(x) - \sum_{j=1}^l n^{-j/2} f_j(x) \right| \leq \frac{C_1}{n^\alpha}, \quad n \in \mathbb{N}.$$

Следующее условие определяет а.р. ф.р. нормированного случайного индекса N_n .

Условие 1.2. *Существуют константы $m \in \mathbb{N}$, $\beta > m/2$, $C_2 > 0$, функция $0 < g(n) \uparrow \infty$, $n \rightarrow \infty$, ф.р. $H(x)$, $H(0+) = 0$, и функции ограниченной вариации $h_i(x)$, $i = 1, \dots, m$, такие что*

$$\sup_{x \geq 0} \left| P\left(\frac{N_n}{g(n)} < x\right) - H(x) - \sum_{i=1}^m n^{-i/2} h_i(x) \right| \leq \frac{C_2}{n^\beta}, \quad n \in \mathbb{N}.$$

В данной работе с помощью а.р. для ф.р. нормированной статистики T_{N_n} , основанной на выборке случайного объема, получена аппроксимация вида

$$P(\sigma g^\gamma(n)(T_{N_n} - \mu) < x) \approx G_n(x), \quad n \rightarrow \infty,$$

где функция $G_n(x)$ имеет вид (см. условия 1.1, 1.2):

$$\begin{aligned} G_n(x) &= \\ &= \int_{1/g(n)}^{\infty} F(xy^\gamma) dH(y) + \sum_{i=1}^m n^{-i/2} \int_{1/g(n)}^{\infty} F(xy^\gamma) dh_i(y) + \\ &\quad + \sum_{j=1}^l g^{-j/2}(n) \int_{1/g(n)}^{\infty} y^{-j/2} f_j(xy^\gamma) dH(y) + \\ &\quad + \sum_{j=1}^l \sum_{i=1}^m n^{-i/2} g^{-j/2}(n) \int_{1/g(n)}^{\infty} y^{-j/2} f_j(xy^\gamma) dh_i(y). \quad (1) \end{aligned}$$

Для пояснения этой формулы, идеи доказательства и удобства дальнейших ссылок приведем ее эвристический вывод.

2 Эвристический вывод основного результата

В идейном плане доказательство основного результата данной работы — теоремы 3.1 — близко к доказательству теорем 6.6.1 и 6.7.3 для случайных сумм из работы [4] и оценкам скорости сходимости распределений случайно индексированных последовательностей из работы [5] (§ 1.3, с. 63).

По формуле полной вероятности имеем:

$$\begin{aligned} P(\sigma g^\gamma(n)(T_{N_n} - \mu) < x) &= \\ &= P\left(\sigma N_n^\gamma(T_{N_n} - \mu) < \left(\frac{N_n}{g(n)}\right)^\gamma x\right) = \\ &= EP\left(\sigma N_n^\gamma(T_{N_n} - \mu) < \left(\frac{N_n}{g(n)}\right)^\gamma x \mid N_n\right) = \\ &= \sum_{k=1}^{\infty} P\left(\sigma k^\gamma(T_k - \mu) < \left(\frac{k}{g(n)}\right)^\gamma x\right) P(N_n = k). \quad (2) \end{aligned}$$

Используя условие 1.1, вероятность под знаком ряда аппроксимируем следующим образом:

$$\begin{aligned} P(\sigma g^\gamma(n)(T_{N_n} - \mu) < x) &\approx \sum_{k=1}^{\infty} \left(F\left(x \left(\frac{k}{g(n)}\right)^\gamma\right) + \right. \\ &\quad \left. + \sum_{j=1}^l k^{-j/2} f_j\left(x \left(\frac{k}{g(n)}\right)^\gamma\right) \right) P(N_n = k) = \end{aligned}$$

$$\begin{aligned} &= E\left(F\left(x \left(\frac{N_n}{g(n)}\right)^\gamma\right) + \right. \\ &\quad \left. + \sum_{j=1}^l N_n^{-j/2} f_j\left(x \left(\frac{N_n}{g(n)}\right)^\gamma\right) \right) = \int_{1/g(n)}^{\infty} \left(F(xy^\gamma) + \right. \\ &\quad \left. + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma) \right) dP\left(\frac{N_n}{g(n)} < y\right). \quad (3) \end{aligned}$$

Теперь, аппроксимируя вероятность под знаком последнего интеграла с помощью условия 1.2, получим формулу (1):

$$\begin{aligned} P(\sigma g^\gamma(n)(T_{N_n} - \mu) < x) &\approx G_n(x) = \int_{1/g(n)}^{\infty} \left(F(xy^\gamma) + \right. \\ &\quad \left. + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma) \right) d\left(H(y) + \right. \\ &\quad \left. + \sum_{i=1}^m n^{-i/2} h_i(y) \right) = \\ &= \int_{1/g(n)}^{\infty} F(xy^\gamma) dH(y) + \sum_{i=1}^m n^{-i/2} \int_{1/g(n)}^{\infty} F(xy^\gamma) dh_i(y) + \\ &\quad + \sum_{j=1}^l g^{-j/2}(n) \int_{1/g(n)}^{\infty} y^{-j/2} f_j(xy^\gamma) dH(y) + \\ &\quad + \sum_{j=1}^l \sum_{i=1}^m n^{-i/2} g^{-j/2}(n) \int_{1/g(n)}^{\infty} y^{-j/2} f_j(xy^\gamma) dh_i(y). \quad (4) \end{aligned}$$

Если в условии 1.1 статистика T_n не нормирована, т. е. $\gamma = 0$, то полученное а.р. приобретает вид:

$$\begin{aligned} G_n(x) &= F(x) \left(1 - H\left(\frac{1}{g(n)}\right) \right) + \sum_{j=1}^l (g(n))^{-j/2} \times \\ &\quad \times f_j(x) \int_{1/g(n)}^{\infty} y^{-j/2} d\left(H(y) + \sum_{i=1}^m n^{-i/2} h_i(y) \right). \end{aligned}$$

Таким образом, у ненормированной статистики T_n исходное а.р.

$$P(\sigma(T_n - \mu) < x) \approx F(x) + \sum_{j=1}^l n^{-j/2} f_j(x)$$

при переходе к выборке случайного объема N_n заменяется на а.р. вида:

$$P(\sigma(T_{N_n} - \mu) < x) \approx F(x) \left(1 - H\left(\frac{1}{g(n)}\right)\right) + \sum_{j=1}^l c_{jn} f_j(x),$$

где

$$c_{jn} = (g(n))^{-j/2} \int_{1/g(n)}^{\infty} y^{-j/2} d\left(H(y) + \sum_{i=1}^m n^{-i/2} h_i(y)\right).$$

3 Формулировка основного результата

Теорема 3.1. Пусть статистика $T_n = T_n(X_1, \dots, X_n)$ удовлетворяет условию 1.1, а случайный объем выборки N_n — условию 1.2. Тогда существует константа $C_3 > 0$ такая, что справедливо неравенство:

$$\sup_x |P(\sigma g^\gamma(n)(T_{N_n} - \mu) < x) - G_n(x)| \leq C_1 E N_n^{-\alpha} + \frac{C_3 + C_2 M_n}{n^\beta},$$

где

$$M_n = \sup_x \int_{1/g(n)}^{\infty} \left| \frac{\partial}{\partial y} (F(xy^\gamma) + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma)) \right| dy$$

и а.р. $G_n(x)$ определено по формуле (1).

Следствие 3.1. Если моменты $E(N_n/g(n))^{-\alpha}$ равномерно по n ограничены, т. е.

$$E\left(\frac{N_n}{g(n)}\right)^{-\alpha} \leq C_4, \quad C_4 > 0, \quad n \in \mathbb{N},$$

то правая часть неравенства в формулировке теоремы 3.1 приобретает вид: $C_1 C_4 / g^\alpha(n) + (C_3 + C_2 M_n) / n^\beta$.

Следствие 3.2. В силу неравенства Гёльдера при $0 < \alpha \leq 1$ справедлива оценка

$$E N_n^{-\alpha} \leq (E N_n^{-1})^\alpha,$$

которая может быть полезной при практическом применении теоремы. В этом случае правая часть неравенства из формулировки теоремы может быть записана в виде: $C_1 (E N_n^{-1})^\alpha + (C_3 + C_2 M_n) / n^\beta$.

Для вычисления $E N_n^{-1}$ можно использовать следующую формулу (см., например, [10], с. 93, задача 40, б). Если неотрицательная целочисленная с.в. N имеет производящую функцию

$$\Psi(s) = E s^N, \quad |s| \leq 1,$$

то из теоремы Фубини непосредственно следует, что

$$E N^{-1} = \int_0^1 \frac{\Psi(s)}{s} ds. \quad (5)$$

Используя это соотношение, оценку из формулировки теоремы можно представить в виде:

$$C_1 \left(\int_0^1 \frac{\Psi_n(s)}{s} ds \right)^\alpha + \frac{C_3 + C_2 M_n}{n^\beta}, \quad (6)$$

где $\Psi_n(s)$ — производящая функция с.в. N_n .

Приведем пример использования формулы (5). Пусть с.в. N_n имеет геометрическое распределение:

$$P(N_n = k) = \frac{1}{n} \left(1 - \frac{1}{n}\right)^{k-1}, \quad k \in \mathbb{N}.$$

В этом случае производящая функция имеет вид:

$$\begin{aligned} \Psi_n(s) = E s^{N_n} &= \frac{s}{n} \left[1 - s \left(1 - \frac{1}{n}\right)\right]^{-1} = \\ &= \frac{s}{n - s(n-1)}, \quad |s| \leq 1, \end{aligned}$$

поэтому

$$E N_n^{-1} = \int_0^1 \frac{\Psi_n(s)}{s} ds = \frac{1}{n-1} \log n, \quad n > 1. \quad (7)$$

С учетом формулы (7) оценка (6) принимает вид: $(C_1 \log^\alpha n) / (n-1)^\alpha + (C_3 + C_2 M_n) / n^\beta, n > 1$.

Замечание 3.1. Заметим, что из условия 1.2, в частности, вытекает, что с.в. $N_n/g(n)$ слабо сходится к с.в. V , имеющей ф.р. $H(x)$. Из определения слабой сходимости с функцией $x^{-\alpha}, x \geq 1$, в случае, если $N_n \geq g(n), n \in \mathbb{N}$, следует, что

$$E\left(\frac{N_n}{g(n)}\right)^{-\alpha} \longrightarrow E \frac{1}{V^\alpha}, \quad n \rightarrow \infty,$$

т. е. моменты $E(N_n/g(n))^{-\alpha}$ равномерно ограничены по n и справедливо утверждение из следствия 3.1.

Случай, когда $N_n \geq g(n)$, возникает, например, если с.в. N_n принимает значения $g(n), 2g(n), \dots, kg(n)$ с равными вероятностями $1/k$ при любом фиксированном $k \in \mathbb{N}$. В этом случае с.в. $N_n/g(n)$ вообще не зависит от n и, значит, слабо сходится к с.в. V , которая принимает значения $1, 2, \dots, k$ с равными вероятностями $1/k$.

4 Доказательство теоремы 3.1

Используя формулы (2)–(4), получаем оценку:

$$\sup_x |\mathbb{P}(\sigma g^\gamma(n)(T_{N_n} - \mu) < x) - G_n(x)| \leq I_{1n} + I_{2n}, \quad (8)$$

где

$$I_{1n} = \sup_x \left| \int_{1/g(n)}^\infty \left(F(xy^\gamma) + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma) \right) d \left(\mathbb{P} \left(\frac{N_n}{g(n)} < y \right) - H(y) - \sum_{i=1}^m n^{-i/2} h_i(y) \right) \right|; \quad (9)$$

$$I_{2n} = \sum_{k=1}^\infty \sup_x \left| \mathbb{P} \left(\sigma k^\gamma (T_k - \mu) < x \left(\frac{k}{g(n)} \right)^\gamma \right) - F \left(x \left(\frac{k}{g(n)} \right)^\gamma \right) - \sum_{j=1}^l k^{-j/2} f_j \left(x \left(\frac{k}{g(n)} \right)^\gamma \right) \right| \mathbb{P}(N_n = k). \quad (10)$$

Для оценки величины I_{1n} используем равенство (9), условие 1.2, формулу интегрирования по частям (см., например, [11], теорема 2.6.11, с. 222 или [12], теорема 18.4, с. 236) и ограниченность функций $f_j(z)$, $j = 1, \dots, l$, получим, что существует константа $C_3 > 0$ такая, что

$$I_{1n} \leq \frac{C_3}{n^\beta} + \sup_x \left| \int_{1/g(n)}^\infty \left(\mathbb{P} \left(\frac{N_n}{g(n)} < y \right) - H(y) - \sum_{i=1}^m n^{-i/2} h_i(y) \right) d \left(F(xy^\gamma) + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma) \right) \right| \leq \frac{C_3}{n^\beta} + \sup_x \int_{1/g(n)}^\infty \left| \mathbb{P} \left(\frac{N_n}{g(n)} < y \right) - H(y) - \sum_{i=1}^m n^{-i/2} h_i(y) \right| \times$$

$$\times \left| \frac{\partial}{\partial y} \left(F(xy^\gamma) + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma) \right) \right| dy \leq \frac{C_3}{n^\beta} + \frac{C_1 M_n}{n^\beta}. \quad (11)$$

Ряд в определении I_{2n} (см. (10)) оценим с помощью условия 1.1 и получим:

$$I_{2n} \leq C_1 \sum_{k=1}^\infty \frac{1}{k^\alpha} \mathbb{P}(N_n = k) = C_1 \mathbb{E} N_n^{-\alpha}. \quad (12)$$

Теперь утверждение теоремы следует из неравенств (8), (11) и (12). Теорема доказана.

5 Примеры

Приведем два примера применения теоремы 3.1 с вполне конкретными предельными функциями распределения статистик, построенных по выборкам случайного объема. Рассмотрим а.р. для ф.р. выборочных средних, построенных по выборкам случайного объема. Аналогичные результаты могут быть получены для статистик, допускающих а.р. типа Эджворта для ф.р. при неслучайном объеме выборки. Например, используя результаты работ [13–18], можно получить а.р. для ф.р. ранговых статистик, L -статистик и U -статистик.

Пусть X_1, X_2, \dots — независимые одинаково распределенные случайные величины с $\mathbb{E}X_1 = \mu$, $0 < \mathbb{D}X_1 = \sigma^2$, $\mathbb{E}|X_1|^{3+2\delta} < \infty$, $\delta \in (0, 1/2)$ и $\mathbb{E}(X_1 - \mu)^3 = \mu_3$. Для натурального n обозначим

$$T_n = \frac{X_1 + \dots + X_n}{n}. \quad (13)$$

Предположим также, что случайная величина X_1 удовлетворяет условию Крамера (C)

$$\limsup_{|t| \rightarrow \infty} |\mathbb{E} \exp\{itX_1\}| < 1,$$

тогда с учетом теоремы 6.3.2 из [19] получаем, что

$$\sup_x \left| \mathbb{P}(\sigma \sqrt{n}(T_n - \mu) < x) - \Phi(x) - \frac{\mu_3 \sigma^3}{6\sqrt{n}} (1 - x^2) \varphi(x) \right| \leq \frac{C_1}{n^{1/2+\delta}}, \quad C_1 > 0, \delta \in \left(0, \frac{1}{2}\right), n \in \mathbb{N}. \quad (14)$$

Таким образом, статистика (13) удовлетворяет условию 1.1 с

$$\gamma = \frac{1}{2}; \quad \alpha = \frac{1}{2} + \delta; \quad l = 1; \quad (15)$$

$$F(x) = \Phi(x); \quad f_1(x) = \frac{\mu_3 \sigma^3}{6} (1 - x^2) \varphi(x). \quad (16)$$

Справедлива следующая лемма.

Лемма 5.1. Пусть $l = 1, 0 < g(n) \uparrow \infty, F(x) = \Phi(x),$

$$f_1(x) = \frac{\mu_3 \sigma^3}{6} (1 - x^2) \varphi(x).$$

Тогда для величины M_n (см. теорему 3.1) справедливо неравенство

$$M_n \leq 2 + \tilde{C} |\mu_3| \sigma^3,$$

где

$$\tilde{C} = \frac{1}{3} \sup_{u \geq 0} \{ \varphi(u)(u^4 + 2u^2 + 1) \} = \frac{16}{3\sqrt{2\pi e^3}} \approx 0,474752293191785 \dots$$

Доказательство. С учетом формул (13)–(16) имеем (см. теорему 3.1):

$$\begin{aligned} M_n &= \sup_x \int_{(g(n))^{-1}}^{\infty} \left| \frac{\partial}{\partial y} \left(F(xy^\gamma) + \sum_{j=1}^l (yg(n))^{-j/2} f_j(xy^\gamma) \right) \right| dy = \\ &= \sup_x \int_{(g(n))^{-1}}^{\infty} \left| \frac{\partial}{\partial y} \left(\Phi(x\sqrt{y}) + \frac{\mu_3 \sigma^3 (1 - x^2 y) \varphi(x\sqrt{y})}{6\sqrt{y g(n)}} \right) \right| dy \leq \\ &\leq \sup_{x \geq 0} \int_{x(g(n))^{-1/2}}^{\infty} \left| \frac{\partial}{\partial u} \left(\Phi(u) + \frac{x \mu_3 \sigma^3 (1 - u^2) \varphi(u)}{6u\sqrt{g(n)}} \right) \right| du + \\ &+ \sup_{x < 0} \int_{-\infty}^{x(g(n))^{-1/2}} \left| \frac{\partial}{\partial u} \left(\Phi(u) + \frac{x \mu_3 \sigma^3 (1 - u^2) \varphi(u)}{6u\sqrt{g(n)}} \right) \right| du = \\ &= \sup_{x \geq 0} \int_{x(g(n))^{-1/2}}^{\infty} \varphi(u) \left| 1 + \frac{x \mu_3 \sigma^3 (u^4 - 2u^2 - 1)}{6\sqrt{g(n)} u^2} \right| du + \end{aligned}$$

$$\begin{aligned} &+ \sup_{x < 0} \int_{-\infty}^{x(g(n))^{-1/2}} \varphi(u) \left| 1 + \frac{x \mu_3 \sigma^3 (u^4 - 2u^2 - 1)}{6\sqrt{g(n)} u^2} \right| du \leq \\ &\leq 2 + \frac{|\mu_3| \sigma^3}{3\sqrt{g(n)}} \sup_{x \geq 0} x \int_{x(g(n))^{-1/2}}^{\infty} \varphi(u) \frac{u^4 + 2u^2 + 1}{u^2} du. \quad (17) \end{aligned}$$

Далее заметим, что справедливо соотношение

$$\sup_{u \geq 0} \{ \varphi(u)(u^4 + 2u^2 + 1) \} = \frac{1}{\sqrt{2\pi}} \sup_{u \geq 0} e^{-u} (2u + 1)^2.$$

Легко видеть, что $(e^{-u}(2u + 1)^2)' = e^{-u}(2u + 1)(3 - 2u) = 0$ при $u = 3/2$. Таким образом,

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \sup_{u \geq 0} e^{-u} (2u + 1)^2 &= \frac{1}{\sqrt{2\pi}} e^{-u} (2u + 1)^2 \Big|_{u=3/2} = \\ &= \frac{16}{\sqrt{2\pi e^3}} \approx 1,42425687951535 \dots, \end{aligned}$$

так что

$$\begin{aligned} \tilde{C} &= \frac{1}{3} \sup_{u \geq 0} \{ \varphi(u)(u^4 + 2u^2 + 1) \} = \frac{16}{3\sqrt{2\pi e^3}} \approx \\ &\approx 0,474752293191785 \dots \quad (18) \end{aligned}$$

При этом из неравенства (17) следует, что

$$\begin{aligned} M_n &\leq 2 + \tilde{C} \frac{|\mu_3| \sigma^3}{\sqrt{g(n)}} \sup_{x \geq 0} x \int_{x(g(n))^{-1/2}}^{\infty} u^{-2} du = \\ &= 2 + \tilde{C} |\mu_3| \sigma^3. \end{aligned}$$

Таким образом, справедливо неравенство:

$$M_n \leq 2 + \tilde{C} |\mu_3| \sigma^3, \quad (19)$$

где постоянная \tilde{C} определена в соотношении (18). Из неравенства (19) следует утверждение леммы. Лемма доказана.

5.1 Распределение Стьюдента

В работе [20] показано, что если случайный объем выборки N_n имеет отрицательно биномиальное распределение с параметрами $p = 1/n$ и $r > 0$, т. е.

$$P(N_n = k) = \frac{(k + r - 2) \dots r}{(k - 1)!} \frac{1}{n^r} \left(1 - \frac{1}{n} \right)^{k-1}, \quad k \in \mathbb{N}$$

(при $r = 1$ имеем геометрическое распределение), то для асимптотически нормальной статистики T_n

справедливо предельное соотношение ([20], следствие 2.1)

$$P(\sigma\sqrt{n}(T_{N_n} - \mu) < x) \longrightarrow G_{2r}(x\sqrt{r}), \quad n \rightarrow \infty, \quad (20)$$

где $G_f(x)$ — функция распределения Стьюдента с параметром $f = 2r$, соответствующая плотности вида

$$p_f(x) = \frac{\Gamma(f+1/2)}{\sqrt{\pi f} \Gamma(f/2)} \left(1 + \frac{x^2}{f}\right)^{-(\gamma+1)/2}, \quad x \in \mathbb{R},$$

где $\Gamma(\cdot)$ — эйлерова гамма-функция, а $f > 0$ — параметр формы (если параметр f натурален, то он называется числом степеней свободы). В рассматриваемой ситуации он может быть произвольно мал, т. е. может иметь место типичное распределение с тяжелыми хвостами. Если $f = 2$, т. е. $r = 1$, то ф.р. $G_2(x)$ выражается в явном виде:

$$G_2(x) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{2+x^2}}\right), \quad x \in \mathbb{R}.$$

При $r = 1/2$ имеем распределение Коши.

