

Информатика и её применения

Том 7 Выпуск 1 Год 2013

СОДЕРЖАНИЕ

Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с разрывными характеристиками И. Н. Сеницын	3
Вероятностно-статистическое моделирование информационных потоков в сложных финансовых системах на основе высокочастотных данных В. Ю. Королев, А. В. Черток, А. Ю. Корчагин, А. К. Горшенин	12
Стационарные характеристики системы обслуживания с инверсионным порядком обслуживания, вероятностным приоритетом и гистерезисной политикой Т. А. Милованова, А. В. Печинкин	22
О сходимости в пространстве L_p максимума процесса нагрузки для одного класса гауссовских систем обслуживания О. В. Лукашенко, Е. В. Морозов	36
Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных Г. И. Рудой, В. В. Стрижов	44
Статистические методы определения запретов вероятностных мер на дискретных пространствах А. А. Грушо, Н. А. Грушо, Е. Е. Тимонина	54
Операции над представлениями кусочно-квазиаффинных функций в виде деревьев С. А. Гуда	58
Методологические основы создания информационных систем для вычисления индикаторов тематических взаимосвязей науки и технологий В. А. Минин, И. М. Зацман, М. Г. Кружков, Т. П. Норемян	70
Стратегии выравнивания параллельных текстов: семантические аспекты Е. Б. Козеренко	82
Построение семантических векторных пространств различных предметных областей Ю. И. Морозова	90
Оценка семантической адекватности текстов информационным методом Л. А. Кузнецов, В. Ф. Кузнецова	94

Информатика и её применения

Том 7 Выпуск 1 Год 2013

СОДЕРЖАНИЕ

Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм Л. М. Закс, В. Ю. Королев	105
Об оценках функций концентрации регулярных статистик, построенных по выборкам случайного объема В. Е. Бенинг, Н. К. Галиева, В. Ю. Королев	116
Об абсолютных константах в неравенстве Берри–Эссеена и его структурных и неравномерных уточнениях И. Г. Шевцова	124
Abstracts	126
Об авторах	131
About Authors	133

Технический редактор *Л. Кокушкина*

Художественный редактор *М. Седакова*

Сдано в набор 11.01.13. Подписано в печать 04.02.13. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 17. Уч.-изд. л. 12. Тираж 100 экз.
Заказ № 2696

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

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

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ С ИНВАРИАНТНОЙ МЕРОЙ В СТОХАСТИЧЕСКИХ СИСТЕМАХ С РАЗРЫВНЫМИ ХАРАКТЕРИСТИКАМИ*

И. Н. Сеницын¹

Аннотация: На базе методов нормальной аппроксимации и статистической линеаризации разработаны точные и приближенные алгоритмы аналитического моделирования плотностей стохастических режимов с инвариантной мерой в гауссовых и негауссовых стохастических системах (СтС) с разрывными характеристиками. Рассмотрены особенности моделирования в СтС с пуассоновскими шумами. На тестовых примерах показана достаточная для многих приложений точность алгоритмов.

Ключевые слова: автокоррелированная помеха; аналитическое моделирование; интегродифференциальные уравнения Пугачёва; метод нормальной аппроксимации; метод статистической линеаризации; нелинейная гауссовская и негауссовская стохастическая система в смысле Ито; пуассоновская стохастическая система; распределение с инвариантной мерой; стохастический режим

1 Введение

Следуя [1, 2], будем рассматривать нестационарный стохастический режим $Z = Z(t)$ в нелинейной дифференциальной СтС, понимаемой в смысле Ито:

$$\dot{Z} = a(Z, t) + b(Z, t)V, \quad Z(t_0) = Z_0. \quad (1)$$

Здесь Z — k -мерный вектор состояния СтС, $Z \in \Delta$ (Δ — многообразии состояний); $a = a(Z, t)$ и $b = b(Z, t)$ — детерминированные $(k \times 1)$ - и $(k \times m)$ -мерные функции отмеченных аргументов; $V = V(t)$ — m -мерный вектор негауссовских (в общем случае) белых шумов с нулевыми математическими ожиданиями и представляющий собой среднеквадратичную (с.к.) производную процесса с независимыми приращениями $W = W(t)$, $V = \dot{W}$. Обозначим через $\chi = \chi(\mu; t)$ логарифмическую производную одномерной характеристической функции $h_1 = h_1(\mu; t)$ процесса $W = W(t)$, определяемую формулой

$$\chi(\mu; t) = \frac{\partial \ln h_1(\mu; t)}{\partial t} = \frac{1}{h_1(\mu; t)} \frac{\partial h_1(\mu; t)}{\partial t}. \quad (2)$$

Начальное состояние Z_0 будем считать случайной величиной (СВ), не зависящей от $W(t)$ для $t > t_0$. Предположим, что стохастический режим $Z(t)$ является сильным решением (1), а функции a, b и χ удовлетворяют известным условиям существования и единственности [1, 2].

Пусть существуют одно- и n -мерные плотности $f_1 = f_1(z; t)$ и $f_n = f_n(z_1, \dots, z_n; t_1, \dots, t_n)$ и характеристические функции $g_1 = g_1(\lambda; t)$ и $g_n = g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)$ ($n \geq 2$), удовлетворяющие интегродифференциальным уравнениям Пугачева [1, 2]:

$$\frac{\partial f_1(z; t)}{\partial t} + \frac{\partial^T}{\partial z} [a(z, t)f_1(z; t)] = \frac{1}{(2\pi)^k} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(b(\xi, t)^T \lambda; t) e^{i\lambda^T(\xi - z)} f_1(z; t) d\xi d\lambda; \quad (3)$$

$$f_1(z; t_0) = f_0(z); \quad (4)$$

$$\begin{aligned} & \frac{\partial f_n(z_1, \dots, z_n; t_1, \dots, t_n)}{\partial t_n} + \\ & + \frac{\partial^T}{\partial z_n} [a(z_n, t_n)f_n(z_1, \dots, z_n; t_1, \dots, t_n)] = \\ & = \frac{1}{(2\pi)^{kn}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(b(\xi_n, t_n)^T \lambda_n; t_n) \times \\ & \times \exp \left\{ i \sum_{l=1}^n \lambda_l^T (\xi_l - z_l) \right\} \times \\ & \times f_n(\xi_1, \dots, \xi_n; t_1, \dots, t_n) d\xi_1 \dots d\xi_n d\lambda_1 \dots d\lambda_n; \\ & f_n(z_1, \dots, z_{n-1}, z_n; t_1, \dots, t_{n-1}, t_n) = \\ & = f_{n-1}(z_1, \dots, z_{n-1}; t_1, \dots, t_{n-1}) \delta(z_n - z_{n-1}); \end{aligned}$$

* Работа выполнена при финансовой поддержке РФФИ (проект № 13-07-00036) и программой «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект 1.7).

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

$$\begin{aligned} & \frac{\partial g_1(\lambda; t)}{\partial t} - \\ & - \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i\lambda^T a(z, t) e^{i(\lambda^T - \mu^T)z} g_1(\mu; t) d\mu dz = \\ & = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(b(z, t)^T \lambda^T; t) e^{i(\lambda^T - \mu^T)z} \times \\ & \quad \times g_1(\mu; t) d\mu dz; \quad (5) \end{aligned}$$

$$g_1(\lambda; t_0) = g_0(\lambda); \quad (6)$$

$$\begin{aligned} & \frac{\partial g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)}{\partial t_n} - \\ & - \frac{1}{(2\pi)^{kn}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} i\lambda^T a(z_n, t_n) \times \\ & \quad \times \exp \left[i \sum_{k=1}^n (\lambda_k^T - \mu_k^T) z_k \right] \times \\ & \times g_n(\mu_1, \dots, \mu_n; t_1, \dots, t_n) d\mu_1 \dots d\mu_n dz_1 \dots dz_n = \\ & = \frac{1}{(2\pi)^{kn}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \chi(b(z_n; t)^T \lambda_n; t_n) \times \\ & \quad \times \exp \left[i \sum_{k=1}^n (\lambda_k^T - \mu_k^T) z_k \right] \times \\ & \times g_n(\mu_1, \dots, \mu_n; t_1, \dots, t_n) d\mu_1 \dots d\mu_n dz_1 \dots dz_n; \end{aligned}$$

$$\begin{aligned} g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_{n-1}, t_{n-1}) &= \\ & = g_{n-1}(\lambda_1, \dots, \lambda_{n-2}, \lambda_{n-1} + \lambda_n; t_1, \dots, t_{n-1}), \end{aligned}$$

$$t_1 \leq t_2 \leq \dots \leq t_n, \quad n = 2, 3, \dots$$

При этом одно- и n -мерные плотности и характеристические функции связаны между собой известными соотношениями:

$$f_1(z; t) = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} e^{-i\mu^T z} g_1(\mu; t) d\mu;$$

$$g_1(\lambda; t) = \int_{-\infty}^{\infty} e^{i\lambda^T z} f_1(z; t) dz;$$

$$\begin{aligned} f_n(z_1, \dots, z_n; t_1, \dots, t_n) &= \\ & = \frac{1}{(2\pi)^{kn}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left\{ -i \sum_{l=1}^n \lambda_l^T z_l \right\} \times \\ & \quad \times g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n) d\lambda_1 \dots d\lambda_n; \quad (7) \end{aligned}$$

$$\begin{aligned} g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n) &= \\ & = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left\{ i \sum_{l=1}^n \lambda_l^T z_l \right\} \times \\ & \quad \times f_n(z_1, \dots, z_n; t_1, \dots, t_n) dz_1 \dots dz_n. \end{aligned}$$

Для нахождения одномерных плотностей $f_1(z, t) = f_1^*(z)$ и характеристических функций $g_1(\lambda; t) = g_1^*(\lambda)$ стохастических режимов в стационарных СтС (1) при

$$a(z, t) = a^*(z); \quad b(z, t) = b^*(z); \quad \chi(\mu; t) = \chi^*(\mu) \quad (8)$$

следует в (3) и (5) положить $\partial f_1 / \partial t = 0$ и $\partial g_1 / \partial t = 0$. В результате получим соответственно

$$\begin{aligned} & \frac{\partial^T}{\partial z} [a^*(z) f_1^*(z)] = \\ & = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^*(b^*(\xi)^T \lambda) e^{i\lambda^T (\xi - z)} f_1^*(\xi) d\xi d\lambda; \\ & - \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i\lambda^T a^*(z) e^{i(\lambda^T - \mu^T)z} g_1^*(\mu) d\mu dz = \\ & = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^*(b^*(z)^T \lambda) e^{i(\lambda^T - \mu^T)z} g_1^*(\mu) d\mu dz. \end{aligned}$$

Поставим задачу разработки точных и приближенных алгоритмов аналитического моделирования распределений (плотностей и характеристических функций) стохастических режимов $Z = Z(t)$ в нелинейных гауссовских и негауссовских СтС (1) с разрывными характеристиками $a = a(z, t)$ и $b = b(z, t)$, обладающих свойством сохранения инвариантной меры, т.е. удовлетворяющих уравнениям (3) и (5) при $\chi = 0$.

Условия сохранения инвариантной меры можно представить в следующем развернутом виде:

$$\left. \begin{aligned} & \frac{\partial f_1(z; t)}{\partial t} + A_a f_1(z; t) = 0; \\ & A_a f_1(z; t) = \frac{\partial^T}{\partial z} [a(z, t) f_1(z; t)] = \text{div } \pi(z; t); \end{aligned} \right\} \quad (9)$$

$$\left. \begin{aligned} & A_a^* f_1^*(z) = 0; \\ & A_a^* f_1^*(z) = \frac{\partial^T}{\partial z} [a^*(z) f_1^*(z)] = \text{div } \pi^*(z); \end{aligned} \right\} \quad (10)$$

$$\frac{\partial g_1(\lambda; t)}{\partial t} - B_a g_1(\lambda; t) = 0;$$

$$\begin{aligned}
 B_a g_1(\lambda; t) &= \\
 &= \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i\lambda^T a(z, t) e^{i(\lambda^T - \mu^T)z} g_1(\mu; t) d\mu dz = \\
 &= \int_{-\infty}^{\infty} i\lambda^T a(z, t) e^{i\lambda^T z} f_1(z; t) dz = \\
 &= \int_{-\infty}^{\infty} e^{i\lambda^T z} i\lambda^T \pi(z; t) dz; \quad (11)
 \end{aligned}$$

$$\left. \begin{aligned}
 B_a^* g_1^*(\lambda) &= 0; \\
 B_a^* g_1^*(\lambda) &= \\
 &= \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} i\lambda^T a^*(z) e^{i(\lambda^T - \mu^T)z} g_1^*(\mu) d\mu dz = \\
 &= \int_{-\infty}^{\infty} i\lambda^T a^*(z) e^{i\lambda^T z} f_1^*(z) dz = \\
 &= \int_{-\infty}^{\infty} e^{i\lambda^T z} i\lambda^T \pi^*(z) dz.
 \end{aligned} \right\} (12)$$

Для гауссовских (нормальных) СтС с гладкими характеристиками точные и приближенные методы и алгоритмы аналитического моделирования рассмотрены в [1–15].

Особое внимание уделим приближенным методам, основанным на методах нормальной аппроксимации и статистической линеаризации. Подробно рассмотрим их применение к пуассоновским СтС.

2 Точные методы и алгоритмы аналитического моделирования распределений с инвариантной мерой

Пусть функция a в СтС (1) допускает представление

$$a = a(z, t) = a_1(z, t) + a_2(z, t) \quad (13)$$

такое, что функция $f_1 = f_1(z; t)$ является плотностью инвариантной меры невозмущенной шумами системы, описываемой векторным обыкновенным дифференциальным уравнением вида

$$\dot{z} = a_1(z, t), \quad (14)$$

т. е. удовлетворяет условию (9):

$$\frac{\partial f_1(z; t)}{\partial t} + \frac{\partial^T}{\partial z} [a_1(z, t) f_1(z; t)] = 0. \quad (15)$$

Для гладких функций $a_1 = a_1(z, t)$ вопросы существования и основные свойства интегральных инвариантов изучены в [16, 17]. При этом в (13) функция $a_2 = a_2(z, t)$ определяется путем решения следующего интегродифференциального уравнения:

$$\begin{aligned}
 \frac{\partial^T}{\partial z} [a_2(z, t) f_1(z; t)] &= \frac{1}{(2\pi)^k} \times \\
 &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(b(\xi, t)^T \lambda; t) e^{i\lambda^T(\xi - z)} f_1(\xi; t) d\xi d\lambda. \quad (16)
 \end{aligned}$$

В общем случае нахождение функций a_1 и a_2 в (13) — такая же трудная задача, как решение уравнений (3) и (4).

Для стационарных СтС, когда выполнены условия (8), уравнения (13)–(16) имеют вид:

$$a(z) = a_1(z) + a_2(z); \quad (17)$$

$$\dot{z} = a_1(z), \quad (18)$$

$$\begin{aligned}
 \frac{\partial^T}{\partial z} [a_2^*(z) f_1^*(z)] &= \\
 &= \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^*(b^*(\xi)^T \lambda) e^{i\lambda^T(\xi - z)} f_1^*(\xi) d\xi d\lambda. \quad (19)
 \end{aligned}$$

В этом случае можно выбирать невозмущенную систему (18) так, чтобы она имела первые интегралы.

В терминах характеристических функций соотношения (15), (16) и (19) могут быть записаны следующим образом:

$$\frac{\partial g_1(\lambda; t)}{\partial t} - B_{a_1} g_1(\lambda; t) = 0; \quad (20)$$

$$B_{a_1}^* g_1^*(\lambda) = 0.$$

Для составляющих $a_2(z, t)$ и $a_2^*(z)$ имеют место уравнения

$$\begin{aligned}
 B_{a_2} g_1(\lambda; t) &= \frac{1}{(2\pi)^k} \times \\
 &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(b(z, t)^T \lambda; t) e^{i(\lambda^T - \mu^T)z} g_1(\mu; t) d\mu dz; \quad (21)
 \end{aligned}$$

$$\begin{aligned}
 B_{a_2}^* g_1^*(\lambda) &= \frac{1}{(2\pi)^k} \times \\
 &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^*(b^*(z)^T \lambda) e^{i(\lambda^T - \mu^T)z} g_1^*(\mu) d\mu dz. \quad (22)
 \end{aligned}$$

Отсюда вытекают конструктивные точные алгоритмы аналитического моделирования распределений с инвариантной мерой. В их основе лежат следующие теоремы.

Теорема 2.1. Функция $f_1 = f_1(z; t)$ будет решением (3) и (4) тогда и только тогда, когда $a = a(z, t)$ допускает представление (13) такое, что $f_1 = f_1(z; t)$ является плотностью инвариантной меры обыкновенного дифференциального уравнения (14), т. е. удовлетворяет условию (15). При этом составляющая a_2 определяется из решения интегродифференциального уравнения (16).

Теорема 2.2. Функция $f_1^* = f_1^*(z)$ будет решением (3) тогда и только тогда, когда $a^* = a^*(z)$ допускает представление (17) такое, что $f_1^* = f_1^*(z)$ является плотностью инвариантной меры (18). При этом составляющая a_2^* определяется из решения уравнения (19).

Теорема 2.3. Функция $g_1 = g_1(\lambda; t)$ будет решением (5), (6) тогда и только тогда, когда недифференцируемая функция $a = a(z, t)$ допускает представление (13) такое, что $g_1 = g_1(\lambda; t)$ является характеристической функцией инвариантной меры уравнения (14), т. е. удовлетворяет условию (20). При этом составляющая a_2 определяется из уравнения (21).

Теорема 2.4. Функция $g_1^* = g_1^*(\lambda)$ будет решением (7) тогда и только тогда, когда недифференцируемая функция $a^* = a^*(z)$ допускает представление (17) такое, что g_1^* является характеристической функцией инвариантной меры (14). При этом a_2^* определяется из решения (22).

Теоремы 2.1–2.4 легко обобщаются на случай многомерных распределений с инвариантной мерой.

3 Приближенные методы и алгоритмы аналитического моделирования распределений с инвариантной мерой, основанные на нормальной аппроксимации и статистической линеаризации

Пусть нелинейная СтС (1) допускает применение метода нормальной аппроксимации (МНА) [1, 2]. Тогда одно- и двумерные нормальные плотности $f_1^{\text{МНА}}, f_2^{\text{МНА}}$ и характеристические функции $g_1^{\text{МНА}}, g_2^{\text{МНА}}$, а также вектор математического

ожидания $m_t = M^{\text{МНА}} Z(t)$, ковариационная матрица $K_t = M^{\text{МНА}} Z^{0T} Z^0(t)$ ($Z^0(t) = Z(t) - m_t$) и матрица ковариационных функций $K(t_1, t_2) = M^{\text{МНА}} Z^{0T}(t_1) Z^0(t_2)$ ($t_1 < t_2$) определяются следующими уравнениями:

$$f_1^{\text{МНА}} = f_1^{\text{МНА}}(z; t, m_t, K_t) = [(2\pi)^k |K_t|]^{-1/2} \times \exp \left\{ -\frac{1}{2} (z^T - m_t^T) K_t^{-1} (z - m_t) \right\}; \quad (23)$$

$$f_2^{\text{МНА}} = f_2^{\text{МНА}}(z_1, z_2; t_1, t_2, m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1, t_2)) = [(2\pi)^k |\bar{K}_2|]^{-1/2} \times \exp \left\{ -([z_1^T z_2^T] - \bar{m}_2^T) \bar{K}_2^{-1} ([z_1^T z_2^T]^T - \bar{m}_2) \right\}; \quad (24)$$

$$g_1^{\text{МНА}}(\lambda; t) = \exp \left\{ i\lambda^T m - \frac{1}{2} \lambda^T K_t \lambda \right\}; \quad (25)$$

$$g_2^{\text{МНА}}(\lambda_1, \lambda_2; t_1, t_2) = \exp \left\{ i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \bar{\lambda} \right\}; \quad (26)$$

$$\bar{\lambda} = [\lambda_1^T \lambda_2^T]^T; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T;$$

$$\bar{K}_2 = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix};$$

$$\dot{m}_t = a_1(t, m_t, K_t) = \int_{-\infty}^{\infty} a(z, t) f_1^{\text{МНА}}(z; t, m_t, K_t) dz; \quad (27)$$

$$\dot{K}_t = a_2(t, m_t, K_t) = a_{21} + a_{12} + a_{22} = \left[\int_{-\infty}^{\infty} a(z, t) (z^T - m_t^T) + (z - m_t) a^T(z, t) + \sigma(z, t) \right] f_1^{\text{МНА}}(z; t, m_t, K_t) dz; \quad (28)$$

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = a_3(t_1, t_2, m_{t_1}, m_{t_2}, K_{t_1}, K_{t_2}, K(t_1, t_2)) = [(2\pi)^{2k} |\bar{K}_2|]^{-1/2} \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{t_1}) a(z_2, t_2) \exp \left\{ -([z_1^T z_2^T] - \bar{m}_2^T) \times \bar{K}_2^{-1} ([z_1^T z_2^T] - \bar{m}_2) \right\} dz_1 dz_2. \quad (29)$$

Здесь введены следующие обозначения:

$$\left. \begin{aligned} z_1 = z_{t_1}; \quad z_2 = z_{t_2}; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T; \\ \bar{K}_2 = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix}, \end{aligned} \right\} \quad (30)$$

$$\sigma(z, t) = b(z, t)\nu(t)b(z, t)^T, \quad (31)$$

где $\nu = \nu(t)$ — интенсивность негауссовского белого шума $V = V(t)$.

Для стационарных СтС при $\dot{m}^* = 0$, $\dot{K}^* = 0$, $K(t_1, t_2) = k(\tau)$ ($\tau = t_1 - t_2$) соотношения (27)–(31) принимают вид:

$$a_1^*(m^*, K^*) = 0; \quad (32)$$

$$a_2^*(m^*, K^*) = 0; \quad (33)$$

$$\frac{dk(\tau)}{d\tau} = a_{11}^{\text{МНА}}(m^*, K^*)k(\tau); \quad (34)$$

$$k(\tau) = k(-\tau^T); \quad k(0) = K.$$

Из уравнения (34) следует, что алгоритм МНА будет устойчивым, если матрица $a_{11}^{\text{МНА}}(m_t, K_t, t)$ будет асимптотически устойчива.

Для m и K уравнения метода статистической линеаризации (МСЛ) в нелинейных СтС при аддитивных шумах, когда $b(z, t) = b_0(t)$, $b^*(z) = b_0^*$ получаются из (27)–(29) и (32)–(34) как частный случай.

Условия наличия нормального распределения с инвариантной мерой (9) и (10), если заменить $a(z, t)$ статистически линеаризованным выражением

$$a(Z, t) \approx a_{10}^{\text{МНА}}(t, m_t, K_t) + a_{11}^{\text{МНА}}(t, m_t, K_t)(Z - m_t),$$

где

$$a_{10}^{\text{МНА}} = a_{10}^{\text{МНА}}(t, m_t, K_t) \equiv a_1;$$

$$\begin{aligned} a_{11}^{\text{МНА}} &= a_{11}^{\text{МНА}}(t, m_t, K_t) = \\ &= \left[\int_{-\infty}^{\infty} a(z, t)(z^T - m_t^T) f_1^{\text{МНА}}(z; t, m_t, K_t) dz \right] \times \\ &\quad \times K_t^{-1} = \left(\frac{\partial}{\partial m_t} a_1^T \right)^T, \end{aligned}$$

приводят к следующим соотношениям:

$$\begin{aligned} \frac{\partial f_1^{\text{МНА}}(z; t, m_t, K_t)}{\partial t} + \frac{\partial^T}{\partial z} \{ [a_{10}^{\text{МНА}}(t, m_t, K_t) + \\ + a_{11}^{\text{МНА}}(t, m_t, K_t)(z - m_t)] \times \\ \times f_1^{\text{МНА}}(z; t, m_t, K_t) \} = 0; \quad (35) \end{aligned}$$

$$\frac{\partial^T}{\partial z} \{ [a_{10}^{\text{МНА}}(m^*, K^*) + a_{11}^{\text{МНА}}(m^*, K^*)(z - m^*)] \times \\ \times f_1^{\text{МНА}}(z; m^*, K^*) \} = 0, \quad (36)$$

где

$$\begin{aligned} f_1^{\text{МНА}}(z; m^*, K^*) &= [(2\pi)^k |K^*|]^{-1/2} \times \\ &\times \exp \left\{ -\frac{1}{2}(z^T - m^{*T})(K^*)^{-1}(z - m^*) \right\}. \end{aligned}$$

Аналогично в развернутом виде выписываются условия (11) и (12):

$$\begin{aligned} \frac{\partial g_1^{\text{МНА}}(\lambda; t)}{\partial t} - \int_{-\infty}^{\infty} i\lambda^T [a_{10}^{\text{МНА}}(m_t, K_t, t) + \\ + a_{11}^{\text{МНА}}(m_t, K_t, t)(z - m_t)] \times \\ \times e^{i\lambda^T z} f_1^{\text{МНА}}(z; m_t, K_t, t) dz = 0; \quad (37) \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} i\lambda^T [a_{10}^{\text{МНА}}(m^*, K^*) + \\ + a_{11}^{\text{МНА}}(m^*, K^*)(z - m^*)] \times \\ \times e^{i\lambda^T z} f_1^{\text{МНА}}(z; m^*, K^*) dz = 0. \quad (38) \end{aligned}$$

Отсюда вытекают следующие теоремы.

Теорема 3.1. Если существуют одно- и двумерные плотности стохастического режима, а матрица $a_{11}^{\text{МНА}}$ коэффициентов статистической (нормальной) линеаризации асимптотически устойчива, то приближенный алгоритм аналитического моделирования МНА нестационарных стохастических режимов в СтС (1) с инвариантной мерой определяется выражениями (23)–(29) и (35).

Теорема 3.2. Если существуют стационарные одно- и двумерные плотности стохастического режима, а матрица $a_{11}^{\text{МНА}}$ коэффициентов статистической (нормальной) линеаризации асимптотически устойчива, то приближенный алгоритм аналитического моделирования стационарных стохастических режимов с инвариантной мерой в стационарной СтС (1) определяется выражениями (32)–(34) и (36).

Как известно [1, 2], одно- и двумерные нормальные распределения определяют и все n -мерные распределения ($n \geq 3$), поэтому МНА и МСЛ дают приближенные алгоритмы для любых многомерных плотностей стохастических режимов, если они существуют. Аналогично формулируются теоремы 3.3 и 3.4 на основе условий (37) и (38).

4 О других приближенных методах и алгоритмах аналитического моделирования распределений с инвариантной мерой

Обобщением МНА являются различные приближенные методы, основанные на параметризации распределений [1, 2]. Аппроксимируя одномерную характеристическую функцию $g_1(\lambda; t)$ и соответствующую плотность $f_1(z, t)$ известными функциями $g_1^*(\lambda; \theta)$, $f_1^*(z; \theta)$, зависящими от конечномерного векторного параметра θ , можно свести задачу приближенного определения одномерного распределения к выводу из уравнения для характеристических функций обыкновенных дифференциальных уравнений, определяющих θ как функцию времени. Это относится и к остальным многомерным распределениям. При аппроксимации многомерных распределений целесообразно выбирать последовательности функций $\{f_n^*(z_1, \dots, z_n; \theta_n)\}$ и $\{g_n^*(\lambda_1, \dots, \lambda_n; \theta_n)\}$, каждая пара которых находилась бы в такой зависимости от векторного параметра θ_n , чтобы при любом n множество параметров, образующих вектор θ_n , включало в качестве подмножества множество параметров, образующих вектор θ_{n-1} . Тогда при аппроксимации n -мерного распределения придется определять только те координаты вектора θ_n , которые не были определены ранее при аппроксимации функций $f_1, g_1, \dots, f_{n-1}, g_{n-1}$.

В зависимости от того, что представляют собой параметры, от которых зависят функции $f_n^*(z_1, \dots, z_n; \theta_n)$ и $g_n^*(\lambda_1, \dots, \lambda_n; \theta_n)$, аппроксимирующие неизвестные многомерные плотности $f_n(z_1, \dots, z_n; t_1, \dots, t_n)$ и характеристические функции $g_n(\lambda_1, \dots, \lambda_n; t_1, \dots, t_n)$, используются различные приближенные методы решения уравнений при условиях (9)–(12), определяющих многомерные распределения вектора состояния системы X_t , в частности методы моментов (ММ), семиинвариантов (МСИ), ортогональных разложений (МОР), квазимоментов (МКМ) и др. [1, 2].

5 Обобщение на случай стохастических систем с автокоррелированными шумами

Пусть СтС описывается нелинейным, в общем случае векторным дифференциальным стохастическим уравнением Ито [1, 2, 15, 18]

$$\left. \begin{aligned} \dot{Z} &= a(Z, t) + b_U(Z, t)U; \\ \sum_{i=0}^l \alpha_i U^{(i)} &= \sum_{j=0}^h \beta_j V^{(j)} \quad (h < l). \end{aligned} \right\} \quad (39)$$

Здесь $U = U(t)$ — векторная помеха размерности $m \times 1$; $V = V(t)$ — негауссовский белый шум с нулевым математическим ожиданием и известной функцией $\chi = \chi(\mu; t)$. В таком случае в зависимости от степени «гладкости» стохастического режима $Z = Z(t)$ и помехи $U = U(t)$ уравнения (39) путем расширения вектора состояния согласно [1, 2] приводятся к виду (1) для расширенного вектора состояния \bar{Z} . Тогда, но уже для расширенного вектора состояния СтС, при решении уравнений (39) могут быть использованы точные (разд. 2) и приближенные (разд. 3) методы и алгоритмы аналитического моделирования нестационарных и стационарных распределений с инвариантной мерой.

6 Особенности аналитического моделирования распределений с инвариантной мерой в пуассоновских стохастических системах

Рассмотрим СтС (1) при $b(z, t) = I_m$ для обобщенного пуассоновского белого шума $V^{\text{OP}} = V^{\text{OP}}(t)$, когда функция (2) определяется формулой

$$\chi^{\text{OP}}(\mu; t) = [g_c^{\text{OP}}(\mu) - 1] \nu^{\text{OP}}(t),$$

где $g_c^{\text{OP}} = g_c^{\text{OP}}(\mu)$ — характеристическая функция скачков; $\nu^{\text{OP}} = \nu^{\text{OP}}(t)$ — интенсивность пуассоновского белого шума $V^{\text{OP}} = V^{\text{OP}}(t)$. Обозначим через $f_c^{\text{OP}} = f_c^{\text{OP}}(z)$ плотность скачков обобщенного пуассоновского процесса. Тогда (3) будет представлять собой известное уравнение Феллера–Колмогорова

$$\begin{aligned} \frac{\partial f_1(z; t)}{\partial t} + \frac{\partial^T}{\partial z} [a(z, t) f_1(z; t)] &= \\ = \nu^{\text{OP}}(t) \left[\int_{-\infty}^{\infty} f_c^{\text{OP}}(z - \xi) f_1(\xi; t) d\xi - f_1(z; t) \right] \end{aligned} \quad (40)$$

с начальным условием (4). В случае простого пуассоновского белого шума с единичными скачками $g_c(\mu) = e^{i\mu}$.

Для стационарной пуассоновской СтС (1) уравнение (40) имеет следующий вид:

$$\begin{aligned} \frac{\partial^T}{\partial z} [a^*(z)f_1^*(z)] = \\ = \nu^{\text{OP}*} \left[\int_{-\infty}^{\infty} f_c^{\text{OP}}(z-\xi)f_1^*(\xi) d\xi - f_1^*(z) \right]. \end{aligned} \quad (41)$$

Пользуясь уравнениями (40), (41) и результатами разд. 1 и 2, нетрудно сформулировать следующие утверждения.

Теорема 6.1. *Функция $f_1 = f_1(z; t)$ будет нестационарным решением (40), (4) тогда и только тогда, когда a допускает представление (13) такое, что f_1 является плотностью инвариантной меры обыкновенного дифференциального уравнения (14), т. е. удовлетворяет условию (15), а составляющая a_2 определяется из решения следующего уравнения:*

$$\begin{aligned} \frac{\partial^T}{\partial z} [a_2(z, t)f_1(z; t)] = \frac{1}{(2\pi)^k} \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^{\text{OP}}(b(\xi, t)^T \lambda; t) e^{i\lambda^T(\xi-z)} f_1(\xi, t) d\xi d\lambda. \end{aligned}$$

Теорема 6.2. *Функция $f_1^* = f_1^*(z)$ будет стационарным решением (41) тогда и только тогда, когда a_2^* допускает представление (17) такое, что f_1^* является плотностью инвариантной меры (18), а составляющая a_2^* определяется из решения следующего уравнения:*

$$\begin{aligned} \frac{\partial^T}{\partial z} [a_2^*(z)f_1^*(z)] = \\ = \frac{1}{(2\pi)^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi^{\text{OP}*}(b(\xi)^T \lambda) e^{i\lambda^T(\xi-z)} f_1^*(\xi) d\xi d\lambda. \end{aligned}$$

При использовании МНА и МСЛ для пуассоновских СтС непосредственно применяются теоремы 3.1–3.4, причем в формулу (31) для $\sigma(z, t)$ входит интенсивность $\nu^{\text{OP}}(t)$ обобщенного пуассоновского белого шума.

7 Тестовые примеры

Пример 1. Рассмотрим осциллятор Дуффинга в обобщенной пуассоновской стохастической среде:

$$\ddot{X} + \omega^2 X - \mu X^3 = -\delta^{\text{OP}} \dot{X} + V^{\text{OP}}(t). \quad (42)$$

Уравнения МСЛ для (42) имеют следующий вид:

$$\dot{m}_X = m_{\dot{X}}; \quad \dot{m}_{\dot{X}} = -\omega_3^2 m_X - \delta^{\text{OP}} m_{\dot{X}}; \quad (43)$$

$$\left. \begin{aligned} \dot{D}_X &= 2K_{X\dot{X}}; \\ \dot{D}_{\dot{X}} &= \nu^{\text{OP}} - 2(\omega_{1\vartheta}^2 K_{X\dot{X}} + \delta^{\text{OP}} D_{\dot{X}}); \\ \dot{K}_{X\dot{X}} &= D_{\dot{X}} - \omega_{1\vartheta}^2 D_X - \delta^{\text{OP}} K_{X\dot{X}}. \end{aligned} \right\} \quad (44)$$

Здесь кубическая функция X^3 была заменена на статистически линеаризованную при гауссовом распределении с дисперсией D_X согласно [1, 2]:

$$X^3 \approx k_0(m_X, D_X)m_X + k_1(m_X, D_X)X^0,$$

где

$$\begin{aligned} k_0(m_X, D_X) &= m_X^2 + 3D_X; \\ k_1(m_X, D_X) &= 3(m_X^2 + D_X); \\ \omega_3^2 &= \omega^2 \left[1 - \frac{\mu(m_X^2 + 3D_X)}{\omega^2} \right]; \\ \omega_{1\vartheta}^2 &= \omega^2 \left[1 - \frac{3\mu(m_X^2 + D_X)}{\omega^2} \right] \quad (\omega_{1\vartheta} > \omega_{1\vartheta}). \end{aligned}$$

Из (43) и (44) в стационарном режиме имеем:

$$\begin{aligned} m_X^* &= 0; \quad m_{\dot{X}}^* = 0; \quad K_{X\dot{X}}^* = 0; \\ D_X^* &= \vartheta; \quad \vartheta = \frac{\nu^{\text{OP}}}{2\delta^{\text{OP}}}, \end{aligned}$$

а D_X^* определяется из уравнения:

$$\omega_{1\vartheta}^2(D_X^*)D_X^* = \vartheta.$$

Условие наличия стационарного распределения с инвариантной мерой (36) требует консерватизма линеаризованной левой части (42). Процесс установления стационарных стохастических колебаний происходит в два этапа: сначала устанавливается D_X^* , а затем $D_{\dot{X}}^*$.

Интересно отметить, что уравнения МСЛ (43) и (44) сохраняют свой вид и для любого белого шума интенсивности $\nu(t)$, представляющего собой с.к., производную от произвольного процесса с независимыми приращениями $W(t)$. Для гауссовского белого шума $\nu = \nu^G$ соответствующие результаты получены в [1, 2, 15]. Как показали вычислительные эксперименты для значений μ , отвечающих стохастическим колебаниям, точность составляет около 10% [15].

Пример 2. Для осциллятора Дуффинга в автокоррелированной пуассоновской среде, когда

$$\ddot{X} + \omega^2 X - \mu X^3 = -\delta^{\text{OP}} \dot{X} + U; \quad \dot{U} + \gamma U = V^{\text{OP}}(t),$$

уравнения МСЛ для $Z = [X \dot{X} U]^T$ имеют вид (27) и (28) при

$$a_1 = \begin{bmatrix} m_{\dot{X}} \\ -\omega_0^2 m_X - \delta^{\text{OP}} m_{\dot{X}} \\ -m_U \end{bmatrix};$$

$$\alpha = \begin{bmatrix} 0 & 1 & 0 \\ -\omega_{13}^2 & -\delta^{\text{OP}} & 0 \\ 0 & 0 & -\gamma \end{bmatrix}; \quad \beta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

$$a_2 = \alpha K_t + K_t \alpha^T + \beta \nu^{\text{OP}} \beta^T.$$

Здесь $\nu^{\text{OP}} = \nu^{\text{OP}}(t)$ — интенсивность белого шума $V^{\text{OP}}(t)$. Отсюда аналитическим моделированием определяются стационарные режимы, а также режимы их установления. Так же, как в случае автокоррелированных гауссовских белых шумов [1, 2, 15], точность МСЛ за счет «профильрованности» помех значительно повышается и достигает 2%–4%. Результат справедлив и для произвольных негауссовских белых шумов.

Пример 3. Для релейного осциллятора в гауссовской стохастической среде

$$\ddot{X} + \omega^2 \text{sgn} X = -\delta^G \dot{X} + V^G + U_0 \quad (45)$$

плотность распределения стационарного режима стохастических колебаний при $U_0 = 0$ определяется формулой Гиббса [1, 2]:

$$f^*(x, \dot{x}) = c \exp \left\{ -\frac{H(x, \dot{x})}{\vartheta^G} \right\}, \quad \vartheta^G = \frac{\nu^G}{2\delta^G}. \quad (46)$$

Здесь через

$$H(x, \dot{x}) = \frac{\dot{x}^2}{2} + \Pi(x), \quad \Pi(x) = \omega^2 |x|,$$

обозначена полная энергия осциллятора.

Для (45) при $U_0 \neq 0$, если заменить релейную характеристику статистически линеаризованной, согласно [1, 2]

$$\text{sgn} X = k_0(m_X, D_X) m_X + k_1(m_X, D_X) (X^0 - m_X);$$

$$k_0(m_X, D_X) = \frac{2}{m_X} \Phi \left(\frac{m_X}{\sqrt{D_X}} \right);$$

$$k_1(m_X, D_X) = \frac{1}{\sqrt{D_X}} \sqrt{\frac{2}{\pi}} \exp \left(-\frac{m_X^2}{2D_X} \right);$$

$$\Phi(\tau) = \frac{1}{2\pi} \int_0^\tau e^{-t^2/2} dt. \quad (47)$$

Тогда уравнения МСЛ будут иметь вид:

$$\left. \begin{aligned} \dot{m}_X &= m_{\dot{X}}; \\ \dot{m}_{\dot{X}} &= U_0 - \omega^2 k_0(m_X, D_X) m_X - \delta m_{\dot{X}}; \end{aligned} \right\} \quad (48)$$

$$\left. \begin{aligned} \dot{D}_X &= 2K_{X\dot{X}}; \\ \dot{D}_{\dot{X}} &= \nu^G - 2[\delta D_{\dot{X}} + \omega^2 k_1(m_X, D_X) K_{X\dot{X}}]; \\ \dot{K}_{X\dot{X}} &= D_{\dot{X}} - \omega^2 k_1(m_X, D_X) D_X - \delta K_{X\dot{X}}, \end{aligned} \right\} \quad (49)$$

где $\delta = \delta^G$, $\nu = \nu^G$. Отсюда для стационарных стохастических колебаний имеем связанную систему уравнений:

$$m_{\dot{X}}^* = 0; \quad \omega^2 k_0(m_X^*, D_X^*) = U_0; \quad (50)$$

$$\left. \begin{aligned} K_{X\dot{X}}^* &= 0; \quad D_X^* = \vartheta = \frac{\nu}{2\delta}; \\ k_1(m_X^*, D_X^*) D_X^* &= \rho = \frac{\vartheta}{\omega^2} = \frac{\nu}{2\delta\omega^2}. \end{aligned} \right\} \quad (51)$$

При $U_0 = 0$ из (47), (50) и (51) находим:

$$m_X^* = 0; \quad m_{\dot{X}}^* = 0; \quad D_X^* = \vartheta; \quad D_{\dot{X}}^* = \frac{\pi}{2} \rho^2.$$

Отсюда видно, что стационарная дисперсия скорости совпадает с точным решением (46). Стационарная дисперсия координаты, найденная согласно МСЛ, отличается от следующего точного решения, полученного согласно (46). При $\rho \leq 1$ относительная ошибка составляет 10%. Стационарные колебания по X и \dot{X} не коррелированы.

Уравнения (48) и (49) показывают, что процесс установления режима стохастических колебаний происходит в две стадии: сначала устанавливается стационарное распределение по скорости \dot{X} , а затем по координате X .

Пример 4. В условиях примера 3, но для пуассоновской среды, когда

$$\ddot{X} + \omega^2 \text{sgn} X = -\delta^{\text{OP}} \dot{X} + V^{\text{OP}} + U_0,$$

уравнения МСЛ имеют вид (48), (49), если принять $\delta = \delta^{\text{OP}}$, $\nu = \nu^{\text{OP}}$, $\vartheta = \vartheta^{\text{OP}} = \nu^{\text{OP}}/(2\delta^{\text{OP}})$, $\rho = \vartheta^{\text{OP}}/\omega^2$. Точного аналитического уравнения Феллера–Колмогорова не обнаружено.

Другие тестовые примеры можно найти в [10, 12–14].

8 Заключение

Дано обобщение точных и приближенных (основанных на параметризации распределений) методов и алгоритмов теории распределений с инвариантной мерой на случай нелинейных дифференциальных гауссовых и негауссовых стохастических систем с гладкими и разрывными характеристиками.

Особое внимание уделено пуассоновским стохастическим системам с разрывными характеристиками.

На тестовых примерах показана достаточная точность для практических приложений в стохастической информатике.

Литература

1. Пугачёв В. С., Синицын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — 2-е изд., доп. — М.: Наука, 1990.
2. Пугачёв В. С., Синицын И. Н. Теория стохастических систем. — 2-е изд. — М.: Логос, 2004.
3. Moshchuk N. K., Sinitsyn I. N. On stationary distributions in nonlinear stochastic differential systems: Preprint. — Coventry, UK: University of Warwick, Mathematics Institute, 1989. 15 p.
4. Moshchuk N. K., Sinitsyn I. N. On stochastic nonholonomic systems: Preprint. — Coventry, UK: University of Warwick, Mathematics Institute, 1989. 32 p.
5. Мошук Н. К., Синицын И. Н. О стохастических неголономных системах // Прикладная механика и математика, 1990. Т. 54. Вып. 2. С. 213–223.
6. Moshchuk N. K., Sinitsyn I. N. On stationary distributions in nonlinear stochastic differential systems // Quart. J. Mech. Appl. Math., 1991. Vol. 44. Pt. 4. P. 571–579.
7. Мошук Н. К., Синицын И. Н. О стационарных и приводимых к стационарным режимам в нормальных стохастических системах // Прикладная механика и математика, 1991. Т. 55. Вып. 6. С. 895–903.
8. Мошук Н. К., Синицын И. Н. Распределения с инвариантной мерой в механических стохастических нормальных системах // Докл. АН СССР, 1992. Т. 322. № 4. С. 662–667.
9. Синицын И. Н. Конечномерные распределения с инвариантной мерой в стохастических механических системах // Докл. РАН, 1993. Т. 328. № 3. С. 308–310.
10. Soize C. The Fokker–Plank equation for stochastic dynamical systems and its explicit steady state solutions. — Singapore: World Scientific, 1994.
11. Синицын И. Н. Конечномерные распределения с инвариантной мерой в стохастических нелинейных дифференциальных системах. — М.: Диалог–МГУ, 1997. С. 129–140.
12. Синицын И. Н., Корепанов Э. Р., Белоусов В. В. Точные методы расчета стационарных режимов с инвариантной мерой в стохастических системах управления // Кибернетика и технологии XXI века: Тр. II Междунар. науч.-техн. конф. С&Т'2002. — Воронеж: Саквое, 2002. С. 124–131.
13. Синицын И. Н., Корепанов Э. Р., Белоусов В. В. Точные аналитические методы в статистической динамике нелинейных информационно-управляющих систем // Системы и средства информатики. Спец. вып. Математическое и алгоритмическое обеспечение информационно-телекоммуникационных систем. — М.: Наука, 2002. С. 112–121.
14. Синицын И. Н. Развитие методов аналитического моделирования распределений с инвариантной мерой в стохастических системах // Современные проблемы прикладной математики, информатики и автоматизации: Тр. Междунар. науч.-техн. семинара. — Севастополь, 2012. С. 24–35.
15. Синицын 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.
16. Немыцкий В. В., Степанов В. В. Качественная теория дифференциальных уравнений. — М.—Л.: Гостехиздат, 1949.
17. Козлов В. В. О существовании интегрального инварианта гладких динамических систем // ПММ, 1987. № 1. С. 538–545.
18. Синицын И. Н. Фильтры Калмана и Пугачёва. — 2-е изд. — М.: Логос, 2007.

ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ ПОТОКОВ В СЛОЖНЫХ ФИНАНСОВЫХ СИСТЕМАХ НА ОСНОВЕ ВЫСОКОЧАСТОТНЫХ ДАННЫХ*

В. Ю. Королев¹, А. В. Черток², А. Ю. Корчагин³, А. К. Горшенин⁴

Аннотация: Предложена микроструктурная модель, описывающая информационные потоки в сложных финансовых системах и случайную природу интенсивностей потоков заявок, определяющих механизм ценообразования финансовых инструментов. При их моделировании поток внешнего информационного фона со случайной интенсивностью рассматривается и аппроксимируется отдельно в рамках предложенной и статистически обоснованной мультипликативной модели. Эта модель позволяет анализировать характеристики, связанные с интенсивностями потоков заявок, а также мгновенное соотношение сил покупателей и продавцов без моделирования внешнего информационного фона, практически не поддающегося прогнозированию. Также предложена модель обобщенного процесса цены, учитывающая всю доступную информацию о потоках заявок и допускающая дальнейшую аналитическую интерпретацию.

Ключевые слова: финансовые рынки; информационные потоки; ценообразование; интенсивности потоков заявок; книга заявок; смесь распределений; обобщенная цена

1 Введение

Финансовый рынок является открытой информационной системой с очень сложной структурой. В последнее десятилетие с развитием электронной торговли на финансовых рынках изучение биржевых высокочастотных данных стало ключевым для более глубокого понимания закономерностей динамики подобных сложных систем и, в частности, для описания механизмов формирования цены.

Как известно, статистические распределения приращений (логарифмов) финансовых индексов и, в частности, биржевых цен имеют более тяжелые хвосты, чем нормальное (гауссово) распределение. В работе [1] этот феномен объяснен с помощью предельных теорем для обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса). В соответствии с подходом, использованным в [1], указанные распределения должны иметь вид смесей нормальных законов. Обоснованием адекватности таких моделей предложено считать стохастический характер интенсивностей хаотических информационных потоков в сложных информационных финансовых системах.

Использование высокочастотных статистических данных, ставших доступными благодаря широкому распространению систем электронной торговли на бирже, позволяет верифицировать указанные выше модели и более детально описать процесс ценообразования.

Финансовые рынки представляют собой пример сложных открытых стохастических информационных систем, в которых можно выделить два основных источника случайности: внутренний и внешний. Внутренний источник случайности порождает неопределенность, обусловленную различием стратегий очень большого числа участников рынка. «Физическим» аналогом такой случайности может служить хаотическое тепловое движение частиц в замкнутых системах. Внешний источник случайности — это плохо поддающийся более или менее полному прогнозированию поток новостей политического и экономического характера (в том числе потоки информации с внешних рынков и инструментов), в соответствии с которыми изменяются интересы и стратегии участников рынка. Эти два источника случайности будут учитываться при описании модели ценообразования.

*Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м, 12-07-00115а).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, vkorolev@cs.msu.su

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Euphoria Group LLC, a.v.chertok@gmail.com

³Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, sasha.korchagin@gmail.com

⁴Институт проблем информатики Российской академии наук, a.k.gorshenin@gmail.com

2 Описание модели

На классических электронных рынках, которые в данный момент принадлежат к числу наиболее активно развивающихся типов рынка, биржевая цена финансового инструмента в ее классическом понимании является результирующей, интегральной характеристикой системы торгов, которая описывается динамикой так называемой *книги заявок* (limit order book), представляющей собой список всех актуальных на данный момент предложений о покупке и продаже инструмента по различным ценам. Динамику книги заявок на электронном рынке определяют три типа заявок:

- (1) *лимитная заявка* обозначает намерение купить или продать фиксированный объем инструмента по определенной цене (купить по цене не выше заданной или продать по цене не ниже заданной), заявка немедленно добавляется в книгу заявок;
- (2) *рыночная заявка* обозначает намерение купить или продать фиксированный объем инструмента по лучшей предложенной цене, после чего немедленно происходит ее сведение с одной из лимитных заявок (при их наличии);
- (3) *заявка на отмену* обозначает намерение отменить существующий лимитный ордер, после чего он удаляется из книги заявок.

Разумеется, лимитная заявка может оказаться рыночной, если заявленная в ней цена позволяет немедленно произвести сведение с одной из лимитных заявок на противоположной стороне книги заявок. Участники рынка, присылающие лимитные ордера, являются *поставщиками ликвидности* (liquidity providers), а те, кто присылают рыночные заявки, — *потребителями ликвидности* (liquidity takers).

Итак, рассмотрим прежде динамику книги заявок на дискретной сетке цен $\Pi = \{1, 2, \dots, M\}$ как процесс с непрерывным временем

$$\mathbf{x}(t) \equiv (\mathbf{V}^a(t); \mathbf{V}^b(t)) \equiv (V_1^a(t), V_2^a(t), \dots, V_M^a(t); V_1^b(t), V_2^b(t), \dots, V_M^b(t)),$$

где $V_p^a(t)$ ($V_p^b(t)$) обозначает количество лимитных заявок на продажу (покупку) с ценой $p \in \Pi$. Так как в один момент не может существовать заявок на покупку и продажу по одной цене (иначе они будут сведены), необходимо потребовать $V_p^a(t) \vee V_p^b(t) = 0$ для всех p и t .

Лучшая цена на продажу $a(t)$ определяется как

$$a(t) = \inf\{p : V_p^a(t) > 0\} \wedge (M + 1),$$

лучшая цена на покупку $b(t)$ определяется как

$$b(t) = \sup\{p : V_p^b(t) > 0\} \vee 0.$$

При этом процесс цены можно, например, определить как

$$P(t) = \frac{a(t) + b(t)}{2}.$$

Таким образом, процесс цены $P(t)$ является результатом процесса эволюции книги заявок, инициированного потоком заявок трех типов.

Предположим вначале, что поток информации, поступающей извне, фиксирован. Тогда при фиксированной информации можно считать, что внутренняя случайность является установившимся хаосом. Как показано, например, в [1, 2], естественными математическими моделями хаотических потоков являются пуассоновские процессы, характеризуемые тем, что интервалы времени между информативными событиями являются независимыми одинаково распределенными случайными величинами с экспоненциальным распределением. Поэтому на первом этапе при построении рассматриваемой модели потоки заявок моделируются с использованием независимых процессов с экспоненциальными распределениями (как это сделано, например, в работах [3, 4]):

- лимитные заявки на покупку (продажу) приходят на ценовой уровень, расположенный на расстоянии i от лучшей котировки противоположного типа, в независимые моменты времени, имеющие экспоненциальное распределение с параметром λ_i^+ (λ_i^-) (эмпирические исследования [5, 6] показывают, что степенный закон $\lambda_i^\pm = k/i^\alpha$ является хорошей аппроксимацией);
- рыночные заявки на покупку (продажу) приходят в независимые моменты времени, имеющие экспоненциальное распределение с параметром μ^+ (μ^-);
- заявки на отмену лимитного ордера на покупку (на продажу), находящегося на дистанции i от лучшей котировки того же типа, приходят с частотой θ_i^+ (θ_i^-).

В данной статье для простоты изложения рассматриваются потоки заявок единичного объема, однако все рассуждения могут быть распространены на более общий случай (см. также [7]). Таким образом, $\mathbf{x}(t)$ является цепью Маркова с непрерывным временем в пространстве состояний $(\mathbb{Z}^+)^{2M}$ и следующими переходами:

$V_i^a(t) \rightarrow V_i^a(t) + 1$ с интенсивностью $\lambda_{i-b(t)}^-$
для $i > b(t)$;

$V_i^a(t) \rightarrow V_i^a(t) - 1$ с интенсивностью $\theta_{i-a(t)}^-$
для $i \geq a(t)$;

$V_i^a(t) \rightarrow V_i^a(t) - 1$ с интенсивностью μ^+
для $i = a(t) > 0$.

$V_i^b(t) \rightarrow V_i^b(t) + 1$ с интенсивностью $\lambda_{a(t)-i}^+$
для $i < a(t)$;

$V_i^b(t) \rightarrow V_i^b(t) - 1$ с интенсивностью $\theta_{b(t)-i}^+$
для $i \leq b(t)$;

$V_i^b(t) \rightarrow V_i^b(t) - 1$ с интенсивностью μ^-
для $i = b(t) < M + 1$.

С учетом вышесказанного можно определить следующие независимые пуассоновские процессы:

- $L_i^\pm(t)$: потоки лимитных ордеров с интенсивностями λ_i^\pm ;
- $M^\pm(t)$: потоки рыночных ордеров с интенсивностями $\mu^+ \mathbb{I}(\mathbf{V}^a \neq 0)$ и $\mu^- \mathbb{I}(\mathbf{V}^b \neq 0)$;
- $C_i^\pm(t)$: потоки заявок на отмену лимитных ордеров с интенсивностями θ_i^\pm ;
- пуассоновский процесс

$$N(t) = M^+(t) + M^-(t) + \sum_{i=1}^M (L_i^+(t) + L_i^-(t)) + \sum_{i=1}^M (C_i^+(t) + C_i^-(t)),$$

описывающий поток всех заявок, поступающих на рынок.

Процессы $L_i^\pm(t)$, $M^\pm(t)$, $C_i^\pm(t)$ полностью определяют процесс цены $P(t)$ и для него, вообще говоря, могут быть выписаны соответствующие стохастические дифференциальные уравнения (см. [8]), однако дальнейшая его аналитическая интерпретация представляется очень сложной или вообще невозможной даже при довольно сильных допущениях о постоянных и независимых интенсивностях потоков заявок разных типов, что никак не соответствует действительности.

3 Обобщенный процесс цены

Справедливо рассмотреть обобщенный процесс цены, в котором учтены не только изменения лучших котировок, но и постановка/снятие заявок в

глубине книги заявок, поскольку каждое такое действие оказывает влияние на текущее распределение сил покупателей и продавцов.

Напомним, что внешний информационный фон пока предполагается неизменным.

Зафиксируем прежде достаточно небольшой интервал времени $[0, t]$, позволяющий считать, что на таком интервале интенсивности описанных событий постоянны. Пусть по-прежнему $N(t)$ — пуассоновский процесс, соответствующий всем событиям в книге заявок и имеющий интенсивность

$$\lambda = \mu^+ + \mu^- + \sum_{i=1}^M (\lambda_i^+ + \lambda_i^-) + \sum_{i=1}^M (\theta_i^+ + \theta_i^-).$$

Расщепим его на два пуассоновских процесса $N_+(t)$ и $N_-(t)$ с интенсивностями соответственно:

$$\lambda_+ = \mu^+ + \sum_{i=1}^M \lambda_i^+ + \sum_{i=1}^M \theta_i^-;$$

$$\lambda_- = \mu^- + \sum_{i=1}^M \lambda_i^- + \sum_{i=1}^M \theta_i^+.$$

Таким образом, $\lambda = \lambda_+ + \lambda_-$, а процессы $N_+(t)$ и $N_-(t)$ характеризуют накопленную силу покупателей и продавцов соответственно (при этом заметим, что снятие заявок на стороне продавцов в данном случае увеличивает силу покупателей и наоборот) и являются условно независимыми при фиксированном потоке информации, поступающем извне за время $[0, t]$.

Теперь рассмотрим процесс обобщенной цены $Q(t)$, приращение которого на интервале $[0, t]$ имеет вид:

$$Q(t) - Q(0) = \sum_{j=1}^{N(t)} X_j,$$

где X_1, X_2, \dots — независимые одинаково распределенные величины, такие что

$$X_j = \begin{cases} +1 & \text{с вероятностью } \frac{\lambda_+}{\lambda_+ + \lambda_-}; \\ -1 & \text{с вероятностью } \frac{\lambda_-}{\lambda_+ + \lambda_-}, \end{cases} \quad j = 1, 2, \dots,$$

причем случайные величины X_1, X_2, \dots стохастически независимы от процесса $N(t)$ (в этом можно убедиться, непосредственно выписав характеристическую функцию случайной величины $N(t)$). При этом

$$EX_j = \frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-};$$

$$DX_j = 1 - \left(\frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-} \right)^2 = \frac{4\lambda_+\lambda_-}{(\lambda_+ + \lambda_-)^2},$$

так что

$$E \sum_{j=1}^{N(t)} X_j = t(\lambda_+ - \lambda_-);$$

$$D \sum_{j=1}^{N(t)} X_j = t(\lambda_+ + \lambda_-) \left[\left(\frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-} \right)^2 + \frac{4\lambda_+\lambda_-}{(\lambda_+ + \lambda_-)^2} \right] = t(\lambda_+ + \lambda_-).$$

В дальнейшем без ограничения общности будем считать, что $Q(0) = 0$. Для удобства временно будем считать, что $t = 1$ (т.е. рассматривается приращение обобщенной цены за единицу времени).

Если $\lambda = \lambda_+ + \lambda_-$ очень велико, т.е. в единицу времени происходит очень много информативных событий, то по центральной предельной теореме для пуассоновских случайных сумм справедливо приближенное соотношение:

$$P(Q(1) < x) \approx \Phi \left(\frac{x - \lambda_+ + \lambda_-}{\sqrt{\lambda_+ + \lambda_-}} \right), \quad x \in \mathbb{R}, \quad (1)$$

где $\Phi(x)$ — стандартная нормальная функция распределения. При этом, используя результаты работы [9], можно выписать довольно аккуратные оценки точности приближения (1).

Теперь вспомним, что выше внешний поток информации считался фиксированным. Это предположение, в частности, широко используется в большинстве работ по моделированию динамики книги заявок и дает возможность использовать аппарат марковских цепей с непрерывным временем, для которых условие марковости в определенном смысле эквивалентно тому, что распределение вероятностей интервалов времени между информативными событиями является экспоненциальным. В реальной практике это условие не выполняется, как видно из рис. 1. На этом рисунке приведены гистограмма интервалов времени между событиями, произошедшими в течение всего рабочего дня, и график плотности гамма-распределения с параметром формы 0,2637 и параметром масштаба 1,2421. Это распределение хорошо согласуется с гистограммой и заметно отличается от экспоненциального.

С другой стороны, хорошее согласие распределения вероятностей интервалов времени между событиями, заметное на рис. 1, с указанным выше гамма-распределением подтверждает правильность рассуждений об условной марковости рассматриваемых процессов, поскольку, как известно, получение безусловного распределения из условного сводится к смешиванию условного распределения

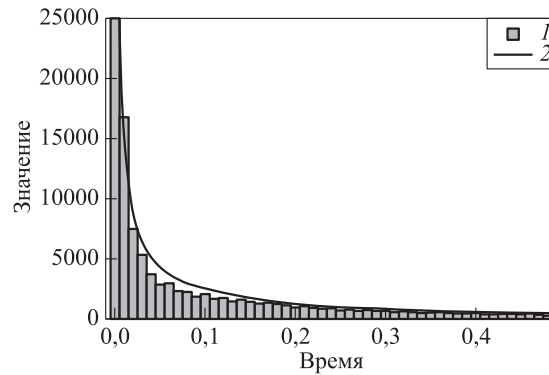


Рис. 1 Гистограмма и распределение вероятностей интервалов времени между информативными событиями: 1 — данные; 2 — гамма-распределение (0,2637, 1,2421)

по распределению вероятностей, соответствующему закону распределения параметра, описывающего фиксированное условие. В то же время гамма-распределение может быть представлено в виде смеси экспоненциальных распределений, только если его параметр формы не превосходит единицы (см. [10]). В той же работе показано, что если параметр формы r гамма-распределения, соответствующего плотности $g(x; r, \mu)$, удовлетворяет условию $0 < r \leq 1$, то плотность $g(x; r, \mu)$ может быть представлена в виде:

$$g(x; r, \mu) = \int_0^{\infty} p_{\mu}(z) z e^{-zx} dz,$$

где $p_{\mu}(z)$ — плотность распределения Снедекора–Фишера:

$$p_{\mu}(z) = \frac{(z - \mu)^{-r} \mu^r}{z \Gamma(1 - r) \Gamma(r)} \mathbb{I}(\mu \leq z).$$

В работе [11] можно найти также результаты статистического анализа эволюции параметра μ модели в течение дня:

$$P(Q(1) < x) \approx \int_{\mathbb{R}^+ \times \mathbb{R}^+} \Phi \left(\frac{x - \lambda_+ + \lambda_-}{\sqrt{\lambda_+ + \lambda_-}} \right) dP(\Lambda_+ < \lambda_+, \Lambda_- < \lambda_-), \quad x \in \mathbb{R}. \quad (2)$$

Тем не менее, рабочий день, за который накапливались исходные данные для рис. 1, — это слишком большой интервал времени, так что практическая ценность этой модели сродни ценности информации о «средней температуре по больнице».

Поэтому для получения более тонких моделей безусловного распределения величины $Q(1)$, в силу непредсказуемости потока внешней информации, следует считать, что λ_+ и λ_- — это некоторые конкретные значения случайных величин Λ_+ и Λ_- . Таким образом, для безусловного распределения приращения $Q(1)$ «форсированно» получается модель (2). Эту модель можно статистически исследовать методом скользящего разделения смесей (СРС-методом), используя конечные аппроксимации для смеси (2)

$$P(Q(1) < x) \approx \sum_{j=1}^k p_j \Phi\left(\frac{x - a_j}{\sigma_j}\right) \quad (3)$$

и оценивая параметры $k, p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k$ модели (3) (см., например, книгу [1], где довольно подробно изложен и сам СРС-метод, и его применение к декомпозиции волатильности). В этом заключается первый, «квазипараметрический» способ анализа. Этот способ непараметрический, потому что методически аналогичен непараметрическим процедурам ядерного оценивания распределений. Он квазипараметрический, потому что выбор нормальных ядер здесь форсирован и обусловлен центральной предельной теоремой для пуассоновских случайных сумм. Но как любой непараметрический метод, этот метод плох тем, что годится только для ретроспективного анализа. Для перспективного анализа (например, прогнозирования) намного удобнее параметрические модели, к построению которых и перейдем.

Предположим, что

$$\Lambda_+ = \Lambda\alpha_+; \quad \Lambda_- = \Lambda\alpha_-,$$

где Λ — неотрицательная случайная величина, имеющая смысл внешнего новостного фона на бирже, а α_+ и α_- — параметры, описывающие тенденции торгов (пока для простоты изложения будем считать параметры α_+ и α_- неслучайными). Тогда модель (2) примет вид

$$P(Q(1) < x) \approx \int_0^\infty \Phi\left(\frac{x - \lambda(\alpha_+ - \alpha_-)}{\sqrt{\lambda(\alpha_+ + \alpha_-)}}\right) dP(\Lambda < \lambda), \quad x \in \mathbb{R}. \quad (4)$$

Заметим, что модель (4) — это хорошо известная дисперсионно-сдвиговая смесь (variance-mean mixture) нормальных законов, в которой смешивание производится как бы и по параметру сдвига, и по параметру масштаба, но фактически смесь является однопараметрической. К такому типу смесей относятся, в частности, обобщенные гиперболические законы, включая дисперсионные гамма-распределения (variance gamma distributions), скошенные

распределения Стьюдента, нормальные\ обратные гауссовские распределения, некоторые устойчивые законы, а также многие другие. Методы исследования и использования таких моделей хорошо известны.

Задача исследования смесей типа (4) также может быть решена непараметрическими методами. Важность этих методов подчеркивается нестационарным характером потока новостей и, стало быть, зависимостью параметров модели (4) от астрономического времени (т. е. от фактического положения окна). По аналогии с моделью (3) можно предложить квазипараметрический подход к приближенному оцениванию параметров модели (4).