В книге [19] (формула (6.112)) приведена следующая оценка скорости сходимости:

$$\sup_{x \geq 0} \left| P\left(\frac{N_n}{EN_n} < x\right) - H_r(x) \right| \leq \begin{cases} \frac{C_r}{n}, & r \geq 1; \\ \frac{C_r}{n^r}, & r \in (0, 1), \end{cases}$$

$$C_r > 0, \quad n \in \mathbb{N}, \quad (21)$$

где $H_r(x)$ — функция гамма-распределения с параметром $r > 0$:

$$H_r(x) = \frac{r^r}{\Gamma(r)} \int_0^x e^{-ry} y^{r-1} dy, \quad x \geq 0, \quad (22)$$

При этом

$$EN_n = r(n-1) + 1. \quad (23)$$

Таким образом, из соотношений (21)–(23) следует, что случайный индекс N_n удовлетворяет условию 1.2 с

$$g(n) = r(n-1) + 1; \quad H(x) = H_r(x); \quad m = 1;$$

$$h_1(x) \equiv 0; \quad C_2 = C_r > 0;$$

$$\beta = \begin{cases} 1, & r \geq 1; \\ r, & r \in (0, 1). \end{cases}$$

Далее, используя равенство

$$(1+x)^\gamma = \sum_{k=0}^{\infty} \frac{\gamma(\gamma-1)\cdots(\gamma-k+1)}{k!} x^k,$$

$$|x| < 1, \quad \gamma \in \mathbb{R},$$

нетрудно получить, что

$$EN_n^{-1} = \frac{1}{(n-1)(1-r)} \left(\frac{1}{n^{r-1}} - 1\right) =$$

$$= O(n^{-r}), \quad r > 0, \quad r \neq 1, \quad n \in \mathbb{N}. \quad (24)$$

Для случая $r = 1$, используя формулу (7), имеем:

$$EN_n^{-1} = \frac{1}{n-1} \log n, \quad n > 1. \quad (25)$$

Таким образом, учитывая теорему 3.1, следствие 3.2, формулы (14)–(16), а также лемму 5.1, соотношения (24), (25) и равенства (справедливые равномерно по x)

$$\int_{(r(n-1)+1)^{-1}}^{\infty} \Phi(x\sqrt{y}) dH_r(y) = \int_0^{\infty} \Phi(x\sqrt{y}) dH_r(y) +$$

$$+ O\left(\frac{1}{n}\right) = G_{2r}(x) + O\left(\frac{1}{n}\right); \quad (26)$$

$$\int_{(r(n-1)+1)^{-1}}^{\infty} \varphi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} dH_r(y) =$$

$$= \int_0^{\infty} \varphi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} dH_r(y) + o(1) \equiv$$

$$\equiv g_r(x) + o(1), \quad (27)$$

получаем следующее утверждение.

Теорема 5.1. Пусть статистика T_n имеет вид (13), где X_1, X_2, \dots — независимые одинаково распределенные с.в. с $EX_1 = \mu, 0 < DX_1 = \sigma^{-2}, E|X_1|^{3+2\delta} < \infty, \delta \in (0, 1/2)$ и $E(X_1 - \mu)^3 = \mu_3$, причем с.в. X_1 удовлетворяет условию Крамера (C). Предположим, что при некотором $r > 0$ случайная величина N_n имеет распределение вида:

$$P(N_n = k) =$$

$$= \frac{(k+r-2)\cdots r}{(k-1)!} \frac{1}{n^r} \left(1 - \frac{1}{n}\right)^{k-1}, \quad k \in \mathbb{N}.$$

Тогда при $r > 1/(1+2\delta)$ для ф.р. нормированной статистики T_{N_n} при $n \rightarrow \infty$ справедливо а.р. вида

$$\sup_x \left| P\left(\sigma\sqrt{r(n-1)+1}(T_{N_n} - \mu) < x\right) - G_{2r}(x) - \right.$$

$$\left. - \frac{\mu_3\sigma^3 g_r(x)}{6\sqrt{r(n-1)+1}} \right| =$$

$$= \begin{cases} O\left(\left(\frac{\log n}{n}\right)^{1/2+\delta}\right), & r = 1; \\ O\left(\frac{1}{n^{\min(1, r(1/2+\delta))}}\right), & r > 1; \\ O\left(\frac{1}{n^{r(1/2+\delta)}}\right), & \frac{1}{1+2\delta} < r < 1, \end{cases}$$

где функции $G_{2r}(x)$ и $g_r(x)$ определены в соотношениях (26) и (27).

5.2 Распределение Лапласа

Рассмотрим распределение Лапласа с ф.р. $\Lambda_\theta(x)$ и плотностью

$$\lambda_\theta(x) = \frac{1}{\theta\sqrt{2}} \exp\left\{-\frac{\sqrt{2}|x|}{\theta}\right\}, \quad \theta > 0, x \in \mathbb{R}.$$

В работе [21] была построена последовательность с.в. $N_n(s)$, зависящая от параметра $s \in \mathbb{N}$, следующего вида. Пусть Y_1, Y_2, \dots — независимые одинаково распределенные с.в., имеющие непрерывную ф.р. Определим с.в.

$$N(s) = \min\left\{i \geq 1 : \max_{1 \leq j \leq s} Y_j < \max_{s+1 \leq k \leq s+i} Y_k\right\}.$$

Хорошо известно, что так определенные с.в. имеют распределение вида

$$P(N(s) \geq k) = \frac{s}{s+k-1}, \quad k \geq 1 \quad (28)$$

(см., например, [22, 23]). Пусть теперь $N^{(1)}(s), N^{(2)}(s), \dots$ — независимые одинаково распределенные с.в., имеющие распределение (28). Определим с.в.

$$N_n(s) = \max_{1 \leq j \leq n} N^{(j)}(s),$$

тогда, как показано в работе [21],

$$\lim_{n \rightarrow \infty} P\left(\frac{N_n(s)}{n} < x\right) = e^{-s/x}, \quad x > 0, \quad (29)$$

и для асимптотически нормальной статистики T_n справедливо соотношение:

$$P\left(\sigma\sqrt{n}(T_{N_n(s)} - \mu) < x\right) \rightarrow \Lambda_{1/s}(x), \quad n \rightarrow \infty, x \in \mathbb{R},$$

где $\Lambda_{1/s}(x)$ — функция распределения Лапласа с параметром $\theta = 1/s$.

В работе [24] была получена следующая оценка скорости сходимости в соотношении (29):

$$\sup_{x \geq 0} \left| P\left(\frac{N_n(s)}{n} < x\right) - e^{-s/x} \right| \leq \frac{C_s}{n}, \quad C_s > 0, \quad n \in \mathbb{N}. \quad (30)$$

Таким образом, из соотношения (30) следует, что случайный индекс $N_n(s)$ удовлетворяет условию 1.2 с

$$g(n) = n; \quad H(x) = e^{-s/x}; \quad m = 1; \quad (31)$$

$$h_1(x) \equiv 0; \quad C_2 = C_s > 0; \quad \beta = 1. \quad (32)$$

Рассмотрим более подробно величину $EN_n^{-1}(s)$. Из определения с.в. $N_n(s)$ и равенства (28) имеем

$$P(N_n(s) = k) = \left(\frac{k}{s+k}\right)^n - \left(\frac{k-1}{s+k-1}\right)^n = sn \int_{k-1}^k \frac{x^{n-1}}{(s+x)^{n+1}} dx,$$

поэтому

$$\begin{aligned} EN_n^{-1}(s) &= \sum_{k=1}^{\infty} \frac{1}{k} P(N_n(s) = k) = \\ &= sn \sum_{k=1}^{\infty} \frac{1}{k} \int_{k-1}^k \frac{x^{n-1}}{(s+x)^{n+1}} dx \leq \\ &\leq sn \sum_{k=1}^{\infty} \int_{k-1}^k \frac{x^{n-2}}{(s+x)^{n+1}} dx = sn \int_0^{\infty} \frac{x^{n-2}}{(s+x)^{n+1}} dx. \end{aligned}$$

Для вычисления последнего интеграла используем формулу (см. [25] формула 856.12, с. 184):

$$\int_0^{\infty} \frac{x^{s-1}}{(a+bx)^{s+n}} dx = \frac{\Gamma(s)\Gamma(n)}{a^n b^s \Gamma(s+n)}, \quad a, b, s, n > 0.$$

Получим

$$EN_n^{-1}(s) \leq sn \frac{\Gamma(n-1)\Gamma(2)}{s^2 \Gamma(n+1)} = \frac{1}{s(n-1)} = O(n^{-1}).$$

Таким образом, учитывая теорему 3.1, следствие 3.2, формулы (14)–(16), а также лемму 5.1, соотношения (31), (32) и равенства (справедливые равномерно по x)

$$\begin{aligned} \int_{n^{-1}}^{\infty} \Phi(x\sqrt{y}) de^{-s/y} &= \int_0^{\infty} \Phi(x\sqrt{y}) de^{-s/y} + O\left(\frac{1}{n}\right) = \\ &= \Lambda_{1/s}(x) + O\left(\frac{1}{n}\right); \quad (33) \end{aligned}$$

$$\begin{aligned} \int_{n^{-1}}^{\infty} \varphi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} de^{-s/y} &= \\ &= \int_0^{\infty} \varphi(x\sqrt{y}) \frac{1-x^2y}{\sqrt{y}} de^{-s/y} + o(1) \equiv \\ &\equiv l_s(x) + o(1), \quad (34) \end{aligned}$$

непосредственно получаем следующую теорему.

Теорема 5.2. Пусть статистика T_n имеет вид (13), где X_1, X_2, \dots — независимые одинаково распределенные с.в. с $EX_1 = \mu$, $0 < DX_1 = \sigma^{-2}$, $E|X_1|^{3+2\delta} < \infty$, $\delta \in (0, 1/2)$ и $E(X_1 - \mu)^3 = \mu_3$, причем с.в. X_1 удовлетворяет условию Крамера (С). Предположим, что при некотором $s \in \mathbb{N}$ с.в. $N_n(s)$ имеет распределение вида:

$$P(N_n(s) = k) = \left(\frac{k}{s+k}\right)^n - \left(\frac{k-1}{s+k-1}\right)^n, \quad k \in \mathbb{N}.$$

Тогда для ф.р. нормированной статистики $T_{N_n(s)}$ справедливо а.р. вида:

$$\sup_x \left| P(\sigma\sqrt{n}(T_{N_n(s)} - \mu) < x) - \Lambda_{1/s}(x) - \frac{\mu_3\sigma^3 l_s(x)}{6\sqrt{n}} \right| = O\left(\frac{1}{n^{1/2+\delta}}\right), \quad n \rightarrow \infty,$$

где функции $\Lambda_{1/s}(x)$ и $l_s(x)$ определены соответственно в соотношениях (33) и (34).

Литература

- Гнеденко Б. В. Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // Тр. Тбилисского математического института, 1989. Т. 92. С. 146–150.
- Гнеденко Б. В., Фахим Х. Об одной теореме переноса // Докл. АН СССР, 1969. Т. 187. С. 15–17.
- Von Ghosy R., Rappl G. Some approximation methods for the distribution of random sums // Insurance: Mathematics and Economics, 1983. Vol. 2. P. 251–270.
- Круглов В. М., Королев В. Ю. Предельные теоремы для случайных сумм. — М.: Изд-во Московского ун-та, 1990.
- Королев В. Ю. Предельные распределения для случайно индексированных последовательностей и их применения: Дисс. . . . докт. физ.-мат. наук. — М.: МГУ, 1993.
- Gnedenko B. V., Korolev V. Yu. Random summation. Limit theorems and applications. — Boca Raton: CRC Press, 1996.
- Bening V. E., Korolev V. Yu. Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002.
- Гнеденко Б. В. Курс теории вероятностей. — М.: Наука, 1988.
- Королев В. Ю. О взаимосвязи обобщенного распределения Стьюдента и дисперсионного гамма-распределения при статистическом анализе выборок случайного объема // Докл. РАН, 2012. Т. 445. Вып. 6. С. 622–627.
- Климов Г. П. Теория вероятностей и математическая статистика. — М.: Изд-во Московского ун-та, 1983.
- Ширяев А. Н. Вероятность. — М.: Наука, 1989.
- Billingsley P. Probability and measure. — John Wiley & Sons, 1995.
- Bickel P. G. Edgeworth expansions in nonparametric statistics // Ann. Stat., 1974. Vol. 2. P. 1–21.
- Albers W. Asymptotic expansions and the deficiency concept in statistics // Mathematical Centre Tracts 58. — Amsterdam: Mathematisch Centrum, 1974.
- Albers W., Bickel P. G., Van Zwet W. R. Asymptotic expansions for the power of distribution free tests in the one-sample problem // Ann. Stat., 1976. Vol. 4. P. 108–156.
- Bickel P. G., Van Zwet W. R. Asymptotic expansions for the power of distribution free tests in the two-sample problem // Ann. Stat., 1978. Vol. 6. P. 947–1004.
- Helmers R. Edgeworth expansions for linear combinations of order statistics // Mathematical Centre Tracts 105. — Amsterdam: Mathematisch Centrum, 1984.
- Bentkus V., Gotze F., Van Zwet W. R. An Edgeworth expansions for symmetric statistics // Ann. Stat., 1997. Vol. 25. P. 851–896.
- Бенинг В. Е., Королев В. Ю., Соколов И. А., Шоргин С. Я. Рандомизированные модели и методы теории надежности информационных и технических систем. — М.: ТОРУС ПРЕСС, 2007.
- Бенинг В. Е., Королев В. Ю. Об использовании распределения Стьюдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения, 2004. Т. 49. Вып. 3. С. 417–435.
- Бенинг В. Е., Королев В. Ю. Некоторые статистические задачи, связанные с распределением Лапласа // Информатика и её применения, 2008. Т. 2. Вып. 2. С. 19–34.
- Wilks S. S. Recurrence of extreme observations // J. Amer. Math. Soc., 1959. Vol. 1. No. 1. P. 106–112.
- Невзоров В. Б. Рекорды. Математическая теория. — М.: Фазис, 2000.
- Лямин О. О. О скорости сходимости распределений некоторых статистик к распределению Лапласа и Стьюдента // Вестник Московского ун-та. Сер. 15: Вычислительная математика и кибернетика, 2011. Вып. 1. С. 39–47.
- Двайт Г. Б. Таблицы интегралов и другие математические формулы. — М.: Наука, 1977.

О СХОДИМОСТИ СЛУЧАЙНЫХ БЛУЖДАНИЙ, ПОРОЖДЕННЫХ ОБОБЩЕННЫМИ ПРОЦЕССАМИ КОКСА, К ПРОЦЕССАМ ЛЕВИ*

В. Ю. Королев¹, Л. М. Закс², А. И. Зейфман³

Аннотация: Доказана функциональная предельная теорема о сходимости случайных блужданий с непрерывным временем, порожденных обобщенными дважды стохастическими пуассоновскими процессами (обобщенными процессами Кокса), к процессам Леви в пространстве Скорохода. В качестве следствий получены теоремы о сходимости указанных случайных блужданий к процессам Леви со смешанными нормальными распределениями, в частности к устойчивым процессам Леви.

Ключевые слова: устойчивое распределение; процесс Леви; α -устойчивый процесс Леви; обобщенный дважды стохастический пуассоновский процесс (обобщенный процесс Кокса); пространство Скорохода; теорема переноса; подчиненный винеровский процесс

1 Введение

В финансовой математике принято моделировать эволюцию логарифмов цен финансовых активов на временных микромасштабах с помощью случайных блужданий. Простейшим примером применения такого подхода является модель Кокса–Росса–Рубинштейна (см., например, [1]). В то же время наиболее успешными (адекватными) моделями динамики логарифмов цен финансовых активов на временных макромасштабах являются подчиненные (subordinated) винеровские процессы (процессы броуновского движения со случайным временем), примерами которых являются обобщенные гиперболические процессы, в частности дисперсионные гамма-процессы (variance gamma processes), нормальные обратные гауссовские процессы (NIG-processes) (см., например, [2]). Естественным теоретическим «мостиком» между случайными блужданиями и подчиненными винеровскими процессами служат функциональные предельные теоремы для случайных блужданий. Операция подчинения (subordination) хорошо описывает наличие тяжелых хвостов у наблюдаемых на практике конечномерных распределений цен финансовых активов и объясняет их возникновение.

Благодаря работам Мандельброта [3], одними из первых успешно применяемых на практике моделей процессов с «тяжелохвостыми» конечномерными распределениями стали устойчивые процессы

Леви. Согласно классическому подходу в функциональных предельных теоремах для случайных блужданий устойчивые процессы Леви могут возникнуть, только если дисперсии элементарных скачков бесконечны. А для этого, в свою очередь, необходимо, чтобы вероятности произвольно больших по абсолютной величине значений скачков были положительны. К сожалению, последнее условие выглядит весьма сомнительным с практической точки зрения. Поэтому в рамках классического подхода теоретическое обоснование адекватности моделей типа устойчивых процессов Леви при описании эволюции финансовых индексов является как минимум небесспорным.

Относительно недавно в работах [4, 5] было показано, что устойчивые законы могут возникнуть в качестве предельных распределений для сумм независимых одинаково распределенных случайных величин с *конечными дисперсиями*, если число слагаемых в суммах случайно и имеет асимптотически устойчивое распределение. Схема случайного суммирования при изучении случайных блужданий является естественным аналогом схемы подчинения случайных процессов. В книге [6] и статьях [7, 8] предложено моделировать эволюцию неоднородных хаотических стохастических процессов, в частности динамику цен финансовых активов, с помощью случайных блужданий, порожденных обобщенными дважды стохастическими пуассоновскими процессами (обобщенными процессами

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 12-07-00115а, 11-01-12026-офи-м).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, victoryukorolev@yandex.ru

² Альфа-банк, отдел моделирования и математической статистики; lily.zaks@gmail.com

³ Вологодский государственный педагогический университет; Институт проблем информатики Российской академии наук, a_zeifman@mail.ru

Кокса). Этот подход, основанный на универсальном принципе неубывания энтропии в замкнутых системах, получил дополнительное обоснование и развитие в [2, 9–11]. В работах [2, 12] этот подход успешно применен к моделированию процессов плазменной турбулентности. В соответствии с указанным подходом поток информативных событий, в результате каждого из которых появляется очередное «наблюденное» значение рассматриваемой характеристики, описывается с помощью точечного случайного процесса вида $N_1(\Lambda(t))$, где $N_1(t)$, $t \geq 0$, — однородный пуассоновский процесс с единичной интенсивностью, а $\Lambda(t)$, $t \geq 0$, — независимый от $N_1(t)$ случайный процесс, обладающий следующими свойствами: $\Lambda(0) = 0$, $P(\Lambda(t) < \infty) = 1$ для любого $t > 0$, траектории $\Lambda(t)$ не убывают и непрерывны справа. Процесс $N_1(\Lambda(t))$, $t \geq 0$, называется дважды стохастическим пуассоновским процессом (процессом Кокса). В частности, если процесс $\Lambda(t)$ допускает представление

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad t \geq 0,$$

в котором $\lambda(t)$ — положительный случайный процесс с интегрируемыми траекториями, то $\lambda(t)$ можно интерпретировать как мгновенную стохастическую интенсивность процесса Кокса.

В работах [13, 14] доказаны функциональные предельные теоремы для обобщенных процессов Кокса с квадратично интегрируемым управляющим процессом $\Lambda(t)$. Однако класс процессов, предельных для обобщенных процессов Кокса с такими управляющими процессами и скачками, имеющими конечные дисперсии, недостаточно широк. В частности, он не может содержать никаких устойчивых процессов Леви кроме винеровского процесса.

Цель данной работы — существенно обобщить упомянутые результаты и восполнить указанный пробел. Здесь будет доказана функциональная предельная теорема, описывающая сходимость обобщенных процессов Кокса со скачками, имеющими конечные дисперсии, к процессам Леви из очень широкого класса, содержащего, в частности, устойчивые процессы Леви.

2 Основные определения и некоторые вспомогательные результаты

Пусть $D = D[0, 1]$ — пространство вещественных функций, определенных на $[0, 1]$, непрерывных

справа и имеющих конечные левосторонние пределы.

Пусть \mathcal{F} — класс строго возрастающих непрерывных отображений отрезка $[0, 1]$ на себя. Пусть f — неубывающая функция на $[0, 1]$, $f(0) = 0$, $f(1) = 1$. Положим

$$\|f\| = \sup_{s \neq t} \left| \log \frac{f(t) - f(s)}{t - s} \right|.$$

Если $\|f\| < \infty$, то функция f непрерывна и строго возрастает и, следовательно, принадлежит классу \mathcal{F} .

Определим расстояние $d_0(x, y)$ в множестве $D[0, 1]$ как нижнюю грань таких положительных чисел ϵ , для которых \mathcal{F} содержит некоторую функцию f такую, что $\|f\| \leq \epsilon$ и

$$\sup_t |x(t) - y(f(t))| \leq \epsilon.$$

Можно показать, что пространство $D[0, 1]$ полно относительно метрики d_0 . Метрическое пространство $\mathcal{D} = (D[0, 1], d_0)$ принято называть *пространством Скорохода*. Всюду в дальнейшем будем рассматривать случайные процессы как случайные элементы в \mathcal{D} .