Чтобы модель (4) можно было использовать для перспективного анализа, ее нужно настроить, т. е. определить распределение случайной величины Λ . С этой целью можно применить уже опробованный и продемонстрировавший хорошие результаты метод аппроксимации к распределению длительностей интервалов между событиями потока изменения цен модели типа конечной смеси гамма-распределения. А именно: если T — случайная величина, равная длине интервала времени между изменениями цены, то ее распределение хорошо согласуется с моделью типа

$$\frac{d}{dx} P(T < x) \approx \sum_{j=1}^k p_j g(x; \theta_j, \mu_j), \quad x \geq 0, \quad (5)$$

где $p_j \geq 0, j = 1, \dots, k, p_1 + \dots + p_k = 1; g(x; \theta, \mu)$ — плотность гамма-распределения с параметром формы $\theta > 0$ и параметром масштаба $\mu > 0$:

$$g(x; \theta, \mu) = \frac{\mu^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\mu x}, \quad x \geq 0;$$

$\Gamma(\theta)$ — эйлерова гамма-функция. Если T — длительность интервала времени между событиями потока, то интенсивность потока — величина, обратно пропорциональная средней длительности. Модель (5) допускает следующую интерпретацию. Имеется k потоков событий, каждый из которых соответствует своей компоненте в модели (5). Всякий раз случайно в соответствии с вероятностями p_j выбирается один из потоков и реализуется случайная величина, распределение которой является соответствующей компонентой модели (5). Как известно, среднее значение величины, имеющей гамма-плотность $g(x; \theta, \mu)$, равно θ/μ . Следовательно, интенсивность соответствующего потока равна μ/θ . Таким образом, в рамках модели (5) распределение случайной интенсивности Λ (см. модель (3)) имеет вид:

$$P\left(\Lambda = \frac{\mu_j}{\theta_j}\right) = p_j, \quad j = 1, \dots, k.$$

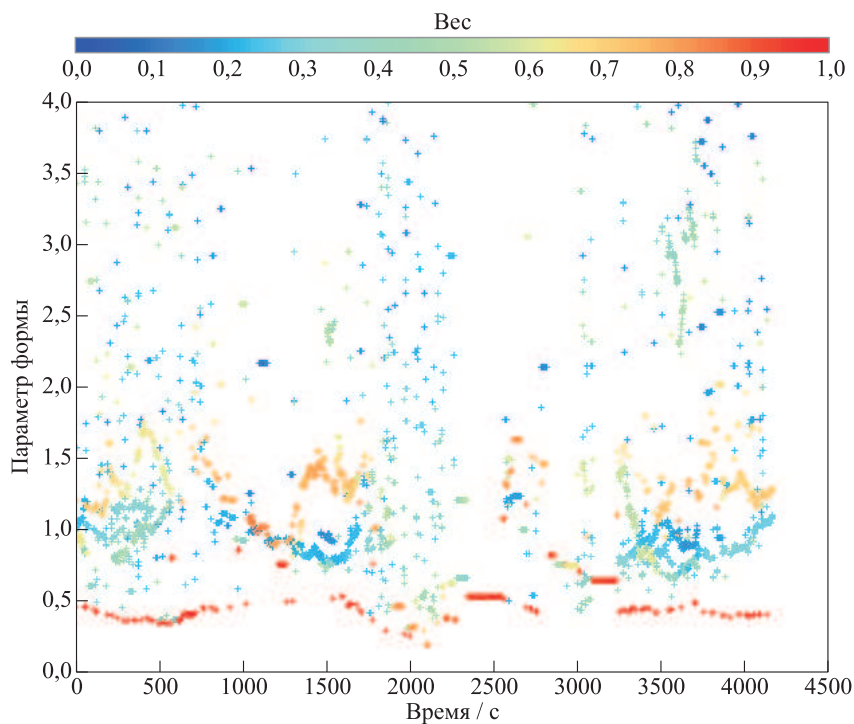


Рис. 2 Эволюция весов и параметров формы компонент модели (5)

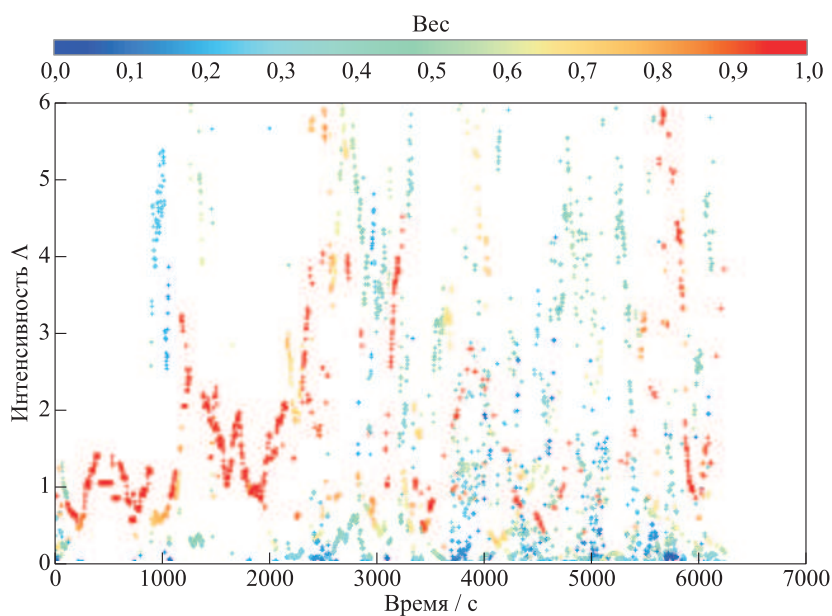


Рис. 3 Эволюция дискретного распределения интенсивности Λ

Это распределение (и его эволюцию во времени при скольжении окна) несложно построить, имея статистически оцененные параметры модели (5).

Примеры применения такого метода приближенного восстановления распределения случайной величины Λ содержатся на рис. 2 и 3.

4 Исследование нестационарности распределения случайной интенсивности

В данном разделе модель, принципиальное устройство которой описано выше, будет адаптирована с учетом нестационарного и стохастического характера интенсивности внешнего информационного потока Λ и параметров α_+ и α_- , описывающих степень реакции на него покупателей и продавцов.

В действительности интенсивности потоков заявок являются нестационарными, поскольку внешний поток информации, определяющий интенсивность этих событий, сам по себе является нестационарным. В таком случае интенсивности потоков заявок разных типов, во-первых, не могут являться независимыми, а во-вторых, определенным образом зависят от некоторого случайного процесса $\Lambda(t)$, определяющего внешний новостной фон. Поэтому можно определить следующие соотношения для мгновенных интенсивностей:

$$\mu^\pm = \mu^\pm(t) = \alpha^{M^\pm}(t)\lambda(t) \text{ (рыночные заявки);}$$

$$\lambda_i^\pm = \lambda_i^\pm(t) = \alpha_i^{L^\pm}(t)\lambda(t) \text{ (лимитные заявки);}$$

$$\theta_i^\pm = \theta_i^\pm(t) = \alpha_i^{C^\pm}(t)\lambda(t) \text{ (заявки на отмену),}$$

где $\lambda(t)$ — мгновенная интенсивность случайного процесса $\Lambda(t)$, определяющего внешний информационный фон (ажиотаж), $\alpha^{M^\pm}(t)$, $\alpha_i^{L^\pm}(t)$, $\alpha_i^{C^\pm}(t)$ — также случайные процессы, характеризующие степень реакции на информационный фон при выставлении заявок данного типа.

В таком случае $N_+(t)$ и $N_-(t)$ являются неоднородными пуассоновскими процессами с мгновенными интенсивностями

$$\lambda_+(t) = \alpha_+(t)\lambda(t); \quad (6)$$

$$\lambda_-(t) = \alpha_-(t)\lambda(t), \quad (7)$$

где

$$\alpha_+(t) = \alpha^{M^+}(t) + \sum_{i=1}^M \alpha_i^{L^+}(t) + \sum_{i=1}^M \alpha_i^{C^-}(t);$$

$$\alpha_-(t) = \alpha^{M^-}(t) + \sum_{i=1}^M \alpha_i^{L^-}(t) + \sum_{i=1}^M \alpha_i^{C^+}(t).$$

При этом по-прежнему обобщенный процесс цены имеет вид:

$$Q(t) = \sum_{i=1}^{N(t)} X_i,$$

где X_i — независимые случайные величины, заданные следующим образом:

$$\begin{aligned} P(X_i = 1) &= \frac{\lambda_+(T_i)}{\lambda_+(T_i) + \lambda_-(T_i)} = \\ &= \frac{\alpha_+(T_i)\lambda(T_i)}{\alpha_+(T_i)\lambda(T_i) + \alpha_-(T_i)\lambda(T_i)} = \\ &= \frac{\alpha_+(T_i)}{\alpha_+(T_i) + \alpha_-(T_i)}; \quad (8) \end{aligned}$$

$$\begin{aligned} P(X_i = -1) &= \frac{\lambda_-(T_i)}{\lambda_+(T_i) + \lambda_-(T_i)} = \\ &= \frac{\alpha_-(T_i)\lambda(T_i)}{\alpha_+(T_i)\lambda(T_i) + \alpha_-(T_i)\lambda(T_i)} = \\ &= \frac{\alpha_-(T_i)}{\alpha_+(T_i) + \alpha_-(T_i)}; \quad (9) \end{aligned}$$

T_i — последовательные моменты скачков процесса $N(t)$.

Таким образом, статистические свойства процесса $Q(t)$ определяются свойствами процессов $\alpha_+(t)$, $\alpha_-(t)$ и $\Lambda(t)$ (характеризующего моменты T_i скачков процесса $N(t)$).

Также следует отметить, что мультипликативные представления мгновенных интенсивностей (6) и (7) хорошо согласуются с реальными данными и дают возможность написать очень важное соотношение:

$$r(t) = \frac{\lambda_+(t)}{\lambda_-(t)} = \frac{\alpha_+(t)}{\alpha_-(t)},$$

что избавляет в дальнейшем от необходимости моделирования внешнего информационного потока $\Lambda(t)$, которое практически нереализуемо с достаточной степенью адекватности в силу его непредсказуемости и при этом дает возможность на основе наблюдаемых значений процесса $r(t) = \lambda_+(t)/\lambda_-(t)$ исследовать процесс относительной реакции покупателей и продавцов на новостной фон $\alpha_+(t)/\alpha_-(t)$, являющийся не чем иным, как мерой их дисбаланса — основным механизмом ценообразования.

5 Анализ реальных данных

Для тестирования некоторых из вышеизложенных концепций были выбраны высокочастотные данные для самого ликвидного инструмента фьючерсного рынка биржи ММВБ-РТС — фьючерса на индекс РТС. Биржа распространяет информацию о полном потоке обезличенных заявок участников рынка, что, в частности, позволяет провести анализ процессов $N_+(t)$ и $N_-(t)$ в рамках описанной модели обобщенного процесса цены.

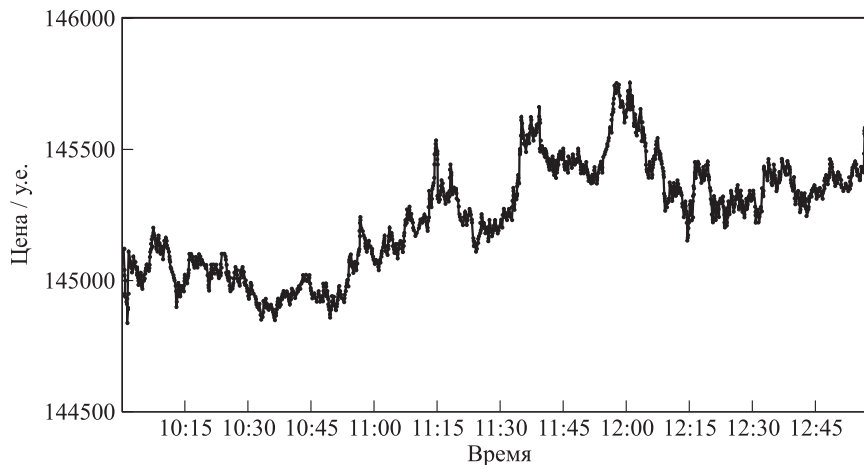
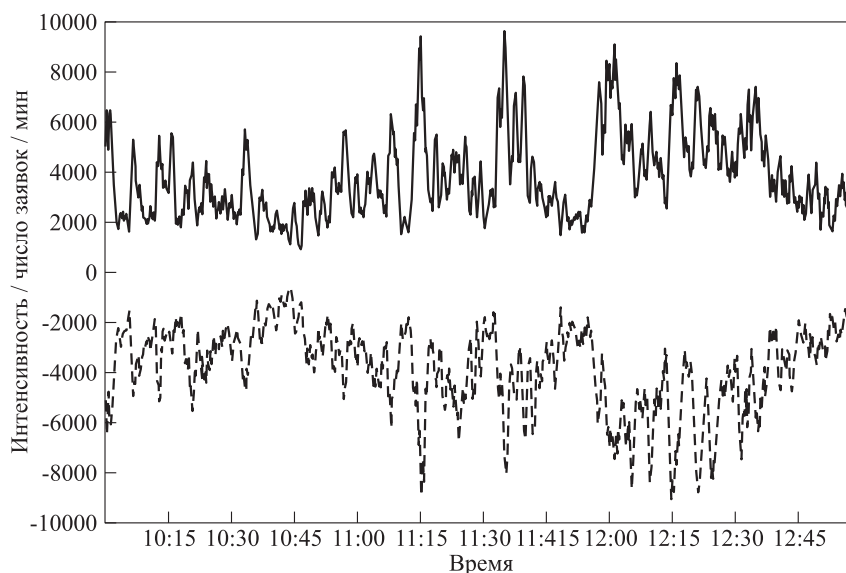


Рис. 4 График цены

Рис. 5 Графики мгновенных интенсивностей потоков заявок покупателей $\lambda_+(t)$ (сплошная линия) и продавцов $-\lambda_-(t)$ (пунктирная линия), размер скользящего окна $\Delta t = 60$ с

Для анализа были выбраны данные за первые три часа торгов данным инструментом за дневную сессию 11 ноября 2012 г.; график цены, построенный по сделкам, изображен на рис. 4.

На рис. 5 представлены графики мгновенных интенсивностей процессов $N_+(t)$ и $N_-(t)$ (количество заявок каждого типа за последнюю минуту). Данный рисунок хорошо подтверждает мультипликативное представление мгновенных интенсивностей в виде $\lambda_+(t) = \alpha_+(t)\lambda(t)$ и $\lambda_-(t) = \alpha_-(t)\lambda(t)$.

На рис. 6 изображены графики тех же процессов, но совмещенные в положительной полуплоскости. Наблюдаемые расхождения графиков означают локальное преобладание покупателей над продавцами ($\lambda_+(t) > \lambda_-(t)$) или продавцов над покупателями

($\lambda_-(t) > \lambda_+(t)$), согласованные же их падение или рост соответствуют общему падению или росту интенсивности торгов без особенной борьбы между покупателями и продавцами и, как следствие, без колебаний цены.

На рис. 7 изображен график процесса $r(t)$, из которого хорошо видно преобладание покупателей над продавцами, что отразилось в поведении цены на протяжении наблюдаемого периода. Другой особенностью графика является наличие уровня поддержки $r = 0,6$, что может означать наличие крупного покупателя, который сдерживал натиск продавцов при достижении данного уровня дисбаланса сил (это также согласуется с поведением цены в эти моменты времени).

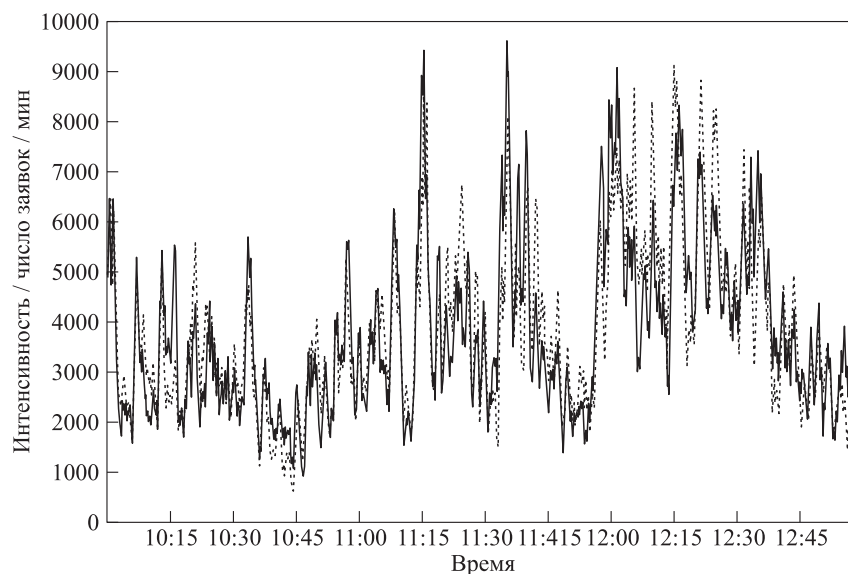


Рис. 6 Графики мгновенных интенсивностей потоков заявок покупателей $\lambda_+(t)$ (сплошная линия) и продавцов $\lambda_-(t)$ (пунктирная линия), размер скользящего окна $\Delta t = 60$ с

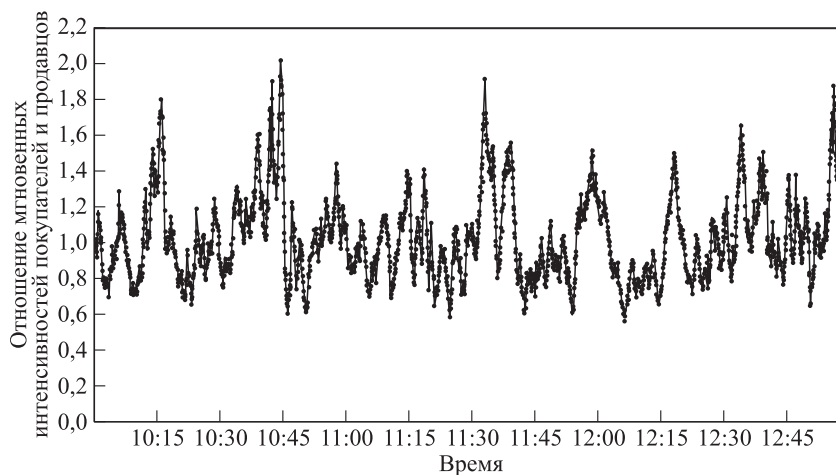


Рис. 7 Отношение мгновенных интенсивностей покупателей и продавцов $r(t) = \lambda_+(t)/\lambda_-(t)$, размер скользящего окна $\Delta t = 60$ с

6 Заключение

В данной статье была построена микроструктурная модель, описывающая информационные потоки в сложных финансовых системах, и смоделирована случайная природа интенсивностей потоков заявок, определяющих механизм ценообразования финансовых инструментов. При их моделировании отдельно был рассмотрен поток внешнего информационного фона со случайной интенсивностью и произведена его аппроксимация.

Также была предложена модель обобщенного процесса цены, учитывающая всю доступную информацию о потоках заявок и допускающая даль-

нейшую аналитическую интерпретацию. Важным результатом являются также предложенные мультипликативные представления интенсивностей потоков заявок (6) и (7), находящие подтверждение на реальных данных и позволяющие анализировать характеристики, связанные с интенсивностями потоков заявок, а также мгновенное соотношение сил покупателей и продавцов без моделирования внешнего информационного фона, практически не поддающегося прогнозированию.

В заключение авторы статьи выражают искреннюю признательность профессору Чикагского университета Юрию Георгиевичу Баласанову за полезное обсуждение вопросов, затронутых в статье.

Литература

1. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011.
2. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011. 620 с.
3. *Cont R., Stoikov S., Talreja R.* A stochastic model for order book dynamics // *Operations Res.*, 2010. Vol. 58. No. 3. P. 549–563.
4. *Cont R., de Larrard A.* Price dynamics in a Markovian limit order market. Working Paper, 2011. Available: <http://ssrn.com/abstract=1735338>.
5. *Bouchaud J.-P., Mezard M., Potters M.* Statistical properties of stock order books: Empirical results and models // *Quantitative Finance*, 2002. Vol. 2. P. 251–256.
6. *Zovko I., Farmer J. D.* The power of patience; A behavioral regularity in limit order placement // *Quantitative Finance*, 2002. Vol. 2. P. 387–392.
7. *Huang H., Kercheval A. N.* A generalized birth–death stochastic model for high-frequency order book dynamics // *Quantitative Finance*, 2012. Vol. 12. No. 4. P. 547–557.
8. *Abergel F., Jedidi A.* A mathematical approach to order book modelling // *Econophysics of order-driven markets* / Eds. F. Abergel, B. K. Chakrabarti, A. Chakraborti, M. Mitra. — New York: Springer, 2011. P. 93–108.
9. *Korolev V., Shevtsova I.* An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // *Scandinavian Actuarial J.*, 2012. No. 2. P. 81–105. Available online since June 4, 2010.
10. *Gleser L. J.* The gamma distribution as a mixture of exponential distributions: Technical Report # 87-28. — West Lafayette: Purdue University, 1987. 6 p.
11. *Gorshenin A., Doynikov A., Korolev V., Kuzmin V.* Statistical properties of the dynamics of order books: Empirical results // *Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems: Abstracts of 4th International Workshop.* — Moscow: IPI RAS, 2012. P. 31–51.

СТАЦИОНАРНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ ОБСЛУЖИВАНИЯ С ИНВЕРСИОННЫМ ПОРЯДКОМ ОБСЛУЖИВАНИЯ, ВЕРОЯТНОСТНЫМ ПРИОРИТЕТОМ И ГИСТЕРЕЗИСНОЙ ПОЛИТИКОЙ*

Т. А. Милованова¹, А. В. Печинкин²

Аннотация: Рассматривается однолинейная система массового обслуживания (СМО) с инверсионным порядком обслуживания, вероятностным приоритетом и простейшим вариантом гистерезисной политики. Найдены основные стационарные показатели функционирования этой системы.

Ключевые слова: система массового обслуживания; инверсионный порядок обслуживания; вероятностный приоритет; гистерезисная политика

1 Введение

Одним из важнейших направлений исследований в теории массового обслуживания является изучение СМО с дисциплинами обслуживания, отличными от обслуживания заявок в порядке поступления, поскольку такие дисциплины часто позволяют практически без каких-либо усовершенствований повысить качество функционирования самых разнообразных технических систем, например информационно-телекоммуникационных систем (ИТС). В частности, дисциплиной такого рода является инверсионный порядок обслуживания с вероятностным приоритетом, введенный в [1] для решения задачи А. Д. Соловьева об оптимальных распределениях для некоторых типов дисциплин обслуживания. Подробное изложение полученных в этом направлении результатов можно найти в [2].

В последнее время значительное внимание уделяется также СМО с гистерезисным управлением, являющимся одним из возможных механизмов предотвращения различного рода перегрузок в ИТС (см., например, [3]). Разновидности гистерезисной политики используются при обнаружении перегрузок как в сетях общеканальной системы сигнализации № 7, так и в сетях, где основой сигнализации является протокол инициации сеансов связи.

В настоящей работе делается попытка связать эти два направления исследования с помощью СМО с инверсионным порядком обслуживания,

вероятностным приоритетом и простейшим вариантом гистерезисной политики, для которой находятся основные стационарные показатели функционирования. Отметим, что некоторые типы системы $M/G/1$ с простейшим вариантом гистерезисной политики при дисциплине обслуживания заявок в порядке поступления изучались в [4–7].

2 Описание системы

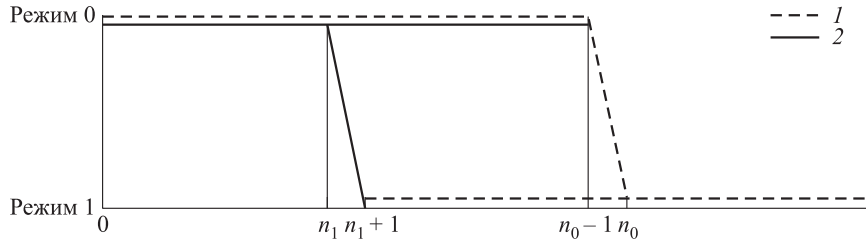
Рассмотрим однолинейную СМО с накопителем бесконечной емкости, инверсионным порядком обслуживания, вероятностным приоритетом и простейшим вариантом гистерезисной политики. Опишем функционирование этой СМО.

Вариант гистерезисной политики заключается в следующем (см. рисунок). Имеется два порога n_0 и n_1 , причем $n_1 < n_0$. Пока число заявок в системе меньше n_0 , система функционирует в режиме 0. Это означает, что заявки поступают с интенсивностью λ_0 и имеют длину, распределенную по закону $B_0(x)$ с плотностью $b_0(x) = B'_0(x)$ и средним значением $\beta_0 = \int_0^{\infty} x b_0(x) dx < \infty$. Но как только число заявок в системе становится равным n_0 , система переходит в режим 1. В этом режиме заявки поступают с интенсивностью λ_1 и имеют длину, распределенную по закону $B_1(x)$ с плотностью $b_1(x) = B'_1(x)$ и средним значением $\beta_1 = \int_0^{\infty} x b_1(x) dx < \infty$. Так продолжа-

* Работа выполнена при поддержке РФФИ (проекты № 11-07-00112 и № 12-07-00108).

¹ Российский университет дружбы народов, tmiivanova77@mail.ru

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



Схематическое изображение функционирования системы: 1 — поступление; 2 — обслуживание

ется до тех пор, пока число заявок в системе не станет равным n_1 . Тогда система снова переходит в режим 0 и т. д.

В системе также реализован инверсионный порядок обслуживания с вероятностным приоритетом. Предполагается, что в любой момент времени известны (остаточные) длины всех заявок в системе. В момент поступления в систему новой заявки ее длина x сравнивается с (остаточной) длиной y заявки на приборе. При этом если система функционирует в режиме 0, то с вероятностью $d_0(x, y)$ на прибор становится вновь поступившая заявка, а находившаяся ранее на приборе занимает первое место в очереди, и наоборот, с вероятностью $\bar{d}_0(x, y) = 1 - d_0(x, y)$ старая заявка продолжает обслуживаться, а новая становится на первое место в очереди. Если же система функционирует в режиме 1, то вероятность постановки на прибор вновь поступившей заявки равна $d_1(x, y)$, а на первое место в очереди — $\bar{d}_1(x, y) = 1 - d_1(x, y)$.

Будем предполагать, что выполнено условие $\lambda b_1 < 1$, необходимое и достаточное для существования стационарного режима функционирования рассматриваемой системы.

Будем считать также, что $n_0 - n_1 \geq 2$. Это предположение вводится только для того, чтобы не рассматривать случаи, которые по записи расчетных формул несколько отличаются от общего вида, и несколько не умаляет общности полученных результатов.

3 Вспомогательные функции

Пусть в некоторый момент система функционирует в режиме 0, в системе находится n , $n_1 < n < n_0$, заявок и в этот момент поступает в систему и становится на прибор новая заявка длины x . Обозначим через $\alpha_n(x)$ вероятность того, что в тот момент, когда в системе впервые снова останется n заявок, она по-прежнему будет пребывать в режиме 0.

Функции $\alpha_n(x)$, $n_1 < n < n_0$, удовлетворяют системе уравнений

$$\alpha_{n_0-1}(x) \equiv 0; \quad (1)$$

$$\begin{aligned} \alpha_n(x) = e^{-\lambda_0 x} + \int_0^x \lambda_0 e^{-\lambda_0 y} dy \times \\ \times \int_0^\infty b_0(z) [d_0(z, x-y)\alpha_{n+1}(z)\alpha_n(x-y) + \\ + \bar{d}_0(z, x-y)\alpha_{n+1}(x-y)\alpha_n(z)] dz, \\ n = \overline{n_1 + 1, n_0 - 2}. \quad (2) \end{aligned}$$

Система уравнений (1), (2) решается последовательно, начиная с $n = n_0 - 1$ и кончая $n = n_1 + 1$.

При решении уравнения (2) удобно привести его к более простому виду. Вводя обозначение

$$a_n(x) = e^{\lambda_0 x} \alpha_n(x), \quad n = \overline{n_1 + 1, n_0 - 2},$$

и производя тривиальные преобразования, получаем из (2):

$$\begin{aligned} a_n(x) = 1 + \\ + \int_0^x \left(\lambda_0 \int_0^\infty b_0(z) d_0(z, y) \alpha_{n+1}(z) dz \right) a_n(y) dy + \\ + \int_0^\infty \left(\lambda_0 b_0(y) e^{-\lambda_0 y} \int_0^x e^{\lambda_0 z} \bar{d}_0(y, z) \alpha_{n+1}(z) dz \right) \times \\ \times a_n(y) dy, \quad n = \overline{n_1 + 1, n_0 - 2}. \quad (3) \end{aligned}$$

Последнее соотношение представляет собой интегральное уравнение

$$a_n(x) = 1 + \int_0^\infty K_n(x, y) a_n(y) dy, \quad n = \overline{n_1 + 1, n_0 - 2}, \quad (4)$$

ядро которого имеет вид:

$$K_n(x, y) = \begin{cases} \lambda_0 \left(\int_0^\infty b_0(z) d_0(z, y) \alpha_{n+1}(z) dz + b_0(y) e^{-\lambda_0 y} \times \right. \\ \left. \times \int_0^x e^{\lambda_0 z} \bar{d}_0(y, z) \alpha_{n+1}(z) dz \right), & y < x; \\ \lambda_0 b_0(y) e^{-\lambda_0 y} \int_0^x e^{\lambda_0 z} \bar{d}_0(y, z) \alpha_{n+1}(z) dz, & y > x. \end{cases}$$

Численное решение уравнения (4) можно произвести итерационным методом. При этом в качестве нулевой итерации удобно выбрать тождественно равную нулю функцию. Тогда итерации будут возрастать, что позволит контролировать сходимость итерационного процесса.

В заключение этого раздела приведем условие на функцию $\bar{d}_0(x, y)$, при котором интегральное уравнение (3) можно свести к системе линейных алгебраических уравнений. А именно: будем предполагать, что

$$\bar{d}_0(x, y) = \sum_{i=1}^I \bar{d}_{0,i}^{(1)}(x) \bar{d}_{0,i}^{(2)}(y), \quad n = \overline{n_1 + 1, n_0 - 2}. \quad (5)$$

Тогда, вводя обозначения

$$c_n(y) = \lambda_0 \int_0^\infty b_0(z) d_0(z, y) \alpha_{n+1}(z) dz, \quad n = \overline{n_1 + 1, n_0 - 2};$$

$$c_{n,i}(x) = e^{\lambda_0 x} \bar{d}_{0,i}^{(2)}(x) \alpha_{n+1}(x), \quad n = \overline{n_1 + 1, n_0 - 2}, \quad i = \overline{1, I};$$

$$a_{n,i} = \int_0^\infty \lambda_0 b_0(y) e^{-\lambda_0 y} \bar{d}_{0,i}^{(1)}(y) a_n(y) dy, \quad n = \overline{n_1 + 1, n_0 - 2}, \quad i = \overline{1, I},$$

получаем из (3):

$$a_n(x) = 1 + \int_0^x c_n(y) a_n(y) dy + \sum_{i=1}^I a_{n,i} \int_0^x c_{n,i}(z) dz, \quad n = \overline{n_1 + 1, n_0 - 2}. \quad (6)$$

Дифференцируя теперь равенство (6), приходим к дифференциальному уравнению

$$a_n'(x) = c_n(x) a_n(x) + \sum_{i=1}^I a_{n,i} c_{n,i}(x), \quad n = \overline{n_1 + 1, n_0 - 2}, \quad (7)$$

начальное условие для которого задается выражением:

$$a_n(0) = 1, \quad n = \overline{n_1 + 1, n_0 - 2}. \quad (8)$$

Решение уравнения (7) с начальным условием (8) имеет вид:

$$a_n(x) = \left(1 + \sum_{i=1}^I a_{n,i} \int_0^x c_{n,i}(y) e^{-C_n(y)} dy \right) e^{C_n(x)}, \quad n = \overline{n_1 + 1, n_0 - 2}, \quad (9)$$

где

$$C_n(x) = \int_0^x c_n(y) dy, \quad n = \overline{n_1 + 1, n_0 - 2}.$$

Для того чтобы найти коэффициенты $a_{n,i}$, $i = \overline{1, I}$, умножим равенство (9) на $\lambda_0 b_0(x) e^{-\lambda_0 x} \bar{d}_{0,j}^{(1)}(x)$ и проинтегрируем в пределах от 0 до ∞ . Тогда

$$a_{n,j} = \int_0^\infty \lambda_0 b_0(x) e^{-\lambda_0 x} \bar{d}_{0,j}^{(1)}(x) e^{C_n(x)} dx + \sum_{i=1}^I a_{n,i} \int_0^\infty \lambda_0 b_0(x) e^{-\lambda_0 x} \bar{d}_{0,j}^{(1)}(x) e^{C_n(x)} dx \times \int_0^x c_{n,i}(y) e^{-C_n(y)} dy, \quad n = \overline{n_1 + 1, n_0 - 2}.$$

Производя эту процедуру при всех j , $j = \overline{1, I}$, получаем систему линейных алгебраических уравнений, решая которую, находим коэффициенты $a_{n,i}$ и соответственно функции $a_n(x)$ и $\alpha_n(x)$.

В дальнейшем будем пользоваться обозначением $\bar{\alpha}_n(x) = 1 - \alpha_n(x)$.

Отметим, что, используя приближение $\bar{d}_0(x, y)$ с помощью представления (5), можно найти функцию $a_n(x)$ с любой степенью точности. Однако повышение точности влечет за собой существенное увеличение числа I коэффициентов $a_{n,i}$ и, значит, размерности системы линейных алгебраических уравнений.

4 Стационарные вероятности состояний

Обозначим через p_0 стационарную вероятность того, что система свободна. При $n = \overline{1, n_1}$ или

$n \geq n_0$ обозначим через $p_n(x_1, \dots, x_n)$ стационарную плотность вероятностей того, что в системе находится n заявок, причем заявка на приборе имеет длину x_1 , первая заявка в очереди — длину x_2 и т.д. Наконец, при $n = \overline{n_1 + 1, n_0 - 1}$ через $p_n(0; x_1, \dots, x_n)$ обозначим стационарную плотность вероятностей того, что система функционирует в режиме 0 и в системе находится n заявок, причем заявка на приборе имеет длину x_1 , первая заявка в очереди — длину x_2 и т.д., а через $p_n(1; x_1, \dots, x_n)$ — аналогичную вероятность, но при этом система функционирует в режиме 1.