Пусть X, X_1, X_2, \dots — случайные элементы пространства \mathcal{D} . Пусть T_X — такое множество точек отрезка $[0, 1]$, что $0 \in T_X$, $1 \in T_X$ и если $0 < t < 1$, то $t \in T_X$ тогда и только тогда, когда $P(X(t) \neq X(t-)) = 0$. Имеет место следующая теорема, устанавливающая достаточные условия слабой сходимости случайных процессов в \mathcal{D} (обозначаемой далее символом \Rightarrow и подразумеваемой при $n \rightarrow \infty$).

Теорема А. Пусть для любого натурального k

$$(X_n(t_1), \dots, X_n(t_k)) \Rightarrow (X(t_1), \dots, X(t_k))$$

для t_1, \dots, t_k , принадлежащих множеству T_X . Пусть $P(X(1) \neq X(1-)) = 0$ и существует неубывающая непрерывная функция F на $[0, 1]$ такая, что для любого $\epsilon > 0$

$$P(|X_n(t) - X_n(t_1)| \geq \epsilon, |X_n(t_2) - X_n(t)| \geq \epsilon) \leq \epsilon^{-2\nu} [F(t_2) - F(t_1)]^{2\gamma} \quad (1)$$

при $t_1 \leq t \leq t_2$ и $n \geq 1$, где $\nu \geq 0$, $\gamma > 1/2$. Тогда $X_n \Rightarrow X$.

Доказательство теоремы А можно найти, например, в книге [15].

Условие (1) может быть заменено более ограничительным моментным условием

$$E[|X_n(t) - X_n(t_1)|^\nu |X_n(t_2) - X_n(t)|^\nu] \leq [F(t_2) - F(t_1)]^{2\gamma}.$$

Под процессом Леви будем понимать случайный процесс $X(t)$, $t \geq 0$, со следующими свойствами:

- 1° $X(0) = 0$ с вероятностью единица.
- 2° $X(t)$ — процесс с независимыми приращениями, т.е. для любых $N \geq 1$ и t_0, t_1, \dots, t_N ($0 \leq t_0 \leq t_1 \leq \dots \leq t_N$) случайные величины $X(t_0)$, $X(t_1) - X(t_0)$, \dots , $X(t_N) - X(t_{N-1})$ независимы в совокупности.
- 3° $X(t)$ — однородный процесс, т.е. для любых $s, t, h > 0$ случайные величины $X(t+h) - X(t)$ и $X(s+h) - X(s)$ имеют одинаковое распределение.
- 4° Процесс $X(t)$ стохастически непрерывен, т.е. для любых $t \geq 0$ и $\epsilon > 0$

$$\lim_{s \rightarrow t} P(|X(t) - X(s)| > \epsilon) = 0.$$

- 5° Траектории процесса $X(t)$ непрерывны справа и имеют конечные пределы слева.

Обозначим характеристическую функцию случайной величины $X(t)$ символом $\psi_t(s)$ ($\psi_t(s) = E e^{isX(t)}$, $s \in \mathbb{R}$). Следующее утверждение описывает хорошо известное свойство процессов Леви.

Лемма 1. Пусть $X = X(t)$, $t \geq 0$, — процесс Леви. При любом $t > 0$ характеристическая функция случайной величины $X(t)$ безгранично делима и имеет вид:

$$\psi_t(s) = [\psi_1(s)]^t = \left[E e^{isX(1)} \right]^t, \quad s \in \mathbb{R}. \quad (2)$$

Обратно, пусть Y — произвольная безгранично делимая случайная величина. Тогда семейство безгранично делимых распределений с характеристическими функциями вида $[E e^{isY}]^t$ полностью определяет конечномерные распределения процесса Леви $X(t)$, $t \geq 0$, причем $X(1) \stackrel{d}{=} Y$.

Свойства процессов Леви подробно описаны в книгах [16, 17]. Книги [18, 19] и обзорная статья [20] посвящены использованию процессов Леви в задачах моделирования динамики финансовых индексов.

Функцию распределения строго устойчивого распределения с характеристическим показателем α и параметром θ , задаваемого характеристической функцией:

$$g_{\alpha, \theta}(s) = \exp \left\{ -|s|^\alpha \exp \left\{ -\frac{i\pi\theta\alpha}{2} \operatorname{sign} s \right\} \right\}, \quad s \in \mathbb{R},$$

где $0 < \alpha \leq 2$, $|\theta| \leq \theta_\alpha = \min\{1, 2/\alpha - 1\}$, будем обозначать $G_{\alpha, \theta}(x)$. Симметричным строго устойчивым распределениям соответствует значение $\theta = 0$. Односторонним устойчивым распределениям соответствуют значения $\theta = 1$ и $0 < \alpha \leq 1$.

Если ξ — случайная величина с функцией распределения $G_{\alpha, \theta}(x)$, $0 < \alpha < 2$, то $E|\xi|^\delta < \infty$ для любого $\delta \in (0, \alpha)$, но моменты порядков, больших или равных α , у случайной величины ξ отсутствуют (см., например, [21]).

Функцию распределения стандартного нормального закона ($\alpha = 2$, $\theta = 0$) будем обозначать $\Phi(x)$:

$$\Phi(x) = \int_{-\infty}^x \phi(z) dz; \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Хорошо известно следующее представление симметричной строго устойчивой функции распределения $G_{\alpha, 0}(x)$ в виде масштабной смеси нормальных законов:

$$G_{\alpha, 0}(x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{u}}\right) dG_{\alpha/2, 1}(u), \quad x \in \mathbb{R} \quad (3)$$

(см., например, теорему 3.3.1 в [21] или задачу 9 на с. 668 в [22]). Представлению (3) соответствует аналогичное представление в терминах характеристических функций:

$$g_{\alpha, 0}(s) = \int_0^\infty \exp\left\{-\frac{s^2 u}{2}\right\} dG_{\alpha/2, 1}(u), \quad s \in \mathbb{R}.$$

Процесс Леви $X(t)$, $t \geq 0$, будем называть α -устойчивым, если

$$P(X(1) < x) = G_{\alpha, \theta}(x), \quad x \in \mathbb{R}.$$

Можно показать, что если $X(t)$, $t \geq 0$, — процесс Леви такой, что $P(X(0) = 0) = 1$, то $X(t)$ является α -устойчивым тогда и только тогда, когда

$$X(t) \stackrel{d}{=} t^{1/\alpha} X(1), \quad t \geq 0 \quad (4)$$

(см., например, [23]).

3 Основные результаты

В дальнейшем без существенного ограничения общности будем рассматривать случайные процессы, определенные при $0 \leq t \leq 1$. Это означает, что по сути рассматривается поведение случайных процессов на конечном временном горизонте. Добиться равенства правой границы временного горизонта единице можно, если соответствующим образом задать единицу измерения времени. Другими словами, внимание будет сосредоточено на изучении пространства Скорохода D .

Поскольку, как уже говорилось, в большинстве приложений нет никаких оснований, чтобы отвергнуть предположение о конечности дисперсий элементарных скачков обобщенных процессов Кокса, рассматриваем только случай конечных дисперсий.

Пусть $Z_n(t)$ — последовательность обобщенных процессов Кокса:

$$Z_n(t) = \sum_{i=1}^{N_1^{(n)}(\Lambda_n(t))} X_{n,i}, \quad t \geq 0, \quad (5)$$

где $\{N_1^{(n)}(t), t \geq 0\}_{n \geq 1}$ — последовательность пуассоновских процессов с единичной интенсивностью; при каждом $n = 1, 2, \dots$ случайные величины $X_{n,1}, X_{n,2}, \dots$ одинаково распределены, причем при каждом $n \geq 1$ величины $X_{n,1}, X_{n,2}, \dots$ и процесс $N_1^{(n)}(t), t \geq 0$, независимы; при каждом $n = 1, 2, \dots$ $\Lambda_n(t), t \geq 0$, — это процесс Леви, независимый от процесса

$$X_n(t) = \sum_{i=1}^{N_1^{(n)}(t)} X_{n,i}, \quad t \geq 0, \quad (6)$$

и такой, что $\Lambda_n(0) = 0$ и существуют $\delta \in (0, 1]$, $\delta_1 \in (0, 1]$ и не зависящие от t числа $C_n \in (0, \infty)$, гарантирующие справедливость неравенства

$$E\Lambda_n^\delta(t) \leq (C_n t)^{\delta_1} \quad (7)$$

для всех $t \in (0, 1]$. Здесь и далее для определенности полагаем $\sum_{i=1}^0 = 0$.

Предположим, что $EX_{n,1} = 0$ и

$$0 < m_n^\beta \equiv E|X_{n,1}|^\beta < \infty \quad (8)$$

для некоторого $\beta \in [1, 2]$.

Из (5) и (6) легко видеть, что $Z_n(t) = X_n(\Lambda_n(t))$. Поскольку при каждом $n \geq 1$ $X_n(t)$ и $\Lambda_n(t)$ — независимые процессы Леви, то суперпозиция $Z_n(t) = X_n(\Lambda_n(t))$ также является процессом Леви (см., например, теорему 3.1.1 в [14]). Отсюда, в частности, вытекает

Лемма 2. Для любых $0 \leq t_1 < t_2 < \infty$ и любого $n \geq 1$

$$Z_n(t_2) - Z_n(t_1) \stackrel{d}{=} Z_n(t_2 - t_1).$$

Также справедлива

Лемма 3. Пусть $Z_n(t)$ — обобщенный процесс Кокса (5), удовлетворяющий условиям (7) и (8). Тогда при каждом $t \in [0, 1]$ для любого $\epsilon > 0$

$$P(|Z_n(t)| \geq \epsilon) \leq \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} (C_n t)^{\delta_1}.$$

Доказательство. С учетом того, что одномерные распределения процесса Кокса (5) являются смешанными пуассоновскими, имеем:

$$\begin{aligned} P(|Z_n(t)| \geq \epsilon) &= P\left(\left|\sum_{j=1}^{N_1^{(n)}(\Lambda_n(t))} X_{n,j}\right| \geq \epsilon\right) = \\ &= \sum_{k=0}^{\infty} P\left(N_1^{(n)}(\Lambda_n(t)) = k\right) P\left(\left|\sum_{j=1}^k X_{n,j}\right| \geq \epsilon\right) = \\ &= \sum_{k=0}^{\infty} \left(\int_0^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} dP(\Lambda_n(t) < \lambda)\right) \times \\ &\times P\left(\left|\sum_{j=1}^k X_{n,j}\right| \geq \epsilon\right) = \int_0^{\infty} \left[\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \times \right. \\ &\left. \times P\left(\left|\sum_{j=1}^k X_{n,j}\right| \geq \epsilon\right)\right] dP(\Lambda_n(t) < \lambda). \quad (9) \end{aligned}$$

Изменение порядка суммирования и интегрирования возможно в силу очевидной равномерной сходимости ряда. Цепочку (9) продолжим, последовательно применяя неравенства Маркова и Иенсена с $\delta \in (0, 1]$, фигурирующим в условии (7), и $\beta \in [1, 2]$, фигурирующим в условии (8), и получим:

$$\begin{aligned} P(|Z_n(t)| \geq \epsilon) &\leq \frac{1}{\epsilon^{\beta\delta}} \times \\ &\times \int_0^{\infty} \left[\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} E\left|\sum_{j=1}^k X_{n,j}\right|^{\beta\delta}\right] dP(\Lambda_n(t) < \lambda) \leq \\ &\leq \frac{1}{\epsilon^{\beta\delta}} \int_0^{\infty} \left[\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \times \right. \\ &\left. \times \left(E\left|\sum_{j=1}^k X_{n,j}\right|^{\beta}\right)^{\delta}\right] dP(\Lambda_n(t) < \lambda), \quad (10) \end{aligned}$$

поскольку для $\delta \in (0, 1]$ функция $f(x) = x^\delta$ является вогнутой при $x \geq 0$. Легко видеть, что при $1 \leq \beta \leq 2$

$$E\left|\sum_{j=1}^k X_{n,j}\right|^{\beta} \leq \sum_{j=1}^k E|X_{n,j}|^{\beta} = km_n^{\beta}.$$

Поэтому, продолжив (10) с учетом неравенства Иенсена для вогнутых функций и условия (7), получим:

$$\begin{aligned}
 P(|Z_n(t)| \geq \epsilon) &\leq \\
 &\leq \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} \int_0^\infty \left(\sum_{k=0}^\infty e^{-\lambda} \frac{k^\delta \lambda^k}{k!} \right) dP(\Lambda_n(t) < \lambda) = \\
 &= \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} \int_0^\infty E[N_1^{(1)}(\lambda)]^\delta dP(\Lambda_n(t) < \lambda) \leq \\
 &\leq \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} \int_0^\infty [EN_1^{(1)}(\lambda)]^\delta dP(\Lambda_n(t) < \lambda) = \\
 &= \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} \int_0^\infty \lambda^\delta dP(\Lambda_n(t) < \lambda) = \\
 &= \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} E\Lambda_n^\delta(t) \leq \frac{m_n^{\beta\delta}}{\epsilon^{\beta\delta}} (C_n t)^{\delta_1}.
 \end{aligned}$$

Лемма доказана.

Чтобы установить слабую сходимость случайных процессов $Z_n(t)$ в пространстве Скорохода \mathcal{D} , сначала необходимо найти предельное распределение случайных величин $Z_n(t)$ при каждом $t > 0$.

Пусть $t = 1$. Обозначим $N_n = N_1^{(n)}(\Lambda_n(1))$. Предположим, что для некоторых $k_n \in \mathbb{N}$ при $n \rightarrow \infty$ имеет место сходимость

$$P(X_{n,1} + \dots + X_{n,k_n} < x) \rightarrow \Phi(x). \quad (11)$$

Обозначим $\sigma_n^2 = DX_{n,1} (= m_n^2)$. Из классической теории предельных теорем хорошо известно, что сходимость (11) имеет место, если при $n \rightarrow \infty$ выполнены условия:

$$k_n \sigma_n^2 \rightarrow 1 \quad (12)$$

$$k_n EX_{n,1}^2 \mathbb{I}(|X_{n,1}| \geq \epsilon) \rightarrow 0 \quad (13)$$

для любого $\epsilon > 0$ (условие Линдеберга).

Предположим, что

$$\frac{\Lambda_n(1)}{k_n} \Rightarrow U, \quad (14)$$

где U — некоторая неотрицательная случайная величина, распределение которой не является вырожденным в нуле. Заметим, что так как $\Lambda_n(t)$ — процесс Леви, то случайная величина U безгранично делима как слабый предел безгранично делимых случайных величин.

Как показано в [6] (также см., например, [9] или [11]), сходимость (14) эквивалентна тому, что

$$\frac{N_n}{k_n} \Rightarrow U. \quad (15)$$

По теореме переноса Гнеденко–Фахима [24] (см. теорему 2.9.1 в [11]) из условий (11) и (15) вытекает, что при $n \rightarrow \infty$

$$Z_n(1) = X_{n,1} + \dots + X_{n,N_n} \Rightarrow Z, \quad (16)$$

где Z — случайная величина с характеристической функцией

$$f(s) = \int_0^\infty \exp\left\{-\frac{us^2}{2}\right\} dP(U < u).$$

Несложно увидеть, что функция распределения $F(x)$ случайной величины Z является масштабной смесью нормальных законов:

$$F(x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{u}}\right) dP(U < u), \quad x \in \mathbb{R}.$$

Поскольку распределение случайной величины U безгранично делимо, то и функция распределения $F(x)$ безгранично делима (см., например, [22]). Поэтому можно определить процесс Леви $Z(t)$, $t \geq 0$, такой, что $Z(1) \stackrel{d}{=} Z$. Так как согласно (16)

$$Z_n(1) = \sum_{i=1}^{N_n} X_{n,i} \Rightarrow Z(1),$$

а процессы $Z_n(t)$ и $Z(t)$ являются процессами Леви, то, используя свойство характеристических функций процессов Леви (2), можно заключить, что для любого $t > 0$ имеет место сходимость

$$Z_n(t) = \sum_{i=1}^{N_{n,1}(\Lambda_n(t))} X_{n,i} \Rightarrow Z(t) \quad (n \rightarrow \infty). \quad (17)$$

Поскольку случайные процессы $Z_n(t)$, $0 \leq t \leq T$, и $Z(t)$, $0 \leq t \leq T$, являются процессами Леви, почти все их траектории принадлежат пространству Скорохода \mathcal{D} .

Рассмотрим вопрос о том, какие дополнительные условия нужны для слабой сходимости обобщенного процесса Кокса $Z_n(t)$ к процессу Леви $Z(t)$ в пространстве \mathcal{D} при $n \rightarrow \infty$. Последовательно рассмотрим каждое из условий теоремы А.

Во-первых, без ограничения общности положим $0 \leq t_1 < t_2 < \dots < t_k \leq 1$. Слабая сходимость

$$(Z_n(t_1), \dots, Z_n(t_k)) \Rightarrow (Z(t_1), \dots, Z(t_k)) \quad (n \rightarrow \infty)$$

эквивалентна сходимости

$$\begin{aligned}
 &(Z_n(t_1), Z_n(t_2) - Z_n(t_1), \dots, Z_n(t_k) - Z_n(t_{k-1})) \Rightarrow \\
 &\Rightarrow (Z(t_1), Z(t_2) - Z(t_1), \dots, Z(t_k) - Z(t_{k-1})) \\
 &\quad (n \rightarrow \infty), \quad (18)
 \end{aligned}$$

поскольку линейное преобразование $(x_1, x_2, \dots, x_{k-1}, x_k) \mapsto (x_1, x_2 - x_1, \dots, x_k - x_{k-1}) \in \mathbb{R}^k$ в \mathbb{R}^k взаимно однозначно и непрерывно в обе стороны. Но сходимость (18) вытекает из (17) и того, что и $Z_n(t)$, и $Z(t)$ являются процессами Леви.

Во-вторых, необходимо проверить условие $\mathbb{P}(Z(1) \neq Z(1-)) = 0$. Это условие выполнено тогда и только тогда, когда

$$\lim_{t \rightarrow 1-} \mathbb{P}(|Z(1) - Z(t)| > \epsilon) = 0$$

для любого $\epsilon > 0$ (см. соотношение (15.16) в [15, с. 177]). Рассмотрим величину $\mathbb{P}(|Z(1) - Z(t)| > \epsilon)$. Так как $Z(t)$ — процесс Леви, то по лемме 2

$$Z(1) - Z(t) \stackrel{d}{=} Z(1-t).$$

Поэтому

$$\mathbb{P}(|Z(1) - Z(t)| > \epsilon) = \mathbb{P}(|Z(1-t)| > \epsilon).$$

Для каждого $\epsilon > 0$ и каждого $t \in [0, 1]$ найдется такое $\epsilon_t \in [\epsilon/2, \epsilon]$, что точки $\pm \epsilon_t$ являются точками непрерывности функции распределения случайной величины $Z(1-t)$. Так как $Z_n(t) \Rightarrow Z(t)$ при каждом $t \in [0, 1]$, то

$$\mathbb{P}(|Z(1-t)| > \epsilon_t) = \lim_{n \rightarrow \infty} \mathbb{P}(|Z_n(1-t)| > \epsilon_t).$$

Таким образом, для каждого $\epsilon > 0$ и каждого $t \in [0, 1]$

$$\begin{aligned} \mathbb{P}(|Z(1-t)| > \epsilon) &\leq \mathbb{P}(|Z(1-t)| > \epsilon_t) = \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(|Z_n(1-t)| > \epsilon_t). \end{aligned} \quad (19)$$

Продолжив (19) с учетом условия (7) и применяя лемму 3, для $\delta \in (0, 1]$, фигурирующего в (7), имеем

$$\begin{aligned} \mathbb{P}(|Z(1-t)| > \epsilon) &\leq \sup_n \mathbb{P}(|Z_n(1-t)| > \epsilon_t) \leq \\ &\leq \sup_n \left(\epsilon_t^{-\beta} m_n^\beta \right)^\delta (C_n |1-t|)^{\delta_1} \leq \\ &\leq (2^{\beta\delta} \epsilon^{-\beta\delta} |1-t|)^{\delta_1} \sup_n m_n^{\beta\delta} C_n^{\delta_1}. \end{aligned} \quad (20)$$

Поэтому если

$$Q \equiv \sup_n C_n^{\delta_1/\delta} m_n^\beta < \infty, \quad (21)$$

то из (20) вытекает, что

$$\begin{aligned} \lim_{t \rightarrow 1-} \mathbb{P}(|Z(1) - Z(t)| > \epsilon) &\leq \\ &\leq 4(Q\epsilon^{-\beta})^\delta \lim_{t \rightarrow 1-} |1-t|^{\delta_1} = 0. \end{aligned}$$

В-третьих, проверим условие (1) в предположении, что выполнены условия (7) и (21). Как уже

отмечалось выше, $Z_n(t)$ — процесс Леви и, стало быть, имеет независимые приращения. Поэтому

$$\begin{aligned} \mathbb{P}(|Z_n(t) - Z_n(t_1)| \geq \epsilon, |Z_n(t_2) - Z_n(t)| \geq \epsilon) &= \\ &= \mathbb{P}(|Z_n(t) - Z_n(t_1)| \geq \epsilon) \times \\ &\times \mathbb{P}(|Z_n(t_2) - Z_n(t)| \geq \epsilon). \end{aligned} \quad (22)$$

Рассмотрим первый сомножитель в правой части (22). По лемме 2 $Z_n(t) - Z_n(t_1) \stackrel{d}{=} Z_n(t - t_1)$. Учитывая условие (21), по лемме 3 получим

$$\begin{aligned} \mathbb{P}(|Z_n(t) - Z_n(t_1)| \geq \epsilon) &= \mathbb{P}(|Z_n(t - t_1)| \geq \epsilon) \leq \\ &\leq (Q\epsilon^{-\beta})^\delta |t - t_1|^{\delta_1}. \end{aligned} \quad (23)$$

Для второго сомножителя в правой части (22) аналогично получим:

$$\begin{aligned} \mathbb{P}(|Z_n(t_2) - Z_n(t)| \geq \epsilon) &= \mathbb{P}(|Z_n(t_2 - t)| \geq \epsilon) \leq \\ &\leq (Q\epsilon^{-\beta})^\delta |t_2 - t|^{\delta_1}. \end{aligned} \quad (24)$$

Таким образом, из (23) и (24) вытекает, что

$$\begin{aligned} \mathbb{P}(|Z_n(t) - Z_n(t_1)| \geq \epsilon, |Z_n(t_2) - Z_n(t)| \geq \epsilon) &\leq \\ &\leq (Q\epsilon^{-\beta})^{2\delta} [(t - t_1)(t_2 - t)]^{\delta_1}. \end{aligned} \quad (25)$$

Несложно видеть, что для любых $t_1 \leq t \leq t_2$

$$(t - t_1)(t_2 - t) \leq \frac{1}{4}(t_2 - t_1)^2.$$

Подставляя эту оценку в (25), получаем

$$\begin{aligned} \mathbb{P}(|Z_n(t) - Z_n(t_1)| \geq \epsilon, |Z_n(t_2) - Z_n(t)| \geq \epsilon) &\leq \\ &\leq \epsilon^{-2\beta\delta} \left[\frac{Q(t_2 - t_1)}{2} \right]^{2\delta_1}. \end{aligned}$$

Следовательно, если выполнены условия (7) и (21), то условие (1) выполнено с $F(t) \equiv (1/2)Qt$, $\nu = 2\delta$ и $\gamma = \delta_1$.