Используя метод исключения состояний (см., например, [8]), можно получить для $p_n(x_1, \dots, x_n)$, $n = \overline{1, n_1}$, уравнения

$$\begin{aligned}
 -p'_1(x) = & -\lambda_0 p_1(x) + \lambda_0 p_1(x) \int_0^\infty b_0(y) d_0(y, x) dy + \\
 & + \lambda_0 b_0(x) \int_0^\infty p_1(y) \bar{d}_0(x, y) dy + \lambda_0 p_0 b_0(x); \quad (10)
 \end{aligned}$$

$$\begin{aligned}
 -p'_n(x_1, \dots, x_n) = & -\lambda_0 p_n(x_1, \dots, x_n) + \\
 & + \lambda_0 p_n(x_1, \dots, x_n) \int_0^\infty b_0(y) d_0(y, x_1) dy + \\
 & + \lambda_0 b_0(x_1) \int_0^\infty p_n(y, x_2, \dots, x_n) \bar{d}_0(x_1, y) dy + \\
 & + \lambda_0 p_{n-1}(x_2, \dots, x_n) b_0(x_1) d_0(x_1, x_2) + \\
 & + \lambda_0 b_0(x_2) p_{n-1}(x_1, x_3, \dots, x_n) \bar{d}_0(x_2, x_1), \quad n = \overline{2, n_1},
 \end{aligned}$$

с начальным условием

$$\lim_{x \rightarrow \infty} p_n(x, x_2, \dots, x_n) = 0, \quad n = \overline{1, n_1}.$$

Можно выписать аналогичные уравнения для остальных функций $p_n(x_1, \dots, x_n)$, $n \geq n_0$, и $p_n(i; x_1, \dots, x_n)$, $n = \overline{n_1 + 1, n_0 - 1}$, $i = 1, 2$, но они ввиду громоздкости здесь не приводятся. Вычисления по этим формулам, хотя теоретически и можно производить на основе решения интегральных уравнений, практически не реализуемы уже при совсем небольших значениях n даже на современной вычислительной технике, поскольку размерность уравнений растет пропорционально n .

Однако для практических расчетов, как правило, достаточно знать только маргинальные стационарные плотности $p_1(x)$,

$$\begin{aligned}
 p_n(x) = & \int_0^\infty \dots \int_0^\infty p_n(x, x_2, \dots, x_n) dx_2 \dots dx_n, \\
 & n = \overline{2, n_1} \text{ или } n \geq n_0,
 \end{aligned}$$

и

$$\begin{aligned}
 p_n(i; x) = & \int_0^\infty \dots \int_0^\infty p_n(i; x, x_2, \dots, x_n) dx_2 \dots dx_n, \\
 & i = 0, 1, \quad n = \overline{n_1 + 1, n_0 - 1}.
 \end{aligned}$$

Для них справедливы соотношения

$$\begin{aligned}
 -p'_n(x) = & -f_0(x) p_n(x) + \int_0^\infty k_0(x, y) p_n(y) dy + \\
 & + g_{0,n}(x), \quad n = \overline{2, n_1}; \quad (11)
 \end{aligned}$$

$$\begin{aligned}
 -p'_n(0; x) = & -f_{0,n}(x) p_n(0, x) + \\
 & + \int_0^\infty k_{0,n}(x, y) p_n(0, y) dy + g_{0,n}(x), \\
 & n = \overline{n_1 + 1, n_0 - 1}; \quad (12)
 \end{aligned}$$

$$\begin{aligned}
 -p'_n(1; x) = & -f_1(x) p_n(1, x) + \\
 & + \int_0^\infty k_1(x, y) p_n(1, y) dy + g_{1,n}(x), \\
 & n = \overline{n_1 + 1, n_0 - 1}; \quad (13)
 \end{aligned}$$

$$\begin{aligned}
 -p'_n(x) = & -f_1(x) p_n(x) + \\
 & + \int_0^\infty k_1(x, y) p_n(y) dy + g_{1,n}(x), \quad n \geq n_0, \quad (14)
 \end{aligned}$$

с начальными условиями

$$\lim_{x \rightarrow \infty} p_n(x) = 0, \quad n = \overline{2, n_1} \text{ или } n \geq n_0, \quad (15)$$

$$\lim_{x \rightarrow \infty} p_n(i; x) = 0, \quad n = \overline{n_1 + 1, n_0 - 1}, \quad i = 0, 1, \quad (16)$$

в которых для сокращения записи введены следующие обозначения:

$$\begin{aligned}
 f_0(x) = & \lambda_0 \left(1 - \int_0^\infty b_0(y) d_0(y, x) dy \right); \\
 k_0(x, y) = & \lambda_0 b_0(x) \bar{d}_0(x, y); \\
 g_{0,1}(x) = & \lambda_0 p_0 b_0(x); \\
 g_{0,n}(x) = & \lambda_0 b_0(x) \int_0^\infty p_{n-1}(y) d_0(x, y) dy + \\
 & + \lambda_0 p_{n-1}(x) \int_0^\infty \bar{d}_0(y, x) b_0(y) dy, \quad n = \overline{2, n_1};
 \end{aligned}$$

$$f_{0,n}(x) = \lambda_0 \left(1 - \int_0^\infty b_0(y) d_0(y, x) \alpha_n(y) dy \right),$$

$$n = \overline{n_1 + 1, n_0 - 1};$$

$$k_{0,n}(x, y) = \lambda_0 b_0(x) \bar{d}_0(x, y) \alpha_n(y),$$

$$n = \overline{n_1 + 1, n_0 - 1};$$

$$g_{0,n_1+1}(x) = \lambda_0 b_0(x) \int_0^\infty p_{n_1}(y) d_0(x, y) dy +$$

$$+ \lambda_0 p_{n_1}(x) \int_0^\infty b_0(y) \bar{d}_0(y, x) dy;$$

$$g_{0,n}(x) = \lambda_0 b_0(x) \int_0^\infty p_{n-1}(0; y) d_0(x, y) dy +$$

$$+ \lambda_0 p_{n-1}(0; x) \int_0^\infty b_0(y) \bar{d}_0(y, x) dy,$$

$$n = \overline{n_1 + 2, n_0 - 1};$$

$$f_1(x) = \lambda_1 \left(1 - \int_0^\infty b_1(y) d_1(y, x) dy \right); \quad (17)$$

$$k_1(x, y) = \lambda_1 b_1(x) \bar{d}_1(x, y), \quad (18)$$

$$g_{1,n_1+1}(x) =$$

$$= \lambda_0 p_{n_1+1}(0; x) \int_0^\infty b_0(y) d_0(y, x) \bar{\alpha}_{n_1+1}(y) dy +$$

$$+ \lambda_0 b_0(x) \int_0^\infty p_{n_1+1}(0; y) \bar{d}_0(x, y) \bar{\alpha}_{n_1+1}(y) dy;$$

$$g_{1,n}(x) = \lambda_0 p_n(0; x) \int_0^\infty b_0(y) d_0(y, x) \bar{\alpha}_n(y) dy +$$

$$+ \lambda_0 b_0(x) \int_0^\infty p_n(0; y) \bar{d}_0(x, y) \bar{\alpha}_n(y) dy +$$

$$+ \lambda_1 b_1(x) \int_0^\infty p_{n-1}(1; y) d_1(x, y) dy +$$

$$+ \lambda_1 p_{n-1}(1; x) \int_0^\infty b_1(y) \bar{d}_1(y, x) dy,$$

$$n = \overline{n_1 + 2, n_0 - 1};$$

$$g_{1,n_0}(x) = \lambda_0 b_0(x) \int_0^\infty p_{n_0-1}(0; y) d_0(x, y) dy +$$

$$+ \lambda_0 p_{n_0-1}(0; x) \int_0^\infty b_0(y) \bar{d}_0(y, x) dy +$$

$$+ \lambda_1 b_1(x) \int_0^\infty p_{n_0-1}(1; y) d_1(x, y) dy +$$

$$+ \lambda_1 p_{n_0-1}(1; x) \int_0^\infty b_1(y) \bar{d}_1(y, x) dy;$$

$$g_{1,n}(x) = \lambda_1 b_1(x) \int_0^\infty p_{n-1}(y) d_1(x, y) dy +$$

$$+ \lambda_1 p_{n-1}(x) \int_0^\infty b_1(y) \bar{d}_1(y, x) dy. \quad n > n_0, \quad (19)$$

Вероятность p_0 вычисляется из условия нормировки

$$p_0 + \sum_{n=1}^{n_1} p_n + \sum_{n=n_1+1}^{n_0-1} [p_{n,0} + p_{n,1}] + \sum_{n=n_0}^{\infty} p_n = 1,$$

где $p_n = \int_0^\infty p_n(x) dx$, $n = \overline{1, n_1}$ или $n \geq n_0$, — стационарная вероятность того, что в системе находится n заявок, а $p_{n,i} = \int_0^\infty p_n(i; x) dx$, $n = \overline{n_1 + 1, n_0 - 1}$, $i = 0, 1$, — стационарная вероятность того, что система функционирует в режиме i и в системе находится n заявок.

Уравнения (10)–(14) легко приводятся к интегральным. Действительно, вводя новые обозначения

$$F_0(x) = \int_0^x f_0(y) dy, \quad n = \overline{1, n_1};$$

$$F_{0,n}(x) = \int_0^x f_{0,n}(y) dy, \quad n = \overline{n_1 + 1, n_0 - 1};$$

$$F_1(x) = \int_0^x f_1(y) dy, \quad n \geq n_1 + 1;$$

$$p_n(x) = \pi_n(x) e^{F_0(x)}, \quad n = \overline{1, n_1};$$

$$p_n(0; x) = \pi_n(0; x) e^{F_{0,n}(x)}, \quad n = \overline{n_1 + 1, n_0 - 1};$$

$$p_n(1; x) = \pi_n(1; x) e^{F_1(x)}, \quad n = \overline{n_1 + 1, n_0 - 1};$$

$$p_n(x) = \pi_n(x) e^{F_1(x)}, \quad n \geq n_0,$$

из (10)–(14) получаем соотношения

$$\begin{aligned}
 -\pi'_n(x) &= e^{-F_0(x)} \int_0^\infty e^{F_0(y)} k_0(x, y) \pi_n(y) dy + \\
 &+ e^{-F_0(x)} g_{0,n}(x), \quad n = \overline{1, n_1}; \quad (20)
 \end{aligned}$$

$$\begin{aligned}
 -\pi'_n(0; x) &= \\
 &= e^{-F_{0,n}(x)} \int_0^\infty e^{F_{0,n}(y)} k_{0,n}(x, y) \pi_n(0; y) dy + \\
 &+ e^{-F_{0,n}(x)} g_{0,n}(x), \quad n = \overline{n_1 + 1, n_0 - 1}; \quad (21)
 \end{aligned}$$

$$\begin{aligned}
 -\pi'_n(1; x) &= e^{-F_1(x)} \int_0^\infty e^{F_1(y)} k_1(x, y) \pi_n(1; y) dy + \\
 &+ e^{-F_1(x)} g_{1,n}(x), \quad n = \overline{n_1 + 1, n_0 - 1}; \quad (22)
 \end{aligned}$$

$$\begin{aligned}
 -\pi'_n(x) &= e^{-F_1(x)} \int_0^\infty e^{F_1(y)} k_1(x, y) \pi_n(y) dy + \\
 &+ e^{-F_1(x)} g_{1,n}(x), \quad n \geq n_0, \quad (23)
 \end{aligned}$$

интегрируя которые в пределах от x до ∞ и учитывая начальные условия (15), (16), имеем

$$\begin{aligned}
 \pi_n(x) &= \\
 &= \int_0^\infty e^{F_0(y)} \left(\int_x^\infty e^{-F_0(u)} k_0(u, y) du \right) \pi_n(y) dy + \\
 &+ \int_x^\infty e^{-F_0(u)} g_{0,n}(u) du; \quad n = \overline{1, n_1}, \quad (24)
 \end{aligned}$$

$$\begin{aligned}
 \pi_n(0; x) &= \\
 &= \int_0^\infty e^{F_{0,n}(y)} \left(\int_x^\infty e^{-F_{0,n}(u)} k_{0,n}(u, y) du \right) \pi_n(0; y) dy + \\
 &+ \int_x^\infty e^{-F_{0,n}(u)} g_{0,n}(u) du, \quad n = \overline{n_1 + 1, n_0 - 1}; \quad (25)
 \end{aligned}$$

$$\begin{aligned}
 \pi_n(1; x) &= \\
 &= \int_0^\infty e^{F_1(y)} \left(\int_x^\infty e^{-F_1(u)} k_1(u, y) du \right) \pi_n(1; y) dy + \\
 &+ \int_x^\infty e^{-F_1(u)} g_{1,n}(u) du, \quad n = \overline{n_1 + 1, n_0 - 1}; \quad (26)
 \end{aligned}$$

$$\begin{aligned}
 \pi_n(x) &= \int_0^\infty e^{F_1(y)} \left(\int_x^\infty e^{-F_1(u)} k_1(u, y) du \right) \pi_n(y) dy + \\
 &+ \int_x^\infty e^{-F_1(u)} g_{1,n}(u) du, \quad n \geq n_0. \quad (27)
 \end{aligned}$$

Соотношения (24)–(27) являются интегральными уравнениями такого же вида, что и (4), и к ним применимы те же методы решения, что и к уравнению (4).

Так же как для функций $\alpha_n(x)$, приведем условия для функций $\bar{d}_0(x, y)$ и $\bar{d}_1(x, y)$, которые позволяют получить решения интегродифференциальных уравнений (20)–(23) с помощью приведения к системе линейных алгебраических уравнений. А именно: будем предполагать, что выполнены условия (5) и

$$\bar{d}_1(x, y) = \sum_{i=1}^{I_1} \bar{d}_{1,i}^{(1)}(x) \bar{d}_{1,i}^{(2)}(y). \quad (28)$$

Тогда, вводя обозначения

$$c_i(x) = \lambda_0 b_0(x) \bar{d}_{0,i}^{(1)}(x) e^{-F_0(x)}, \quad i = \overline{1, I};$$

$$q_{n,i} = \int_0^\infty e^{F_0(y)} \bar{d}_{0,i}^{(2)}(y) \pi_n(y) dy, \quad n = \overline{1, n_1}, \quad i = \overline{1, I};$$

$$q_n(x) = e^{-F_0(x)} g_{0,n}(x), \quad n = \overline{1, n_1};$$

$$\begin{aligned}
 c_{0;n,i}(x) &= \lambda_0 b_0(x) \bar{d}_{0,i}^{(1)}(x) e^{-F_{0,n}(x)}, \\
 &n = \overline{n_1 + 1, n_0 - 1}, \quad i = \overline{1, I};
 \end{aligned}$$

$$\begin{aligned}
 q_{0;n,i} &= \int_0^\infty e^{F_{0,n}(y)} \bar{d}_{0,i}^{(2)}(y) \alpha_n(y) \pi_n(0; y) dy, \\
 &n = \overline{n_1 + 1, n_0 - 1}, \quad i = \overline{1, I};
 \end{aligned}$$

$$q_{0;n}(x) = e^{-F_{0,n}(x)} g_{0,n}(x), \quad n = \overline{n_1 + 1, n_0 - 1};$$

$$c_{1;i}(x) = \lambda_1 b_1(x) \bar{d}_{1,i}^{(1)}(x) e^{-F_1(x)}, \quad i = \overline{1, I_1};$$

$$\begin{aligned}
 q_{1;n,i} &= \int_0^\infty e^{F_1(y)} \bar{d}_{1,i}^{(2)}(y) \pi_n(1; y) dy, \\
 &n = \overline{n_1 + 1, n_0 - 1}, \quad i = \overline{1, I_1};
 \end{aligned}$$

$$q_{1;n_0}(x) = e^{-F_1(x)} g_{1,n}(x), \quad n \geq n_1 + 1;$$

$$q_{1;n,i} = \int_0^\infty e^{F_1(y)} \bar{d}_{1,i}^{(2)}(y) \pi_n(y) dy, \quad n \geq n_0, \quad i = \overline{1, I_1},$$

получаем из (20)–(23) после интегрирования в пределах от x до ∞ с учетом начальных условий (15), (16):

$$\pi_n(x) = \sum_{i=1}^I C_i(x) q_{n,i} + Q_n(x), \quad n = \overline{1, n_1}; \quad (29)$$

$$\pi_n(0; x) = \sum_{i=1}^I C_{0;n,i}(x) q_{0;n,i} + Q_{0;n}(x),$$

$$n = \overline{n_1 + 1, n_0 - 1}; \quad (30)$$

$$\pi_n(1; x) = \sum_{i=1}^{I_1} C_{1;i}(x) q_{1;n,i} + Q_{1;n}(x),$$

$$n = \overline{n_1 + 1, n_0 - 1}; \quad (31)$$

$$\pi_n(x) = \sum_{i=1}^{I_1} C_{1;i}(x) q_{1;n,i} + Q_{1;n}(x), \quad n \geq n_0, \quad (32)$$

где

$$C_i(x) = \int_x^\infty c_i(y) dy, \quad n = \overline{1, n_1}, \quad i = \overline{1, I};$$

$$C_{0;n,i}(x) = \int_x^\infty c_{0;n,i}(y) dy,$$

$$n = \overline{n_1 + 1, n_0 - 1}, \quad i = \overline{1, I};$$

$$C_{1;i}(x) = \int_x^\infty c_{1,i}(y) dy, \quad n \geq n_1, \quad i = \overline{1, I_1};$$

$$Q_n(x) = \int_x^\infty q_n(y) dy, \quad n = \overline{1, n_1};$$

$$Q_{0;n}(x) = \int_x^\infty q_{0;n}(y) dy, \quad n = \overline{n_1 + 1, n_0 - 1};$$

$$Q_{1;n}(x) = \int_x^\infty q_{1;n}(y) dy, \quad n \geq n_1.$$

Для определения постоянных $q_{n,i}$, $q_{0;n,i}$ и $q_{1;n,i}$ умножим равенства (29)–(32) на $\bar{d}_{0,j}^{(2)}(y)e^{F_0(y)}$, $\bar{d}_{0,j}^{(2)}(y)\alpha_n(y)e^{F_{0,n}(y)}$ и $\bar{d}_{1,j}^{(2)}(y)e^{F_1(y)}$ соответственно и проинтегрируем в пределах от 0 до ∞ . Тогда

$$q_{n,j} = \sum_{i=1}^I \int_0^\infty \bar{d}_{0,j}^{(2)}(y)e^{F_0(y)} C_i(y) dy q_{n,i} +$$

$$+ \int_0^\infty \bar{d}_{0,j}^{(2)}(y)e^{F_0(y)} Q_n(y) dy, \quad n = \overline{1, n_1}, \quad j = \overline{1, I}; \quad (33)$$

$$q_{0;n,j} =$$

$$= \sum_{i=1}^I \int_0^\infty \bar{d}_{0,j}^{(2)}(y)\alpha_n(y)e^{F_{0,n}(y)} C_{0;n,i}(y) dy q_{0;n,i} +$$

$$+ \int_0^\infty \bar{d}_{0,j}^{(2)}(y)\alpha_n(y)e^{F_{0,n}(y)} Q_{0;n}(y) dy,$$

$$n = \overline{n_1 + 1, n_0 - 1}, \quad j = \overline{1, I}; \quad (34)$$

$$q_{1;n,j} = \sum_{i=1}^{I_1} \int_0^\infty \bar{d}_{1,j}^{(2)}(y)e^{F_1(y)} C_{1;i}(y) dy q_{1;n,i} +$$

$$+ \int_0^\infty \bar{d}_{1,j}^{(2)}(y)e^{F_1(y)} Q_{1;n}(y) dy,$$

$$n \geq n_1 + 1, \quad j = \overline{1, I_1}. \quad (35)$$

Каждое из соотношений (33)–(35) представляет собой систему линейных алгебраических уравнений, что позволяет легко находить коэффициенты $q_{n,i}$, $q_{0;n,i}$ и $q_{1;n,i}$ и в конечном счете плотности $p_n(x)$, $p_n(0; x)$ и $p_n(1; x)$.

В заключение этого раздела приведем выражение для суммарной стационарной интенсивности λ входящего потока:

$$\lambda = \lambda_0 p_0 + \lambda_0 \sum_{n=1}^{n_1} p_n +$$

$$+ \lambda_0 \sum_{n=n_1+1}^{n_0-1} p_{n,0} + \lambda_1 \sum_{n=n_1+1}^{n_0-1} p_{n,1} + \lambda_1 \sum_{n=n_0}^\infty p_n. \quad (36)$$

5 Применение производящих функций

Для вычисления моментов стационарного распределения числа заявок в системе можно воспользоваться производящей функцией (ПФ):

$$p(z, x) = \sum_{n=n_0}^\infty z^n p_n(x).$$

Правда, для того чтобы определить ПФ $p(z, x)$, необходимо знать плотности вероятностей $p_{n_0-1}(0; x)$ и $p_{n_0-1}(1; x)$, а для этого предварительно вычислить $p_n(x)$, $n = \overline{1, n_1}$, $p_n(0; x)$, $n = \overline{n_1 + 1, n_0 - 2}$, и $p_n(1; x)$, $n = \overline{n_1 + 1, n_0 - 2}$.

Умножая соотношения (14) на z^n и суммируя по n , получаем после простейших преобразований с учетом (17)–(19)

$$-p'_x(z, x) = -(1-z)f_1(x)p(z, x) +$$

$$+ \lambda_1 b_1(x) \int_0^\infty p(z, y) [\bar{d}_1(x, y) + z d_1(x, y)] dy +$$

$$+ z^{n_0} g_{1, n_0}(x) \quad (37)$$

с начальным условием

$$\lim_{x \rightarrow \infty} p(z, x) = 0. \quad (38)$$

Уравнение (37) с начальным условием (38) легко приводится к интегральному уравнению

$$q(z, x) = \int_0^{\infty} K(x, y) q(z, y) dy + z^{n_0} R(x),$$

где

$$q(z, x) = e^{-(1-z)F_1(x)} p(z, x);$$

$$K(x, y) = \lambda_1 \int_x^{\infty} e^{(1-z)[F_1(y) - F_1(u)]} b_1(u) [\bar{d}_1(u, y) + z d_1(u, y)] du;$$

$$R(x) = \int_x^{\infty} e^{-(1-z)F_1(u)} g_{1, n_0}(u) du.$$

Последнее уравнение имеет такой же вид, как и (4), с теми же замечаниями относительно решения, что и раньше. Кроме того, если выполнено условие (28), то решение этого уравнения, как и прежде, сводится к решению системы линейных алгебраических уравнений.

Производящая функция $P(z)$ стационарного распределения числа заявок в системе без учета их длин и режима функционирования определяется формулой:

$$P(z) = p_0 + \sum_{n=1}^{n_1} z^n p_n + \sum_{n=n_1+1}^{n_0-1} z^n [p_{n,0} + p_{n,1}] + \int_0^{\infty} e^{(1-z)F_1(x)} q(z, x) dx.$$

Моменты стационарного распределения числа заявок в системе вычисляются с помощью дифференцирования ПФ $P(z)$ в точке $z = 1$ и последующего решения получившихся уравнений.

6 Стационарное распределение времени пребывания заявки в системе

Обозначим через $u(s; x)$ преобразование Лапласа–Стилтьеса (ПЛС) для открываемого заявкой длины x периода занятости (ПЗ) обычной СМО $M/G/1/\infty$ с интенсивностью λ_1 входящего потока и функцией распределения $B_1(x)$ времени обслуживания заявки, а через $u(s)$ — то же самое ПЛС,

но для ПЗ, открываемого заявкой произвольной длины. Тогда

$$u(s; x) = e^{-[s + \lambda_1 - \lambda_1 u(s)]x};$$

$$u(s) = \beta_1 (s + \lambda_1 - \lambda_1 u(s)).$$

Предположим теперь, что в начальный момент рассматриваемая СМО функционирует в режиме 0 и в ней находится n , $n = \overline{1, n_0 - 1}$, заявок. Обозначим через $u_n(s; x)$, $n = \overline{1, n_0 - 1}$, ПЛС времени до того момента, когда в системе впервые останется $n - 1$ заявок и при этом система по-прежнему будет функционировать в режиме 0, при условии что на приборе начала обслуживаться заявка длины x , а через $u_n^*(s; x)$, $n = \overline{n_1 + 2, n_0 - 1}$, — функцию, подобную $u_n(s; x)$, но при этом система перейдет в режим 1.

Справедливы уравнения

$$u'_{n_0-1}(s; x) = -[s + \lambda_0] u_{n_0-1}(s; x); \quad (39)$$

$$u'_n(s; x) = -(s + \lambda_0) u_n(s; x) + \lambda_0 \int_0^{\infty} b_0(y) [d_0(y, x) u_{n+1}(s; y) u_n(s; x) + \bar{d}_0(y, x) u_{n+1}(s; x) u_n(s; y)] dy,$$

$$n = \overline{n_1 + 2, n_0 - 2},$$

с начальным условием

$$u_n(s; 0) = 1, \quad n = \overline{n_1 + 2, n_0 - 1},$$

уравнения

$$u'_{n_0-1}(s; x) = -[s + \lambda_0] u_{n_0-1}^*(s; x) + \lambda_0 \int_0^{\infty} b_0(y) [d_0(y, x) u(s; y) u(s; x) + \bar{d}_0(y, x) u(s; x) u(s; y)] dy; \quad (40)$$

$$u'_n(s; x) = -[s + \lambda_0] u_n^*(s; x) + \lambda_0 \int_0^{\infty} b_0(y) [d_0(y, x) u_{n+1}^*(s; y) u(s; x) + \bar{d}_0(y, x) u_{n+1}^*(s; x) u(s; y)] dy + \lambda_0 \int_0^{\infty} b_0(y) [d_0(y, x) u_{n+1}(s; y) u_n^*(s; x) + d_0(y, x) u_{n+1}(s; x) u_n^*(s; y)] dy,$$

$$n = \overline{n_1 + 2, n_0 - 2}, \quad (41)$$

с начальным условием

$$u_n(s; 0) = 0, \quad n = \overline{n_1 + 2, n_0 - 1},$$

и уравнения

$$\begin{aligned}
 u'_{n_1+1}(s; x) = & -[s + \lambda_0] u_{n_1+1}(s; x) + \\
 & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x) u_{n_1+2}^*(s; y) u(s; x) + \\
 & + \bar{d}_0(y, x) u_{n_1+2}^*(s; x) u(s; y)] dy + \\
 & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x) u_{n_1+2}(s; y) u_{n_1+1}(s; x) + \\
 & + \bar{d}_0(y, x) u_{n_1+2}(s; x) u_{n_1+1}(s; y)] dy; \quad (42)
 \end{aligned}$$

$$\begin{aligned}
 u'_n(s; x) = & -(s + \lambda_0) u_n(s; x) + \\
 & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x) u_{n+1}(s; y) u_n(s; x) + \\
 & + \bar{d}_0(y, x) u_{n+1}(s; x) u_n(s; y)] dy, \quad n = \overline{1, n_1},
 \end{aligned}$$

с начальным условием

$$u_n(s; 0) = 1, \quad n = \overline{1, n_1}.$$

Решения уравнений (39) и (40) имеют вид:

$$\begin{aligned}
 u_{n_0-1}(s; x) = & e^{-(s+\lambda_0)x}; \\
 u_{n_0-1}^*(s; x) = & \lambda_0 \int_0^x e^{(s+\lambda_0)(z-x)} dz \times \\
 & \times \int_0^\infty b_0(y) [d_0(y, z) u(s; y) u(s; z) + \\
 & + \bar{d}_0(y, z) u(s; z) u(s; y)] dy. \quad (43)
 \end{aligned}$$

Остальные уравнения являются интегродифференциальными и подобны уравнениям, полученным в предыдущих разделах.

Пусть в начальный момент в системе находится $n, n \geq n_1 + 1$, заявок, система функционирует в режиме 1, на приборе обслуживается заявка длины y и в этот момент в систему поступает заявка длины x . Обозначим через $w(s; x, y)$ ПЛС времени ожидания начала обслуживания этой заявки. Тогда

$$w(s; x, y) = d_1(x, y) + \bar{d}_1(x, y) u(s; y).$$

Пусть в начальный момент в системе находится $n, n = \overline{1, n_0 - 1}$, заявок, система функционирует в режиме 0, на приборе обслуживается заявка длины y и в этот момент в систему поступает заявка длины x . Обозначим через $w_n(s; x, y)$ ПЛС времени ожидания начала обслуживания этой заявки, причем в момент начала обслуживания система по-прежнему будет функционировать в режиме 0. Имеем:

$$w_{n_0-1}(s; x, y) = 0;$$

$$w_n(s; x, y) = d_0(x, y) + \bar{d}_0(x, y) u_{n+1}(s; y), \quad n = \overline{1, n_0 - 2}.$$

Наконец, пусть в начальный момент в системе находится $n, n = \overline{n_1 + 1, n_0 - 1}$, заявок, система функционирует в режиме 0, на приборе обслуживается заявка длины y и в этот момент в систему поступает заявка длины x . Обозначим через $w_n^*(s; x, y)$ ПЛС времени ожидания начала обслуживания этой заявки, причем в момент начала обслуживания система окажется в режиме 1. В этом случае

$$w_{n_0-1}^*(s; x, y) = d_0(x, y) + \bar{d}_0(x, y) u(s; y); \quad (44)$$

$$w_n^*(s; x, y) = \bar{d}_0(x, y) u_{n+1}^*(s; y), \quad n = \overline{n_1 + 1, n_0 - 2}.$$

Стационарное распределение времени ожидания начала обслуживания имеет ПЛС

$$\begin{aligned}
 w(s) = & \frac{1}{\lambda} \left[\lambda_0 p_0 + \right. \\
 & + \lambda_0 \int_0^\infty \sum_{n=1}^{n_1} p_n(y) dy \int_0^\infty b_0(x) w_n(s; x, y) dx + \\
 & + \lambda_0 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(0; y) dy \times \\
 & \times \int_0^\infty b_0(x) [w_n(s; x, y) + w_n^*(s; x, y)] dx + \\
 & + \lambda_1 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(1; y) dy \int_0^\infty b_1(x) w(s; x, y) dx + \\
 & \left. + \lambda_1 \int_0^\infty \sum_{n=n_0}^\infty p_n(y) dy \int_0^\infty b_1(x) w(s; x, y) dx \right].
 \end{aligned}$$

Обозначим через $t(s; x)$ ПЛС времени от момента первого попадания заявки длины x на прибор до момента ухода ее из системы при условии, что в момент первого попадания на прибор система функционировала в режиме 1. Для $t(s; x)$ справедливо дифференциальное уравнение

$$\begin{aligned}
 t'(s; x) = & -t(s; x) \left(s + \right. \\
 & \left. + \lambda_1 \left[1 - \int_0^\infty b_1(y) [\bar{d}_1(y, x) + d_1(y, x) u(s; y)] dy \right] \right)
 \end{aligned}$$

с начальным условием

$$t(s; 0) = 1,$$

решение которого имеет вид:

$$t(s; x) = \exp \left\{ - (s + \lambda_1)x + \lambda_1 \int_0^x dz \int_0^\infty b_1(y) [\bar{d}_1(y, z) + d_1(y, z)u(s; y)] dy \right\}.$$

Обозначим через $t_n(s; x)$, $n = \overline{0, n_0 - 2}$, ПЛС времени от момента первого попадания заявки длины x на прибор до момента ухода ее из системы при условии, что в момент первого попадания на прибор в очереди было еще n заявок и система функционировала в режиме 0. Тогда из дифференциальных уравнений

$$t'_{n_0-2}(s; x) = -(s + \lambda_0)t_{n_0-2}(s; x) + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)u(s; y) + \bar{d}_0(y, x)] dy t(s; x);$$

$$t'_n(s; x) = - \left(s + \lambda_0 - \lambda_0 \int_0^\infty b_0(y)d_0(y, x)u_{n+2}(s; y) dy \right) t_n(s; x) + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)u_{n+2}^*(s; y)t(s; x) + \bar{d}_0(y, x)t_{n+1}(s; x)] dy, \quad n = \overline{n_1, n_0 - 3};$$

$$t'_n(s; x) = - \left(s + \lambda_0 - \lambda_0 \int_0^\infty b_0(y)d_0(y, x)u_{n+2}(s; y) dy \right) t_n(s; x) + \lambda_0 t_{n+1}(s; x) \int_0^\infty b_0(y)\bar{d}_0(y, x) dy, \quad n = \overline{0, n_1 - 1},$$

с начальным условием

$$t_n(s; 0) = 1, \quad n = \overline{0, n_0 - 2},$$

имеем:

$$t_{n_0-2}(s; x) = e^{-(s+\lambda_0)x} \left(1 + \lambda_0 \int_0^x e^{(s+\lambda_0)z} t(s; z) dz \times \int_0^\infty b_0(y) [d_0(y, z)u(s; y) + \bar{d}_0(y, z)] dy \right); \quad (45)$$

$$t_n(s; x) = e^{-\int_0^x (s+\lambda_0-\lambda_0 \int_0^\infty b_0(y) d_0(y, z) u_{n+2}(s; y) dy) dz} \left(1 + \lambda_0 \int_0^x e^{\int_0^v (s+\lambda_0-\lambda_0 \int_0^\infty b_0(y) d_0(y, z) u_{n+2}(s; y) dy) dz} dv \times \int_0^\infty b_0(y) [d_0(y, v)u_{n+2}^*(s; y)t(s; v) + \bar{d}_0(y, v)t_{n+1}(s; v)] dy \right), \quad n = \overline{n_1, n_0 - 3}; \quad (46)$$

$$t_n(s; x) = e^{-\int_0^x (s+\lambda_0-\lambda_0 \int_0^\infty b_0(y) d_0(y, z) u_{n+2}(s; y) dy) dz} \left(1 + \lambda_0 \int_0^x e^{\int_0^v (s+\lambda_0-\lambda_0 \int_0^\infty b_0(y) d_0(y, z) u_{n+2}(s; y) dy) dz} dv \times \int_0^\infty b_0(y)\bar{d}_0(y, v)t_{n+1}(s; v) dy \right), \quad n = \overline{0, n_1 - 1}.$$

Обратимся к общему времени пребывания заявки в системе.