Суммируя приведенные выше рассуждения, связанные с проверкой условий теоремы А, приходим к следующему утверждению.

Теорема 1. Пусть обобщенные процессы Кокса $Z_n(t)$ (см. (5)) управляются процессами Леви $\Lambda_n(t)$, удовлетворяющими условиям (7) и (14) с некоторыми $\delta_1 \in (0, 1]$ и $k_n \in \mathbb{N}$. Предположим, что случайные величины $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$, — скачки обобщенного процесса Кокса $Z_n(t)$ — удовлетворяют условиям (12) и (13) с теми же самыми k_n . Предположим также, что для некоторого $\beta \in [1, 2]$ выполнено условие (21).

Тогда случайные блуждания, порожденные указанными обобщенными процессами Кокса, слабо сходятся в пространстве Скорохода \mathcal{D} к процессу Леви $Z(t)$ с

$$P(Z(1) < x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{u}}\right) dP(U < u), \quad x \in \mathbb{R}. \quad (26)$$

Пример 1. Предположим, что $\Lambda_n(0) = 0$, $\Lambda_n(1) \stackrel{d}{=} k_n U_{\alpha,1}^{(n)}$, где $\{k_n\}_{n \geq 1}$ — неограниченно возрастающая последовательность натуральных чисел, а $U_{\alpha,1}^{(1)}, U_{\alpha,1}^{(2)}, \dots$ — последовательность одинаково распределенных почти наверное положительных случайных величин, имеющих одностороннее устойчивое распределение с параметрами $\alpha \in (0, 1]$ и $\theta = 1$. Из сказанного выше вытекает, что

$$E\Lambda_n^\delta(1) < \infty$$

для любого $\delta < \alpha$ и с учетом (4)

$$\begin{aligned} \Lambda_n(t) &\stackrel{d}{=} t^{1/\alpha} \Lambda_n(1) \stackrel{d}{=} t^{1/\alpha} k_n U_{\alpha,1}^{(n)} \stackrel{d}{=} \\ &\stackrel{d}{=} t^{1/\alpha} k_n U_{\alpha,1}^{(1)}, \quad t \geq 0, \end{aligned} \quad (27)$$

так что

$$E\Lambda_n^\delta(t) = t^{\delta/\alpha} k_n^\delta E\left(U_{\alpha,1}^{(1)}\right)^\delta$$

и условие (7) выполняется с $\delta_1 = \delta/\alpha$ и $C_n = \left(k_n \left[E\left(U_{\alpha,1}^{(1)}\right)^\delta\right]^{1/\delta}\right)^\alpha$. Предположим, что (8) выполнено при $\beta = 2$, т. е.

$$0 < \sigma_n^2 \equiv m_n^2 < \infty.$$

Тогда $\delta_1/\delta = \alpha$ и $C_n^{\delta_1/\delta} \sigma_n^2 = C_n^{1/\alpha} \sigma_n^2 = \left[E\left(U_{\alpha,1}^{(1)}\right)^\delta\right]^{1/\delta} k_n \sigma_n^2$. Поэтому если выполнено условие (12), то условие (21) выполняется автоматически.

Более того, в силу (27) очевидно, что

$$\frac{\Lambda_n(1)}{k_n} \stackrel{d}{=} \frac{k_n U_{\alpha,1}^{(1)}}{k_n} = U_{\alpha,1}^{(1)}.$$

Поэтому условие (14) также выполняется автоматически с $U = U_{\alpha,1}^{(1)}$. При этом в силу представления (3) соотношение (26) принимает вид:

$$P(Z(1) < x) = G_{2\alpha,0}(x), \quad x \in \mathbb{R}.$$

Таким образом, для рассматриваемой ситуации из теоремы 1 вытекает

Следствие 1. Пусть $\alpha \in (0, 1]$ и обобщенные процессы Кокса $Z_n(t)$ (см. (5)) управляются процессами Леви

$\Lambda_n(t)$ такими, что $\Lambda_n(1) \stackrel{d}{=} k_n U_{\alpha,1}^{(n)}$, $\{k_n\}_{n \geq 1}$ — неограниченно возрастающая последовательность натуральных чисел, а $U_{\alpha,1}^{(1)}, U_{\alpha,1}^{(2)}, \dots$ — последовательность одинаково распределенных почти наверное положительных случайных величин, имеющих одностороннее устойчивое распределение с параметрами α и $\theta = 1$. Предположим, что случайные величины $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$, — скачки обобщенного процесса Кокса $Z_n(t)$ — удовлетворяют условиям (12) и (13) с теми же самыми числами k_n . Тогда случайные блуждания, порожденные указанными обобщенными процессами Кокса, слабо сходятся в пространстве Скорохода \mathcal{D} к 2α -устойчивому процессу Леви $Z(t)$ с $P(Z(1) < x) = G_{2\alpha,0}(x)$.

Пример 2. Предположим, что $\Lambda_n(t)$ — процесс Леви с $E\Lambda_n(1) < \infty$, $n \in \mathbb{N}$. В таком случае, используя представление (2) и известное соотношение между математическими ожиданиями случайных величин и производными соответствующих характеристических функций, легко убедиться, что

$$E\Lambda_n(t) = tE\Lambda_n(1), \quad t \in [0, 1], \quad n \in \mathbb{N}.$$

Поэтому в рассматриваемой ситуации условие (7) выполнено с $\delta = \delta_1 = 1$ и $C_n = E\Lambda_n(1)$, так что теореме 1 можно придать следующий вид.

Следствие 2. Пусть обобщенные процессы Кокса $Z_n(t)$ (см. (5)) управляются процессами Леви $\Lambda_n(t)$, удовлетворяющими условию (14) с некоторыми $k_n \in \mathbb{N}$. Предположим дополнительно, что семейство случайных величин $\{k_n^{-1} \Lambda_n(1)\}_{n \geq 1}$ равномерно интегрируемо. Предположим, что случайные величины $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$, — скачки обобщенного процесса Кокса $Z_n(t)$ — удовлетворяют условиям (12) и (13) с теми же самыми k_n . Тогда случайные блуждания, порожденные указанными обобщенными процессами Кокса, слабо сходятся в пространстве Скорохода \mathcal{D} к процессу Леви $Z(t)$, удовлетворяющему соотношению (26).

Чтобы убедиться в справедливости следствия 2, достаточно заметить, что условие (14) в совокупности с равномерной интегрируемостью семейства $\{k_n^{-1} \Lambda_n(1)\}_{n \geq 1}$ означает, что $EU < \infty$ и

$$\frac{E\Lambda_n(1)}{k_n} \longrightarrow EU$$

(см., например, [25, с. 196]). Поэтому существует конечное положительное число A такое, что

$$E\Lambda_n(1) \leq Ak_n \quad (28)$$

для всех n начиная с некоторого. Поэтому условие (12) и соотношение (28) автоматически влекут

$$\sup_n \sigma_n^2 E\Lambda_n(1) < \infty,$$

т. е. условие (21) выполнено с $\beta = 2$, $\delta = \delta_1 = 1$ и $C_n = \sigma_n^2$. Теперь требуемое утверждение вытекает из теоремы 1.

Следующее утверждение содержит немного более ограничительное, но, возможно, более удобно проверяемое условие, нежели равномерная интегрируемость семейства $\{k_n^{-1}\Lambda_n(1)\}_{n \geq 1}$.

Следствие 3. Пусть обобщенные процессы Кокса $Z_n(t)$ (см. (5)) управляются процессами Леви $\Lambda_n(t)$, удовлетворяющими условию (14) с некоторыми $k_n \in \mathbb{N}$, причем для некоторого $r > 1$

$$\sup_n \mathbb{E} \left[\frac{\Lambda_n(1)}{k_n} \right]^r < \infty. \quad (29)$$

Предположим, что случайные величины $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$, — скачки обобщенного процесса Кокса $Z_n(t)$ — удовлетворяют условиям (12) и (13) с теми же самыми k_n . Тогда случайные блуждания, порожденные указанными обобщенными процессами Кокса, слабо сходятся в пространстве Скорохода \mathcal{D} к процессу Леви $Z(t)$, удовлетворяющему соотношению (26).

Чтобы убедиться в справедливости следствия 3, достаточно заметить, что из условия (29) вытекает равномерная интегрируемость семейства $\{k_n^{-1}\Lambda_n(1)\}_{n \geq 1}$ (см., например, [25, с. 198]), и сойтись на следствие 2.

Частный случай следствия 3 для $r = 2$ доказан в [13, 14] прямым методом.

Литература

1. Ширяев А. Н. Основы стохастической финансовой математики. Т. 1: Факты. Модели. — М.: Фазис, 1998.
2. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского университета, 2011.
3. Mandelbrot B. B. The variation of certain speculative prices // J. Business, 1963. Vol. 36. P. 394–419.
4. Королев В. Ю. О сходимости распределений случайных сумм независимых случайных величин к устойчивым законам // Теория вероятностей и ее применения, 1997. Т. 42. Вып. 4. С. 818–820.
5. Королев В. Ю. О сходимости распределений обобщенных процессов Кокса к устойчивым законам // Теория вероятностей и ее применения, 1998. Т. 43. Вып. 4. С. 786–792.
6. Gnedenko B. V., Korolev V. Yu. Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996.
7. Королев В. Ю. Построение моделей распределений биржевых цен при помощи методов асимптотической теории случайного суммирования // Обозрение промышленной и прикладной математики, 1997. Т. 4. Вып. 1. С. 86–102.
8. Королев В. Ю. Асимптотические свойства экстремумов обобщенных процессов Кокса и их применение к некоторым задачам финансовой математики // Теория вероятностей и ее применения, 2000. Т. 45. Вып. 1. С. 182–194.
9. Bening V., Korolev V. Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002.
10. Королев В. Ю., Соколов И. А. Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008.
11. Королев В. Ю., Бенинг В. Е., Шоргин С. Я. Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011.
12. Stochastic models of structural plasma turbulence / Eds. V. Korolev, N. Skvortsova N. — Utrecht: VSP, 2006.
13. Кащеев Д. Е. Функциональные предельные теоремы для сложных процессов Кокса // Обозрение прикладной и промышленной математики, 2000. Т. 7. Вып. 2. С. 494–495.
14. Кащеев Д. Е. Моделирование динамики финансовых временных рядов и оценивание производных ценных бумаг: Дисс. . . . канд. физ.-мат. наук. — Тверь: Тверской гос. ун-т, 2001.
15. Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977.
16. Bertoin J. Lévy processes // Cambridge Tracts in Mathematics. Vol. 121. — Cambridge: Cambridge University Press, 1996.
17. Sato K. Lévy processes and infinitely divisible distributions. — Cambridge: Cambridge University Press, 1999.
18. Barndorff-Nielsen O. E., Mikosch T., Resnick S. I. Lévy processes: Theory and applications. — Boston: Birkhäuser, 2001.
19. Schoutens W. Lévy processes in finance: Pricing financial derivatives. — New York: Wiley, 2003.
20. Geman H. Pure jump Lévy processes for asset price modelling // J. Banking Finance, 2002. Vol. 26. No. 7. P. 1297–1316.
21. Золотарев В. М. Одномерные устойчивые распределения. — М.: Наука, 1983.
22. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. — М.: Мир, 1984.
23. Embrechts P., Maejima M. Selfsimilar processes. — Princeton: Princeton University Press, 2002.
24. Гнеденко Б. В., Фахим Х. Об одной теореме переноса // Докл. АН СССР, 1969. Т. 187. Вып. 1. С. 15–17.
25. Лозэ М. Теория вероятностей. — М.: ИЛ, 1962.

СТАТИСТИЧЕСКИЕ МЕХАНИЗМЫ ФОРМИРОВАНИЯ АССОЦИАТИВНЫХ ПОРТРЕТОВ ПРЕДМЕТНЫХ ОБЛАСТЕЙ НА ОСНОВЕ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ БОЛЬШИХ ОБЪЕМОВ ДЛЯ СИСТЕМ ИЗВЛЕЧЕНИЯ ЗНАНИЙ*

М. М. Шарнин¹, Н. В. Сомин², И. П. Кузнецов³, Ю. И. Морозова⁴, И. В. Галина⁵,
Е. Б. Козеренко⁶

Аннотация: Ассоциативные связи между терминами, понятиями и другими элементами естественного языка (ЕЯ) играют важную роль в решении широкого класса прикладных задач, среди которых интеллектуальная обработка текстов, извлечение знаний, их обработка с формированием баз знаний и организация различных видов поиска, в том числе семантических. Предложены методы автоматизированного выявления ассоциативных связей в текстах из Интернета и построения ассоциативных портретов различных предметных областей, ориентированных на решение перечисленных задач. Ассоциативный портрет предметной области (АППО) представляет собой словарь значимых терминов и словосочетаний, элементы которого связаны ассоциативными связями. Ассоциативный портрет предметной области создается автоматическим на базе статистического анализа больших объемов текстов. Теоретическая значимость предлагаемого подхода заключается в использовании методов статистики, корпусной лингвистики и дистрибутивной семантики для обработки больших текстовых массивов на ЕЯ (постоянно пополняемых и обновляемых в Интернете) с целью построения модели предметной области в виде АППО.

Ключевые слова: автоматическая обработка корпусов текстов; статистические методы; интеллектуальные интернет-технологии; лексико-семантический анализ; извлечение знаний из текстов; семантический поиск; семантические векторы; семантическое контекстное пространство

1 Постановка задачи

Для компьютерной лингвистики актуальной является фундаментальная научная проблема создания лингвистической и предметной базы для систем извлечения и обработки знаний. Создание вручную универсальных хорошо выверенных тезаурусов требует колоссальных трудовых и временных затрат, поэтому в последние годы ведется поиск новых путей создания лингвистической предметно-ориентированной базы знаний. Одним из таких направлений является создание АППО, на основе которых можно с достаточной надежностью решать большинство важнейших проблем, связанных с информационным поиском, извлечением структур знаний и построением классификаций.

Ассоциативный портрет — это совокупность наиболее характерных предметных и лингвистических знаний, свойственных определенной пред-

метной области. Под предметными знаниями понимаются присущие предметной области термины, понятия, связанные различного рода ассоциативными связями. Лингвистические знания — это варианты словарного выражения понятий.

В данной работе основное внимание обращено на статистические механизмы и обоснование методики автоматизированного выявления ассоциативных связей между значимыми словосочетаниями.

Суть предлагаемого подхода состоит в обеспечении возможности создавать и использовать АППО автоматически, с минимальным вмешательством лингвиста. В связи с этим трудоемкость разработки АППО будет на несколько порядков ниже разработки традиционных тезаурусов. Это позволит АППО стать реальным дополнением тезаурусов в новом поколении интеллектуальных интернет-технологий, которые будут отличаться, как предполагается, большей полнотой охвата терминологии

* Работа выполнена при частичной поддержке РФФИ (проект 13-07-00272-а).

¹ Институт проблем информатики Российской академии наук, l@keywen.com

² Институт проблем информатики Российской академии наук, somin@post.ru

³ Институт проблем информатики Российской академии наук, igor-kuz@mtu-net.ru

⁴ Институт проблем информатики Российской академии наук, judez@yandex.ru

⁵ Институт проблем информатики Российской академии наук, irn_gl@mail.ru

⁶ Институт проблем информатики Российской академии наук, kozerenko@mail.ru

лишь при незначительном снижении точности. Тем самым преодолевается главная проблема — трудоемкости ручного труда лингвистов, не позволявшая проводить разработки с достаточной скоростью.

Настоящая работа ставит своей целью описать методы построения АППО и их возможные применения для создания аналитических систем обработки текстов.

Решение проблемы связано с анализом больших массивов текстов русского языка на основе моделей векторных пространств и методов дистрибутивной семантики, ориентированных на автоматический анализ текстов (что отличает данный проект от широкомасштабных проектов типа WordNet и RusNet).

Отметим, что для русского языка подобные исследования проводятся впервые.

2 Существующие модели

В настоящее время известно значительное количество работ по автоматическому извлечению семантических связей из больших массивов текстов на ЕЯ [1–15]. Наиболее успешные подходы используют метод дистрибутивной семантики и модели семантических векторных пространств (СВП). В основе всех вариантов этого метода лежат количественные оценки, которые характеризуют совместную встречаемость языковых единиц текста в контекстах определенной величины. Основная гипотеза метода состоит в том, что слова, встречающиеся вместе в пределах некоторого текстового интервала, как-то связаны между собой. Для оценки связанности вводится коэффициент «силы связи», который рассчитывается по некоторой формуле. Вне зависимости от вида формулы в ней обычно используются характеристики совместной встречаемости пар слов и одиночной встречаемости каждого из слов.

Величина контекста, в рамках которого осуществляются подсчеты коэффициентов «силы связи», как показывают результаты исследований, позволяет наиболее вероятно устанавливать:

- (а) при малых размерах контекста, ограниченного одним или двумя соседними словами, — контактные синтагматические связи словосочетаний;
- (б) при размере 5–10 слов — дистантные синтагматические связи и парадигматические отношения;
- (в) дальнейшее увеличение ширины контекста до 50–100 слов (размер предложения, сверхфразового единства, абзаца) — тематические связи между словами.

Тематические связи могут оказаться доминирующими, если принять размер контекста величиной с сам текст [1, с. 120].

Модели векторных пространств находят все более широкое применение в исследованиях, связанных с семантическими моделями ЕЯ, и имеют разнообразный спектр потенциальных и действующих приложений [6, 7]. Данная область в настоящее время является одной из наиболее актуальных. Следует отметить модели Word-Space Model [16] и Semantic Space Model [17]. В основе этих работ лежит пространство лексем и термов. Такое пространство базируется на распределении слов в корпусе текстов с целью представления их семантической связанности путем оценки пространственной близости.

Отметим, что концепция СВП впервые была реализована в информационно-поисковой системе SMART [18]. SMART был пионером многих концепций, которые успешно используются современными поисковиками [19]. Идея СВП состоит в представлении каждого документа из коллекции в виде точки в пространстве, т. е. вектора в векторном пространстве. Точки, расположенные ближе друг к другу в этом пространстве, считаются более близкими по смыслу. Пользовательский запрос рассматривается как псевдодокумент и тоже представляется как точка в этом же пространстве. Документы сортируются в порядке возрастания расстояния, т. е. в порядке уменьшения семантической близости от запроса, и в таком виде предоставляются пользователю. В настоящее время большинство поисковиков используют СВП для измерения степени близости запроса и найденных документов [19].

Varoni и Lenci [9] предложили обобщенную модель, названную «дистрибутивная память», которая является обобщением ранее известных моделей векторных пространств (vector spaces), семантических пространств (semantic spaces), пространств слов (word spaces), семантических моделей корпусной статистики (corpus-based semantic models) и дистрибутивных семантических моделей (distributional semantic models).

Успех СВП в решении информационно-поисковых задач направил исследователей на применение СВП для других семантических задач. Например, Rapp [2] использовал контекстное векторное пространство для оценки семантической близости слов. Его система достигла результата 92,5% на тесте по выбору наиболее подходящего синонима из стандартного теста английского языка TOEFL, в то время как средний результат людей был 64,5%.

В настоящее время ведутся активные исследования по унификации модели СВП и выработке общего подхода к различным задачам выявления семантических связей из корпусов текстов [7].

В предлагаемой работе используется более общий термин — семантическое контекстное пространство (СКП), где точки пространства соответствуют контекстным векторам не отдельных слов или термов, а значимых словосочетаний (см. разд. 3). Отметим, СКП является абстрактной моделью, не связанной с предметной областью. Приложение модели связано с выбором такой области и ее компонент: знаковых словосочетаний, ассоциативных связей и, соответственно, контекстов для их выделения. В результате образуется упоминавшееся ранее АППО (см. разд. 1).

Более того, разработчики СВП отмечают, что основная проблема известных семантических пространств заключается в трудностях учета порядка слов, составляющих контексты. В рамках данного проекта (СКП) эта проблема решается путем перехода от контекста слов к контексту значимых словосочетаний. Отметим также, что технология СВП развивалась для английского языка. Проект СКП предполагает работу как с русским, так и с английским языком. Возможно в дальнейшем включение других языков. Таким образом, модель СКП является расширением моделей СВП.

Предлагаемые методы, разрабатываемые в рамках модели СКП, ориентированы на решение следующих задач:

- выявление синонимии и семантической близости слов и словосочетаний путем оценки их встречаемости в различных контекстах;
- автоматическая кластеризация слов по степени их близости в СКП;
- автоматическая классификация слов путем использования лексико-семантических форм;
- автоматическая генерация тезаурусов методами статистической обработки терминов;
- разрешение неоднозначности слов путем использования контекста;
- расширение запросов за счет ассоциативных связей;
- извлечение знаний из текстов с использованием статистических методов и лингвистических моделей;
- оценка степени сходства лексических конструкций на основе их лексико-семантического анализа;
- поиск отношений с помощью лексико-синтаксических форм;
- выявление близких по смыслу отношений и их классификация методами статистического анализа контекстных зависимостей.

3 Значимые словосочетания и определение семантического контекстного пространства

Для автоматического построения СКП предлагаются методики автоматического выявления значимых словосочетаний (ЗС), которые в общем случае рассматриваются как лексические последовательности, имеющие самостоятельную значимость, и которые определяются по абсолютной частоте их встречаемости в текстах предметной области [4, 8]. Предполагается также выявление ассоциативных связей между ЗС в определенной предметной области. При этом для выделения ЗС используются методы статистического ранжирования, а для расчета силы ассоциативной связи ЗС используется косинусная мера между контекстными векторами (компонентами вектора ЗС являются частоты совместной встречаемости данного ЗС с другими ЗС в одном и том же контексте). Такие векторы образуют СКП. Следует отметить, что посредством ЗС могут быть представлены как отдельные слова, словосочетания и термины, так и более сложные конструкции — объекты и именованные сущности.

Для вычисления косинусной меры между контекстными векторами используется следующая формула:

$$\frac{xy}{|x||y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}.$$

В методе СКП существенное значение имеет выбор контекста. Вообще под контекстом в данном случае понимается некий алгоритм идентификации текстовых фрагментов, связанных с данным ЗС. Всевозможные контексты и являются координатами векторов СКП. В зависимости от того, какие контексты считаются идентичными, различают типы контекстов. Классификация и даже перечисление типов контекстов — проблема, которая в силу своей новизны требует особого рассмотрения [15], поэтому ограничимся лишь примерами. Так, статья (абзац) может считаться контекстом данного ЗС (если данное ЗС не входит в эту статью, то можно говорить о нулевом контексте). В этом случае координатами вектора СКП являются все статьи, относящиеся к определенной предметной области. Другой пример: все предложения, содержащие ЗС, считаются контекстом данного ЗС.

Именно на использовании такого контекста построен пример использования методики построения СКП на основе следующего текстового фрагмента:

Искусственный интеллект — наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ.

Компьютерная лингвистика — направление искусственного интеллекта, которое ставит своей целью использование математических моделей для описания естественных языков.

Дискретная математика — область математики, занимающаяся изучением дискретных структур, которые возникают как в пределах самой математики, так и в ее приложениях.

Конструктивная математика — близкое к интуиционизму течение в математике, изучающее конструктивные построения.

Построим контекстные векторы для ЗС «искусственный интеллект», «компьютерная лингвистика», «дискретная математика», «конструктивная математика» и слов, встречающихся в текстовом фрагменте более одного раза.

Применив формулу вычисления косинусной меры между контекстными векторами, получим следующие коэффициенты семантической близости между рассматриваемыми ЗС:

«дискретная математика» и «конструктивная математика» — 0,95;

«искусственный интеллект» и «компьютерная лингвистика» — 0,7;

«компьютерная лингвистика» и «дискретная математика» — 0,52;

«компьютерная лингвистика» и «конструктивная математика» — 0,4;

«искусственный интеллект» и «дискретная математика» — 0,36;

«искусственный интеллект» и «конструктивная математика» — 0,29.

Отметим, что в реальных приложениях такого рода матрицы будут очень большими — порядка миллионов столбцов (строк). Однако благодаря тому, что подавляющее большинство элементов будет заполнено нулями, реально хранимая информация будет вполне обозримой.

Для нахождения ассоциативных связей, которые могут войти в ассоциативный портрет, необходимо выбрать из этой матрицы пары терминов с самыми большими коэффициентами семантической близости (косинусной меры). В данном случае это будут пары («дискретная математика», «конструктивная математика») и («искусственный интеллект», «компьютерная лингвистика»).

4 Ассоциативные портреты и методы их получения

Ассоциативный портрет предметной области — это множество ассоциативных связей между значимыми терминами предметной области. Формально АППО определяется как граф $G = (V, E)$ с узлами v из V , представляющими значимые термины/словосочетания и дугами графа $(v_i, v_j, \text{Link}, w_{ij})$ из E , описывающие отношения/связи между словосочетаниями, где w_{ij} — это вес, выражающий силу связи, а Link — тип связи, определяемый типом контекста. Тип контекста определяется параметрами алгоритма расчета контекстных векторов, такими как размер контекстного окна или тип лексико-синтаксического шаблона/конструкции, связывающей словосочетания.

Идеология АППО базируется на дистрибутивной гипотезе, утверждающей, что семантически близкие (или связанные) лексемы имеют похожий контекст и, наоборот, при похожем контексте лексемы семантически близки. В предлагаемой модели используется расширенная гипотеза, предполагающая, что при сходстве контекстов близкими признаются не только отдельные лексемы, но и произвольные многолексемные фрагменты — значимые словосочетания.

Получение АППО предполагает реализацию совокупности методов, включающих:

- методы выявления в Интернете текстов определенных предметных областей;
- методы выявления в текстах ЗС и их ранжирования;
- методы выявления и ранжирования ассоциативных связей между ЗС.

Методы основаны на прохождении предварительного обучения на текстах, в том числе взятых из различных интернет-ресурсов.

Обработка больших массивов текстов, постоянно пополняемых в сети Интернет, позволяет собирать необходимые статистические данные для формирования достаточно полной картины о предметной области, представленной в виде СКП. Возможность проводить машинное обучение на большом числе примеров придает системе определенную гибкость и улучшает результаты.

Отображение многомерных векторов на плоскость является удобным средством визуализации связей. В результате образуются визуальные карты ЗС. На таких картах расстояние между ЗС тем меньше, чем больше сила их ассоциативной связи, что позволяет выделять сильно связанные и близкие по смыслу ЗС.

Подход предусматривает исследование различных типов и источников контекста, а также различных методов выделения контекста и оценки силы ассоциативной связи по контекстным векторам. Важным представляется исследование таких типов контекста, как простая совместная встречаемость ЗС в тексте, а также совместная встречаемость ЗС в текстах в рамках заданных лексико-синтаксических шаблонов.

Предлагаемые в данном проекте методы и модели позволяют учитывать порядок слов в лексических последовательностях, что необходимо для более качественного решения вышеперечисленных задач, в том числе для выявления ассоциативных связей словосочетаний, объектов и именованных сущностей. Предлагается также разработать средства визуализации, которые позволят точно отобразить координаты выявленных объектов в двумерном семантическом контекстном пространстве и использовать эвристические методы для установления степени близости объектов.

Дистрибутивно-статический метод позволяет на основе частотной информации о ЕЯ-единицах получать по некоторой заданной формуле количественную характеристику их связанности. Философия данного метода состоит в том, «что семантическую классификацию значимых элементов языка можно с большим основанием индуктивно извлечь из анализа текста, чем получить ее с некоторой точки зрения, внешней по отношению к структуре языка. Следует ожидать, что такая классификация даст более надежные ответы на проблемы синонимии и выражения смысла, чем существующие тезаурусы и списки синонимов, основанные, главным образом, на интуитивно ощущаемых сходствах

без адекватной эмпирической проверки» [1, с. 115, 116].

5 Энциклопедия ключевых понятий и другие разработки авторов

Элементы предлагаемого подхода уже были частично апробированы в работах по созданию энциклопедии ключевых понятий KEYWEN (содержит 260 000 статей и 5 000 000 ключевых фраз, число которых постоянно растет) и в ряде работ других участников проекта.

Система Keywen представляет собой средство построения больших энциклопедий по материалам Интернета и на их основе составления рефератов и аналитических статей. Имеется опыт построения корпуса английских текстов из Интернета размером более 1 ТБ, проведены эксперименты по построению русских корпусов текстов для ряда предметных областей.

На рис. 1 приведен фрагмент статьи Artificial Intelligence из Keywen (<http://keywen.com/en/ARTIFICIAL-INTELLIGENCE>) с автоматически построенным набором ключевых слов, выбранных из ассоциативных связей термина Artificial Intelligence. В состав ключевых слов входят: MIT, MACHINE, COMPUTER SCIENCE, COGNITIVE SCIENCE, EXPERT SYSTEMS, TURING, MINSKY, REASONING.

На рис. 2 представлен абзац Categories из той же статьи Artificial Intelligence, демонстрирующий автоматический выбор иерархических связей и до-

The screenshot shows the Keywen website interface. At the top left is the logo 'Key-1' with 'Key-phrases KeyWen .COM' below it. To the right is a search bar and a 'Keywen Search' button. The main heading is 'Artificial intelligence' with a 'History' link. Below the heading is a breadcrumb trail: 'Encyclopedia of Keywords > Information > Information Technology > Computer Science > Artificial Intelligence'. A section titled 'Keywords and Sections' contains a list of keywords: ARTIFICIAL INTELLIGENCE, MIT, MACHINE, CONFERENCE, COMPUTER SCIENCE, COGNITIVE SCIENCE, PROCEEDINGS, HAYES, LABORATORY, EXPERT SYSTEMS, MCCARTHY, TURING, MINSKY, HANDBOOK, RESEARCHERS, REASONING. At the bottom is a section titled 'Review of Short Phrases and Links'.

Рис. 1 Фрагмент статьи Artificial Intelligence из Keywen

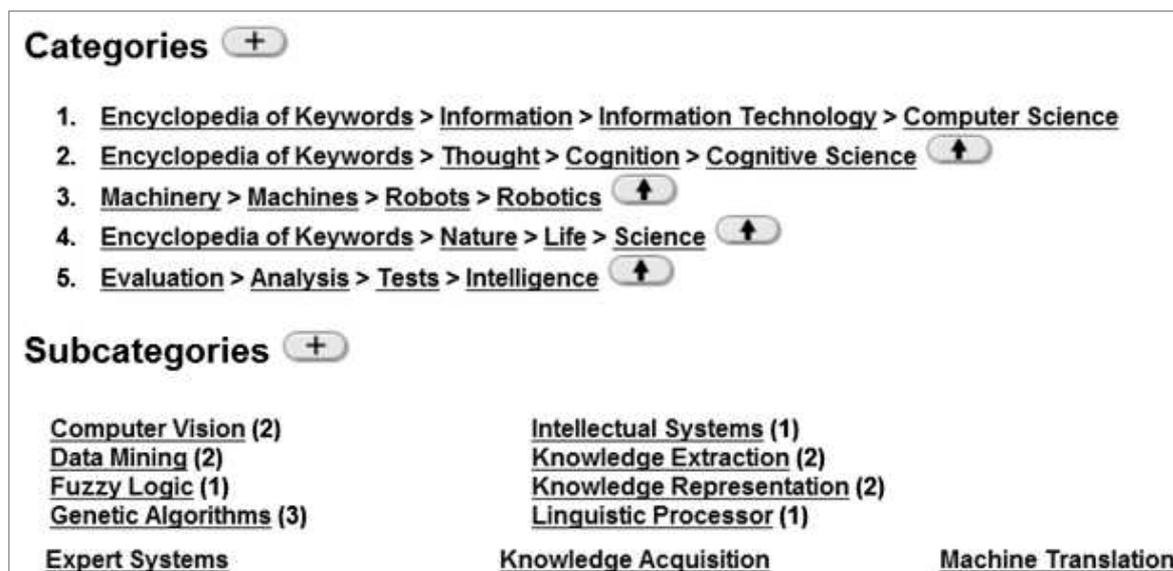


Рис. 2 Категории и подкатегории статьи «Искусственный интеллект»

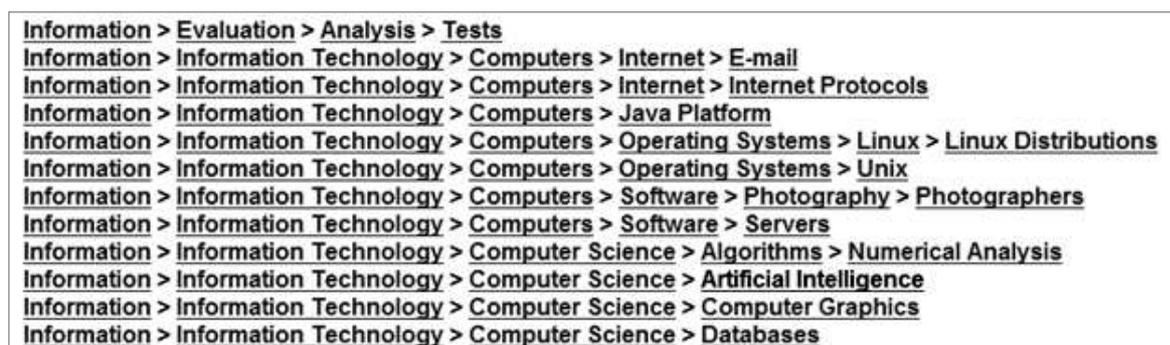


Рис. 3 Фрагмент общего дерева категорий

минирующей категории для статьи из набора ассоциативных связей, т. е. из набора ключевых слов статьи.

На рис. 2 также показаны доминирующие категории (Computer Science, Robotics, Science, Intelligence, Cognitive Science), которые система автоматически выбрала для статьи Artificial Intelligence. Для выбора доминирующей категории рассчитывались ассоциативные связи, найденные с использованием английских аналогов для следующих лексико-синтаксических шаблонов: «относиться к», «включать в себя», «классифицировать», «различать», «подразделять», «разделяться на», «входит», «составлять». Возможные шаблоны для русского и немецкого языка описаны в статье [15].

Из доминирующих категорий строится дерево всех категорий (<http://keywen.com/Category-Structure>), фрагмент которого представлен на рис. 3.

На основе проведенных исследований оформлены следующие патентные заявки в США:

1. System and method for generating hierarchical categories from collection of related terms. U.S. Provisional Patent Application No. 61/096,255, filed 2008.12.22 (Charnine).
2. USPTO Application #20100161671, filed 2010.06.24 (Charnine).

В основу заявок положено изобретение, описанное в монографии [8], которое было использовано для построения одной из крупнейших иерархий категорий для электронной энциклопедии Keywen. В монографии также опубликована информация о верхних уровнях иерархии категорий. Система Keywen не имеет аналогов как в России, так и за рубежом.

Статистические механизмы формирования АППО в настоящее время встраиваются в техно-

логии аналитической обработки текстов и извлечения информационных объектов (именованных сущностей) и их связей, в области которых авторский коллектив уже много лет ведет исследования и разработки: создан ряд прикладных систем, основанных на глубинном анализе текстов [10–14]:

- «Криминал» и «Аналитик» — системы, извлекающие информацию из милицейских сводок для логико-аналитических решений, которые являются основой для проведения следственных действий;
- «Semantix» — многоязычный лингвистический процессор, обеспечивающий выделение объектов (сущностей) и их связей из текстов на английском и русском языке;
- «Антитеррор» — система, извлекающая из текстов информацию о террористах и террористических акциях с автоматическим формированием базы знаний;
- «Резюме» — система семантической обработки резюме и их отображения в формат сайта крупной рекрутинговой компании;
- «Памятники» — обеспечивает извлечение информации из текстов описания памятников культуры и формирования ролевых функций относящихся к ним лиц;
- «ДИЕС» — диалоговая среда, обеспечивающая построение конкретных систем языкового взаимодействия с пользователем.

6 Направления дальнейших исследований

Авторской группой в первую очередь предполагается использование моделей СКП в двух областях приложений: поиске информации и построении классификаций. С помощью СКП можно существенно повысить гибкость поиска, построив реальные механизмы переформулирования запроса, позволяющие заменять части запроса иными — близкими к ним по контексту. Автоматическое построение классификаций может быть достигнуто путем выявления ассоциативных связей между ЗС.

Авторами разработана программная среда ДЕ-КЛАР, включающая в себя систему представления знаний и язык программирования задач, связанных с обработкой знаний. В рамках этой среды разработаны системы лексико-морфологического и синтактико-семантического анализа русских и английских текстов, которые используются при извлечении знаний. Программная среда ДЕ-КЛАР будет также применяться для разработки новых при-

кладных систем извлечения и обработки знаний, интегрирующих АППО.

Литература

1. *Москович В. А.* Информационные языки. — М.: Наука, 1971. 144 с.
2. *Rapp R.* Word sense discovery based on sense descriptor dissimilarity // 9th MT Summit Proceedings. — New Orleans, LA, 2003. P. 315–322.
3. *Шарнин М. М., Кузнецов И. П.* Автоматическое формирование электронных энциклопедий и справочных пособий по информации из сети Интернет // Системы и средства информатики. — М.: ИПИ РАН, 2004. Вып. 14. С. 210–223.
4. *Charnine M. M., Kuznetsov I. P., Kozerenko E. B.* Semantic navigator for Internet search // MLMTA'05: Conference (International) on Machine Learning Proceedings. — Las Vegas: CSREA Press, 2005. P. 60–68.
5. *Кузнецов И. П., Сомин Н. В.* Англо-русская система извлечения знаний из потоков информации в среде Интернет // Системы и средства информатики. — М.: ИПИ РАН, 2007. Вып. 17. С. 236–254.
6. *Lenci A.* Distributional semantics in linguistic and cognitive research // Rivista di Linguistica, 2008. Vol. 1. P. 1–30.
7. *Turney P.* A uniform approach to analogies, synonyms, antonyms and associations // Proceedings of COLING. — Manchester, 2008. P. 905–912.
8. *Charnine M., Charnine V.* Keywen category structure. — Wordclay, USA, 2008. 60 p.
9. *Baroni M., Lenci A.* Distributional memory: A general framework for corpus-based semantics // Comput. Linguistics, 2010. Vol. 36. Iss. 4. P. 673–721.
10. *Кузнецов И. П., Сомин Н. В.* Выявление имплицитной информации из текстов на естественном языке: проблемы и методы // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 48–57.
11. *Kuznetsov I. P., Kozerenko E. B., Charnin M. M.* Technological peculiarity of knowledge extraction for logical-analytical systems // WORLDCOMP'12: ICAI'12 Proceedings. — Las Vegas: CSREA Press, USA, 2012. Vol. II. P. 762–768.
12. *Шарнин М. М., Кузнецов И. П.* Особенности семантического поиска информационных объектов на основе технологии баз знаний // Информатика и её применения, 2012. Т. 6. Вып. 2. С. 47–56.
13. *Kuznetsov I. P., Charnine M. M., Kozerenko E. B., et al.* Intelligent tools for the semantic Internet navigator design // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. XIV Всеросс. научн. конф. RCDL'2012. — Переславль-Залесский: Университет города Переславля, 2012. С. 274–283.

14. Кузнецов И. П., Шарнин М. М., Мацкевич А. Г. Технология извлечения структур знаний с использованием аппарата расширенных семантических сетей // Искусственный интеллект: Журнал НАН Украины, 2012. Т. 4. С. 190–203.
15. Schumann A. Towards the automated enrichment of multilingual terminology databases with knowledge-rich contexts // Conference (International) Dialogue 2012 Proceedings. Vol. 1. P. 559–567.
16. *Sahlgren M.* The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. Ph.D. Thesis. — Department of Linguistics, Stockholm University, 2006.
17. *Sahlgren M.* Towards pertinent evaluation methodologies for word-space models // LREC 2006: 5th Conference (International) on Language Resources and Evaluation Proceedings. — Genoa, Italy, 2006.
18. The SMART retrieval system: Experiments in automatic document processing / Ed. G. M. Salton. — Prentice-Hall, 1971.
19. *Manning C., Raghavan P., Schütze H.* Introduction to information retrieval. — Cambridge: Cambridge University Press, 2008.

INFORMATION TECHNOLOGIES FOR CREATING THE DATABASE OF EQUIVALENT VERBAL FORMS IN THE RUSSIAN-FRENCH MULTIVARIANT PARALLEL CORPUS*

S. Loiseau¹, D. V. Sitchinava², A. A. Zalizniak³, and I. M. Zatsman⁴

Abstract: The Russian-French parallel corpus as a part of the Russian National Corpus (RNC) is being transformed into a multivariant corpus with several translations corresponding to each original texts. Concurrently, a Database of functionally equivalent lexicogrammatical verbal forms is being created using the multivariant corpus. The main purpose of database creation is to calculate the statistical estimates of the equivalences between Russian and French verbal forms. The paper discusses an information technology for creating the Russian-French multivariant parallel corpus and the database simultaneously.