Обозначим через $v(s; x, y)$ ПЛС времени пребывания в системе заявки длины x при условии, что эта заявка застала систему в режиме 1, причем заявка на приборе имела длину y . Тогда

$$v(s; x, y) = w(s; x, y)t(s; x).$$

Обозначим через $v_n(s; x, y)$, $n = \overline{1, n_0 - 1}$, ПЛС времени пребывания в системе заявки длины x при условии, что эта заявка застала в системе n других заявок, причем заявка на приборе имела длину y , а система пребывала в режиме 0. Имеем:

$$v_{n_0-1}(s; x, y) = d_0(x, y)t(s; x) + \bar{d}_0(x, y)u(s; x, y)t(s; x); \quad (47)$$

$$v_n(s; x, y) = d_0(x, y)t_n(s; x) + \bar{d}_0(x, y) [u_{n+1}(s; x, y)t_{n-1}(s; x) + u_{n+1}^*(s; x, y)t(s; x)], \quad n = \overline{n_1 + 1, n_0 - 2}; \quad (48)$$

$$v_n(s; x, y) = d_0(x, y)t_n(s; x) + \bar{d}_0(x, y)u_{n+1}(s; x, y)t_{n-1}(s; x), \quad n = \overline{1, n_1}.$$

Стационарное распределение общего времени пребывания заявки в системе имеет ПЛС

$$v(s) = \frac{1}{\lambda} \left[\lambda_0 p_0 \int_0^\infty b_0(x) t_0(s; x) dx + \lambda_0 \int_0^\infty \sum_{n=1}^{n_1} p_n(y) dy \int_0^\infty b_0(x) v_n(s; x, y) dx + \lambda_0 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(0; y) dy \int_0^\infty b_0(x) v_n(s; x, y) dx + \lambda_1 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(1; y) dy \int_0^\infty b_1(x) v(s; x, y) dx + \lambda_1 \int_0^\infty \sum_{n=n_0}^\infty p_n(y) dy \int_0^\infty b_1(x) v(s; x, y) dx \right].$$

Дифференцируя $w(s)$ и $v(s)$ в точке $s = 0$, можно найти моменты стационарных распределений времен ожидания начала обслуживания и пребывания заявки в системе.

7 Накопитель конечной емкости

В этом разделе будет показано, какие изменения нужно произвести в полученных формулах для случая накопителя конечной емкости. Заметим, что к формулам, остающимся без изменений, комментарии приводиться не будут.

Итак, будем предполагать, что максимальное число заявок, находящихся в системе, равно n^* , $n^* \geq n_0$, (емкость накопителя $n^* - 1$).

Для конечного накопителя необходимо также задать дисциплину принятия заявок в систему при отсутствии в нем свободных мест. В соответствии с рассматриваемой СМО естественно такую дисциплину определить с помощью функции $d^*(x, y)$ следующим образом: поступающая заявка длины x , застающая на приборе заявку длины y , с вероятностью $d^*(x, y)$ сразу же покидает систему, не оказывая на нее никакого воздействия, и с дополнительной вероятностью $\bar{d}^*(x, y) = 1 - d^*(x, y)$ становится на прибор, вытесняя заявку на приборе из системы. Для всех СМО с такой дисциплиной принятия заявок в систему при отсутствии в накопителе свободных мест стационарные вероятности $p_n(x_1, \dots, x_n)$ при $n < n^*$ совпадают с точностью до постоянной с аналогичными вероятностями для

системы с бесконечным накопителем, различие заключается только в вероятностях $p_{n^*}(x_1, \dots, x_{n^*})$. Однако несколько более сложно вычисляются стационарные распределения, связанные с временем пребывания заявки в системе. Более того, заявки, принятые в систему, могут покидать ее недообслуженными.

Здесь для простоты изложения будет рассмотрен только случай $d^*(x, y) = 1$, т. е. тот случай, когда поступающая в заполненную систему заявка теряется. Заметим, что в этом случае принятая в систему заявка будет обязательно обслужена полностью. Общий случай нетрудно исследовать с помощью результатов, полученных в [9, 10].

Далее будем предполагать, что $n^* \geq n_0 + 2$, поскольку при $n^* = n_0$ и $n^* = n_0 + 1$ расчетные формулы будут несколько отличаться от приведенных выше.

Как уже говорилось, стационарные вероятности $p_n(x)$, $n = \overline{1, n_1}$ или $n = \overline{n_0, n^* - 1}$, и $p_n(i; x)$, $n = n_1 + 1, n_0 - 1$, $i = 1, 2$, с точностью до вероятности p_0 можно определить из тех же самых уравнений (10)–(14), что и раньше. Вероятность $p_{n^*}(x)$ удовлетворяет дифференциальному уравнению

$$-p'_{n^*}(x) = g_{1, n^*}(x)$$

с начальным условием

$$\lim_{x \rightarrow \infty} p_{n^*}(x) = 0,$$

где

$$g_{1, n^*}(x) = \lambda_1 b_1(x) \int_0^\infty p_{n^*-1}(y) d_1(x, y) dy + \lambda_1 p_{n^*-1}(x) \int_0^\infty b_1(y) \bar{d}_1(y, x) dy.$$

Решение этого уравнения определяется выражением:

$$p_{n^*}(x) = \int_x^\infty g_{1, n^*}(y) dy.$$

Вероятность p_0 вычисляется из условия нормировки, которое в данном случае имеет вид:

$$p_0 + \sum_{n=1}^{n_1} p_n + \sum_{n=n_1+1}^{n_0-1} [p_{n,0} + p_{n,1}] + \sum_{n=n_0}^{n^*} p_n = 1.$$

Стационарная интенсивность λ входящего в систему потока задается формулой (36), в которой, естественно, верхний индекс ∞ в последней сумме заменен на n^* .

В системах с конечным накопителем важной характеристикой является стационарная вероятность π_{loss} потери заявки, определяемая формулой:

$$\pi_{\text{loss}} = \frac{\lambda_1}{\lambda} p_{n^*}.$$

Для того чтобы найти показатели функционирования СМО, связанные с временем пребывания в системе, нужно прежде всего изменить некоторые формулы для ПЛС ПЗ.

Предположим, что в начальный момент рассматриваемая СМО функционирует в режиме 1 и в ней находится n , $n = \overline{n_1 + 1, n^*}$, заявок. Обозначим через $\tilde{u}_n(s; x)$, $n = \overline{n_1 + 1, n^*}$, ПЛС времени до того момента, когда в системе впервые останется $(n - 1)$ заявок, при условии что на приборе начала обслуживаться заявка длины x (очевидно, что в этот момент система по-прежнему будет функционировать в режиме 1). Преобразования Лапласа—Стилтьеса $\tilde{u}_n(s; x)$ удовлетворяют уравнениям

$$\tilde{u}_{n^*}(s; x) = e^{-sx};$$

$$\begin{aligned} \tilde{u}'_n(s; x) = & -(s + \lambda_1)\tilde{u}_n(s; x) + \\ & + \lambda_1 \int_0^\infty b_1(y) [d_1(y, x)\tilde{u}_{n+1}(s; y)\tilde{u}_n(s; x) + \\ & + \bar{d}_1(y, x)\tilde{u}_{n+1}(s; x)\tilde{u}_n(s; y)] dy, \\ & n = \overline{n_1 + 1, n^* - 1}, \end{aligned}$$

с начальным условием

$$\tilde{u}_n(s; 0) = 1, \quad n = \overline{n_1 + 1, n^* - 1}.$$

Уравнения (40)–(42) принимают следующий вид:

$$\begin{aligned} u_{n_0-1}^{*'}(s; x) = & -[s + \lambda_0]u_{n_0-1}^*(s; x) + \\ & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)\tilde{u}_{n_0}(s; y)\tilde{u}_{n_0-1}(s; x) + \\ & + \bar{d}_0(y, x)\tilde{u}_{n_0}(s; x)\tilde{u}_{n_0-1}(s; y)] dy; \\ u_n^{*'}(s; x) = & -[s + \lambda_0]u_n^*(s; x) + \\ & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)u_{n+1}^*(s; y)\tilde{u}_n(s; x) + \\ & + \bar{d}_0(y, x)u_{n+1}^*(s; x)\tilde{u}_n(s; y)] dy + \\ & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)u_{n+1}(s; y)u_n^*(s; x) + \\ & + \bar{d}_0(y, x)u_{n+1}(s; x)u_n^*(s; y)] dy, \\ & n = \overline{n_1 + 2, n_0 - 2}; \end{aligned}$$

$$\begin{aligned} u'_{n_1+1}(s; x) = & -[s + \lambda_0]u_{n_1+1}(s; x) + \\ & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)u_{n_1+2}^*(s; y)\tilde{u}_{n_1+1}(s; x) + \\ & + \bar{d}_0(y, x)u_{n_1+2}^*(s; x)\tilde{u}_{n_1+1}(s; y)] dy + \\ & + \lambda_0 \int_0^\infty b_0(y) [d_0(y, x)u_{n_1+2}(s; y)u_{n_1+1}(s; x) + \\ & + \bar{d}_0(y, x)u_{n_1+2}(s; x)u_{n_1+1}(s; y)] dy. \end{aligned}$$

Соответственно изменится и формула (43).

Пусть в начальный момент в системе находится n , $n = \overline{n_1 + 1, n^* - 1}$, заявок, система функционирует в режиме 1, на приборе обслуживается заявка длины y и в этот момент в систему поступает заявка длины x . Обозначим через $\tilde{w}_n(s; x, y)$ ПЛС времени ожидания начала обслуживания этой заявки. Имеет место равенство:

$$\begin{aligned} \tilde{w}_n(s; x, y) = & d_1(x, y) + \bar{d}_1(x, y)\tilde{u}_{n+1}(s; y), \\ & n = \overline{n_1 + 1, n^* - 1}. \end{aligned}$$

Формула (44) принимает вид:

$$w_{n_0-1}^*(s; x, y) = d_0(x, y) + \bar{d}_0(x, y)\tilde{u}_{n_0}(s; y),$$

а ПЛС стационарного распределения времени ожидания начала обслуживания принятой в систему заявки определяется формулой:

$$\begin{aligned} w(s) = & \frac{1}{\lambda(1 - \pi_{\text{loss}})} \left[\lambda_0 p_0 + \right. \\ & + \lambda_0 \int_0^\infty \sum_{n=1}^{n_1} p_n(y) dy \int_0^\infty b_0(x) w_n(s; x, y) dx + \\ & + \lambda_0 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(0; y) dy \int_0^\infty b_0(x) [w_n(s; x, y) + \\ & \quad \left. + w_n^*(s; x, y)] dx + \right. \\ & + \lambda_1 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(1; y) dy \int_0^\infty b_1(x) \tilde{w}_n(s; x, y) dx + \\ & \left. + \lambda_1 \int_0^\infty \sum_{n=n_0}^{n^*-1} p_n(y) dy \int_0^\infty b_1(x) \tilde{w}_n(s; x, y) dx \right]. \end{aligned}$$

Обозначим через $\tilde{t}_n(s; x)$, $n = \overline{n_1, n^* - 1}$, ПЛС времени от момента первого попадания заявки длины x на прибор до момента ухода ее из системы при условии, что в момент первого попадания на прибор

в очереди было еще n заявок и система функционировала в режиме 1. Тогда

$$\begin{aligned} \tilde{t}_{n^*-1}(s; x) &= e^{-sx}; \\ \tilde{t}_n(s; x) &= \exp \left\{ - (s + \lambda_1)x + \right. \\ &+ \lambda_1 \int_0^x dz \int_0^\infty b_1(y) [\bar{d}_1(y, z) + \\ &+ d_1(y, z)\tilde{u}_{n+2}(s; y)] dy \left. \right\}, \quad n = \overline{n_1, n^* - 2}. \end{aligned}$$

При этом формулы (45) и (46) записываются в виде:

$$\begin{aligned} t_{n_0-2}(s; x) &= e^{-(s+\lambda_0)x} \left(1 + \lambda_0 \int_0^x e^{(s+\lambda_0)z} dz \times \right. \\ &\times \int_0^\infty b_0(y) [d_0(y, z)\tilde{u}_{n_0}(s; y)\tilde{t}_{n_0-2}(s; z) + \\ &\quad \left. + \bar{d}_0(y, z)\tilde{t}_{n_0-1}(s; z)] dy \right); \\ t_n(s; x) &= \\ &= e^{-\int_0^x (s+\lambda_0-\lambda_0 \int_0^\infty b_0(y)d_0(y, z)u_{n+2}(s; y) dy) dz} \left(1 + \right. \\ &+ \lambda_0 \int_0^x e^{\int_0^v (s+\lambda_0-\lambda_0 \int_0^\infty b_0(y)d_0(y, z)u_{n+2}(s; y) dy) dz} dv \times \\ &\times \int_0^\infty b_0(y) [d_0(y, v)u_{n+2}^*(s; y)\tilde{t}_n(s; v) + \\ &\quad \left. + \bar{d}_0(y, v)t_{n+1}(s; v)] dy \right), \quad n = \overline{n_1, n_0 - 3}. \end{aligned}$$

Наконец, перейдем к общему времени пребывания заявки в системе. Обозначим через $\tilde{v}_n(s; x, y)$, $n = \overline{n_1 + 1, n^* - 1}$, ПЛС времени пребывания в системе заявки длины x при условии, что эта заявка застала в системе n других заявок, причем заявка на приборе имела длину y , а система пребывала в режиме 1. Справедливо соотношение:

$$\begin{aligned} \tilde{v}_n(s; x, y) &= d_1(x, y)\tilde{t}_n(s; x) + \\ &+ \bar{d}_1(x, y)\tilde{u}_{n+1}(s; x, y)\tilde{t}_{n-1}(s; x), \\ &\quad n = \overline{n_1 + 1, n^* - 1}. \end{aligned}$$

Формулы (47) и (48) преобразуются следующим образом:

$$\begin{aligned} v_{n_0-1}(s; x, y) &= d_0(x, y)\tilde{t}_{n_0-1}(s; x) + \\ &+ \bar{d}_0(x, y)\tilde{u}_{n_0}(s; x, y)\tilde{t}_{n_0-2}(s; x); \\ v_n(s; x, y) &= d_0(x, y)t_n(s; x) + \\ &+ \bar{d}_0(x, y) [u_{n+1}(s; x, y)t_{n-1}(s; x) + \\ &+ u_{n+1}^*(s; x, y)\tilde{t}_{n-1}(s; x)], \quad n = \overline{n_1 + 1, n_0 - 2}, \end{aligned}$$

а ПЛС стационарного распределения общего времени пребывания в системе принятой к обслуживанию заявки имеет вид:

$$\begin{aligned} v(s) &= \frac{1}{\lambda(1 - \pi_{\text{loss}})} \left[\lambda_0 p_0 \int_0^\infty b_0(x)t_0(s; x) dx + \right. \\ &+ \lambda_0 \int_0^\infty \sum_{n=1}^{n_1} p_n(y) dy \int_0^\infty b_0(x)v_n(s; x, y) dx + \\ &+ \lambda_0 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(0; y) dy \int_0^\infty b_0(x)v_n(s; x, y) dx + \\ &+ \lambda_1 \int_0^\infty \sum_{n=n_1+1}^{n_0-1} p_n(1; y) dy \int_0^\infty b_1(x)\tilde{v}_n(s; x, y) dx + \\ &\quad \left. + \lambda_1 \int_0^\infty \sum_{n=n_0}^{n^*-1} p_n(y) dy \int_0^\infty b_1(x)\tilde{v}_n(s; x, y) dx \right]. \end{aligned}$$

8 Заключение

В настоящей статье рассмотрена возможность применения аналитических методов для вычисления основных стационарных характеристик функционирования СМО, в которых одновременно имеется два отличия от классических СМО: инверсионный порядок обслуживания с вероятностным приоритетом и гистерезисная политика. На примере однолинейной СМО с простейшим вариантом гистерезисной политики показано, что полученные вычислительные алгоритмы основаны на интегральных и дифференциальных уравнениях, которые могут быть решены на современной вычислительной технике. Приведены условия, при которых интегральные уравнения могут быть сведены к системам линейных алгебраических уравнений.

Полученные результаты могут служить основой для продолжения работ в части математического моделирования технических систем, использующих как элементы «нестандартных» дисциплин

обслуживания, так и сложные варианты гистерезисного механизма предотвращения различного рода перегрузок в ИТС.

Литература

1. Печинкин А. В. Об одной инвариантной системе массового обслуживания // Math. Operationsforsch. und Statist. Ser. Optimization, 1983. Vol. 14. No. 3. S. 433–444.
2. Печинкин А. В., Стальченко И. В. Система $MAR/G/1/\infty$ с инверсионным порядком обслуживания и вероятностным приоритетом, функционирующая в дискретном времени // Вестник Российского ун-та дружбы народов. Сер. Математика. Информатика. Физика, 2010. № 2. С. 26–36.
3. Абаев П. О., Гайдамака Ю. В., Самуйлов К. Е. Гистерезисное управление сигнальной нагрузкой в сети SIP-серверов // Вестник Российского ун-та дружбы народов. Сер. Математика. Информатика. Физика, 2011. № 4. С. 54–71.
4. Nishimura S., Jiang Y. An $M/G/1$ vacation model with two service modes // Prob. Eng. Informational Sci., 1995. Vol. 9. No. 3. P. 355–374.
5. Dudin A. Optimal control for an $M^x/G/1$ queue with two operation modes // Prob. Eng. Informational Sci., 1997. Vol. 11. No. 2. P. 255–265.
6. Nobel R. D., Tijms H. C. Optimal control for an $M^X/G/1$ queue with two service modes // Eur. J. Operational Res., 1999. Vol. 113. No. 3. P. 610–619.
7. Жерновы К. Ю., Жерновы Ю. В. Система $M^0/G/1/m$ с двухпороговой гистерезисной стратегией переключения интенсивности обслуживания // Информационные процессы, 2012. Т. 12. № 2. С. 127–140.
8. Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S. Queueing theory. — Utrecht, Boston: VSP, 2004.
9. Нагоненко В. А. О характеристиках одной нестандартной системы массового обслуживания. I // Изв. АН СССР. Технич. кибернет., 1981. № 1. С. 187–195.
10. Нагоненко В. А. О характеристиках одной нестандартной системы массового обслуживания. II // Изв. АН СССР. Технич. кибернет., 1981. № 3. С. 91–99.

О СХОДИМОСТИ В ПРОСТРАНСТВЕ L_p МАКСИМУМА ПРОЦЕССА НАГРУЗКИ ДЛЯ ОДНОГО КЛАССА ГАУССОВСКИХ СИСТЕМ ОБСЛУЖИВАНИЯ*

О. В. Лукашенко¹, Е. В. Морозов²

Аннотация: Рассматривается класс систем обслуживания, на вход которых поступает поток, содержащий линейную детерминированную компоненту и случайную компоненту, описываемую центрированным гауссовским процессом. Дисперсия входного процесса является правильно меняющейся функцией с показателем $V \in (0, 2)$. Найден условия, при которых максимум стационарного процесса нагрузки (незавершенной работы) на интервале $[0, t]$ сходится при $t \rightarrow \infty$ (и при соответствующей нормировке) в пространстве L_p к явно выписанной константе. Также найдена асимптотика максимума процесса нагрузки в нестационарном режиме. Получена асимптотика минимального времени достижения процессом нагрузки растущего значения b .

Ключевые слова: гауссовская система обслуживания; максимум процесса нагрузки; дробное броуновское движение; асимптотический анализ; правильное изменение

1 Введение

В работе [1] был осуществлен асимптотический анализ процесса нагрузки в жидкостной системе с постоянной скоростью обслуживания C и входным гауссовским процессом, дисперсия которого правильно меняется на бесконечности с показателем $V \in (0, 2)$. Рассмотренный класс входных процессов включает, в частности, сумму независимых дробных броуновских движений (ДБД) с соответствующими значениями индекса Херста. В [1] показано, что при соответствующей нормировке максимум процесса нагрузки на интервале $[0, t]$ сходится по вероятности при $t \rightarrow \infty$ к явно выписанной константе. Этот результат существенно обобщает результат из работы [2], в которой рассмотрена жидкостная система с единственным входным процессом ДБД.

В данной статье, которая опирается на методы работы [2], а также на результаты статьи [1], продолжен асимптотический анализ описанной жидкостной системы. Основным результатом данной работы состоит в том, что при некоторых дополнительных условиях на асимптотическое поведение дисперсии входного гауссовского процесса доказанная в [1] сходимости усилена до сходимости (к той же константе) в пространстве L_p при любом $p \geq 1$. Более того, при соответствующей нормировке найдена

асимптотика максимума процесса нагрузки в нестационарном режиме. Кроме того, с использованием полученной асимптотики максимума процесса нагрузки найдена асимптотика времени достижения процессом нагрузки растущего уровня b .

Опишем рассматриваемую систему и используемые ниже результаты из [1] более подробно. Рассмотрим жидкостную систему с одним обслуживающим устройством и постоянной скоростью обслуживания C , на вход которой поступает процесс $A(t)$, заданный в следующем виде:

$$A(t) = mt + X(t), \quad (1)$$

где m — средняя интенсивность входного потока, а $X := \{X(t), t \geq 0\}$ — центрированный гауссовский процесс со стационарными приращениями, $X(0) = 0$. Такая система обслуживания называется гауссовской [3]. Будем считать, что выполнено условие $r := C - m > 0$. Обозначим также $W(t) = X(t) - rt$, и пусть $Q(t)$ есть величина нагрузки (незавершенная работа в системе) в момент времени t . Будем предполагать, что $Q(0) = 0$. Тогда имеет место соотношение [4]:

$$Q(t) = \sup_{0 \leq s \leq t} (W(t) - W(s)). \quad (2)$$

* Работа выполняется 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, lukashenko-oleg@mail.ru

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

Условие $r > 0$ обеспечивает существование стационарного процесса нагрузки, который определяется следующим образом [3]:

$$Q = \sup_{t \geq 0} W(t). \quad (3)$$

Для пояснения заметим, что величина $-r < 0$ есть средний снос процесса W , являющегося аналогом случайного блуждания (с независимыми приращениями), максимум которого, по аналогии с (3), определяет стационарное время ожидания в классических системах обслуживания [5].

Основное предположение, принятое в работе [1], а также в данной статье, состоит в том, что функция $v(t)$ правильно меняется на бесконечности с индексом $0 < V < 2$, т. е. представима в виде

$$v(t) = t^V L(t), \quad (4)$$

где функция L медленно меняется на бесконечности [6]. Обозначим

$$\beta = \frac{1}{2 - V},$$

а также выберем и зафиксируем любое $\varepsilon \in (0, 2 - V)$. Будем считать, что функция L является дважды дифференцируемой на \mathbb{R}_+ и выполнены следующие условия (при $t \rightarrow \infty$):

$$L(tL^\beta(t)) \sim L(t); \quad (5)$$

$$L''(t) = o\left(\frac{1}{t^{V+\varepsilon}}\right). \quad (6)$$

Нетрудно проверить, что из условия (6) следует сходимость

$$v''(t) \ln t \rightarrow 0, \quad t \rightarrow \infty. \quad (7)$$

Как показано в [1], условия (4)–(6) на самом деле выполнены для широкого класса гауссовских систем обслуживания, включающего, например, системы, на вход которых поступает сумма нескольких независимых ДБД. В работе [7] показано, что на одном вероятностном пространстве можно задать процесс $W(t) = X(t) - rt$ и стационарный процесс $Q^* := \{Q^*(t), t \geq 0\}$ таким образом, что одновременно выполнены условия:

$$\begin{aligned} Q^*(t) &= {}_d Q \text{ для всех } t \geq 0; \\ Q^*(t) &= W(t) + \max\{Q^*(0), L^*(t)\}, \quad t \geq 0, \end{aligned}$$

где $=_d$ означает равенство по распределению, а $L^*(t) := -\min_{0 \leq s \leq t} \{W(s)\}$. Обозначим

$$M(t) = \max_{0 \leq s \leq t} Q(s), \quad M^*(t) = \max_{0 \leq s \leq t} Q^*(s).$$

Для удобства обозначим далее

$$\gamma(t) = L \left[(\ln t)^\beta \right] \ln t.$$

Ниже будем опираться на результаты следующей теоремы, доказанной в работе [1].

Теорема 1.1. Пусть дисперсия гауссовской компоненты X входного процесса (1) удовлетворяет условиям (5) и (6), а также $r > 0$. Тогда

$$\frac{M^*(t)}{\gamma^\beta(t)} \Rightarrow \left(\frac{1}{\theta}\right)^\beta, \quad t \rightarrow \infty; \quad (8)$$

$$\frac{M(t)}{\gamma^\beta(t)} \Rightarrow \left(\frac{1}{\theta}\right)^\beta, \quad t \rightarrow \infty, \quad (9)$$

где \Rightarrow означает сходимость по вероятности, а параметр θ удовлетворяет соотношению:

$$\theta = \frac{2}{(2 - V)^{2-V}} \left(\frac{r}{V}\right)^V. \quad (10)$$

Как отмечено выше, этот результат обобщает работу [2], где процесс $X = B_H$ является ДБД с параметром Херста $H \in (1/2, 1)$.

2 Сходимость в пространстве L_p

В данном разделе будет доказан основной результат, состоящий в том, что при дополнительных условиях на функцию L из (4) сходимость по вероятности в (8), (9) можно усилить до сходимости в пространстве L_p , где $p \in [1, \infty)$.

Теорема 2.1. Пусть дополнительно к условиям теоремы 1.1

$$\liminf_{t \rightarrow \infty} L(t) > 0; \quad \limsup_{t \rightarrow \infty} L(t) < \infty. \quad (11)$$

Тогда в (8), (9) имеет место сходимость в пространстве L_p , $p \in [1, \infty)$.

Доказательство. Зафиксируем $p \in [1, \infty)$. Для доказательства теоремы достаточно доказать равномерную интегрируемость семейства случайных величин

$$\left\{ \left(\frac{M^*(t)}{\gamma^\beta(t)} \right)^p, \quad t \geq T \right\},$$

где T — некоторое (конечное) положительное число. Для этого, в свою очередь, достаточно, чтобы было выполнено (см., например, [8]) условие

$$\sup_{t \geq T} \mathbb{E} \left[\frac{M^*(t)}{\gamma^\beta(t)} \right]^{p+1} < \infty. \quad (12)$$

(Значение $\mathbb{E}(\cdot)$ при $t < T$ может быть произвольным.) Выберем далее некоторое число $K > 0$. Значения величин K и T будут уточняться в процессе доказательства. Кроме того, всюду далее предполагается, что $t \geq T$. Имеют место соотношения:

$$\begin{aligned} \mathbb{E} \left[\frac{M^*(t)}{\gamma^\beta(t)} \right]^{p+1} &= \\ &= \int_0^\infty (p+1)y^p \mathbb{P}(M^*(t) > y\gamma^\beta(t)) dy = \\ &= \int_0^K (p+1)y^p \mathbb{P}(M^*(t) > y\gamma^\beta(t)) dy + \\ &+ \int_K^\infty (p+1)y^p \mathbb{P}(M^*(t) > y\gamma^\beta(t)) dy \leq \\ &\leq K^{p+1} + (p+1)(I_t + R_t), \end{aligned} \quad (13)$$

где использованы обозначения:

$$\begin{aligned} I_t &= \int_K^\infty y^p [t] \mathbb{P} \left(Q^*(0) > \frac{y\gamma^\beta(t)}{2} \right) dy; \\ R_t &= \int_K^\infty y^p [t] \times \\ &\times \mathbb{P} \left(\max_{0 \leq s \leq 1} W(s) - \min_{0 \leq s \leq 1} W(s) > \frac{y\gamma^\beta(t)}{2} \right) dy. \end{aligned}$$

Отметим, что при получении выражения (13) использовано неравенство:

$$\begin{aligned} \mathbb{P}(M^*(t) > x) &\leq \\ &\leq [t] \mathbb{P} \left(Q^*(0) + \max_{0 \leq s \leq 1} W(s) - \min_{0 \leq s \leq 1} W(s) > x \right). \end{aligned}$$

(См. доказательство теоремы 1.1 в [1].) Оценим вначале интеграл I_t . В работе [9] показано, что если дисперсия $v(t)$ центрированного гауссовского процесса со стационарными приращениями правильно меняется на бесконечности с индексом $0 < V < 2$, то справедлива такая (логарифмическая) асимптотика:

$$\lim_{b \rightarrow \infty} \frac{v(b)}{b^2} \ln \mathbb{P}(Q^* > b) = -\theta, \quad (14)$$

где параметр θ удовлетворяет соотношению (10). Определим число K_1 следующим образом:

$$K_1 = \inf \left\{ y > 0 : \frac{L(x) \ln \mathbb{P}(Q^*(0) > x)}{x^{1/\beta}} \leq -\frac{\theta}{2}, \forall x > y \right\}.$$

Напомним, что $\beta = 1/(2 - V)$. Поэтому ввиду (4) из (14) следует, что $K_1 < \infty$. Тогда при $x > K_1$ справедливо неравенство:

$$\mathbb{P}(Q^*(0) > x) \leq \exp \left(-\frac{\theta}{2} \frac{x^{1/\beta}}{L(x)} \right). \quad (15)$$

Заметим, что $\gamma^\beta(t) \rightarrow \infty, t \rightarrow \infty$. Следовательно, существует такое число t_0 , что $\gamma^\beta(t)/2 > 1$ при $t \geq t_0$. Поэтому, если $K > K_1, t \geq t_0$, то из (15) вытекает неравенство:

$$\begin{aligned} I_t &= \int_K^\infty y^p [t] \mathbb{P} \left(Q^*(0) > \frac{y\gamma^\beta(t)}{2} \right) dy \leq \int_K^\infty y^p [t] \times \\ &\times \exp \left[-\frac{\theta}{2^{1/\beta+1}} \frac{\ln t \cdot L[(\ln t)^\beta]}{L(y\gamma^\beta(t)/2)} y^{1/\beta} \right] dy. \end{aligned} \quad (16)$$

Из условия (11) следует, что существуют такие числа $0 < A_1 \leq A_2 < \infty$ и $K_2, t_1 \geq 0$, что при $y > K_2, t > \max(t_0, t_1)$ выполняются неравенства:

$$L \left(\frac{y\gamma^\beta(t)}{2} \right) \leq A_2; \quad (17)$$

$$L((\ln t)^\beta) \geq A_1. \quad (18)$$

Выберем теперь и временно зафиксируем в (16) некоторое $K > \max\{K_1, K_2\}$, и пусть пока $T := \max(t_0, t_1)$. Обозначим также

$$\alpha = \frac{\theta A_1}{2^{1/\beta+1} A_2}.$$

Последовательно применяя (17), (18), а также принимая во внимание, что $[t] \leq 2t$, можно получить из (16) (при $t \geq T$) следующую оценку сверху интеграла I_t :

$$\begin{aligned} I_t &\leq 2 \int_K^\infty y^p t \exp \left(-\alpha \ln t \cdot y^{1/\beta} \right) dy = \\ &= 2 \int_K^\infty y^p \exp \left[\ln t \left(1 - \alpha y^{1/\beta} \right) \right] dy. \end{aligned} \quad (19)$$

Заметим, что при $y > (2/\alpha)^\beta := K_3$ справедливо неравенство:

$$1 - \alpha y^{1/\beta} < -\frac{\alpha}{2} y^{1/\beta}. \quad (20)$$

Если теперь выбрать в (19) $K > K_3$, то ввиду (20) получим:

$$I_t \leq 2 \int_K^\infty y^p \exp \left(-\frac{\alpha}{2} \ln t \cdot y^{1/\beta} \right) dy. \quad (21)$$

Заметим, что $\ln t \geq \ln T$ при $t \geq T$, и обозначим $\gamma_1 = (\alpha/2) \ln T$. Тогда из (21) получим:

$$\begin{aligned} I_t &\leq 2 \int_K^\infty y^p \exp(-\gamma_1 y^{1/\beta}) dy = \\ &= \frac{2\beta}{\gamma_1^{\beta+\beta p}} \int_{\gamma_1 K^{1/\beta}}^\infty u^{\beta p + \beta - 1} e^{-u} du = \\ &= \frac{2\beta}{\gamma_1^{\beta+\beta p}} \Gamma(\gamma_1 K^{1/\beta}, \beta p + \beta), \quad (22) \end{aligned}$$

где $\Gamma(w, z)$ — неполная гамма-функция:

$$\Gamma(w, z) := \int_w^\infty u^{z-1} e^{-u} du, \quad w \geq 0, \quad z \geq 0.$$

Теперь оценим интеграл R_t в (13). Напомним соотношение $W(t) = X(t) - rt$ и заметим, что

$$\max_{0 \leq s \leq 1} W(s) = \max_{0 \leq s \leq 1} (X(s) - rs) \leq \max_{0 \leq s \leq 1} X(s). \quad (23)$$

Кроме того,

$$\begin{aligned} - \min_{0 \leq s \leq 1} W(s) &= \max_{0 \leq s \leq 1} (rs - X(s)) = d \\ &= d \max_{0 \leq s \leq 1} (rs + X(s)) \leq r + \max_{0 \leq s \leq 1} X(s). \quad (24) \end{aligned}$$

Неравенства (23) и (24) после несложных преобразований приводят, в свою очередь, к неравенству:

$$\begin{aligned} \mathbb{P} \left(\max_{0 \leq s \leq 1} W(s) - \min_{0 \leq s \leq 1} W(s) > \frac{y\gamma^\beta(t)}{2} \right) &\leq \\ &\leq 2\mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) > \frac{y\gamma^\beta(t)}{4} - r \right). \quad (25) \end{aligned}$$

Применяя соотношения (25) и (18), можно получить следующую цепочку неравенств:

$$\begin{aligned} R_t &= \int_K^\infty y^p [t] \times \\ &\times \mathbb{P} \left(\max_{0 \leq s \leq 1} W(s) - \min_{0 \leq s \leq 1} W(s) > \frac{y\gamma^\beta(t)}{2} \right) dy \leq \\ &\leq 2 \int_K^\infty y^p [t] \mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) > \frac{y\gamma^\beta(t)}{4} - r \right) dy \leq \\ &\leq 4 \int_K^\infty y^p t \mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) > \frac{yA_1^\beta (\ln t)^\beta}{4} - r \right) dy. \quad (26) \end{aligned}$$

Теперь используем следующее неравенство Бореля–Судакова–Цирельсона для максимума центрированного гауссовского процесса со стационарными приращениями на конечном интервале [10, 11]:

$$\mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) > x \right) \leq e^{-(1/(2v))(x-a)^2}, \quad x > a, \quad (27)$$

где $v := \mathbb{D}X(1)$, $a := \mathbb{E} \max_{0 \leq s \leq 1} X(s) < \infty$. Положим

$$K_4 = \inf \left\{ y : \frac{yA_1^\beta (\ln T)^\beta}{4} - r > a, \quad \forall x \geq y \right\}.$$

Тогда при $x \geq K_4$ неравенство (27) выполнено для всех $z := xA_1^\beta (\ln T)^\beta / 4 - r$, причем $z > a$.