Keywords: parallel multivariant corpora; Russian National Corpus; information technologies; XML marking up Russian-French parallel texts; lexicogrammatical form; functional equivalence; statistical estimates of equivalences

1 Introduction

Creating databases for cross-linguistic studies is a time-consuming process, but it is justified by the richness of linguistic data, including information on lexical units and grammatical structures they provide. Data can enter into such databases from off-line sources or digital resources such as parallel corpora [1]. The use of parallel corpora to create databases for cross-linguistic and contrastive studies allows to save time for actual research [2–5]. Contrastive study is very productive for linguistic representations design. Functional motivation of cross-linguistic models for machine translation and parallel texts alignment was presented in the works [6–9].

The parallel corpora within the RNC have been developed since 2005 [2, 3]. At the present time, within the RNC, there are eight bilingual parallel corpora, with Russian as the source language and eight other languages (English, German, French, Spanish, Italian, Polish, Ukrainian, Belarusian) as target languages, and one multilingual parallel corpus. The RNC includes Russian-French and French-Russian parallel corpora with a single translation for each textual pair. In 2013, the Russian-French parallel corpus will be transformed into a multivariant corpus with one Russian original and its several French translations. The multivariant corpus of parallel texts will be formed on the basis of the pre-

existing monovariant parallel corpus. In this corpus the sentence-alignment association between an original text and translated texts should be maintained.

The purpose of this paper is to outline the research questions related to the creation of the *Database of functionally equivalent lexicogrammatical verbal forms* using the Russian-French multivariant parallel corpus. One of the main objectives of the database creation is to obtain the statistical estimates of different types of equivalences between Russian and French verbal forms. Now, there is a need of an information technology for creating the Russian-French multivariant parallel corpus and the database simultaneously.

The information technology for creating the corpus and the database should support the following basic functions:

- alignment of the parallel texts with several variants of translation;
- morphological annotation and lemmatization of the parallel texts⁵;
- including the parallel texts with several variants of translation into the corpus;
- constructing the database of the equivalent lexicogrammatical verbal forms (Russian-French);
- calculating the statistical estimates of these equivalences between Russian and French verbal forms.

*The reported study was partially supported by RFBR, research project No. 12-06-33038 and project No. 13-06-00403.

¹Université Paris 13, Sorbonne Paris Cité, Laboratoire LDI (Lexiques, dictionnaires, informatique), CNRS, UMR 7187, sylvain.loiseau@univ-paris13.fr

²Institute of the Russian Language of the Russian Academy of Sciences, mitrius@gmail.com

³Institute of Linguistics of the Russian Academy of Sciences; Institute of Informatics Problems of the Russian Academy of Sciences, anna.zalizniak@gmail.com

⁴Institute of Informatics Problems of the Russian Academy of Sciences, iz_ipi@a170.ipi.ac.ru

⁵For the solution of task 2 and 3 it is planned to use software of the Russian National Corpus.

The main research question is how to align and mark up the parallel texts in order to construct equivalences between Russian and French verbal forms. In this paper, the authors outline their experience with the development of information technologies for creating the Russian-French multivariant parallel corpus and the database simultaneously.

2 Different Translations of the Same Text: Methodological Tasks

A framework for aligning several French translations of the same Russian text within the French-Russian parallel corpora is suggested. The Russian prose has been translated into French for more than 150 years, and the translations of different periods greatly vary as to their accuracy, choice of lexicon and grammar, and style. A comparison of these translations can be a valid task per se. Alongside with this, lexical and grammatical studies performed on parallel corpora may use alternative options of translations, envisaging all possible correspondences. They can reflect objective variability in the target language and can be a valuable resource for contrastive grammatical description of both languages. For instance, one can expect to observe greater variability amongst the translations when there are structural differences between the two languages (a Russian “signifié” with no clear correspondence in the French lexicon is more likely to exhibit variation in its translations). Several fruitful avenues of research are opened:

- it may provide valuable insight into phenomena of short diachronical variation: some variation in the translation of the same source texts may be related to diachronic variation in the target language;
- when lexical and grammatical values differ between the two languages, the inventory of the choices made by the translators may help uncovering the values of the elements translated. For instance, how the Russian aspectual system is reflected into the French system? Which are the various options, and how are they conditioned by the context? This line of research will bring new elements for a contrastive grammar of Russian and French; and
- the variability of the translations may help discern the intricate parameters — cultural, stylistic, lexical, and grammatical — at work in the translation process, since local choice may be interpreted in the wider context of the choices made by the translator.

Since the corpus is built on texts for which several translations are available, the corpus will rely mainly on the classical Russian fiction writers of the last two

centuries. It will be based on the larger Russian-French corpora previously built for the RNC, which also contains law texts, subtitles, and newspapers. In the literary classics, special attention will be given to some axes of variation, in particular, the ones of diachrony and genres (novel, short story, children’s literature).

Quantitative methodologies for contrastive grammatical description based on aligned corpora are being elaborated. The quantitative methods may be used for identifying the most frequent cases of variability. If there are some well known cases of discrepancies between the Russian and the French linguistic systems, a systematic outline of the importance of variation in the translation according to grammatical and lexical features is still to be done and may prove to be a valuable resource. This quantitative investigation will provide an empiric basis for building a typology of translation discrepancies.

Quantitative analyses may also be used for analyzing the regular correspondences between the two linguistic systems. When a variability of translation is observed for a given lexical or grammatical feature, statistical analysis will allow for discovering which contextual features are correlated with each variants. These contextual features help describing the values of each variant. Factorial analysis, and, in particular, Correspondences analysis, is well suited for this task of identifying the stable groups of contextual features across numerous instances of variants. Statistical analysis is used as a mean for providing a summary of numerous translations under scrutiny. While each single context is of no use for drawing conclusions and while it is also impossible to figure out the big picture when dealing with numerous collected contexts, statistical analysis is useful for mapping variants.

3 Equivalences Between Russian and French Verbal Forms

The entry of the database is an *equivalence*, i. e., a pair of functionally equivalent Russian and French forms. Within the information technology for creating the Russian-French parallel corpus, two types of equivalences should be distinguished:

- “translation models” — the set of French translations $\{F_1, \dots, F_n\}$ for a Russian verbal form R ;
- “translation stimuli” — the set of Russian “stimuli” $\{R_1, \dots, R_n\}$, to which a French form F is a “reaction.”

The equivalences of the first type (including their statistical characteristics) provide relevant information on the structure of the Russian language, whereas the equivalences of the second type specify characteristic features of the French language. Since the main object of the

interest is Russian, on the first stage, only the equivalences of the first type, i. e., the “translation models,” have been taken into consideration. A further expansion of the database is possible in the direction of including data from the parallel French-Russian corpus; they will answer the question: “What are the French ‘stimuli’ that provoke as ‘reaction’ such or such a Russian form?”

In the created database, the cluster of grammatical categories known in typological studies as TAM (Tense-Aspect-Mood, see [10–14]) has been explored. The following restrictions have been accepted: only finite verbal forms were included and the category of voice and other divergences due to the conceptualization modes were not considered. The set of TAM values for the languages in question is generally well established (although there is a number of controversial issues). The list was created on the basis of the existing Russian and French grammars and the researches on Russian grammar, including corpus-based ones [15–22].

The conceptual category of modality is generally considered to be composed of two parts, viz. objective and subjective modality. Objective modality is represented by the grammatical category ‘mood’ and enters the TAM cluster. The set of values of the category of subjective modality constitutes another cluster

(Mod): it was elaborated especially for the purposes of creating the database on the basis of grammatical descriptions [15, 16, 23, 24].

The notion “lexical-grammatical form” (LGF) was introduced; it is understood as a combination of grammatical features (TAM for Russian and Tense-Mood for French) or ascribed to a class of lexemes or even to an individual lexeme (cf. the term “construction” in Construction Grammar [25–27]). For an *equivalence* as a database entry, the set of Russian LGFs constitutes the Source domain, the set of French LGFs — the Target domain.

A database entry is conceived as having a three-level structure. The upper level is designed to search for information objects that is called “hyperequivalences,” the medium level — search for “polyequivalences,” and the lower one — for “monoequivalences.”

Monoequivalence is a binomial tuple (an ordered pair) of the form $\langle R_n(i); F_m(j) \rangle$, where the first position is occupied by the i th occurrence of an LGF R_n . Each occurrence $R_n(i)$ is provided with its “address” in the Russian part of the database (all occurrences of each form, in both Russian and French parts of the database, are indexed). The second position is occupied by the j th occurrence of the French LGF F_m , equivalent to $R_n(i)$

Таблица 1 Monoequivalence types and examples of Russian sentences and their French translations

Monoequivalence type	Russian sentence	French translation
$\langle \text{Fut-PF; Pr} \rangle$	Я <i>вернусь</i> через 10 минут; Что я ему <i>скажу</i> ?	Je <i>reviens</i> dans 10 minutes; Qu’est ce que je lui <i>dis</i> ?
$\langle \text{Pres-PF; Pr} \rangle$	не <i>пойму</i>	Je n’y <i>comprends</i> rien
$\langle \text{Past-PF(= “fut”); Pr} \rangle$	Ну я <i>пошел</i>	Eh bien, je <i>pars</i>
$\langle \text{Past-PF; Pr} \rangle$	Он только что <i>вышел</i>	Il <i>sort</i> à l’instant
$\langle \text{Imperat; Pr} \rangle$	А ты <i>помолчи!</i>	Toi, tu <i>te tais!</i>
$\langle \text{Fut-PF; Pr} \rangle$	Если я его <i>увиджу</i> , я ему это скажу	Si je le <i>vois</i> , je le lui <i>dis</i>
$\langle \text{Fut-PF; Pr} \rangle$	Если я его <i>увиджу</i> , я ему это <i>скажу</i>	Si je le <i>vois</i> , je le lui <i>dis</i>
$\langle \text{Pr; Imparf} \rangle$	Он понял, что она его <i>любит</i>	Il a compris qu’elle l’ <i>aimait</i>
$\langle \text{Pst-PF; PqParf} \rangle$	Он понял, что она его <i>любила</i>	Il a compris qu’elle l’ <i>avait aimait</i>
$\langle \text{Conj; Imparf} \rangle$	<i>если бы я знал</i> , я бы сказал	si <i>je savais</i> , je dirais
$\langle \text{Fut-PF; FutAnt} \rangle$	я <i>закончу</i> через час	<i>J’aurai fini</i> dans une heure
$\langle \text{Fut-PF; PasCom} \rangle$	Если в понедельник ничего от меня не <i>получишь</i> , возвращайся	Si lundi <i>tu n’as rien reçu</i> de moi, reviens
$\langle \text{Vfin, что Past-PF; Vfin + InfPas} \rangle$	сожалею, <i>что не сказал</i>	je regrette <i>de n’avoir pas dit</i>
$\langle \text{Vfin что/как; Acc.c.Inf} \rangle$	Я <i>видел, что</i> он вернулся	je l’ <i>ai vu revenir</i>
$\langle \text{Vfin + Inf; Vfin + Inf} \rangle$	Майор Ковалев <i>имел обыкновение</i> каждый день <i>прохаживаться</i> по Невскому проспекту	Le major Kovaliov <i>avait l’habitude de se promener</i> chaque jour dans l’avenue Nevski
$\langle \text{Vfin + Inf; Vfin} \rangle$	Он <i>начал цунпать</i> рукою, чтобы узнать: не спит ли он?	<i>il tâta</i> avec la main, se pinça pour se convaincre qu’il ne dormait pas
$\langle \text{Vfin + DET; Vfin + Inf} \rangle$	Послушай, голубушка, — <i>говорил</i> он <i>обыкновенно</i> , встретивши на улице бабу, продававшую манишки	Écoute un peu, ma chère, <i>avait-il coutume de dire</i> lorsqu’il rencontrait une de ces femmes qui vendent des chemises dans les rues
$\langle \text{Vfin + DET; Vfin + Inf} \rangle$	Бедный Ковалев <i>чуть не сошел с ума</i>	Le pauvre Kovaliov <i>faillit devenir fou</i>

Таблица 2 Polyequivalence type and an example of Russian sentence and two French translations

Polyequivalence type	Russian sentence	French translations
$\langle \text{Conj}; \{ \text{CondPas}, \text{VDisp}(\text{CondPas}) + \text{Inf} \} \rangle$	То есть Иван Яковлевич хотел бы и того и другого , но знал, что было совершенно невозможно требовать двух вещей разом, ибо Прасковья Осиповна очень не любила таких прихотей.	<i>First translation:</i> . . . Ivan Iakovlévitch se serait volontiers régaté de café et de pain frais [. . .] <i>Second translation:</i> . . . Ivan aurait bien voulu goûter de l'un et de l'autre [. . .]

in one of the French translations. Table 1 contains monoequivalence types, and examples of Russian sentences and their French translations (some of them are taken from comparative Russian-French grammars [28–30]).

Polyequivalence is a binomial tuple of the form $\langle \text{Rn}(i); \{ \text{Fm}(j), \text{Fk}(r), \dots \} \rangle$, which represents a combination of several monoequivalences with the identical first position ($\langle \text{Rn}(i); \text{Fm}(j) \rangle$, $\langle \text{Rn}(i); \text{Fk}(r) \rangle$, etc.), i. e., the set of French translations of the given Russian form in the given Russian sentence. So, for the i th occurrence of the Russian LGF Rn the French form Fm appears as its functional equivalent in the first translation, the form Fk appears in the second French translation, and so on. The index j for the form Fm indicates that the i th entry of the form Rn corresponds to the j th occurrence of the form Fm in the first French translation, index r denotes that to the i th occurrence of the Russian LGF, Rn corresponds to the r th occurrence of the French form Fk in the second translation, and so on; cf. a polyequivalence type and an example of Russian sentence and two French translations in Table 2.

Hyperequivalence is a binomial tuple of the form $\langle \text{Rn}; \{ F \} \rangle$, which comprises one Russian LGF and a multitude of its functionally equivalent French LGFs which enter into one or more monoequivalences with Rn : $\langle \text{Rn}(i); \text{Fm}(j) \rangle$, $\langle \text{Rn}(i); \text{Fk}(r) \rangle$, etc. So, a hyperequivalence represents an aggregation of all the possible translation equivalents, i. e., the *functionally equivalent fragments*¹ of a given Russian form that exist in the present database. It is important to note that if a new monoequivalence type appears in the database (when a new translation or a new pair of parallel texts enter into the corpus), the content of the corresponding hyperequivalence is automatically updated.

Three-level structure of the database serves to provide the possibility of a cross-identification on all three levels, each of them being sources of relevant linguistic data. The database structure is designed in a way to allow separate statistical estimates of mono- and polyequivalences.

¹The notion “functionally equivalent fragment — FEF” was introduced in [2] for resolving the tasks of computer linguistics within the domains of contrastive lexicology and bilingual lexicography.

4 Information Technologies for Database Creating

On the basis of the database architecture described above, a technology of developing the database using the parallel Russian-French texts was set up. The technology includes the following stages:

- selecting multivariant parallel texts (multiple French translations for each single Russian text);
- optical character recognition (OCR) for the French texts with not available digitalized version;
- aligning and XML (eXtensible Markup Language) marking (see section 5);
- metatextual alignment;
- morphological annotation of the texts and uploading them into the corpus (to provide an access to the corpus online via a search engine);
- creating the *Database of functionally equivalent lexicogrammatical verbal forms* using the multivariant parallel Russian-French corpus.

On the stage of selecting multivariant parallel texts the existing translations of Russian fiction into the French language are analyzed. For the multivariant corpus, only texts that exist in not less than two French versions are selected. Moreover, these translations should be created not earlier than in the middle of the XX century. The earlier French translations, and particularly those created in the XIX century, are obsolete linguistically and, which is even more important, include a great deal of mistakes and abridgements.

Many Russian literary works (by Gogol, Goncharov, Leo Tolstoy, Dostoevsky and others) exist in different modern French versions translated by skilful experts. They exhibit both different approaches to translating fiction and different selection of linguistic means including those relevant to the grammar.

While the texts are digitalized, the French translations are scanned, recognized, and checked (the OCR process). The Russian original texts are normally already available in a digital version (some of them are included in the monolingual RNC). The recent French translations have no public digital versions. To process pocket books, they are unstitched and then scanned on a workstation with simultaneous recognition by OCR software. A recognized text is compared to its printed version.

The alignment algorithm provides a separate alignment for each pair of a text and its translation. After that, the markup expert checks and corrects the results of alignment and moves, merges, or separates sentences in order to make the chunks of the translated text correspond to the respective chunks of the original text.

The same pattern is followed for each new translation, and afterwards, all the translation pairs are merged automatically into a multivariant array of three or more texts. The resulting texts are saved into an XML file. The technology of alignment and markup is described in the next section of this paper in more detail.

At the stage of the metatextual markup of the corpus, the expert adds a piece of relevant metatextual information to each aligned text. This information consists of the following fields:

- the name of the author in the original (source) language;
- the title of the text in the source language;
- the date of the author's birth;
- the title of the text in any of the translations;
- the name of the author in the target language of the translation (as specified in each translation);
- the names of all the translators in the target language;
- the dates of all the translations; and
- the indication of the source and target languages.

At the stage of the morphological markup the French and Russian texts are automatically marked with means of the morphological analyzers for these two languages. Each word is tagged by a set of grammatical features. The parallel texts with a grammatical markup are saved into an XML file and uploaded to the corpus.

To grant an access to the corpus online, the aligned texts with metatextual and morphological markup are made searchable within the multivariant Russian-French parallel corpus. The parallel texts are made available via a search interface that supports the standard search functions of the RNC by lexical and/or grammatical features and their combinations. A subcorpus can be also customized by metatextual attributes,

the texts can be searched within a subcorpus; and the query result can be sorted and exported.

As it has been already stated, the database is built on the basis of the bilingual multivariant aligned texts of the corpus. The main *raison d'être* of this database is the possibility of a bilingual grammatical search for mono- and polyequivalences and finding statistical estimates for their types composed of Russian and French LGFs listed in Table 3.

The methods used are based on a formalized description of each monoequivalence type $\langle R_n; F_m \rangle$ (see Table 1). This description consists in the ascribing to each LGF R_n and its F_m a corresponding combination of grammatical features, which appear during the morphological markup, when both Russian and French texts are automatically morphologically analyzed. The database is built in multiple iterations. During each iteration, one search query is built in terms of Russian and French grammatical features of a monoequivalence type $\langle R_n; F_m \rangle$. Each search query is implemented in bilingual multivariant aligned texts of the corpus.

The following monoequivalences from the corpus (*Скучная история / L'histoire banale* by Anton Chekhov) may be cited as an example:

$\langle \text{Pres-Ipf}; \text{Pr} \rangle$ corresponds to the grammatical tags *praes, ipf* (present imperfective) in Russian and *praes* (*présent*) in French:

Все это и многое, что еще можно было бы сказать, составляет (*praes, ipf = Pres-Ipf*) то, что называется (*praes, ipf = Pres-Ipf*) моим именем.

Tout cela, et beaucoup d'autres choses que l'on pourrait ajouter, constitue (*praes = Pr*) ce qu'on appelle (*praes = Pr*) "mon nom."

$\langle \text{Past-Pf}; \text{PasCom} \rangle$ corresponds to the tags *praet* (*past*) and *pf* (*perfective*) in the Russian text, and the auxiliary verb *avoir* or *être* in present (*praes*) plus the passive participle (*partcp, pf*) in French (the compound tense *passé composé*):

Наука, слава богу, отжила (*praet, pf = Past-Pf*) свой век, — говорит Михаил Федорович с расстановкой. — Ее песня уже спета. Да-с.

La science a fait (*avoir, praes + partcp, pf = PasCom*) son temps, Dieu merci, dit Fiodorovitch en détachant les mots. N-i-ni, c'est fini! Oui!

The retrieved fragments of the multivariant aligned texts are analyzed in order to identify and mark all monoequivalences of the type $\langle R_n; F_m \rangle$ in these fragments. After the search of all types of monoequivalences from Table 1 is completed, in the corpus, there remain unmarked the equivalences that are absent in this table. The search for those equivalences is performed using the queries including grammatical features only of the Russian LGFs listed in Table 3. On this stage, all types of monoequivalences, listed in Table 1 which contain the searched Russian LGFs are

Таблица 3 Lists of LGFs in Russian and in French selected for database creation

Identifier	Verbal form, comments
Russian	
Pres-IPF	Present
Pres-PF	Present Perfective (e. g., <i>не припомню</i>)
Past-IPF	Past Imperfective
Past-PF	Past Perfective
Fut-PF	Simple Future
Fut-IPF	Compound Future
Imperat-PF	Imperative Perfective
Imperat-IPF	Imperative Imperfective
Conj-PF	Subjunctive Perfective
Conj-IPF	Subjunctive Imperfective
Past + <i>bylo</i>	Cancelled Result/Frame Past (e. g., <i>хотел было, пошел было</i>), “Russian Pluperfect,” cf. [31]
Vfin + Inf	Embedded infinitive construction
Vfin + DET	Construction with a modal determinant
Dupl(V-IPF)	Implicit conative (e. g., дед <i>бил, бил</i> — не разбил) cf. [32]
French	
Pr	Présent
PasCom	passé compose
PasSim	passé simple
Imparf	imparfait
PqParf	plus-que-parfait
PasAnt	passé antérieur
PasIm	passé immédiat
Fut	Futur
FutAnt	futur antérieur
FutIm	futur immédiat
Imperat	impératif
SubjPres	subjonctif présent
SubjPas	subjonctif passé
SubjImparf	subjonctif imparfait
SubjPqParf	subjonctif plus-que-parfait
CondPr	conditionnel présent
CondPas	conditionnel passé
Vfin + Inf	embedded infinitive construction
Vfin + InfPas	embedded past infinitive construction
Vfin + Acc.c.Inf	Accusativus cum infinitivo (e. g., <i>je la vois venir</i>)
Vfin + Nom.c.Inf	Nominativus cum infinitivo (e. g., <i>elle paraît comprendre</i>)

excluded. If during the analysis of the found fragments of aligned texts new types of monoequivalences appear, they are added to Table 1. After completing the search, all marked monoequivalences are automatically grouped into polyequivalences according to their Russian LGFs $R_n(i)$.