Введем обозначения:

$$c_1 = \frac{4(r+a)}{A_1^\beta}; \quad c_2 = \frac{32v}{A_1^{2\beta}}.$$

Пусть теперь $K > K_4$ в (26). Тогда с учетом (27) после несложных преобразований можно получить, что

$$R_t \leq 4 \int_K^\infty y^p \exp \left[\ln t - \frac{(y(\ln t)^\beta - c_1)^2}{c_2} \right] dy. \quad (28)$$

Рассмотрим функцию

$$f(t, y) := \ln t - \frac{(y(\ln t)^\beta - c_1)^2}{c_2}.$$

Нетрудно убедиться, что

$$\frac{\partial f(t, y)}{\partial t} = \frac{1}{t} \left[1 - \frac{2\beta}{c_2} (y(\ln t)^\beta - c_1) (\ln t)^{\beta-1} \right] \quad (29)$$

и что при любом $y > K$

$$\frac{\partial f}{\partial t}(t, y) < \frac{\partial f}{\partial t}(t, K).$$

Анализ правой части выражения (29) показывает, что существует такое число t_2 , что функция $f(t, y)$ (при каждом $y > K$) убывает (по t) при $t \geq t_2$. Это, в свою очередь, означает, что

$$f(t, y) \leq f(t_2, y), \quad t \geq t_2, \quad y > K. \quad (30)$$

Обозначим $T = \max\{t_0, t_1, t_2\}$. Заметим, что $f(T, y) \rightarrow -\infty, y \rightarrow \infty$, и, как нетрудно проверить,

$$\lim_{y \rightarrow \infty} y^{p+2} e^{f(T, y)} = 0.$$

Поэтому найдется такое число $K_5 > 0$, что

$$e^{f(T, y)} < y^{-p-2}, \quad y > K_5. \quad (31)$$

Теперь, используя соотношения (30) и (31), получим из (28) при $K > K_5$ (и $t \geq T$)

$$R_t \leq 4 \int_K^\infty y^p e^{f(T,y)} dy \leq 4 \int_K^\infty \frac{dy}{y^2} = \frac{4}{K}. \quad (32)$$

Выберем окончательно в (13) (и далее, где требуется) $K = \max\{K_1, K_2, K_3, K_4, K_5\}$. Тогда из (22) и (32) следует неравенство

$$\mathbb{E} \left[\frac{M^*(t)}{\gamma^\beta(t)} \right]^{p+1} \leq K^{p+1} + \frac{2\beta(p+1)}{\gamma_1^{\beta+\beta p}} \Gamma(\gamma_1 K^{1/\beta}, \beta p + \beta) + \frac{4(p+1)}{K}, \quad (33)$$

правая часть которого не зависит от t . Беря в левой части неравенства (33) супремум по $t \geq T$, получаем требуемое условие (12). \square

Замечание. Если существует предел

$$\lim_{t \rightarrow \infty} L(t) = A \in (0, \infty), \quad (34)$$

то условие (11) автоматически выполнено.

Приведем важные для практических применений примеры, когда условия теоремы 2.1 выполнены.

Пример 1. Пусть стохастическая компонента входного процесса является суммой независимых ДБД, т. е.

$$X(t) = \sum_{i=1}^n B_{H_i}(t), \quad t \geq 0,$$

где параметры Херста $H_i \in (0, 1)$. Подробная мотивировка такого входного потока обсуждается в работе [12]. Без ограничения общности будем считать, что $H_1 > \max_{i>1} H_i$. Тогда дисперсия $v(t)$ процесса $\{X(t)\}$ имеет вид:

$$v(t) = \sum_{i=1}^n t^{2H_i} = t^{2H_1} L(t),$$

где медленно меняющаяся функция $L(t) = 1 + \sum_{i>1} t^{2(H_i - H_1)} \rightarrow 1, t \rightarrow \infty$. Таким образом, условия теоремы 2.1 выполнены.

Пример 2. Пусть стохастическая компонента входного процесса есть так называемый интегральный гауссовский процесс, т. е.

$$X(t) = \int_0^t Z(s) ds, \quad (35)$$

где Z — центрированный стационарный гауссовский процесс с ковариационной функцией $R(u) :=$

$:= \text{Cov}(Z(0), Z(u))$. Входные потоки такого типа рассматривались в работах [14, 13]. Нетрудно проверить, что для дисперсии $v(t)$ процесса X справедливо соотношение:

$$v(t) = 2 \int_0^t \int_0^s R(u) dud s. \quad (36)$$

Отсюда следует, что $v''(t) = 2R(t)$, а значит условие (7) влечет сходимость

$$R(t) \ln t \rightarrow 0, \quad t \rightarrow \infty.$$

Если дополнительно к условию (7) потребовать существования таких $A \in (0, \infty), V \in (0, 2)$, что

$$\frac{\int_0^t \int_0^s R(u) dud s}{t^V} \rightarrow A, \quad t \rightarrow \infty, \quad (37)$$

то условия теоремы 2.1 оказываются выполненными. Например, пусть Z — процесс Орнштейна–Уленбека, который по определению является центрированным стационарным гауссовским процессом. Поскольку его ковариационная функция имеет вид $R(t) = \lambda e^{-\alpha t}, \lambda, \alpha > 0$, то из (36) несложно получить, что

$$v(t) = \frac{2\lambda}{\alpha} t + \frac{2\lambda}{\alpha^2} (e^{-\alpha t} - 1).$$

Следовательно, условие (37) выполнено для $V = 1, A = \lambda/\alpha$. Отметим, что если Z — процесс Орнштейна–Уленбека, то формула (35) определяет *интегральный процесс Орнштейна–Уленбека*. Заметим, что этот процесс является гауссовским аналогом (т. е. гауссовским процессом с соответствующей корреляционной структурой) модели Аника–Митра–Сонди [15], описывающей динамику некоторых сетевых ресурсов (см. также [16]).

На самом деле в обоих примерах выше выполнено более сильное, чем (11), условие (34). Однако утверждение теоремы 2.1 верно и в случае, когда функция L не имеет предела при $t \rightarrow \infty$.

3 Дополнительные асимптотические результаты

В данном разделе получены два важных асимптотических результата для максимума процесса нагрузки $M(t)$, дополняющие анализ, проведенный в разд. 2. Вначале рассмотрим случай, когда

параметр $r < 0$. В соответствии с замечанием, сделанным после формулы (3), в этом случае система находится в нестационарном режиме и величина процесса нагрузки должна неограниченно возрастать.

Теорема 3.1. *Если $r < 0$, то имеет место следующая сходимость по распределению:*

$$\frac{M(t) + rt}{\sqrt{v(t)}} \xrightarrow{d} \mathcal{N}(0, 1), \quad t \rightarrow \infty. \quad (38)$$

Доказательство. Напомним, что процесс нагрузки в момент времени s определяется соотношением

$$Q(s) = W(s) - \min_{0 \leq u \leq s} W(u),$$

где $W(u) = X(u) - ru$. Поскольку с вероятностью 1 (с в. 1)

$$\frac{X(t)}{t} \rightarrow 0,$$

то также $W(t) \rightarrow +\infty$ с в. 1. Пусть $\Psi := \min_{t \geq 0} W(t)$.

Тогда существует случайный момент $T_0 < \infty$ (с в. 1) такой, что $\min_{t \geq 0} W(t) = \min_{0 \leq t \leq T_0} W(t)$. Поэтому справедлива следующая цепочка соотношений:

$$\begin{aligned} M(t) &= \max_{0 \leq s \leq t} \left[W(s) - \min_{0 \leq u \leq s} W(u) \right] \leq \\ &\leq \max_{0 \leq s \leq t} W(s) - \Psi = \\ &= \max_{0 \leq s \leq t} W(s) - \min_{0 \leq s \leq T_0} W(s). \end{aligned}$$

Отсюда следует неравенство:

$$\begin{aligned} M(t) - W(t) &\leq \\ &\leq \max_{0 \leq s \leq t} W(s) - W(t) - \min_{0 \leq s \leq T_0} W(s). \quad (39) \end{aligned}$$

Поскольку

$$\begin{aligned} M(t) + rt &= M(t) + X(t) - W(t); \\ \frac{X(t)}{\sqrt{v(t)}} &\xrightarrow{d} \mathcal{N}(0, 1), \end{aligned}$$

то сходимость (38) эквивалентна сходимости

$$\frac{M(t) - W(t)}{\sqrt{v(t)}} \Rightarrow 0, \quad t \rightarrow \infty. \quad (40)$$

Докажем справедливость (40). Для этого достаточно показать, что для любого $\varepsilon > 0$ справедливо соотношение:

$$\mathbb{P} \left(\frac{M(t) - W(t)}{\sqrt{v(t)}} > \varepsilon \right) \rightarrow 0, \quad t \rightarrow \infty. \quad (41)$$

В силу (39)

$$\begin{aligned} \mathbb{P} \left(\frac{M(t) - W(t)}{\sqrt{v(t)}} > \varepsilon \right) &\leq \\ &\leq \mathbb{P} \left(\frac{\max_{0 \leq s \leq t} W(s) - W(t)}{\sqrt{v(t)}} > \frac{\varepsilon}{2} \right) + \\ &+ \mathbb{P} \left(\frac{-\min_{0 \leq s \leq T_0} W(s)}{\sqrt{v(t)}} > \frac{\varepsilon}{2} \right) := \mathbb{P}_1(t) + \mathbb{P}_2(t). \end{aligned}$$

В силу стационарности приращений процесса X

$$\begin{aligned} \max_{0 \leq s \leq t} W(s) - W(t) &= \\ &= \max_{0 \leq s \leq t} [X(s) - X(t) + r(t - s)] =_d \\ &= \max_{0 \leq s \leq t} [X(t - s) + r(t - s)] = \\ &= \max_{0 \leq u \leq t} [X(u) + ru] := \tilde{Q}(t). \quad (42) \end{aligned}$$

Поскольку $r < 0$, то существует стационарный предел $\tilde{Q}(t) \xrightarrow{d} \tilde{Q}$ (при $t \rightarrow \infty$), причем $\tilde{Q} < \infty$ с в. 1 (см. замечание после формулы (2)). Поскольку $v(t) \rightarrow \infty$, то из (42) следует, что

$$\mathbb{P}_1(t) = \mathbb{P} \left(\max_{0 \leq s \leq t} W(s) - W(t) > \frac{\varepsilon}{2} \sqrt{v(t)} \right) \rightarrow 0, \quad t \rightarrow \infty.$$

Рассмотрим вероятность $\mathbb{P}_2(t)$ и заметим, что

$$\begin{aligned} -\min_{0 \leq s \leq T_0} W(s) &= \max_{0 \leq s \leq T_0} [rs - X(s)] \leq \\ &\leq \max_{0 \leq s \leq T_0} [-X(s)] =_d \max_{0 \leq s \leq T_0} X(s). \end{aligned}$$

Поэтому

$$\begin{aligned} \mathbb{P}_2(t) &= \mathbb{P} \left(-\min_{0 \leq s \leq T_0} W(s) > \frac{\varepsilon}{2} \sqrt{v(t)} \right) \leq \\ &\leq \mathbb{P} \left(\max_{0 \leq s \leq T_0} X(s) > \frac{\varepsilon}{2} \sqrt{v(t)} \right) \rightarrow 0, \quad t \rightarrow \infty, \end{aligned}$$

где учитывается, что случайная величина T_0 , а значит и $\max_{0 \leq s \leq T_0} X(s)$, конечны с в. 1. Таким образом, соотношение (41), а значит и (40), выполнено. \square

Следующий результат касается асимптотики времени достижения стационарным процессом нагрузки $Q^*(t)$ растущего порога b , т.е. асимптотики величины

$$T(b) = \inf\{t \geq 0 : Q^*(t) \geq b\}$$

при $b \rightarrow \infty$. Распределение максимума стационарного процесса нагрузки $M^*(t)$ определяет распределение случайной величины $T(b)$ в силу очевидного соотношения

$$\{T(b) \leq t\} = \{M^*(t) \geq b\}, \quad t \geq 0. \quad (43)$$

Напомним обозначение $\gamma(t) = \ln t \cdot L((\ln t)^\beta)$.

Теорема 3.2. Пусть в дополнение к условиям теоремы 1.1 функция $\gamma(t)$ монотонно возрастает на некотором луче $[t_0, \infty)$. Тогда имеет место сходимость

$$\frac{\gamma(T(b))}{b^{1/\beta}} \Rightarrow \theta, \quad b \rightarrow \infty, \quad (44)$$

где параметр θ удовлетворяет соотношению (10).

Доказательство. В силу теоремы 1.1 для любого $\delta > 0$

$$\mathbb{P} \left(M^*(t) > \left(\frac{1+\delta}{\theta} \gamma(t) \right)^\beta \right) \rightarrow 0, \quad t \rightarrow \infty. \quad (45)$$

Ввиду монотонного возрастания функции γ , обратная ей функция также монотонно возрастает. В частности, при любом $\delta > 0$ функция

$$t(b) := \gamma^{-1} \left(b^{1/\beta} \frac{\theta}{1+\delta} \right) \rightarrow \infty, \quad b \rightarrow \infty. \quad (46)$$

Подставляя (46) в (45) и учитывая (43), получим

$$\begin{aligned} \mathbb{P} \left(M^*(t(b)) \geq \left(\frac{1+\delta}{\theta} \gamma(t(b)) \right)^\beta \right) &= \\ &= \mathbb{P} (M^*(t(b)) \geq b) = \mathbb{P} (T(b) \leq t(b)) = \\ &= \mathbb{P} \left(T(b) \leq \gamma^{-1} \left(b^{1/\beta} \frac{\theta}{1+\delta} \right) \right) = \\ &= \mathbb{P} \left(\frac{\gamma(T(b))}{b^{1/\beta}} \leq \frac{\theta}{1+\delta} \right) \rightarrow 0, \quad b \rightarrow \infty. \end{aligned} \quad (47)$$

Снова используя монотонность функции γ , получим (для любого фиксированного $\delta > 0$):

$$\hat{t}(b) := \gamma^{-1} \left(b^{1/\beta} \frac{\theta}{1-\delta} \right) \rightarrow \infty, \quad b \rightarrow \infty.$$

С учетом того, что по теореме 1.1

$$\mathbb{P} \left(M^*(t) > \left(\frac{1-\delta}{\theta} \gamma(t) \right)^\beta \right) \rightarrow 1, \quad t \rightarrow \infty,$$

как и выше, получим:

$$\begin{aligned} \mathbb{P} \left(M^*(\hat{t}(b)) > \left(\frac{1-\delta}{\theta} \gamma(\hat{t}(b)) \right)^\beta \right) &= \\ &= \mathbb{P} \left(\frac{\gamma(T(b))}{b^{1/\beta}} \leq \frac{\theta}{1-\delta} \right) \rightarrow 1, \quad b \rightarrow \infty. \end{aligned}$$

Ввиду произвольности δ отсюда и из (47) следует (44). \square

4 Заключение

В данной статье продолжен (начатый в работе [1]) асимптотический анализ максимума процесса нагрузки в системе обслуживания, в которой дисперсия гауссовской компоненты входного процесса правильно меняется на бесконечности с показателем $V \in (0, 2)$. В частности, показано, что при некотором дополнительном условии доказанная в [1] сходимость по вероятности указанного процесса имеет место и в пространстве L_p при любом $p \in [1, \infty)$.

Также найдена асимптотика максимума процесса нагрузки в нестационарном режиме (при соответствующей нормировке). Кроме того, с использованием полученной асимптотики максимума найдена асимптотика времени достижения стационарным процессом нагрузки растущего порога b .

Литература

1. Лукашенко О. В., Морозов Е. В. Асимптотика максимума процесса нагрузки для некоторого класса гауссовских очередей // Информатика и её применения, 2012. Т. 6. Вып. 3. С. 81–89.
2. Zeevi A., Glynn P. On the maximum workload in a queue fed by fractional Brownian motion // Ann. Appl. Prob., 2000. Vol. 10. P. 1084–1099.
3. Mandjes M. Large deviations of Gaussian queues. — Chichester: Wiley, 2007. 339 p.
4. Reich E. On the integrodifferential equation of Takacs I // Ann. Math. Stat., 1958. Vol. 29. P. 563–570.
5. Asmussen S. Applied probability and queues. — New York: Springer, 2002. 440 p.
6. Сенета Е. Правильно меняющиеся функции. — М.: Наука, 1985. 143 с.
7. Konstantopoulos T., Zazanis M., De Veciana G. Conservation laws and reflection mappings with application to multiclass mean value analysis for stochastic fluid queues // Stochastic Processes and Their Applications, 1996. Vol. 65. P. 139–146.
8. Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977. 352 с.
9. Duffy K., Lewis J. T., Sullivan W. G. Logarithmic asymptotics for the supremum of a stochastic process // Ann. Appl. Prob., 2003. Vol. 13. No. 2. P. 430–445.
10. Adler R. J. An introduction to continuity, extrema, and related topics for general Gaussian processes. — Hayward, CA: Institute of Mathematical Statistics, 1990. 170 p.

11. *Лифшиц М. А.* Гауссовские случайные функции. — Киев: ТвиМС, 1995. 248 с.
12. *Taqqu M. S., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modeling // *Computer Comm. Rev.*, 1997. Vol. 27. P. 5–23.
13. *Kulkarni V., Rolski T.* Fluid model driven by an Ornstein–Uhlenbeck process // *Probability Engineering Information Sci.*, 1994. Vol. 8. P. 403–417.
14. *Debicki K., Rolski T.* A Gaussian fluid model // *Queueing Syst.*, 1995. Vol. 20. P. 433–452.
15. *Anick D., Mitra D., Sondhi M. M.* Stochastic theory of a data handling system with multiple resources // *Bell Syst. Techn. J.*, 1982. Vol. 61. P. 1871–1894.
16. *Addie R., Mannersalo P., Norros I.* Most probable paths and performance formulae for buffers with Gaussian input traffic // *Eur. Trans. Telecommunications*, 2002. Vol. 13. P. 183–196.

АЛГОРИТМЫ ИНДУКТИВНОГО ПОРОЖДЕНИЯ СУПЕРПОЗИЦИЙ ДЛЯ АППРОКСИМАЦИИ ИЗМЕРЯЕМЫХ ДАННЫХ*

Г. И. Рудой¹, В. В. Стрижов²

Аннотация: Исследуется алгоритм индуктивного порождения допустимых существенно нелинейных моделей. Предлагается алгоритм, порождающий все возможные суперпозиции заданной сложности за конечное число шагов. Приводятся результаты вычислительного эксперимента по выбору оптимальной модели, аппроксимирующей синтетический набор данных.

Ключевые слова: символьная регрессия; нелинейные модели; индуктивное порождение; сложность моделей

1 Введение

В ряде приложений [1, 2] возникает задача восстановления некоторой функциональной зависимости по набору известных данных. При этом предполагается, что эксперт должен иметь возможность проинтерпретировать полученную модель в контексте предметной области.

Одним из методов, позволяющих получать интерпретируемые модели, является символьная регрессия [3–7], согласно которой известные данные приближаются некоторой математической формулой, например $\sin x^2 + 2x$ или $\log x - e^x/x$. Эти формулы являются произвольными суперпозициями функций из некоторого заданного набора. Одна из возможных реализаций описываемого метода предложена Джоном Коза [8, 9], использовавшим эволюционные алгоритмы для реализации символьной регрессии. Зелинка с соавторами предложили дальнейшее развитие этой идеи [10], получившее название аналитического программирования.

Алгоритм построения требуемой математической модели в аналитическом программировании выглядит следующим образом: дан набор примитивных функций, из которых можно строить различные формулы (например, степенная функция, $+$, \sin , \tan). Начальный набор формул строится либо произвольным образом, либо на базе некоторых предположений эксперта. Затем на каждом шаге производится оценка каждой из формул согласно функции ошибки либо другого функционала качества [11]. На базе этой оценки у некоторой части формул случайным образом заменяется одна элементарная функция на другую (например, \sin на \cos или $+$ на \times), а у некоторой другой части происходит взаимный попарный обмен подвыражениями.

Получаемая формула является математической моделью исследуемого процесса или явления, т. е. представляет собой математическое отношение, описывающее основные закономерности, присущие этому явлению [12].

Алгоритм индуктивного порождения моделей, предложенный в настоящей работе, свободен от некоторых типичных проблем известных методов, упомянутых, например, в [10]. Вот главные из них:

- порождение рекурсивных суперпозиций, суперпозиций, содержащих несоответствующее используемым функциям число аргументов, и т. д. (в предложенном алгоритме эти проблемы не возникают по построению);
- несовпадение области определения некоторой примитивной функции и области значений ее аргументов (возможно, тоже некоторых суперпозиций);
- порождение слишком сложных суперпозиций.

Для любой выборки можно построить такой многочлен, который пройдет через все точки выборки, но при этом число параметров такого многочлена линейно растет с объемом выборки. Кроме того, такой многочлен неинтерпретируем экспертами. Предложенный в настоящей работе алгоритм решает проблему порождения слишком сложных суперпозиций введением дополнительного штрафа за сложность. Кроме того, так как используемые признаки объектов выборки учитываются при расчете сложности, применение подобного штрафа обеспечивает выбор суперпозиций, использующих меньшее число признаков, т. е. проводит отбор признаков.

* Работа выполнена при поддержке РФФИ, грант № 12-07-13118.

¹Московский физико-технический институт, rudoy@forecsys.ru

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

Во второй части работы формально поставлена задача построения алгоритма индуктивного порождения моделей. Затем, в третьей части, строится искомый алгоритм для частного случая беспараметрических моделей и доказывается его корректность, а затем алгоритм обобщается на случай моделей, имеющих параметры. В четвертой части оценивается количество порожденных предложенным алгоритмом моделей на каждой итерации. В пятой части предлагается метод выбора допустимых моделей из множества всех порожденных моделей. В седьмой части описывается адаптированный стохастический алгоритм порождения моделей, результаты работы которого на синтетических данных приведены в восьмой части настоящей работы.

2 Постановка задачи

Пусть дана выборка

$$D = \{(\mathbf{x}_i, y_i) \mid i \in \{1, \dots, N\}, \\ \mathbf{x}_i \in \mathbb{X} \subset \mathbb{R}^n, y_i \in \mathbb{Y} \subset \mathbb{R}\},$$

где N — число элементов выборки, \mathbf{x}_i — вектор значений свободных переменных для i -го элемента выборки, y_i — значение зависимой переменной для i -го элемента выборки, \mathbb{X} — множество значений независимых переменных, лежащее в \mathbb{R}^n , \mathbb{Y} — множество значений зависимой переменной.

Требуется выбрать параметрическую функцию $f : \Omega \times \mathbb{X} \rightarrow \mathbb{R}$ из порождаемого множества $\mathcal{F} = \{f_r\}$, где Ω — пространство параметров, доставляющую минимум некоторому заданному функционалу качества Q , зависящему от функционала ошибки S на данной выборке D и сложности суперпозиции $C(f)$.

Таким образом, для множества всех суперпозиций

$$\mathcal{F} = \{f_r \mid f_r : (\omega, \mathbf{x}) \mapsto y \in \mathbb{Y}, r \in \mathbb{N}\}$$

требуется найти такой индекс \hat{r} , при котором функция $f_{\hat{r}}$ среди всех $f \in \mathcal{F}$ доставляет минимум функционалу качества Q при данной выборке D :

$$\hat{r} = \arg \min_{r \in \mathbb{N}} Q(f_r \mid \hat{\omega}_r, D),$$

где $\hat{\omega}_r$ — оптимальный вектор параметров функции f_r для каждой $f \in \mathcal{F}$ при данной выборке D :

$$\hat{\omega}_r = \arg \min_{\omega \in \Omega} S(\omega \mid f_r, D).$$

Сформулируем также постановку теоретической задачи. Для этого сначала введем понятие суперпозиции функций.

Если множество значений \mathbb{Y}_i функции f_i содержится в области определения \mathbb{X}_{i+1} функции f_{i+1} , т. е.

$$f_i : \mathbb{X}_i \rightarrow \mathbb{Y}_i \subset \mathbb{X}_{i+1}, \quad i = 1, 2, \dots, \theta - 1,$$

то функция

$$f_\theta \circ f_{\theta-1} \circ \dots \circ f_1, \quad \theta \geq 2,$$

определяемая равенством

$$(f_\theta \circ f_{\theta-1} \circ \dots \circ f_1)(\mathbf{x}) = f_\theta(f_{\theta-1}(\dots(f_1(\mathbf{x}))))), \quad x \in \mathbb{X}_1,$$

называется *сложной функцией* [13] или *суперпозицией функций* $f_1, f_2, \dots, f_\theta$.

Таким образом, получаем

Определение 1. *Суперпозиция функций — функция, представленная как композиция нескольких функций.*

Пусть $G = \{g_1, \dots, g_l\}$ — множество данных порождающих функций, а именно: для каждой $g_i \in G$ заданы

- сама функция g_i (например, \sin, \cos, \times);
- аргументы функции и порядок следования аргументов;
- домен ($\text{dom } g_i$) и кодомен ($\text{cod } g_i$) функции;
- область определения $\mathcal{D}g_i \subset \text{dom } g_i$ и область значений $\mathcal{E}g_i \subset \text{cod } g_i$.

Требуется построить упомянутую функцию f как суперпозицию порождающих функций из заданного множества G .

Поясним различие между последними двумя пунктами. Например, $\text{dom } f$ показывает, значения из какого множества принимает функция f (целые числа, действительные числа, декартово произведение целых чисел и $\{0, 1\}$, и т. п.). Область определения же показывает, на каких значениях из $\text{dom } f$ функция f определена и имеет смысл. Так, для функции $f(x_1, x_2) = \log_{x_1} x_2$:

$$\text{dom } f = \mathbb{R} \times \mathbb{R}, \quad \text{cod } f = \mathbb{R};$$

$$\mathcal{D}f = \{(x_1, x_2) \mid x_1 \in (0; 1) \cup (1; +\infty), x_2 \in (0; +\infty)\};$$

$$\mathcal{E}f = (-\infty; +\infty).$$

Требуется также:

- построить алгоритм \mathcal{A} , за конечное число итераций порождающий любую конечную суперпозицию данных примитивных функций;
- указать способ проверки изоморфности двух суперпозиций.

Заметим, что для примитивных функций не требуются свойства их непорождаемости в наиболее общей формулировке типа принципиальной невозможности породить в ходе работы искомого алгоритма суперпозицию, изоморфную некоторой функции из G . Такое требование является слишком ограничивающим. В частности, невозможно было бы иметь в G одновременно, например, функции id , exp и log , так как $\text{id} \equiv \text{log} \circ \text{exp}$.

В дальнейшем будем также считать, что суперпозиция, соответствующая единственной свободной переменной ($f(x) = x_i$), эквивалентна функции вида $\text{id } x_i$.

3 Алгоритм индуктивного порождения допустимых суперпозиций

Условимся считать, что каждой суперпозиции f сопоставлено дерево Γ_f , эквивалентное этой суперпозиции и строящееся следующим образом:

- в вершинах V_i дерева Γ_f находятся соответствующие порождающие функции $g_s, s = s(i)$;
- число дочерних вершин у некоторой вершины V_i равно арности соответствующей функции g_s ;
- порядок смежных некоторой вершине V_i вершин соответствует порядку аргументов соответствующей функции $g_s(i)$;
- в листьях дерева Γ_f находятся свободные переменные x_i либо числовые параметры ω_i ;
- порядок вершин V_i в смысле уровня вершин определяет порядок вычисления примитивных функций: дерево вычисляется снизу вверх, т. е. сначала подставляются конкретные значения свободных переменных, затем вычисляются значения в вершинах, все дочерние вершины которых — свободные переменные, и так далее до тех пор, пока не останется единственная вершина, бывшая корнем дерева. Она и содержит результат соответствующего выражения.

Таким образом, вычисление значения выражения f в некоторой точке с данным вектором параметров $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ эквивалентно подстановке соответствующих значений свободных переменных x_i и параметров ω_i в дерево Γ_f , где x_i — компоненты вектора признакового описания объекта x .

Заметим важное свойство таких деревьев: каждое поддерево Γ_f^i дерева Γ_f , соответствующее вершине V_i , также соответствует некоторой суперпози-

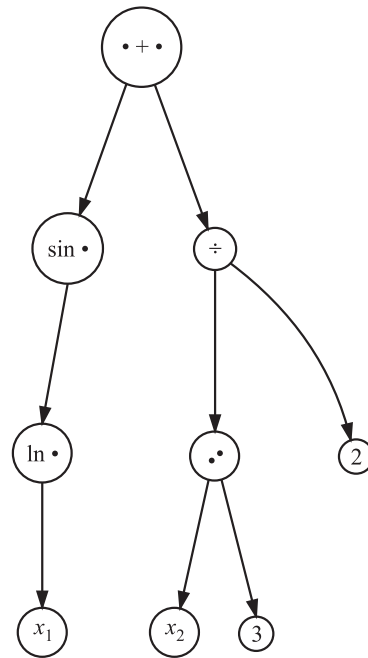


Рис. 1 Дерево выражения $\sin(\ln x_1) + x_2^3/2$

ции, являющейся составляющей исходной суперпозиции f .

Для примера рассмотрим дерево, соответствующее суперпозиции $f = \sin(\ln x_1) + x_2^3/2$ (рис. 1). Здесь точками обозначены аргументы функций. Как видно, корнем дерева является вершина, соответствующая операции сложения, которая должна быть выполнена в последнюю очередь. Операция сложения имеет два различных поддерева, соответствующих двум аргументам этой операции. Заметим также, что здесь не использованы операции типа «разделить на два» или «возвести в куб». Вместо этого используются операции деления и возведения в степень в общем виде, а в данном конкретном дереве соответствующие аргументы зафиксированы соответствующими константами.

Алгоритм порождения суперпозиций. Сначала определим понятие *глубины суперпозиции*:

Определение 2. Глубина суперпозиции f — максимальная глубина дерева Γ_f .

Теперь опишем итеративный алгоритм \mathcal{A}^* , порождающий суперпозиции, не содержащие параметров. Описанный алгоритм породит любую суперпозицию конечной глубины за конечное число шагов.

Пусть дано множество примитивных функций $G = \{g_1, \dots, g_l\}$ и множество свободных переменных $X = \{x_1, \dots, x_n\}$. Для удобства будем исходить из предположения, что множество G состоит только

из унарных и бинарных функций, и разделим его соответствующим образом на два подмножества: $G = G_b \cup G_u$ | $G_b = \{g_{b_1}, \dots, g_{b_k}\}$, $G_u = \{g_{u_1}, \dots, g_{u_l}\}$, где G_b — множество всех бинарных функций, а G_u — множество всех унарных функций из G . Потребуем также наличия id в G_b .

Алгоритм 1. Алгоритм \mathcal{A}^* итеративного порождения суперпозиций.

1. Перед первым шагом зададим начальные значения множества \mathcal{F}_0 и вспомогательного индексного множества \mathcal{I} , служащего для запоминания, на какой итерации впервые встречена каждая суперпозиция:

$$\mathcal{F}_0 = X; \quad \mathcal{I} = \{(x, 0) \mid x \in X\}.$$

2. Для множества \mathcal{F}_i построим вспомогательное множество U_i , состоящее из суперпозиций, полученных в результате применения функций $g_u \in G_u$ к элементам \mathcal{F}_i :

$$U_i = \{g_u \circ f \mid g_u \in G_u, f \in \mathcal{F}_i\}.$$

3. Аналогичным образом построим вспомогательное множество B_i для бинарных функций $g_b \in G_b$:

$$B_i = \{g_b \circ (f, h) \mid g_b \in G_b, f, h \in \mathcal{F}_i\}.$$

4. Обозначим $\mathcal{F}_{i+1} = \mathcal{F}_i \cup U_i \cup B_i$.
5. Для каждой суперпозиции f из \mathcal{F}_{i+1} добавим пару $(f, i+1)$ в множество \mathcal{I}_f , если суперпозиция f еще там не присутствует.
6. Перейдем к следующей итерации, п. 2.

Тогда $\mathcal{F} = \cup_{i=0}^{\infty} \mathcal{F}_i$ — множество всех возможных суперпозиций конечной длины, которые можно построить из данного множества примитивных функций.

Вспомогательное множество \mathcal{I} позволяет запоминать, на какой итерации была впервые встречена каждая суперпозиция. Это необходимо, так как каждая суперпозиция, впервые порожденная на i -й итерации, будет порождена также и на любой итерации после i . Одной из возможностей избежать необходимости в этом множестве является построение \mathcal{F}_{i+1} как $\mathcal{F}_{i+1} = U_i \cup B_i$ (без \mathcal{F}_i), а множества U_i и B_i строить следующим образом:

$$U_i = \{g_u \circ f \mid g_u \in G_u, f \in \cup_{j=0}^i \mathcal{F}_j\};$$

$$B_i = \{g_b \circ (f, h) \mid g_b \in G_b, f, h \in \cup_{j=0}^i \mathcal{F}_j\}.$$

Алгоритм \mathcal{A}^* очевидным образом обобщается на случай, когда множество G содержит функции произвольной (но конечной) арности. Действительно, для такого обобщения достаточно строить

аналогичным образом вспомогательные множества для этих функций, а именно: для множества функций G_n арности n построить вспомогательное множество H_i^n вида

$$H_i^n = \{g \circ (f_1, f_2, \dots, f_n) \mid g \in G_n, f_j \in \mathcal{F}_i\}.$$

В этих обозначениях $U_i \equiv H_i^1$, а $B_i \equiv H_i^2$.

Тогда множество $\mathcal{F}_{i+1} = \mathcal{F}_i \cup_{n=0}^{n_{\max}} H_i^n$, где n_{\max} — максимальное значение арности функций из G .

Теорема 1. Алгоритм \mathcal{A}^* действительно породит любую конечную суперпозицию за конечное число шагов.

Доказательство. Чтобы убедиться в этом, найдем номер итерации, на которой будет порождена некоторая произвольная конечная суперпозиция f . Чтобы найти этот номер, пронумеруем вершины графа Γ_f по следующим правилам:

- если это вершина со свободной переменной, то она имеет номер 0;
- если вершина V соответствует унарной функции, то она имеет номер $i+1$, где i — номер дочерней для этой функции вершины;
- если вершина V соответствует бинарной функции, то она имеет номер $i+1$, где $i = \max(l, r)$, а l и r — номера соответственно первой и второй дочерней вершины.

Нумеруя вершины графа Γ_f таким образом, можно получить номер вершины, соответствующей корню графа. Это и будет номером итерации, на которой получена суперпозиция f .

Иными словами, для любой суперпозиции можно указать конкретный номер итерации, на которой она будет получена, что и требовалось.

В предложенных ранее методах построения суперпозиций [10] необходимо было самостоятельно следить за тем, чтобы в ходе работы алгоритма не возникало «зацикленных» суперпозиций типа $f(x, y) = g(f(x, y), x, y)$. Заметим, что в предложенном алгоритме \mathcal{A}^* такие суперпозиции не могут возникнуть по построению.

Порождение моделей с параметрами. Алгоритм \mathcal{A}^* , описанный выше, не позволяет получать выражения, содержащие численные параметры ω суперпозиции $f(\omega, \mathbf{x})$. Покажем, однако, на примере конструирования множеств U_i и B_i , как исходный алгоритм \mathcal{A}^* может быть расширен путем введения параметров

$$U_i = g_u \circ (\alpha f + \beta);$$

$$B_i = g_b \circ (\alpha f + \beta, \psi h + \phi).$$

Будем обозначать этот расширенный алгоритм как \mathcal{A} . Здесь параметры α, β зависят только от

комбинации g_a, f (или g_b, f, h для $\alpha, \beta, \psi, \phi$). Соответственно, для упрощения их индексы опущены. Иными словами, предполагается, что каждая суперпозиция, полученная на предыдущих итерациях, входит в порождаемую на следующей итерации, будучи умноженной на некоторый коэффициент и с константной поправкой.

Очевидно, при таком добавлении параметров $\alpha, \beta, \psi, \phi$ не происходит изменения мощности получившегося множества суперпозиций, поэтому алгоритм и выводы из него остаются корректными. В частности, исходный алгоритм является частным случаем данного при $\alpha \equiv \psi \equiv 1, \beta \equiv \phi \equiv 0$.

Переменные $\alpha, \beta, \psi, \phi$ являются параметрами модели. В практических приложениях можно оптимизировать значения этих параметров у получившихся суперпозиций, например, алгоритмом Левенберга–Марквардта [14, 15].

Заметим также, что такая модификация алгоритма позволяет получить единицу, например, для построения суперпозиций типа $1/x$: $1 = \alpha \text{ id } x + \beta \mid \alpha = 0, \beta = 1$.

Отдельно подчеркнем, что параметры ω у разных суперпозиций различны. Однако, так как каждый из параметров зависит только от соответствующей комбинации функций, к которым он относится, конкретные значения параметров не учитываются при поиске одинаковых суперпозиций. Иными словами, при тестировании суперпозиций на равенство сравниваются лишь структуры соответствующих им деревьев и значения в узлах, соответствующих функциям и свободным переменным.

Заметим, что и этот алгоритм очевидным образом обобщается на случай множества G , содержащего функции произвольной арности.

4 Число возможных суперпозиций

Оценим число суперпозиций, получаемых после каждой итерации алгоритма \mathfrak{A} . Очевидно, с учетом вышеупомянутых оговорок касательно сравнения параметризованных суперпозиций, это число равно аналогичному числу для алгоритма \mathfrak{A}^* .

Итак, пусть дано n независимых переменных: $|X| = n$, а мощность множества G распишем через мощности его подмножеств функций соответствующей арности: $|G_1| = l_1, |G_2| = l_2, \dots, |G_p| = l_p$. На нулевой итерации имеем $P_0 = n$ суперпозиций.

На первой итерации дополнительно порождается

$$P_1 = l_1 n + l_2 n^2 + \dots + l_p n^p = \sum_{i=1}^p l_i P_0^i,$$

и суммарное число суперпозиций после первой итерации

$$\hat{P}_1 = P_1 + P_0 = \sum_{i=1}^p l_i P_0^i + P_0.$$

Как было замечено ранее, суперпозиции, порожденные на k -й итерации, будут также порождены и на любой следующей после k итерации, поэтому суммарное число суперпозиций после второй итерации будет равно

$$\hat{P}_2 = \sum_{i=1}^p l_i \hat{P}_1^i.$$

И вообще, после k -й итерации будет порождено

$$\hat{P}_k = \sum_{i=1}^p l_i \hat{P}_{k-1}^i.$$

Оценим порядок роста количества функций, порожденных после k -й итерации.

Теорема 2. Пусть в множестве примитивных функций G содержится l_p функций арности $p > 1$ и ни одной функции арности $p + k \mid k > 0$ и имеется $n > 1$ независимых переменных. Тогда справедлива следующая оценка числа суперпозиций, порожденных алгоритмом \mathfrak{A} после k -й итерации:

$$|\mathcal{F}_k| = \mathcal{O} \left(l_p^{\sum_{i=0}^{k-1} p^i} n^{p^k} \right).$$

Доказательство. Оценим сначала порядок роста для случая, когда есть лишь одна m -арная функция и n свободных переменных.

После первой итерации алгоритма будет порождено $n^m + n$ суперпозиций. После второй — $(n^m + n)^m + n^m + n$, что можно оценить как $(n^m)^m = n^{m^2}$. И вообще, после k -й итерации число суперпозиций можно оценить как n^{m^k} .

Видно, что для оценки скорости роста количества порожденных суперпозиций можно учитывать только функции с наибольшей арностью.

Рассмотрим теперь случай, когда имеется не одна функция арности m , а l_m таких функций. Тогда на первой итерации порождается $l_m n^m + n$ суперпозиций, на второй:

$$l_m (l_m n^m + n)^m + l_m n^m + n \approx l_m^{m+1} n^{m^2},$$

на третьей, с учетом этого приближения:

$$l_m (l_m^{m+1} n^{m^2})^m = l_m l_m^{m(m+1)} n^{m^3} = l_m^{m^2+m+1} n^{m^3}.$$

И вообще, скорость роста количества порожденных суперпозиций можно оценить как:

$$|\mathcal{F}_k| = \mathcal{O} \left(l_m^{\sum_{i=0}^{k-1} m^i} n^{m^k} \right).$$

Таким образом, получаем оценку в общем случае, когда в множестве G содержится l_p функций арности p и ни одной функции арности $p+k \mid k > 0$:

$$|\mathcal{F}_k| = \mathcal{O} \left(l_p^{\sum_{i=0}^{k-1} p^i} n^{p^k} \right).$$

5 Множество допустимых суперпозиций

Предложенный выше алгоритм позволяет получить действительно все возможные суперпозиции, однако не все они будут пригодны в практических приложениях: например, $\ln x$ имеет смысл только при $x > 0$, а $x/0$ не имеет смысла вообще никогда. Выражения типа $x/\sin x$ имеют смысл только при $x \neq \pi k$.

Таким образом, необходимо введение понятия множества *допустимых* суперпозиций, т.е. таких суперпозиций, которые в условиях данной задачи корректны.

Определение 3. *Допустимая суперпозиция f — такая суперпозиция, значение которой определено для любой комбинации значений свободных переменных, область значений \mathbb{X} которых определяется конкретной задачей, $\mathbb{X} \subset \mathbb{R}^n$, где n — число свободных переменных.*

Одним из способов построения только допустимых суперпозиций является модификация предложенного алгоритма таким образом, чтобы отслеживать совместность областей определения и областей значений соответствующих функций в ходе построения суперпозиций. Для свободных переменных это, в свою очередь, означает необходимость задания областей значений \mathbb{X} пользователем при решении конкретных задач.

Таким образом, можно сформулировать очевидное *достаточное условие недопустимости* суперпозиции:

Определение 4. *Достаточное условие недопустимости суперпозиции f : в соответствующем дереве Γ_f хотя бы одна вершина V_i имеет хотя бы одну дочернюю вершину V_j такую, что область значений функции $g_{s(j)}$ шире, чем область определения функции $g_{s(i)}$:*

$$\exists i, j : V_i \in \Gamma_f, V_j \in \Gamma_f \wedge \exists \kappa : \kappa \in \mathcal{E}g_{s(j)} \wedge \kappa \notin \mathcal{D}g_{s(i)}.$$

Говоря, что область значений функции f шире области определения функции g , имеем в виду,

что существует, по крайней мере, одно значение функции f , не входящее в область определения функции g .

Подчеркнем, что, хотя свободные переменные могут принимать, например, все значения из \mathbb{R} , выбором множества \mathbb{X} можно обеспечить возможность использования их в качестве аргументов функций с более узкой, чем \mathbb{R} , но не менее узкой, чем \mathbb{X} , областью определения, если это не противоречит данной выборке.

Для построения множества допустимых суперпозиций достаточно построить множество всех возможных суперпозиций при помощи алгоритма \mathcal{A} , а затем удалить из этого множества все суперпозиции, не удовлетворяющие сформулированному признаку.

6 Алгоритм итеративного стохастического порождения суперпозиций

Несмотря на то что построенный ранее итеративный алгоритм \mathcal{A} порождения суперпозиций позволяет получить за конечное число шагов произвольную суперпозицию, для практических применений он непригоден в связи с чрезмерной вычислительной сложностью, как и любой алгоритм, реализующий полный перебор. Вместо него предлагается использовать стохастические алгоритмы и ряд эвристик, позволяющих на практике получать за приемлемое время результаты, удовлетворяющие заранее заданным условиям. В данном разделе описывается практически реализуемый вариант алгоритма \mathcal{A} , который и был использован в вычислительном эксперименте. Опишем вспомогательный алгоритм случайного порождения суперпозиции.

Алгоритм 2. Алгоритм случайного порождения суперпозиции $\mathcal{R}\mathcal{F}$.

Вход:

- набор пороговых значений $0 < \xi_1 < \xi_2 < \xi_3 < 1$;
- максимальная глубина порождаемой суперпозиции Td .

Алгоритм работает следующим образом. Генерируется случайное число ξ на интервале $(0; 1)$ и рассматриваются следующие случаи:

- $\xi \leq \xi_1$: результатом алгоритма является некоторая случайно выбранная свободная переменная;
- $\xi_1 < \xi \leq \xi_2$: результатом алгоритма является числовой параметр;

- $\xi_2 < \xi \leq \xi_3$: результатом алгоритма является некоторая случайно выбранная унарная функция, для определения аргумента которой данный алгоритм рекурсивно запускается еще раз;
- $\xi_3 < \xi$: результатом алгоритма является некоторая случайно выбранная бинарная функция, аргументы которой порождаются аналогичным образом.

При этом порождение тривиальных суперпозиций (свободных переменных и параметров) запрещено: на самом первом шаге пороговые значения масштабируются таким образом, чтобы всегда порождалась унарная или бинарная функция. Аналогично при превышении значения Td пороговые значения масштабируются таким образом, чтобы был порожден узел, соответствующий свободной переменной или параметру, и алгоритм завершился.

Каждой порожденной суперпозиции f ставится в соответствие ее качество Q_f , рассчитываемое исходя из функционала ошибки S_f этой суперпозиции на выборке D и ее сложности C_f — числа узлов в соответствующем графе Γ_f . Функционал Q_f выбирается эвристически с учетом следующих естественных соображений:

- из двух суперпозиций одинаковой сложности C_f выбирается обеспечивающая более оптимальное значение функционала ошибки S_f ;
- из двух суперпозиций, имеющих одно и то же значение функционала ошибки S_f , выбирается суперпозиция, обладающая меньшей сложностью C_f .

Алгоритм 3. Итеративный алгоритм стохастического порождения суперпозиций.

Вход:

- множество порождающих функций G , состоящее только из унарных и бинарных функций;
- регрессионная выборка D ;
- N_{\max} — максимальное число одновременно рассматриваемых суперпозиций;
- I_{\max} — максимальное число итераций алгоритма;
- \hat{Q} — минимальное значение функционала Q_f :

$$Q_f = \frac{1}{1 + S_f} \left(\alpha + \frac{1 - \alpha}{1 + \exp(C_f/\beta - \tau)} \right), \quad (1)$$

где α — некоторый коэффициент влияния штрафа за сложность, $0 \ll \alpha < 1$, β — коэффициент строгости штрафа за сложность, $\beta > 0$, а τ — коэффициент, характеризующий желаемую сложность модели;

- γ_{mut} — доля суперпозиций, подверженных случайной замене узлов их деревьев;
- γ_{cross} — доля суперпозиций, для которых выполняется случайный обмен поддеревьями;
- прочие параметры, используемые в (1) и алгоритме 2.

1. Инициализируется упорядоченный набор \mathcal{X}_f суперпозиций, а именно: порождается N_{\max} суперпозиций алгоритмом 2.
2. Оптимизируются параметры ω суперпозиций из \mathcal{X}_f алгоритмом Левенберга–Марквардта.
3. Выполняются простейшие преобразования, упрощающие суперпозицию: например, выражения вида $0 \cdot x$ заменяются на 0.
4. Вычисляется значение Q_f для каждой еще не оцененной суперпозиции f из \mathcal{X}_f : для нее рассчитывается значение функционала ошибки S_f на выборке D и ставится в соответствие значение Q_f . Для суперпозиций, при вычислении Q_f которых была хотя бы раз получена ошибка вычислений из-за несовпадения областей определений и значений, принимается $Q_f = -\infty$.
5. Набор суперпозиций \mathcal{X}_f сортируется согласно значениям функционала Q_f .
6. Суперпозиции с наименьшими значениями Q_f удаляются из массива \mathcal{X}_f до тех пор, пока его размер не станет равен N_{\max} .
7. Отбирается некоторая часть γ_{mut} суперпозиций с наименьшими значениями Q_f из \mathcal{X}_f . У этой части происходит случайная замена одной функции или свободной переменной на другую: генерируются две случайные величины, одна из которых служит для выбора вершины дерева Γ_f , которую предстоит изменить, а другая — для выбора нового элемента для этой вершины. Замена такова, что сохраняется структура суперпозиции, т. е. в случае замены функции сохраняется арность, а свободная переменная заменяется только на другую свободную переменную. Исходные суперпозиции сохраняются в массиве \mathcal{X}_f .
8. Повторяются шаги 4–5.
9. Производится случайный обмен поддеревьями у γ_{cross} суперпозиций с наибольшими значениями Q_f . Вершины, соответствующие этим поддеревьям, выбираются случайным образом. При этом исходные суперпозиции сохраняются в \mathcal{X}_f .
Таким образом, чем лучше результаты суперпозиции и чем она проще, тем ближе значение функционала Q_f к 1.

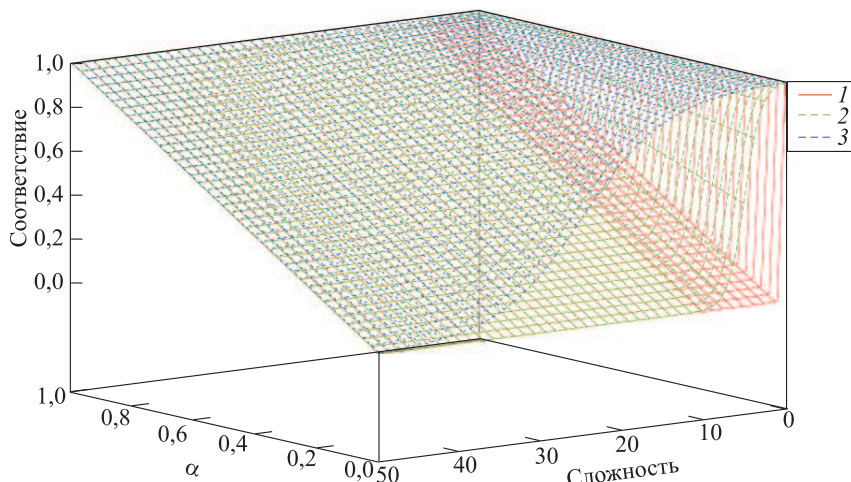


Рис. 2 Поверхности функции Q_f для некоторых β ($1 - \beta = 0,1; 2 - 1; 3 - \beta = 5$) и фиксированного $\tau = 5$

10. Повторяются шаги 2–5.
11. Проверяются условия останова: если либо число итераций превышает I_{\max} , либо в массиве \mathcal{X}_f находится суперпозиция со значением Q_f , большим \hat{Q} , то алгоритм останавливается и результатом считается суперпозиция с наибольшим значением Q_f , иначе осуществляется переход к шагу 2.

7 Вычислительный эксперимент

В вычислительном эксперименте восстанавливается функциональная зависимость $y = 2 \cosh \sqrt{(x_1^2 + x_2^2)/2}$, соответствующая фигуре вращения цепной линии. При этом значения зависимой переменной y были искусственно зашумлены аддитивной добавкой из распределения $\mathcal{N}(0, 0,1)$ и соответствующая ей переменная присутствовала во множестве используемых свободных переменных.

В качестве функционала ошибки S используется сумма квадратов регрессионных остатков для данной суперпозиции f с вектором параметров ω при регрессионной выборке D :

$$S(\omega, f, D) = \sum_{i=1}^N (y_i - f(\omega, \mathbf{x}_i))^2. \quad (2)$$

Значение функционала ошибки S при подстановке исходной незашумленной функциональной зависимости составляет $\approx 4,29$, сложность исходной суперпозиции — 14.

В данной работе используется функционал Q_f вида (1). Значения параметров α , β и τ выбираются экспертно исходя из предположений о виде искомой суперпозиции и моделируемом явлении.

Второй множитель в (1) выполняет роль штрафа за слишком большую сложность суперпозиции, что позволяет выбирать более простые модели, избегая эффекта переобучения и экстремальных случаев вроде порождения интерполяционных полиномов. На рис. 2 приведены поверхности Q_f для различных значений β при фиксированном $\tau = 5$.

Использованные параметры алгоритма 3: $N_{\max} = 200, I_{\max} = 50, \hat{Q} = 0,95, \tau = 20, \alpha = 0,05, \beta = 1, \gamma_{\text{mut}} = 1/3, \gamma_{\text{cross}} = 1/3$. При отсутствии улучшения результатов в течение нескольких итераций подряд алгоритм 3 также завершался.

Результаты вычислительного эксперимента приведены в табл. 1. Указан номер итерации i , на которой суперпозиция была впервые получена, сама суперпозиция, среднеквадратичная ошибка (2) и сложность в смысле числа узлов в соответствующем графе выражения. Числовые коэффициенты в приведенных формулах и значения функционала S_f искусственно округлены до нескольких значащих цифр.

Алгоритм запускался для двух разных наборов элементарных функций. В обоих случаях элементарные функции включали в себя стандартные арифметические операции и операцию возведения в степень. Для удобства возведение в степень $1/2$ (и близкие ей) заменено в таблице на операцию извлечения корня.

В первом случае в наборе отсутствовала функция \cosh . При этом по результатам 10 запусков наилучшей суперпозицией, полученной предложенным алгоритмом, оказалась функция за номером 1 из табл. 1. Видно, что выражение в скобках близко определению $\cosh x = (e^x + e^{-x})/2$, однако разные значения оснований степенных функций

Таблица 1 Результаты вычислительного эксперимента для предложенного алгоритма

N	i	Суперпозиция	S_f	C_f	Q_f
1	13	$1,0002 \left(2,72^{\sqrt{x \cdot x + y \cdot y} / 2} + 2,56^{\sqrt{x \cdot x + y \cdot y} / -1,93} \right)$	$\approx 4,10$	29	$\approx 0,010$
2	9	$2,001 \cosh \frac{\sqrt{x \cdot x + y \cdot y}}{1,999}$	$\approx 4,25$	14	$\approx 0,188$

Таблица 2 Результаты вычислительного эксперимента для алгоритма [10]

i	Суперпозиция	S_f	C_f	Q_f
29	$2,66^{\sqrt{x^2+y^2}/2,23} - \frac{x^2+y^2}{3,03} + \frac{x^2 \cdot x^2 + y^2 \cdot y^2}{6,3} + 0,93$	$\approx 6,2$	43	$\approx 0,007$

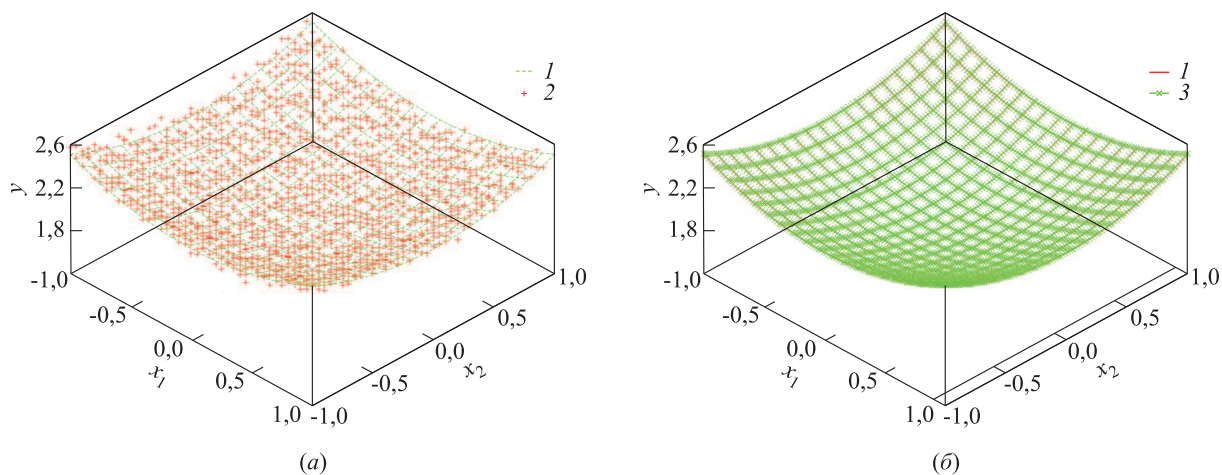


Рис. 3 Первая порожденная суперпозиция (1) и зашумленные точки выборки (2) (а) и исходная зависимость (3) (б)

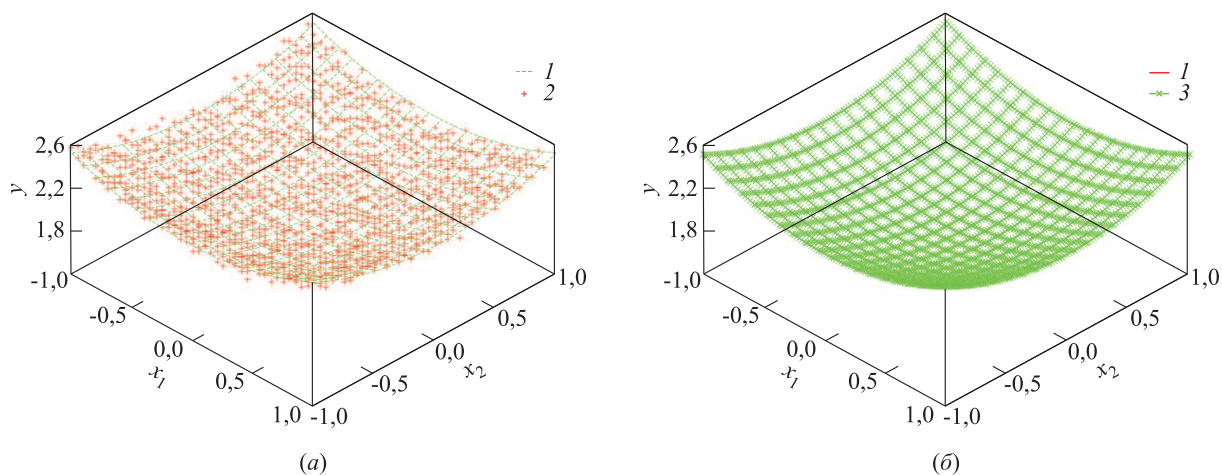


Рис. 4 Вторая порожденная суперпозиция (1) и зашумленные точки выборки (2) (а) и исходная зависимость (3) (б)

могут затруднить экспертный анализ полученного выражения, которое само по себе является достаточно громоздким.

Во втором случае набор элементарных функций также включал в себя функцию \cosh , результату этого выражения соответствует суперпозиция за номером 2. Включение \cosh в G позволило существенно быстрее подобрать искомую функцию, и сложность получившейся суперпозиции также существенно меньше.

Кроме того, предложенный алгоритм сравнивался с алгоритмом [10], в котором суперпозиции кодировались бинарной строкой и применялись стандартные генетические алгоритмы на получавшихся строках; во множестве используемых функций также отсутствовала функция \cosh .

Наилучшая суперпозиция, полученная алгоритмом [10] по результатам 10 запусков, приведена в табл. 2. Полученная суперпозиция имеет существенно более высокую сложность, чем суперпозиции, перечисленные в табл. 1.

На рис. 3 отображены изометрические проекции первой из приведенных в табл. 1 суперпозиций. На рис. 3а данная суперпозиция сравнивается с точками синтезированной зашумленной выборки, на рис. 3б она же приведена вместе с исходной незашумленной зависимостью. Аналогичные проекции приведены для второй суперпозиции на рис. 4.

8 Заключение

В работе исследованы индуктивные алгоритмы порождения допустимых существенно нелинейных суперпозиций. Предложен переборный алгоритм, порождающий все возможные суперпозиции заданной сложности за конечное число шагов. Сформулированный алгоритм решает некоторые типичные проблемы предложенных ранее методов. Описан стохастический алгоритм индуктивного порождения существенно нелинейных суперпозиций и приведены результаты вычислительного эксперимента на синтетических данных. Описанный алгоритм выбирает менее точные, но более простые модели, что позволяет избежать переобучения и выполнить простейший отбор признаков.

Литература

1. *Duffy J., Engle-Warnick J.* Using symbolic regression to infer strategies from experimental data // *Evolutionary Computation in Economics and Finance*, 2002. Vol. 100. P. 61–84.
2. *Barmpalexis P., Kachrimanis K., Tsakonas A., Georgarakis E.* Symbolic regression via genetic programming in the optimization of a controlled release pharmaceutical formulation // *Chemometrics and Intelligent Laboratory Systems*, 2011. Vol. 107. No. 1. P. 75–82.
3. *Davidson J. W., Savic D. A., Walters G. A.* Symbolic and numerical regression: Experiments and applications // *Developments in Soft Computing*, 2001. Vol. 6. P. 175–182.
4. *Стрижов В. В.* Поиск параметрической регрессионной модели в индуктивно заданном множестве // *Вычислительные технологии*, 2007. Т. 1. С. 93–102.
5. *Стрижов В. В.* Методы индуктивного порождения регрессионных моделей. — М.: ВЦ РАН, 2008.
6. *Sammot C., Webb G. I.* Symbolic regression // *Encyclopedia of Machine Learning*. — Berlin: Springer, 2010.
7. *Strijov V. V., Weber G. W.* Nonlinear regression model generation using hyperparameter optimization // *Computers and Mathematics with Applications*, 2010. Vol. 60. No. 4. P. 981–988.
8. *Koza J. R.* Genetic programming // *Encyclopedia of Computer Science and Technology*, 1998. Vol. 39. No. 24. P. 29–43.
9. *Koza J. R.* Introduction to genetic algorithms. — Cambridge: MIT Press, 1998.
10. *Zelinka I., Oplatkova Z., Nolle L.* Analytic programming and symbolic regression by means of arbitrary evolutionary algorithms // *Int. J. Simulation Syst. Sci. Technol.*, 2005. Vol. 6. No. 9. P. 44–56.
11. *Тырсин А. Н.* Об эквивалентности знакового и наименьших модулей методов построения линейных моделей // *Обзорные прикладной и промышленной математики*, 2005. Т. 12. № 4. С. 879–880.
12. *Павловский Ю. Н.* Имитационные модели и системы. — М.: Фазис, 2000.
13. *Битюцков В. И., Войцеховский М. И., Иванов А. Б.* Математическая энциклопедия. Т. 4. — М.: Советская энциклопедия, 1984.
14. *Marquardt D. W.* An algorithm for least squares estimation of nonlinear parameters // *J. Soc. Ind. Appl. Math.*, 1963. Vol. 11. No. 2. P. 431–441.
15. *More J. J.* The Levenberg–Marquardt algorithm: Implementation and theory // *Lecture Notes in Mathematics 630: Numerical Analysis*. — Berlin: Springer-Verlag, 1978. P. 105–116.

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ ЗАПРЕТОВ ВЕРОЯТНОСТНЫХ МЕР НА ДИСКРЕТНЫХ ПРОСТРАНСТВАХ*

А. А. Грушо¹, Н. А. Грушо², Е. Е. Тимонина³

Аннотация: Предложен метод статистического определения запретов вероятностных мер на дискретных пространствах. Показана состоятельность сделанных оценок. Построена схема применения полученных оценок для проверки статистических гипотез в дискретных пространствах. Показано, что в некотором смысле оценки запретов могут порождать состоятельные последовательности критериев (СПК).

Ключевые слова: состоятельные последовательности критериев; запреты вероятностных мер в дискретных пространствах; состоятельность оценок

1 Введение

Конечные вероятностные пространства играют существенную роль при решении различных проблем сетевой и компьютерной безопасности, криптографии и др. Рассмотрим задачу проверки последовательности простых гипотез $H_{0,n}$ против сложных альтернатив $H_{1,n}$ в конечных пространствах. Пусть $X = \{x_1, \dots, x_m\}$ — конечное множество, X^n — декартово произведение множества X , X^∞ — множество бесконечных последовательностей с элементами из X , \mathcal{A} — минимальная σ -алгебра, порожденная всеми цилиндрическими множествами [1–3]. Если на X рассматривается дискретная топология, то тихоновское произведение X^∞ является компактным топологическим пространством со счетной базой [2, 4]. В этом случае борелевская σ -алгебра \mathcal{B} совпадает с σ -алгеброй \mathcal{A} .

Пусть P_0 — вероятностная мера на (X^∞, \mathcal{A}) . Для каждого n $P_{0,n}$ — проекция P_0 на первые n координат последовательностей из X^∞ .

Ранее было определено [5, 6] понятие запрета конечной вероятностной меры. Под запретами понимаются начальные участки последовательностей из X^∞ , вероятности которых равны нулю. В статистических задачах существуют критерии, критические множества которых полностью определяются через запреты. В работах [5, 6] доказаны необходимые и достаточные условия существования СПК, критические множества которых полностью определяются запретами.

Статистические критерии, определяемые с помощью запретов, обладают тем свойством, что при нулевой гипотезе $P_{0,n}(S_n) = 0$, где S_n , $n = 1, 2, \dots$, — критические множества критериев. В случае явного задания вероятностных мер $P_{0,n}$ иногда удается довольно легко определить запреты этих мер.

Если последовательность критериев определяется с помощью запретов, то множество запретов является бесконечным. Поэтому необходимо определять алгоритмы, порождающие бесконечное множество запретов. Это удается сделать для класса мер, у которых все запреты порождаются минимальными запретами. Неформально минимальные запреты определяются следующим образом.