After the mono- and polyequivalences' search is over, this database can be used for calculating type-frequencies. The frequency of monoequivalences of the type $\langle R_n; F_m \rangle$ are found by the formula:

$$\text{Freq}(t, n, m) = \frac{100Q(t, n, m)}{Q(t)}$$

where $\text{Freq}(t, n, m)$ is the percentage of type $\langle R_n; F_m \rangle$ at the time moment t ; $Q(t, n, m)$ is the number of monoequivalences of the type $\langle R_n; F_m \rangle$ at the moment of t ;

and $Q(t)$ is the overall number of monoequivalences of all the types in the database at the moment of t .

The frequency of polyequivalences of the type with two FEFs (functionally equivalent fragments) $\langle R_n; \{F_m, F_k\} \rangle$ which is called the second-order type is calculated according to the following formula:

$$\text{Freq}(t, n, m, k) = \frac{100Q(t, n, m, k)}{Q(t)}$$

where $\text{Freq}(t, n, m, k)$ is the percentage of second-order type $\langle R_n; \{F_m, F_k\} \rangle$ at the moment of t ; $Q(t, n, m, k)$ is the number of polyequivalences of the type $\langle R_n; \{F_m, F_k\} \rangle$ at the moment of t ; and $Q(t)$ is the overall number of polyequivalences of all the second-order types in the database at the moment of t .

The three- and n -order polyequivalences are also determined. Their frequencies are found by analogous formulae with the variable of t , and there should be $n + 1$ indices in the formula for the n th order of polyequivalences. All the type-frequencies depend on t , as the corpus and the database are constantly updated. Consequently, the frequencies of all the mono- and polyequivalence types can fluctuate accordingly.

5 Aligning and XML Marking

The technology of developing the database described above includes aligning and XML marking processes. The XML markup currently in use allows for building bi- or polylingual corpora with multiple translations of original texts (including the option of multiple translations into the same language); presently, a multilingual corpus featuring different other Slavic languages is available for search within the RNC.

The parallel texts in the RNC are aligned sentence-by-sentence. A lot of texts kindly offered for the use in the RNC by Adrian Barentsen and included into his multilingual Amsterdam Slavic Parallel Aligned Corpus (ASPAC) are already aligned paragraph-by-paragraph [33]. This segmentation has been additionally refined by the present authors semiautomatically, introducing boundaries between the sentences within these predefined paragraphs. For the majority of the texts, the alignment is made completely by the RNC team. The original text's sentence segmentation overrides the segmentation in the translation. So, each single element of parallel tagging corresponds to the original sentence, whilst its translated counterpart(s) may be either a part of a sentence starting with a space or a comma or more than one sentence.

Multiple alignment tools have been used for the RNC parallel corpora. It is evident that the procedure of alignment consists of two stages: introducing sentence boundaries into texts and the alignment in the narrow sense of the word. There exist programs that do not have an embedded sentence-splitting algorithm (HunAlign by Andras Farkas, <http://mokk.bme.hu/resources/hunalign>; LeoBilingua by Leonid Brodsky, www.hot.ee/bclogic/) and those who enable sentence-boundaries markup like TextAlign (<http://www.englishhelp.ru/soft/soft-for-translator/151-textalign.html>) or *Parallelnye Texty*, a program developed for the RNC by A. A. Kretov's team in the Voronezh University and used for markup of some English-Russian and German-Russian texts.

The algorithm of breaking the text into sentences is straightforward in programs of both types; it uses the punctuation marks like exclamation mark, quotation mark, and full stop without taking into consideration the initial letters, abbreviation, quotation, and paren-

thesis marks or the rules of direct speech (for problems in using TextAlign in an Ukrainian-Polish parallel corpus, see [34]). The segmentation in programs of both types can be corrected manually, although the algorithm itself cannot be corrected, and some general mistakes are to be treated each time they occur. In the TextAlign program, additionally, the automatic sentence-breaking is obligatory, and one cannot escape it by creating a standalone program for this purpose. For LeoBilingua and HunAlign, this is the only option, and it is possible to elaborate rules of sentence-breaking and change them as the new texts bring new challenges.

The alignment proper for all the four tools is automatic with a possible manual control. This is further divided, however, into two possible modes — step-by-step (with corrections possible in the middle of the consequent alignment) and total alignment with post-correction. The first approach is embraced by LeoBilingua and *Parallelnye teksty*, while the other is chosen by TextAlign and HunAlign. The last two programs, therefore, call for a rereading of an already aligned text with correction of the wrongly aligned sentences. While in TextAlign a GUI interface is provided for this (however a single correction calls for realigning of the whole text), in HunAlign, only a manual editing of the output file is possible.

The choice of parallel sentences may be additionally controlled for the point of view of sentence length and/or lexical contents; this feature is supported by LeoBilingua (one sentence should not be twice or more longer than the other) and HunAlign that uses a statistical mechanism evaluating the probability of a good alignment using the sentence length and, optionally, a bilingual dictionary. If the evaluation count in HunAlign is below zero, the alignment is usually mismatched; these places are to be corrected manually.

So, LeoBilingua and HunAlign seem to be the best choice for the RNC and both are used in it currently, both allowing for user-defined sentence splitting and using statistical mechanisms of alignment. Both have their advantages. While LeoBilingua allows for a slow well controlled process, with a possibility to split sentences manually in all tricky places and correct possible text misprints in a GUI (graphical user interface), sending the results directly into a Unicode XML file, HunAlign aligns the whole text quickly with very few mismatches, marked and easily discerned. The latter is currently used by Ruprecht von Waldenfels in ParaSol, a project close to the tasks of the RNC parallel corpora (earlier aka Regensburg Parallel Corpus, <http://www-korpus.uni-r.de/ParaSol/>, see [35]). However, the material of Slavic corpora offers some challenges, including the dictionary problems, as the languages with rich inflection demand including most forms of the paradigm into the dictionaries used in alignment.

In order to deal with some of the issues pointed above, another strategy is currently explored. It involves using a part-of-speech tagger for annotating the corpora in both languages before the alignment step. A part-of-speech tagger such as TreeTagger [36] is available for both Russian and French languages and provides, beside categorization, sentence-breaking as well as lemmatization. The sentence-breaking is based on linguistic rules and is more elaborated than *ad-hoc* scripts mentioned above. However, it may be used only with alignment tools allowing for external sentence-breaking. The lemmatization helps reducing the morphological diversity: each original sentence (made of inflected forms) can be transformed into a sentence made of lemma. The task of aligning sentences of lemma is easier, since a bilingual dictionary of lemma is easier to produce. Once aligned, sentences of lemma can, of course, be replaced with original, inflected sentences.

While the quality and accuracy of the alignments rely heavily on the quality and richness of the bilingual lexicon, it was found that a simple, word-by-word general lexicon was not available for Russian and French. Fortunately, the HunAlign program may produce, after an alignment task, the bilingual lexicon it built in memory while aligning the sentences. The acquisition of a bilingual lexicon was thus an incremental feedback process: the first lexicon produced by HunAlign has been corrected; in turn, it allows for a better alignment in a new run of HunAlign and, thus, new and better suggestions of word pairs were produced by HunAlign. These word pair suggestions were incorporated in the lexicon, entailing a new run of HunAlign.

6 Format and Morphological Tagging

The parallel texts in the RNC are presented in an XML format where sentences are paired by the `<para></para>` tag. Each sentence has an attribute indicating the language (this may be changed when bi- or polylingual texts are inaugurated). If a sentence is not translated, an omission is marked by three dashes.

The texts are automatically annotated using the morphological analyzers designed by Yandex search engine. The lexical and grammatical annotation is included into `<ana></ana>` tag. The tags are not currently disambiguated: however, some Russian texts selected for parallel corpora are already manually disambiguated for the monolingual corpus and may be later included into parallel corpora as well.

There should be also a possibility for tagging translation discrepancies, i. e., adding of new text or omitting of the source fragments in the translated text. This kind

of discrepancies is not a rare thing in translating fiction. They should be taken into consideration in every serious linguistic study based on parallel texts as the translation accuracy can never be taken for granted. These discrepancies are also a valid object of study in their own right, concerning the theory of translation as such.

Let quote an example of aligned sentences in the Russian-French parallel corpora in XML. Note the tags for loose translation: “add,” “omit,” and “change” mark, respectively, meaningful change of the source as rendered into the target text.

```
<para id = "292">
  <se lang = "ru"> Нужно знать, что одно значительное
  лицо недавно сделался значительным лицом, а до того
  времени он был незначительным лицом.</se>
  <se lang = "fr" loose = "omit"> Il y a lieu d'indiquer
  que ce personnage n'était devenu important que depuis
  peu;</se>
</para>
<para id = "105">
  <se lang = "ru"> Рисунок: рожа Момуса.</se>
  <se lang = "fr" loose = "add"> Un dessin représentant
  la trogne hilare du dieu Momus,</se>
</para>
<para id = "338">
  <se lang = "ru"> Он позволял это себе потому, что
  чувствовал в себе силу всегда, когда ему понадобится,
  опять выделить одно служебное и откинуть человеческое.</se>
  <se lang = "fr" loose = "change"> Il le faisait
  seulement parce qu'il se sentait de force à rétablir à
  n'importe quel moment les barrières fatidiques.</se>
</para>
```

The parallel texts are made available for search online at the www.ruscorpora.ru website. Due to copyright reasons, no text is available for full view; the search results are always presented in the form of separate sentences with minimal context (so-called “snippets”). The following parameters are searchable:

- any combination of lemmata, exact word forms, and morphological tags within a 10-word combination (e. g., “avoir” in the Imparfait tense + Past Participle yields the French Pluperfect); and
- names of the author and the translator, language of the original text, language of the translation text.

These parameters are available by selecting a sub-corpus for further textual or grammatical search. New software is under preparation, providing for semiautomatic alignment of more than two translation texts. The result should be an XML document with multiple French phrases (marked as belonging to a given translation) corresponding to each Russian phrase of the source text. By the end of 2013, the first “multivariant” (as they

are called by the team) aligned texts will be included into the Russian-French parallel corpus.

7 Concluding Remarks

A user working with a multivariant parallel corpus has as his/her typical task retrieving all the occurrences of an original LGF and its FEFs in all the translation variants present in the corpus. To solve such tasks, a technology is set up for creating a database of binomial tuples $\langle \text{LGF}; \{\text{FEF}\} \rangle$ based on the multivariant corpus. This database should offer an opportunity to search LGFs and their respective FEFs in parallel texts.

The main difference between this database and the parallel corpus itself is that the former enables the search by LGFs, FEFs, and the types of mono- or polyequivalences (see Tables 1–3). The corpus can be searched only by the grammatical features of words. The query results for the corpus consist of the aligned text fragments including the requested grammatical features. The query results for the database consist of the LGFs, FEFs, mono-, polyequivalences and their types, as well as the fragments of the parallel texts including them. Moreover, the database allows for building hyperequivalences based on the retrieved monoequivalences.

While creating the database, the expert has to select and to mark up the monoequivalences relevant for the monoequivalence type in question. This selection is personalized in the database by the expert's ID. The grammatical features used by him/her are also recorded in the database. Currently, the selection of the monoequivalences relevant for the type in question is done by one expert only. The present authors' plans provide that later multiple experts should work collaboratively, and the results of their collaboration should be processed using the methods and models of coordinating personal expert knowledge [37–40].

References

1. The use of databases in cross-linguistic studies / Eds. M. Everaert, S. Musgrave, A. Dimitriadis. — Belin—New York: Walter de Gruyter GmbH & Co., 2009.
2. *Dobrovolsky D. O., Kretov A. A., Sharoff S. A.* Corpus of parallel texts // Scientific and Technical Information. Ser. 2: Information Processes and Systems, 2005. No. 6. P. 16–27.
3. *Dobrovolsky D. O., Kretov A. A., Sharoff S. A.* Corpus of parallel texts: Architecture and usage // Russian National Corpus: 2003–2005. — Moscow: Indrik, 2005. P. 263–296. [In Russian.]
4. *Andreeva E. G., Kasevich V. B.* Grammar and lexicon in the English-Russian corpus of parallel texts // Russian National Corpus: 2003–2005. — Moscow: Indrik, 2005. P. 297–307. [In Russian.]
5. *Dobrovolsky D. O.* A Corpus of parallel texts and studying culture-specific lexicon // Russian National Corpus: 2006–2008. New results and prospects. — St. Petersburg: Nestor-Istoriya, 2009. P. 383–401.
6. *Kozerenko E. B.* Cognitive approach to language structure segmentation for machine translation algorithms // Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA: CSREA Press, 2003. P. 49–55.
7. *Kozerenko E. B.* Linguistic filters in statistical machine translation models // Informatics and Applications, 2010. Vol. 4. No. 2. P. 83–92.
8. *Kozerenko E. B.* Syntactic transformations modelling for hybrid machine translation // ICAI'11, WORLD-COMP'11 Proceedings. — Las Vegas, Nevada, USA: CRSEA Press, 2011. P. 875–881.
9. *Kozerenko E. B.* Parallel texts alignment strategies: The semantic aspects // Informatics and Applications, 2013. Vol. 7. No. 1. P. 82–89.
10. *Comrie B.* Aspect. An introduction to the study of verbal aspect and related problems. — Cambridge: Cambridge Univ. Press, 1976.
11. *Comrie B.* Tense. — Cambridge: Cambridge Univ. Press, 1985.
12. *Dahl Ö.* Tense and aspect systems. — Oxford: Blackwell, 1985.
13. *Bybee J. L., Perkins R., Pagliuca W.* The evolution of grammar: Tense, aspect and modality in the languages of the world. — Chicago: University of Chicago Press, 1994.
14. Tense and aspect in the languages of Europe / Ed. Ö. Dahl. — Belin—New York: Mouton de Gruyter, 2000.
15. Grammar of contemporary Russian. — In 2 vols. — Moscow, 1954. [In Russian.]
16. Russian grammar / Ed. N. Ju. Shvedova. — In 2 vols. — Moscow, 1980. [In Russian.]
17. *Gak V. G.* Theoretical grammar of French. — Moscow, 2000. [In Russian.]
18. *Vinogradov V. V.* Russian language. — Moscow, 2001. [In Russian.]
19. Russian National Corpus: 2003–2005. — Moscow: Indrik, 2005.
20. Russian National Corpus: 2006–2008. — St. Petersburg: Nestor-Istoriya, 2009. [In Russian.]
21. Corpus studies on Russian grammar. — Moscow: Probel-2000, 2009. [In Russian.]
22. Construction linguistics / Ed. E. V. Rakhilina. — Moscow: Azbukovnik, 2010. [In Russian.]
23. *Bondarko A. V., Beliaeva E. I., Biriulin L. A., et al.* The theory of functional grammar. Temporality. Modality. — Leningrad: Nauka, 1990. [In Russian.]
24. *Paducheva E. V.* Modality. <http://rusgram.ru>.
25. *Goldberg A.* Constructions: A construction grammar approach to argument structure. — Chicago: Univ. of Chicago Press, 1995.

26. *Tomasello M.* Constructing a language. A usage-based theory of language acquisition. — Cambridge—London: Harvard Univ. Press, 2003.
27. *Goldberg A.* Constructions at work. The nature of generalization in grammar. — Oxford: Oxford Univ. Press, 2006.
28. *Gak V.G.* Russian language compared to French. — Moscow: URSS, 2006. [In Russian.]
29. *Kouznetsova I. N.* Grammaire contrastive du français et du russe. — M.: Nestor Academic Publs., 2009.
30. *Gak V. G.* Comparative typology of French and Russian. — Moscow: URSS, 2010. [In Russian.]
31. *Sitchinava D. V.* Seek to prevent at the root: Russian construction with *bylo*: corpus-based analysis // Corpus studies on Russian grammar. — Moscow: Probel-2000, 2009. P. 362–396.
32. *Plungian V. A., Rakhilina E. V.* Tushat-tushat — ne potushat: Grammar of one verb construction // Construction linguistics / Ed. E. V. Rakhilina. — Moscow: Azbukovnik, 2010. P. 83–94.
33. Information on ASPAC — Amsterdam Slavic Parallel Aligned Corpus. <http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/b/a/a.a.barentsen/a.a.barentsen.html>.
34. *Kotsyba N.* The current state of work on the Polish-Ukrainian Parallel Corpus (PolUKR) // Problems of Slavic Lexicography: Workshop (International) within MONDILEX Project Proceedings. — Kyiv, 2009. http://www.domeczek.pl/~natko/papers/NKotsyba_Kyiv2009.pdf.
35. *Von Waldenfels R.* Compiling a parallel corpus of slavic languages. Text strategies, tools and the question of lemmatization in alignment // Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV) / Eds. B. Brehmer, V. Zdanova, R. Zimny. — München, 2006. P. 123–138 (available at: <http://www-nw.uni-regensburg.de/%7E.war05297.slavistik.sprachlit.uni-regensburg.de/pub/WaldenfelsParallelCorpora2006.pdf>).
36. *Schmid H.* Probabilistic part-of-speech tagging using decision trees // Conference (International) on New Methods in Language Processing Proceedings / Ed. D. Jones. — Manchester, UK: UMIST, 1994. P. 44–49.
37. *Zatsman I.* Time-dependent semiotic model of computer coding of concepts, information objects and denotata // Informatics and Applications, 2009. Vol. 3. No. 4. P. 87–101.
38. *Zatsman I., Durnovo A.* Modelling of processes for creation of expert knowledge for monitoring of goal-oriented programme activities // Informatics and Its Applications, 2011. Vol. 5. No. 4. P. 84–98.
39. *Zatsman I.* Denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation // 11th IEEE Conference (International) on Cognitive Informatics and Cognitive Computing Proceedings. — Los Alamitos, CA: IEEE Computer Society Press, 2012. P. 27–34.
40. *Zatsman I.* Tracing emerging meanings by computer: Semiotic framework // 13th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing Intern. Ltd., 2012. Vol. 2. P. 1298–1307.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ СОЗДАНИЯ БАЗ ДАННЫХ ЭКВИВАЛЕНТНЫХ ГЛАГОЛЬНЫХ ФОРМ В РУССКО-ФРАНЦУЗСКОМ ПОЛИВАРИАНТНОМ ПАРАЛЛЕЛЬНОМ КОРПУСЕ

С. Луазо¹, Д. В. Сичинава², А. А. Зализняк³, И. М. Зацман⁴

¹Университет Париж-13, лаборатория лексики, словарей и информатики НЦНИ, sylvain.loiseau@univ-paris13.fr

²Институт русского языка РАН, mitrius@gmail.com

³Институт ИЯз РАН; ИПИ РАН, anna.zalizniak@gmail.com

⁴ИПИ РАН, iz_ipi@al70.ipi.ac.ru

Аннотация: Русско-французский параллельный корпус как часть Национального корпуса русского языка в настоящее время преобразовывается в поливариантный корпус с одним оригинальным и несколькими вариантами переводов этого текста. Одновременно создается база данных функционально эквивалентных лексико-грамматических глагольных форм на основе текстов поливариантного корпуса. Главная цель создания базы данных состоит в том, чтобы вычислить статистические оценки соответствий между русскими и французскими глагольными формами. В статье рассматриваются вопросы разработки информационной технологии для одновременного создания русско-французского поливариантного параллельного корпуса и этой базы данных.

Ключевые слова: параллельные поливариантные корпуса; Национальный корпус русского языка; информационные технологии; XML разметка русско-французских параллельных текстов; лексико-грамматические формы; функциональные соответствия; статистические оценки соответствий

PARAMETRICAL STATISTICAL AND ANALYTICAL MODELING OF DISTRIBUTIONS IN NONLINEAR STOCHASTIC SYSTEMS ON MANIFOLDS

I. N. Sinitsyn

IPI RAN, sinitsin@dol.ru

Discrete parametrical statistical and analytical modeling methods in nonlinear Ito stochastic systems on manifolds with Wiener and Poisson noises have been developed. For one- and multidimensional densities parametrization, the coefficients in orthogonal expansions of different orders are taken. Special attention is paid to nonlinear correlational theory of statistical and analytical modeling.

Keywords: analytical modeling; method of normal approximation; nonlinear correlational theory; nonlinear Ito stochastic system on manifold; orthogonal expansions method; parametrization of one- and multidimensional densities; statistical linearization method; statistical modeling

EVALUATION METHODS FOR EFFICIENCY AND DIRECTIVE TERMS OF PERFORMANCE OF RESOURCE-INTENSIVE COMPUTING TASKS

I. K. Kupalov-Yaropolk¹, Yu. E. Malashenko², I. A. Nazarova³, and A. F. Ronzhin⁴

¹Lebedev Institute of Precision Mechanics and Computer Engineering, Russian Academy of Sciences, kupyar@rambler.ru

²Dorodnicyn Computing Center, Russian Academy of Sciences, malash09@ccas.ru

³Dorodnicyn Computing Center, Russian Academy of Sciences, irina-nazar@yandex.ru

⁴Dorodnicyn Computing Center, Russian Academy of Sciences, raf-zao-zt@yandex.ru

The problem of effective use of the heterogeneous computing system is considered at parallel processing of diverse tasks. In a case of date violation of works completion, the processor spent time belongs to production losses. Planning and optimization controls are exercised on the basis of the guaranteed estimates constructed for the worst case.

Keywords: parallel computing; multiprocessor systems; optimization; the principle of guaranteed result

ON ESTIMATION OF THE EFFECTIVE BANDWIDTHS IN A SYSTEM WITH REGENERATIVE INPUT

A. V. Borodina¹ and E. V. Morozov²

¹Institute of Applied Mathematical Research, Karelian Research Center, Russian Academy of Sciences, borodina@krc.karelia.ru

²Institute of Applied Mathematical Research, Karelian Research Center, Russian Academy of Sciences, emorozov@karelia.ru

The effective bandwidths (EB) of a communication system are considered. The EB guarantees that the stationary workload overflow/loss probability to exceed a threshold is limited by a (small) quantity. It is shown how to calculate EB in a fluid queue fed by an input with the independent increments. Then, a fluid system with regenerative input is considered. Using heuristic arguments, an approximation of the limiting logarithmic exponential moment generating function of the input was deduced. Numerical simulations show a satisfactory accuracy of the approximation applied to a few systems with regenerative input.

Keywords: effective bandwidths; fluid queue; workload process; overflow/loss probability; effective bandwidths approximation; regenerative estimation

STATIONARY WAITING TIME DISTRIBUTION IN QUEUEING SYSTEM WITH NEGATIVE CUSTOMERS AND BUNKER FOR OUSTED CUSTOMERS UNDER FIRST–FIFO–FIFO SERVICE DISCIPLINE

R. V. Razumchik

IPI RAN, rrazumchik@ieee.org

Consideration is given to the single server queueing system with Poisson input flows of ordinary and negative customers. An arriving ordinary customer occupies one place in an infinite buffer. Negative customer upon arrival pushes out one ordinary customer from the queue in the buffer to another queue (bunker) and leaves the system. Customers from bunker are served with relative priority. Service times of customers from buffer and bunker are both exponentially distributed but with different rates. It is assumed that negative customer always pushes out the first customer in the queue in the buffer and after service completion the first customer in the queue in the buffer enters server or, if buffer is empty, the first customer in the queue in the bunker. Stationary waiting time distribution of an arriving ordinary customer in terms of Laplace–Stieltjes transform is obtained.

Keywords: queueing system; negative customers; waiting time

ON THE RATE OF CONVERGENCE TO THE NORMAL LAW OF RISK ESTIMATE FOR WAVELET COEFFICIENTS THRESHOLDING WHEN USING ROBUST VARIANCE ESTIMATES

O. V. Shestakov

Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, oshestakov@cs.msu.su

The asymptotic properties of risk estimate for thresholding wavelet coefficients of signal function are analyzed. Some estimates for rate of convergence to the normal law are obtained.

Keywords: wavelets; thresholding; risk estimate; normal distribution; rate of convergence

STATISTICAL TESTING FOR THE NONEXECUTABILITY OF FRAGMENTS OF THE CODE OF A LINEAR PROGRAM

V. Yu. Korolev¹, R. L. Smelyansky², T. R. Smelyansky³, and A. V. Shalimov⁴

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, victoryukorolev@yandex.ru

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, smel@sc.msu.ru

³Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, smelyanskiy.t@bk.ru

⁴Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, ashalimov@lvk.cs.msu.su

A problem of statistical testing for the nonexecutiveness of fragments of the code of a linear program is considered. Methods based on the minimization of prior error probabilities are considered as well as those based on the minimization of posterior error probabilities.

Keywords: testing statistical hypotheses; geometric distribution; probability of the error of the first kind; probability of the error of the second kind; Neyman–Pearson lemma; posterior error probability

BAYESIAN RECURRENT MODEL OF RELIABILITY GROWTH: UNIFORM DISTRIBUTION OF PARAMETERS

A. A. Kudriavtsev¹, I. A. Sokolov², and S. Ya. Shorgin³

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, nubigena@hotmail.com

²IPI RAN, isokolov@ipiran.ru

³IPI RAN, sshorgin@ipiran.ru

The paper is devoted to justification of expediency of Bayesian approach at the solution of the tasks connected with determination of reliability of complex modifiable systems. As illustration, average value of reliability of system is presented in a case where characteristics of “defectiveness” and “efficiency” of the means correcting imperfections of system are uniformly distributed.

Keywords: modifiable information systems; reliability theory; Bayesian approach

THE ERROR-IN-VARIABLES MODEL IDENTIFICATION ON THE BASIS OF DEMING'S APPROACH

V. S. Timofeev¹, V. Yu. Schekoldin², and A. Yu. Timofeeva³

¹Novosibirsk State Technical University, netsc@rambler.ru

²Novosibirsk State Technical University, raix@ngs.ru

³Novosibirsk State Technical University, supernasty@mail.ru

Some approaches to regression models constructing with stochasticity for both endogenous and exogenous variables are considered. The original geometric interpretation of particular cases of Deming regression for parameter estimation's functional is suggested. The proposition of mutual arrangement for straight, inverse, diagonal, and orthogonal regressions is proved. The bias and standard deviation for regression parameter estimators in dependence of weight's coefficients ratio in Deming model has been obtained.

Keywords: least square estimation; Deming regression; geometric interpretation; dispersion ellipse

ASYMPTOTIC NORMALITY OF THE ESTIMATION OF THE MULTIVARIATE LOGISTIC REGRESSION

A. Yu. Khaplanov

M. V. Lomonosov Moscow State University, Khaplanova@gmail.com

Estimation of characteristics of the multivariate logistic regression with a diverging number of covariates has been made. The convergence rate for estimate of characteristics of the multivariate logistic regression coefficients has been obtained. Asymptotic normality of its rejection has been proved.

Keywords: logistic regression; convergence rate; asymptotic normality; high-dimensional covariates

ASYMPTOTIC EXPANSIONS FOR THE DISTRIBUTION FUNCTIONS OF STATISTICS CONSTRUCTED FROM SAMPLES WITH RANDOM SIZES

V. E. Bening¹, N. K. Galieva², and V. Yu. Korolev³

¹Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, bening@yandex.ru

²Kazakhstan Branch of the M. V. Lomonosov Moscow State University, nurgul_u@mail.ru

³Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, victoryukorolev@yandex.ru

A general transfer theorem is proved making it possible to construct asymptotic expansions for the distribution function of a statistic constructed from the sample with a random size from the asymptotic expansion for the distribution function of the random sample size and the asymptotic expansion for the distribution function of the same statistic constructed from the samples with a nonrandom size.

Keywords: sample with a random size; asymptotic expansion; transfer theorem; mixture of probability distributions; Laplace distribution; Student distribution

ON CONVERGENCE OF RANDOM WALKS GENERATED BY COMPOUND COX PROCESSES TO LEVY PROCESSES

V. Yu. Korolev¹, L. M. Zaks², and A. I. Zeifman³

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, victoryukorolev@yandex.ru

²Department of Modeling and Mathematical Statistics, Alpha-Bank, lily.zaks@gmail.com

³Vologda State Pedagogical University; IPI RAN, a_zeifman@mail.ru

A functional limit theorem is proved establishing weak convergence of random walks generated by compound doubly stochastic Poisson processes to Lévy processes in the Skorokhod space. As corollaries, theorems on convergence of random walks with jumps having finite variances to Lévy processes with mixed normal distributions, in particular, to stable Lévy processes have been proved.

Keywords: stable distribution; Lévy process; stable Lévy process; compound doubly stochastic Poisson process (compound Cox process); Skorokhod space; transfer theorem

STATISTICAL MECHANISMS OF THE SUBJECT DOMAINS ASSOCIATIVE PORTRAITS FORMATION ON THE BASIS OF BIG NATURAL LANGUAGE TEXTS FOR THE SYSTEMS OF KNOWLEDGE EXTRACTION

M. M. Charnine¹, N. V. Somin², I. P. Kuznetsov³, Yu. I. Morozova⁴, I. V. Galina⁵, and E. B. Kozerenko⁶

¹IPI RAN, 1@keywen.com

²IPI RAN, somin@post.ru

³IPI RAN, igor-kuz@mtu-net.ru

⁴IPI RAN, judez@yandex.ru

⁵IPI RAN, irn_gl@mail.ru

⁶IPI RAN, kozerenko@mail.ru

Associative relations between terms, concepts and other elements of natural language play an important role in decision of a wide variety of application tasks including intelligent texts processing, knowledge extraction, and management comprising the formation of knowledge bases and semantic information retrieval. The paper presents the methods of automatic establishment of the associative relations between terms and concepts in the texts from Internet and creation of subject domains associative portraits designed for the tasks of intelligent systems development. An associative portrait of a subject domain (APSD) is a dictionary of the meaningful support terms and word combinations interconnected by associative relations. It is essential that the APSD are constructed automatically on the basis of statistical analysis of big volumes of texts. The theoretical impact of the proposed method consists in the use of statistics, corpus linguistics, and distributional semantics for processing big volumes of natural language texts which are dynamically updated and enriched in the Internet for constructing the model of a subject domain in the form of APSD.

Keywords: automatic processing of text corpora; statistical methods; intelligent Internet technologies; lexical semantic analysis; knowledge extraction from texts; semantic retrieval; semantic vectors; semantic context space

INFORMATION TECHNOLOGIES FOR CREATING THE DATABASE OF EQUIVALENT VERBAL FORMS IN THE RUSSIAN-FRENCH MULTIVARIANT PARALLEL CORPUS

S. Loiseau¹, D. V. Sitchinava², A. A. Zalizniak³, and I. M. Zatsman⁴

¹Université Paris 13, Sorbonne Paris Cité, Laboratoire LDI (Lexiques, dictionnaires, informatique), CNRS, UMR 7187, sylvain.loiseau@univ-paris13.fr

²Institute of the Russian Language of the Russian Academy of Sciences, mitrius@gmail.com

³Institute of Linguistics of the Russian Academy of Sciences; IPI RAN, anna.zalizniak@gmail.com

⁴IPI RAN, iz_ipi@i170.ipi.ac.ru

The Russian-French parallel corpus as a part of the Russian National Corpus is being transformed into a multivariant corpus with several translations corresponding to each original texts. Concurrently, a database of functionally equivalent lexicogrammatical verbal forms is being created using the multivariant corpus. The main purpose of database creation is to calculate the statistical estimates of the equivalences between Russian and French verbal forms. The paper discusses an information technology for creating the Russian-French multivariant parallel corpus and the database simultaneously.

Keywords: parallel multivariant corpora; Russian National Corpus; information technologies; XML marking up Russian-French parallel texts; lexicogrammatical form; functional equivalence; statistical estimates of equivalences

Бенинг Владимир Евгеньевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Бородина Александра Валентиновна (р. 1980) — младший научный сотрудник Института прикладных математических исследований КарНЦ РАН

Галиева Нургуль Кадыржановна (р. 1985) — преподаватель Казахстанского филиала Московского государственного университета им. М. В. Ломоносова

Галина Ирина Владимировна (р. 1967) — старший научный сотрудник ИПИ РАН

Закс Лилия Михайловна (р. 1989) — главный специалист отдела моделирования и математической статистики Альфа-банка

Зализняк Анна Андреевна (р. 1959) — доктор филологических наук, ведущий научный сотрудник ИЯЗ РАН и ИПИ РАН

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом ИПИ РАН

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, декан факультета прикладной математики и компьютерных технологий Вологодского государственного педагогического университета; старший научный сотрудник ИПИ РАН, ведущий научный сотрудник ВНКЦ ЦЭМИ РАН

Козеренко Елена Борисовна (р. 1959) — кандидат филологических наук, заведующая лабораторией ИПИ РАН

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; ведущий научный сотрудник ИПИ РАН

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, старший преподаватель кафедры математической статистики факультета вычислительной математики и кибернетики

Московского государственного университета им. М. В. Ломоносова

Кузнецов Игорь Петрович (р. 1938) — доктор технических наук, главный научный сотрудник ИПИ РАН

Купалов-Ярополк Игорь Константинович (р. 1977) — заместитель начальника отдела Института точной механики и вычислительной техники им. С. А. Лебедева РАН

Луазо Сильвен (р. 1977) — доктор философии (PhD), доцент, Университет Париж-13, лаборатория лексики, словарей и информатики НЦНИ

Малашенко Юрий Евгеньевич (р. 1946) — доктор физико-математических наук, заведующий сектором Вычислительного центра им. А. А. Дородницына РАН

Морозов Евсей Викторович (р. 1947) — доктор физико-математических наук, профессор, ведущий научный сотрудник Института прикладных математических исследований КарНЦ РАН; профессор Петрозаводского государственного университета

Морозова Юлия Игоревна (р. 1984) — научный сотрудник ИПИ РАН

Назарова Ирина Александровна (р. 1966) — кандидат физико-математических наук, научный сотрудник Вычислительного Центра им. А. А. Дородницына РАН

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН

Ронжин Александр Федорович (р. 1954) — доктор физико-математических наук, профессор, главный научный сотрудник Вычислительного Центра им. А. А. Дородницына РАН

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом ИПИ РАН

Сичинава Дмитрий Владимирович (р. 1980) — кандидат филологических наук, старший научный сотрудник ИРЯ РАН

Смелянский Руслан Леонидович (р. 1950) — доктор физико-математических наук, член-корреспондент

РАН, профессор кафедры автоматизации систем вычислительных комплексов, заведующий лабораторией факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Смелянский Тимофей Русланович (р. 1990) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Соколов Игорь Анатольевич (р. 1954) — академик (действительный член) Российской академии наук; доктор технических наук, директор ИПИ РАН

Сомин Николай Владимирович (р. 1947) — кандидат физико-математических наук, ведущий научный сотрудник ИПИ РАН

Тимофеев Владимир Семенович (р. 1972) — доктор технических наук, профессор Новосибирского государственного технического университета

Тимофеева Анастасия Юрьевна (р. 1984) — кандидат экономических наук, научный сотрудник Новосибирского государственного технического университета

Хапланов Арсений Юрьевич (р. 1988) — аспирант кафедры теории вероятностей механико-математического факультета Московского государственного университета им. М. В. Ломоносова

Шалимов Александр Владиславович (р. 1985) — программист лаборатории вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Шарнин Михаил Михайлович (р. 1959) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора ИПИ РАН

Щеколдин Владислав Юрьевич (р. 1976) — кандидат технических наук, доцент Новосибирского государственного технического университета

ABOUT AUTHORS

Bening Vladimir E. (b. 1954) — Doctor of Science in physics and mathematics; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Borodina Alexandra V. (b. 1980) — junior scientist, Institute of Applied Mathematical Research, Karelian Research Center, Russian Academy of Sciences

Charnine Mikhail M. (b. 1959) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Galieva Nurgul K. (b. 1985) — lecturer, Kazakhstan Branch of M. V. Lomonosov Moscow State University

Galina Irina V. (b. 1967) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Khaplanov Arseniy Yu. (b. 1988) — PhD student, Department of Probability Theory, Faculty of Mechanics and Mathematics, M. V. Lomonosov Moscow State University

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Kozerenko Elena B. (b. 1959) — Candidate of Science (PhD) in linguistics, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

Kudriavtsev Alexei A. (b. 1978) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Kupalov-Yaropolk Igor K. (b. 1977) — Deputy Head of Department, Lebedev Institute of Precision Mechanics and Computer Engineering, Russian Academy of Sciences

Kuznetsov Igor P. (b. 1938) — Doctor of Science in technology, professor, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences

Loiseau Sylvain (b. 1977) — PhD, maitre de conférences, Université Paris-13, Laboratoire LDI — Lexiques, Dictionnaires, Informatique, CNRS, UMR 7187

Malashenko Yuri E. (b. 1946) — Doctor of Science in physics and mathematics, Head of Laboratory, Dorodnitsyn Computing Center, Russian Academy of Sciences

Morozov Evsey V. (b. 1947) — Doctor of Science in physics and mathematics, professor, leading scientist, Institute of Applied Mathematical Research, Karelian Research Center, Russian Academy of Sciences; professor, Petrozavodsk State University

Morozova Yulia I. (b. 1984) — scientist, Institute of Informatics Problems, Russian Academy of Sciences

Nazarova Irina A. (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Dorodnitsyn Computing Center, Russian Academy of Sciences

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Ronzhin Alexander F. (b. 1954) — Doctor of Science in physics and mathematics, professor, principal scientist, Dorodnitsyn Computing Center, Russian Academy of Sciences

Schekoldin Vladislav Yu. (b. 1976) — Candidate of Science (PhD) in technology, associate professor, Novosibirsk State Technical University

Shalimov Alexander V. (b. 1985) — programmer, Computer Systems Laboratory, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Shestakov Oleg V. (b. 1976) — Doctor of Science in physics and mathematics, assistant professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences

Sinityn Igor N. (b. 1940) — Doctor of Science in technology, professor, honored scientist of RF, Head of

Department, Institute of Informatics Problems, Russian Academy of Sciences

Sitchinava Dmitri V. (b. 1980) — Candidate of Science (PhD) in philology, senior scientist, Institute of the Russian Language, Russian Academy of Sciences

Smelyansky Ruslan L. (b. 1950) — Doctor of Science in physics and mathematics, Corresponding Member of the Russian Academy of Sciences, professor, Department of Computing Systems and Automation, Head of Laboratory, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Smelyansky Timofey R. (b. 1990) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences

Somin Nicolay V. (b. 1947) — Candidate of Science (PhD) in physics and mathematics, leading scientist,

Institute of Informatics Problems, Russian Academy of Sciences

Timofeev Vladimir S. (b. 1972) — Doctor of Science in technology, professor, Novosibirsk State Technical University

Timofeeva Anastasia Yu. (b. 1984) — Candidate of Science (PhD) in economics, scientist, Novosibirsk State Technical University

Zaks Lily M. (b. 1989) — principal officer, Department of Modeling and Mathematical Statistics, Alpha-Bank

Zalizniak Anna A. (b. 1959) — Doctor of Science in philology, leading scientist, Institute of Linguistics and Institute of Informatics Problems, Russian Academy of Sciences

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

Zeifman Alexander I. (b. 1954) — Doctor of Science in physics and mathematics, professor, Dean of the Faculty of Applied Mathematics and Computer Technologies, Vologda State Pedagogical University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; leading scientist, VSCC CEMI RAS

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 20 страниц указанного формата.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления, представлены в Интернете:

<http://www.ipiran.ru/journal/template.doc>.

8. Статья должна содержать следующую информацию на **русском и английском языках:**

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием города и страны и электронного адреса каждого автора;

- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных; Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком.
- ключевые слова, желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.

9. Литература. По включенным в список литературы работам на русском языке информация в списке представляется как в кириллице, так и с использованием латинской транслитерации, а по работам, написанным латиницей, — на языке оригинала.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в списке литературы в порядке упоминания.

В списке литературы не должно быть позиций, на которые нет ссылки в тексте статьи.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции: Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)

Requirements for manuscripts submitted to Journal “Informatics and Applications”

1. The Journal publishes original articles which have not been published before and are not intended for publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

«We, the undersigned authors of the manuscript “. . . ,” pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date)».

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article does not have secret data prohibited for publication.

2. A submitted article should be attached with **the data on the author(s)** (see p. 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If authors receive their article for correction after reviewing it does not mean that the article is approved to be published. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats and, additionally, in the .pdf format. All documents may be sent by e-mail or on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.
Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.
Text: font — Times New Roman, font size — 14, paragraph — 0.5, line spacing — 1.5, justified alignment.
The recommended manuscript size: no more than 20 pages of the specified format.
Word abbreviations are not allowed except the standard ones.
Abbreviations should be minimal. All pages of the manuscript should be numbered.
The templates for the manuscript typesetting are presented on site:
<http://www.ipiran.ru/journal/template.doc>.
8. Articles should enclose data both in **Russian and English**:
 - title;
 - author(s) name(s) and surname(s);
 - affiliation — organization, its address with ZIP code, city, country, and e-mail address;
 - data on authors according to the format (see site):
http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
 - abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately from the article. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required;

- indexing is performed on the basis of key words. The use of key words from the internationally accepted thematic Thesauri is recommended.

Important! Key words must not be sentences.

9. References. Russian references have to be presented both in Cyrillic and Latin transliteration.

References in Latin transcript are presented in original language.

References in the text are numbered according to the order of the appearance and the number is placed in square bracket. References absent in the text should not be included into the list of references.

10. Manuscripts and additional materials are not returned to authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author(s) name(s) in the “Subject” field;
- an article and additional materials have to be attached using the “attach” function;
- an electronic version of the article should contain the file with the text and (a) separate file(s) with figures.

12. “Informatics and Applications” Journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address: 119333, IPIRAN, Vavilova St., 44, block 2, Moscow, Russia

Ph.: +7(499) 135 8692, Fax: +7 (495) 930 4505

e-mail: rust@ipiran.ru (To Prof. Rustem Seyfoul-Mulyukov)