Пусть \mathcal{N} — конечное множество векторов и пусть все запреты меры P_0 содержат хотя бы один вектор из \mathcal{N} . В этом случае можно говорить, что \mathcal{N} порождает множество запретов. Для этого класса вероятностных мер множество \mathcal{N} можно определить статистически.

Предположим, что можно сколь угодно раз независимо реализовать векторы из X^n с вероятностными мерами $P_{0,n}$ для любых конечных n . Тогда можно построить состоятельную оценку множества \mathcal{N} .

В разд. 2 построены оценки множества \mathcal{N} и доказана их состоятельность. В разд. 3 рассмотрен алгоритм применения метода запретов для проверки гипотез $H_{0,n}$ против $H_{1,n}$ с учетом возможности неточного определения множества \mathcal{N} . В заключении обсуждаются полученные результаты.

* Работа выполнена при поддержке РФФИ, гранты № 10-01-00480 и № 11-07-00112.

¹Институт проблем информатики Российской академии наук; Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, grusho@yandex.ru

²Институт проблем информатики Российской академии наук, info@itake.ru

³Институт проблем информатики Российской академии наук, eltimon@yandex.ru

2 Оценки множества минимальных запретов

Пусть на (X^∞, \mathcal{A}) определена вероятностная мера P_0 . Очевидно, что для любого $B_n \in X^n$

$$P_{0,n}(B_n) = P_0(B_n \times X^\infty).$$

Пусть $D_{0,n}$ — носитель меры $P_{0,n}$:

$$D_{0,n} = \{\vec{x}_n \in X^\infty, P_{0,n}(\vec{x}_n) > 0\}.$$

Обозначим $\Delta_{0,n} = D_{0,n} \times X^\infty$. Последовательность $\Delta_{0,n}$, $n = 1, 2, \dots$, невозрастающая и

$$\Delta_0 = \lim_{n \rightarrow \infty} \Delta_{0,n} = \bigcap_{n=1}^{\infty} \Delta_{0,n}.$$

Множество Δ_0 является замкнутым и носителем меры P_0 .

Рассматривается также множество вероятностных мер $\{P_\theta, \theta \in \Theta\}$ на (X^∞, \mathcal{A}) , для которых определены $P_{\theta,n}, D_{\theta,n}, \Delta_{\theta,n}, \Delta_\theta$.

Если $\vec{\omega}_k \in X^k$, то $\vec{\omega}_{k-1}$ получена из $\vec{\omega}_k$ отбрасыванием последней координаты.

Определение 1. Запретом в мере $P_{0,n}$ называется вектор $\vec{\omega}_k \in X^k$, $k \leq n$, такой, что

$$P_{0,n}(\vec{\omega}_k \times X^{n-k}) = 0.$$

Если $P_{0,k-1}(\vec{\omega}_{k-1}) > 0$, то $\vec{\omega}_k$ называется наименьшим запретом.

Если $\vec{\omega}_k$ является запретом в мере $P_{0,n}$, то для любых $k \leq s \leq n$ и для любых векторов $\vec{\omega}_s$, начинающихся с $\vec{\omega}_k$, имеем:

$$P_{0,s}(\vec{\omega}_s) = 0.$$

Пусть $\vec{\omega}_s \in X^s$ и для любого $n \geq s$ и любого $\vec{\omega}_n \in X^n$, если $\vec{\omega}_s$ является частью $\vec{\omega}_n$, то $P_{0,n}(\vec{\omega}_n) = 0$. Если существует n такое, что $\vec{\omega}_n$ содержит $\vec{\omega}_s$ или содержит вектор, отличающийся от $\vec{\omega}_s$ только первой координатой, и $P_{0,n}(\vec{\omega}_n) > 0$, то вектор $\vec{\omega}_s$ назовем *минимальным запретом*. Далее под \mathcal{N} будем понимать множество минимальных запретов.

Пусть все запреты меры P_0 определяются минимальными запретами, т.е. любой запрет содержит хотя бы один элемент из множества \mathcal{N} . Предположим, что множество \mathcal{N} конечно. Тогда существует s_0 такое, что длины всех минимальных запретов не больше s_0 . Допустим, что известна оценка s_0 , хотя не известно, какие элементы входят в множество \mathcal{N} и мощность множества \mathcal{N} . Построим состоятельную оценку множества \mathcal{N} в указанных предположениях.

Пусть $\vec{\omega}_n^{(1)}, \vec{\omega}_n^{(2)}, \dots, \vec{\omega}_n^{(N)}$ — выборка из распределения $P_{0,n}$ объема N . Для простоты изложения считаем, что n — фиксированный параметр, хотя дальнейшие оценки можно провести для случайного набора n_1, \dots, n_N . В полученной выборке считаем частоты $\nu_n^{(i)}$, $i = 1, 2, \dots, m$, встречаемости i -го символа в конце каждого вектора из $\vec{\omega}_n^{(1)}, \vec{\omega}_n^{(2)}, \dots, \vec{\omega}_n^{(N)}$. Выделяем символы из X , частоты которых равны нулю. Выбранные символы являются потенциальными минимальными запретами длины 1. Считаем частоты $\nu_n^{(i,r)}$, $i, r = 1, 2, \dots, m$, последних биграмм в векторах выборки $\vec{\omega}_n^{(1)}, \vec{\omega}_n^{(2)}, \dots, \vec{\omega}_n^{(N)}$. Биграммы, имеющие нулевые частоты, — это кандидаты на минимальные запреты длины 2. Аналогично строим частоты $\nu_n^{(i_1, \dots, i_s)}$, $i_1, \dots, i_s = 1, 2, \dots, m$, s -цепочек. s -цепочки, частоты которых равны нулю, являются кандидатами на минимальные запреты длины s . Число шагов в приведенном алгоритме не превосходит s_0 .

Построенное в результате указанного алгоритма множество потенциальных запретов \mathcal{N}_1 может отличаться от истинного множества \mathcal{N} . Множество \mathcal{N}_1 является оценкой множества \mathcal{N} . Возможны следующие ошибки. s -цепочка, которая не является минимальным запретом, случайно ни разу не встретилась в просмотренной выборке и попадает в множество \mathcal{N}_1 . Тогда при замене \mathcal{N} оценкой \mathcal{N}_1 получаются лишние минимальные запреты. Возможно, что мера P_0 такова, что истинный минимальный запрет может встретиться в цепочках, длины которых больше n . В этом случае частота встречаемости соответствующей s -цепочки также равна нулю и она входит в \mathcal{N}_1 . Это следует из определения минимального запрета (любая встреча минимального запрета обращает вероятность содержащей ее цепочки в 0). Таким образом, \mathcal{N}_1 содержит \mathcal{N} .

Рассмотрим вероятность того, что s -цепочка не встретилась в выборке случайно. Обозначим вероятности появления всех s -цепочек в концах элементов выборки $\vec{\omega}_n^{(1)}, \vec{\omega}_n^{(2)}, \dots, \vec{\omega}_n^{(N)}$ через $p_1^{(s)}, p_2^{(s)}, \dots, p_m^{(s)}$. Вероятность того, что i -я цепочка имеет нулевую частоту встречаемости, равна $(1 - p_i^{(s)})^N$. Среднее число не встретившихся цепочек равно $\sum_{i=1}^{m^s} (1 - p_i^{(s)})^N$. Тогда математическое ожидание числа элементов \mathcal{N}_1 равняется $\sum_{s=1}^{s_0} \sum_{i=1}^{m^s} (1 - p_i^{(s)})^N$.

Обозначим

$$\varepsilon = \min_{1 \leq i \leq m^s, 1 \leq s \leq s_0, p_i^{(s)} > 0} p_i^{(s)}.$$

Тогда математическое ожидание мощности множества \mathcal{N}_1 равняется $|\mathcal{N}| + E\xi$, где ξ — случайная

величина, равная числу случайно не встретившихся s -цепочек, $s = 1, \dots, s_0$. Очевидно, что

$$E\xi = \sum_{s=1; p_i^{(s)} > 0}^{s_0} \sum_{i=1}^{m^s} (1 - p_i^{(s)})^N.$$

Для $E\xi$ справедлива следующая оценка:

$$E\xi \leq s_0 m^{s_0} (1 - \varepsilon)^N.$$

$E\xi$ стремится к 0. Обозначим через λ_N вероятность того, что $|\mathcal{N}_1| > |\mathcal{N}|$. Используя неравенство Маркова, получаем, что при $N \rightarrow \infty$ $\lambda_N \rightarrow 0$.

3 Состоятельные последовательности критериев с использованием минимальных запретов

Перейдем к построению критериев, зависящих от запретов. Ранее предполагалось, что существует семейство распределений $\{P_\theta, \theta \in \Theta\}$. Согласно [5] необходимым и достаточным условием существования СПК, определяемой запретами, является условие: для всех $\theta \in \Theta$

$$P_\theta(\Delta_0) = 0.$$

Для того чтобы проверить это условие, достаточно доказать, что для всех $\theta \in \Theta$ существует минимальный запрет из \mathcal{N} такой, что вероятность в мере P_θ появления этой цепочки из \mathcal{N} стремится к 1.

Пусть условие существования СПК, определяемой запретами \mathcal{N}_1 , выполняется. Тогда последовательность критериев с критическими множествами, содержащими все последовательности, оканчивающиеся цепочками из \mathcal{N}_1 , является состоятельной последовательностью критериев для проверки гипотез $H'_{0,n} : P'_{0,n}, n = 1, 2, \dots$, против альтернатив $H_{1,n}$, где $P'_{0,n}, n = 1, 2, \dots$, — некоторые меры с запретами \mathcal{N}_1 .

В работе [6] было доказано выполнение следующего соотношения между мощностями носителей $P_{0,n}, n = 1, 2, \dots$, и числами наименьших запретов в этих мерах:

$$v_1 m^{n-1} + \dots + v_{n-1} m + v_n + |D_{0,n}| = m^n,$$

где v_i — число наименьших запретов длины i . Из этого соотношения видно, что если увеличивать число наименьших запретов (увеличивая число минимальных запретов), то, вообще говоря, уменьшаются размеры носителей $D_{0,n}$. Это означает, что

вместо исходной меры P_0 с множеством минимальных запретов \mathcal{N} для определения нулевых гипотез можно рассматривать некоторую меру P'_0 . Для этой меры P'_0 носитель Δ'_0 , вообще говоря, меньше, чем носитель исходной меры P_0 , т. е. $\Delta'_0 \subseteq \Delta_0$. Поэтому проверка условия существования СПК, зависящей от запретов, для множества \mathcal{N}_1 (т. е. для меры P'_0) не гарантирует выполнения этого условия для меры P_0 с множеством запретов \mathcal{N} .

Определим корректно меру P'_0 и ее носитель Δ'_0 . Рассмотрим множество $B_{0,n} \subseteq D_{0,n}$, которое получается исключением из $D_{0,n}$ всех цепочек, содержащих запреты из $\mathcal{N}_1 \setminus \mathcal{N}$. Определим замкнутое множество $\Delta'_{0,n} = B_{0,n} \times X^\infty$. Очевидно, что $\Delta'_{0,n} \supseteq \Delta'_{0,n+1}$. Тогда существует предел

$$\Delta'_0 = \bigcap_{n=1}^{\infty} \Delta'_{0,n}.$$

Очевидно, что $\Delta_0 \setminus \Delta'_0$ — измеримое множество. Обозначим $P_0(\Delta_0 \setminus \Delta'_0) = \mu$. Тогда P'_0 определяется следующим образом:

$$P'_0(A) = \frac{1}{1 - \mu} P_0(A \cap \Delta'_0).$$

Для проекций P'_0 множество минимальных запретов совпадает с \mathcal{N}_1 . Пусть СПК, зависящая от запретов, для проверки $H'_{0,n}$ против $H_{1,n}$ (выше предполагалось, что такая последовательность существует) определяется критическими множествами $S_n, n = 1, 2, \dots$. Тогда

$$\begin{aligned} P_{0,n}(S_n) &= P_0(S_n \times X^\infty) = \\ &= P_0\left((S_n \times X^\infty) \cap (\Delta_0 \setminus \Delta'_0)\right) + \\ &\quad + P_0\left((S_n \times X^\infty) \cap (\Delta_0 \cap \Delta'_0)\right). \end{aligned}$$

Второе слагаемое в последней сумме равно нулю, так как критическое множество S_n в мере $P'_{0,n}$ имеет вероятность 0.

Используя определение μ , получаем оценку:

$$P_0\left((S_n \times X^\infty) \cap (\Delta_0 \setminus \Delta'_0)\right) \leq \mu.$$

Отсюда следует, что последовательность критериев с критическими множествами S_n при проверке $H_{0,n}$ против $H_{1,n}$ имеет уровень значимости μ . При этом для всех $\theta \in \Theta$ $P_{\theta,n}(S_n) \rightarrow 1$.

Вспомним, что μ является случайной величиной в вероятностной схеме, связанной со статистическим определением минимальных запретов. Было показано, что при $N \rightarrow \infty$ с вероятностью, стремящейся к 1 (в схеме оценки минимальных запретов), $\mathcal{N}_1 = \mathcal{N}$. В этом случае $\mu = 0$ и $P_0 = P'_0$. Таким образом, получается, что при $N \rightarrow \infty$ СПК, постро-

енная для проверки $H'_{0,n}$ против $H_{1,n}$, будет также состоятельной для проверки $H_{0,n}$ против $H_{1,n}$.

4 Заключение

Знание того факта, что число минимальных запретов конечно, достаточно для того, чтобы определить последовательность критериев, зависящую от запретов, которая асимптотически (в определенном смысле) будет состоятельной. Этот результат позволяет строить алгоритмы статистического выявления скрытых каналов, в которых вставки осуществляются с помощью некоторых функциональных соотношений. Кроме того, в задачах контроля последовательностей на предмет реализации некоторых событий подобный подход может использоваться при непараметрическом определении нулевой гипотезы.

Литература

1. Леман Э. Проверка статистических гипотез. — М.: Наука, 1964.
2. Бурбаки Н. Общая топология. Основные структуры. — М.: Наука, 1968.
3. Неве Ж. Математические основы теории вероятностей. — М.: Мир, 1969.
4. Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. — М.: Наука, 1993.
5. Грушо А. А., Тимонина Е. Е. Запреты в дискретных вероятностно-статистических задачах // Дискретная математика, 2011. Т. 23. № 2. С. 53–58.
6. Grusho A., Timonina E. Statistical tests based on bans // 1st Symposium (International) and 10th Balkan Conference on Operational Research Proceedings. — Thessaloniki, Greece, 2011. Vol. 1. P. 234–241.

ОПЕРАЦИИ НАД ПРЕДСТАВЛЕНИЯМИ КУСОЧНО-КВАЗИАФФИННЫХ ФУНКЦИЙ В ВИДЕ ДЕРЕВЬЕВ*

С. А. Гуда¹

Аннотация: Введены понятия кусочно заданной квазиаффинной (ККА) функции и ККА-множества. Определены представления ККА-множеств и ККА-функций в виде деревьев и понятие сложности представления. Описаны алгоритмы выполнения операций над древовидными представлениями, в частности объединение, пересечение, проверка ККА-множества на пустоту, сумма, вычисление образа/прообраза, обращение, суперпозиция, сравнение ККА-функций. Даны оценки сложности получающихся в результате объектов. Доказана теорема о виде и сложности лексикографического экстремума в ККА-множестве, зависящем от параметров.

Ключевые слова: кусочно-квазиаффинная функция; выпуклый Z -многогранник; лексикографический экстремум

1 Введение

При анализе программ на этапе компиляции, например во время автоматического распараллеливания, важную роль играют информационные зависимости. Решетчатый граф информационных зависимостей гнезда циклов (или граф алгоритма, см. [1, 2]) связывает итерации, на которых происходят обращения к одной и той же ячейке памяти. Для хранения решетчатого графа не подходят такие традиционные методы, как матрица смежностей или список дуг. Это связано с тем, что множество вершин графа совпадает с итерационным пространством программы. Затраты памяти на хранение такого количества вершин или дуг для реальных программ не приемлемы. Кроме того, время просмотра такого графа будет сопоставимо со временем исполнения программы, что также недопустимо. В [1, 3–5] было предложено описывать решетчатый граф набором функций. Для каждой дуги $I \rightarrow J$ графа найдется функция Φ из набора такая, что $I = \Phi(J)$ или $J = \Phi(I)$. Функции, описывающие информационные зависимости в гнезде циклов, являются кусочно заданными и на каждой части области определения представляют собой суперпозицию некоторого числа аффинных функций и операций целочисленного деления на константу. Анализ ККА-функций позволяет применять оптимизирующие или распараллеливающие преобразования [2, 6, 7], размещать данные в параллельной памяти [8], получать детальное описание потока данных в программе [4, 9], оценивать количество кеш-промахов [10], вычислять объем динамически

выделяемой памяти, генерировать код для многоконвейерных архитектур [11], делать прогноз возможности эффективного распараллеливания программы [12]. Важно, чтобы время данного анализа не зависело от времени работы программы или количества итераций в гнезде циклов. Все алгоритмы работают и для случая ККА-функций, зависящих от параметров.

Существует два основных подхода к представлению ККА-функций в компьютере: предложенные П. Фотрье деревья вложенных друг в друга операторов `if` [3] и описание каждой части области определения системой неравенств и возвращаемым значением функции для данной части [1, 13, 14]. Так как оба представления математически связаны с одним и тем же объектом, у них много общего. Деревья позволяют сгенерировать более быстрый программный код, тогда как системы неравенств допускают обобщение на случай нелинейных зависимостей [14] (остальные недостатки деревьев легко преодолимы). В данной работе рассматривается линейный класс программ [3], от первого представления ККА-функции легко перейти ко второму, к тому же автор считает, что нелинейные выражения следует анализировать сводя к линейным. По этим причинам принято решение использовать деревья.

Для исследования ККА-функций используются различные библиотеки, предназначенные для работы с многогранниками. В Страсбургском университете создана библиотека PolyLib [15], выполняющая операции с отдельными многогранниками, их объединениями и решетками; в проекте автораспараллеливателя GCC Graphite [16] используется Parma

*Работа выполнена при поддержке программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы, госконтракты № 02.740.11.0208 от 7 июля 2009 г.; № 14.740.11.0006 от 1 сентября 2010 г.

¹Южный федеральный университет, gudasergey@gmail.com

Polyhedra Library [17]; недавно начала развиваться библиотека Integer Set Library [18], позволяющая работать с целочисленными отношениями. В данной работе вводится обобщающее понятие ККА-множества, определяемого своей характеристической ККА-функцией. Пользователи библиотек PolyLib и PPL сталкиваются с множеством трудностей: объединение или разность выпуклых многогранников не является выпуклым многогранником; если в соотношениях, описывающих множество, встречаются операции деления нацело или взятия остатка, приходится рассматривать решетки или расширять пространство, вводя новые переменные; при вычислении образа многогранника при линейном отображении (например, проекции) также возникают соотношения с операциями деления нацело, описывающие «дырявые» многогранники (содержащие не все целые точки внутри своей выпуклой оболочки).

В данной работе показано, что класс ККА-множеств является замкнутым: объединение, разность, образ ККА-множества при ККА-отображении принадлежат тому же классу. Приведены несложные алгоритмы построения ККА-множеств, получающихся в результате данных операций.

Тема ККА-функций не нова. Все исследователи, занимающиеся автоматическим анализом программ, так или иначе с ними сталкиваются, преобразуют их древовидные представления или семейства соответствующих систем неравенств. К сожалению, описание алгоритмов выполнения операций зачастую остается «за кадром», быть может, ввиду своей громоздкости и рутинности. Цель статьи — представить известные операции в унифицированной, удобной для программирования форме.

Работа имеет следующую структуру. В разд. 2 определены основные понятия: ККА-множеств и ККА-функций. Раздел 3 посвящен представлению ККА-функций в виде деревьев, вводится понятие сложности представления ККА-функций. В разд. 4 описаны алгоритмы выполнения основных операций над представлениями ККА-функций и ККА-множеств. В частности, для древовидных представлений ККА-множеств даны алгоритмы построения пересечения, объединения, разности, дополнения и выведены оценки сложности получающихся в результате объектов. Описан алгоритм, определяющий, является ли заданное ККА-множество пустым; алгоритм поиска лексикографического максимума в ККА-множестве, зависящем от параметров. Для представлений ККА-функций построены необходимые для автоматического распараллеливания и конвейеризации операции вычисления образа/прообраза ККА-множества, выведена оценка сложности получающегося в результате объекта, приведен алгоритм сравнения двух ККА-функций. Заметим, что всюду в статье все множества подразумеваются состоящими из целочисленных точек. Большинство вышперечисленных алгоритмов реализовано в Открытой распараллеливающей системе и Диалоговом высокоуровневом оптимизирующем распараллеливателе, разрабатываемых на факультете математики, механики и компьютерных наук Южного федерального университета под руководством Б. Я. Штейнберга [19–23].

В качестве базисных можно взять следующие четыре операции:

- (1) суперпозицию ККА-функций,
- (2) составление из нескольких координатных ККА-функций одной,

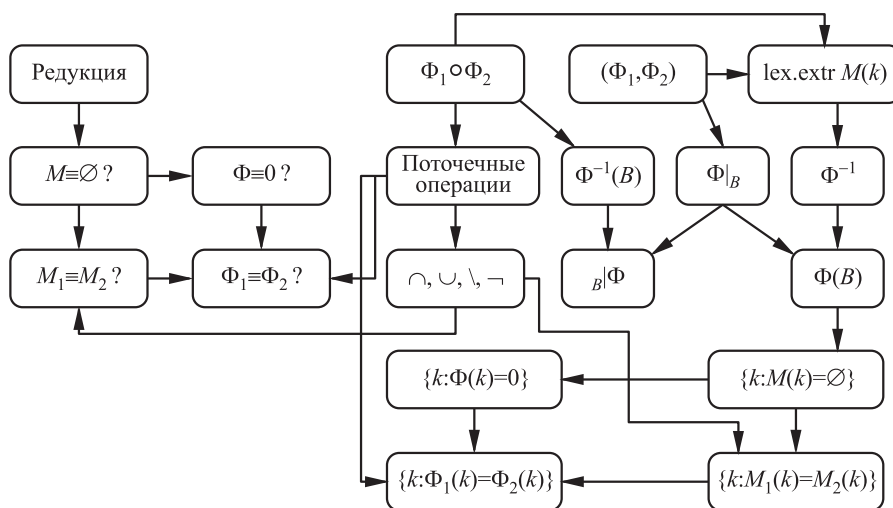


Рис. 1 Зависимости между операциями

добавочными неравенствами, определяющими параметры p_1, p_2, \dots, p_k :

$$0 \leq \left(c^{(i)} \cdot I + d_{i1}p_1 + \dots + d_{i,i-1}p_{i-1} + e_i \right) - \ell_i p_i \leq \leq \ell_i - 1, \quad i = 1, 2, \dots, k. \quad (6)$$

Каждое i -е неравенство данной системы эквивалентно i -му равенству (3). Является ли проекция квазивыпуклым \mathbb{Z} -многогранником в общем случае? В библиотеке операций над многогранниками PolyLib \mathbb{Z} -многогранником называют пересечение выпуклого многогранника и решетки. Аффинный образ этих объектов не принадлежит тому же классу множеств [24]. Это заставило авторов библиотеки ISL [18] определить \mathbb{Z} -многогранник как множество $\{x \in \mathbb{Z}^n \mid \exists z \in \mathbb{Z}^m Rx + Qz + t \geq 0\}$. Автор считает вероятным, что для квазивыпуклых \mathbb{Z} -многогранников из определения 2.4 или эквивалентной ему теоремы 2.2 справедливо утверждение: образ выпуклого \mathbb{Z} -многогранника, а также образ квазивыпуклого \mathbb{Z} -многогранника при аффинном отображении $\Phi(I) = AI + b$, $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ являются квазивыпуклыми \mathbb{Z} -многогранниками. Если доказать данную гипотезу, то окажется, что образы выпуклых \mathbb{Z} -многогранников при аффинных отображениях и только они являются квазивыпуклыми \mathbb{Z} -многогранниками. Сложность доказательства можно ощутить, попытавшись представить в виде (5) образы нескольких простых многогранников:

- 1) $M = \{(I_1, I_2) \in \mathbb{Z}^2 \mid I_{1,2} \geq 0\}$, $\Phi : \mathbb{Z}^2 \rightarrow \mathbb{Z}^1$, $\Phi(I_1, I_2) = 10I_1 + 11I_2$;
- 2) $M = \{(I_1, I_2, I_3) \in \mathbb{Z}^3 \mid I_{1,2,3} \geq 0, I_1 \geq 10I_3, I_2 \geq 11I_3, 13I_3 \geq I_1 + I_2\}$, $\Phi : \mathbb{Z}^3 \rightarrow \mathbb{Z}^2$, $\Phi(I_1, I_2, I_3) = (I_1, I_2)$.

Автору удалось доказать квазивыпуклость следующих типов множеств: конечное множество, периодическое множество (т.е. с периодической по всем переменным характеристической функцией), разность квазивыпуклого и конечного множества, разность квазивыпуклого и периодического множества. Задача поиска образа легко решается, если расширить класс рассматриваемых множеств (см. ниже).

Определение 2.5. Функцию $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ с областью определения $\mathcal{D}(\Phi) \subset \mathbb{Z}^m$ будем называть кусочно-аффинной, если область $\mathcal{D}(\Phi)$ представима в виде объединения конечного числа непересекающихся выпуклых \mathbb{Z} -многогранников, на каждом из которых функция Φ аффинная.

Произвольная функция с конечной областью определения $\mathcal{D}(\Phi)$ является кусочно-аффинной в смысле данного определения. Однако число частей

$\mathcal{D}(\Phi)$ может оказаться очень большим, сравнимым с количеством элементов в $\mathcal{D}(\Phi)$.

Определение 2.6. Функцию $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ с областью определения $\mathcal{D}(\Phi) \subset \mathbb{Z}^m$ будем называть кусочно-квазиаффинной, если область $\mathcal{D}(\Phi)$ представима в виде объединения конечного числа непересекающихся квазивыпуклых \mathbb{Z} -многогранников, на каждом из которых функция Φ квазиаффинная.

Замечание. Пусть область определения кусочно-квазиаффинной функции $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ состоит из ν кусочков; $p^{(1)}, p^{(2)}, \dots, p^{(\nu)}$ — наборы параметров (3), соответствующие каждой квазиаффинной функции, заданной на своем кусочке. Тогда Φ можно рассматривать как кусочно-аффинную функцию аргумента $(I, p^{(1)}, \dots, p^{(\nu)}) \in \mathbb{Z}^{m+k_1+\dots+k_\nu}$.

В приложениях многогранники обычно зависят от параметров. В связи с этим дадим еще два определения. Здесь и далее все параметры считаются целочисленными.

Определение 2.7. Будем говорить, что выпуклый/квазивыпуклый \mathbb{Z} -многогранник $D(r) \subset \mathbb{Z}^n$ линейно зависит от вектора параметров $r \in \mathbb{Z}^s$, если D для каждого $r = r_0$ является сечением некоторого фиксированного выпуклого/квазивыпуклого \mathbb{Z} -многогранника в расширенном пространстве \mathbb{Z}^{n+s} переменных (I, r) гиперплоскостью $r = r_0$.

Теорема 2.3. Выпуклый \mathbb{Z} -многогранник D линейно зависит от вектора параметров $r \in \mathbb{Z}^s$ тогда и только тогда, когда он описывается системой неравенств, в которой от параметров зависят только свободные члены, причем зависят линейно:

$$I \in D \Leftrightarrow AI + t + Br \geq 0,$$

где A и B — постоянные матрицы; t — постоянный вектор.

Теорема 2.4. Квазивыпуклый \mathbb{Z} -многогранник D линейно зависит от вектора параметров $r \in \mathbb{Z}^s$ тогда и только тогда, когда он описывается системой неравенств

$$I \in D \Leftrightarrow AI + t + Br + Qp \geq 0,$$

где A, B и Q — постоянные матрицы, t — постоянный вектор, p — нелинейные функции вектора переменных I и вектора параметров r следующего вида:

$$\left. \begin{aligned} p_1 &= (c^{(1)} \cdot I + g^{(1)} \cdot r + e_1) \div \ell_1; \\ p_2 &= (c^{(2)} \cdot I + g^{(2)} \cdot r + d_{21}p_1 + e_2) \div \ell_2, \\ p_3 &= (c^{(3)} \cdot I + g^{(3)} \cdot r + d_{31}p_1 + d_{32}p_2 + e_3) \div \\ &\quad \div \ell_3; \\ &\dots\dots\dots \\ p_k &= (c^{(k)} \cdot I + g^{(k)} \cdot r + d_{k1}p_1 + \dots \\ &\quad \dots + d_{k,k-1}p_{k-1} + e_k) \div \ell_k, \end{aligned} \right\} \quad (7)$$

$I, c^{(i)} \in \mathbb{Z}^n, r, g^{(i)} \in \mathbb{Z}^s, l_i \in \mathbb{N}, d_{i,j}, e_i \in \mathbb{Z}, s$ — число параметров.

Определение 2.8. Будем говорить, что ККА-функция Φ линейно зависит от вектора параметров r , если от r линейно зависят значения функции и части области определения $\mathcal{D}(\Phi)$.

Приведем теорему, которая следует из алгоритма [3].

Теорема 2.5 (П. Фотрье). Пусть D — выпуклый \mathbb{Z} -многогранник, линейно зависящий от вектора параметров r . Тогда функции $\Phi_{\max}(r) = \text{lex.max}_{I \in D(r)} I$ и $\Phi_{\min}(r) = \text{lex.min}_{I \in D(r)} I$ являются кусочно-квазиаффинными в смысле определения 2.6. Области определения $\mathcal{D}(\Phi_{\max})$ и $\mathcal{D}(\Phi_{\min})$ состоят из тех точек, для которых многогранник $D(r)$ не пуст и ограничен с соответствующей стороны.

При работе с ККА-функциями часто возникают структуры в виде объединений конечного числа непересекающихся квазिवыпуклых \mathbb{Z} -многогранников. Например, к числу таких структур относятся области определения ККА-функций.

Определение 2.9. Кусочно-квазиаффинным множеством будем называть объединение конечного числа квазिवыпуклых \mathbb{Z} -многогранников.

Пример 2.1. Рассмотрим цикл

```
for (i = 0; i < N; i++)
{
    x[i+a] = ...
    x[2*i] = ...
    ... = ... x[i] ...
}
```

Пространство итераций цикла зависит от параметра N . В одно из индексных выражений входит другой параметр a . Предполагается, что $a \geq 0$ и $N \geq 1$. Граф истинной информационной зависимости [1, 2], соответствующий вхождению массива x , состоит из дуг $j \rightarrow i$, соединяющих итерации цикла так, что на итерации i произошло чтение из ячейки массива $x[i]$, а на итерации $j \leq i$ — последняя перед итерацией i запись в эту же ячейку, т.е. $j = \text{lex.max}_{j' \leq i, j'+a=i \text{ или } 2j'=i} j'$. Область, в которой

ищется лексикографический максимум, является ККА-множеством (объединением двух выпуклых \mathbb{Z} -многогранников), линейно зависящим от параметров a и i . Ниже будет доказано обобщение теоремы 2.5, согласно которому дуги такого графа можно задать ККА-функцией $j = \Phi(i)$. В данном случае

$$\Phi(i) = \begin{cases} i - a, & \text{если } i \geq a, i \% 2 = 0, i - a > i \div 2, \\ & 0 \leq i < N; \\ i \div 2, & \text{если } i \geq a, i \% 2 = 0, i - a \leq i \div 2, \\ & 0 \leq i < N; \\ i - a, & \text{если } i \geq a, i \% 2 \neq 0, 0 \leq i < N; \\ i \div 2, & \text{если } i < a, i \% 2 = 0, 0 \leq i < N. \end{cases}$$

Символом «%» обозначена операция взятия остатка от деления ($0 \leq k \% m < m$). Область определения функции Φ : $\mathcal{D}(\Phi) = \{i \mid 0 \leq i < N, i \geq a \text{ или } i \% 2 = 0\}$ представима в виде объединения четырех непересекающихся квазिवыпуклых \mathbb{Z} -многогранников: $\mathcal{D}(\Phi) = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3 \cup \mathcal{D}_4$, где

$$\begin{aligned} \mathcal{D}_1 &= \{i \mid i \geq a, i \% 2 = 0, i - a > i \div 2, 0 \leq i < N\}; \\ \mathcal{D}_2 &= \{i \mid i \geq a, i \% 2 = 0, i - a \leq i \div 2, 0 \leq i < N\}; \\ \mathcal{D}_3 &= \{i \mid i \geq a, i \% 2 \neq 0, 0 \leq i < N\}; \\ \mathcal{D}_4 &= \{i \mid i < a, i \% 2 = 0, 0 \leq i < N\}. \end{aligned}$$

Выпишем неравенства (5) для \mathbb{Z} -многогранников $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ и \mathcal{D}_4 :

$$\begin{aligned} & i - a \geq 0; & i - a \geq 0; \\ & -i + 2p \geq 0; & -i + 2p \geq 0; \\ \mathcal{D}_1: & i - a - p - 1 \geq 0; & \mathcal{D}_2: & -i + a + p \geq 0; \\ & i \geq 0; & & i \geq 0; \\ & -i + N - 1 \geq 0; & & -i + N - 1 \geq 0; \\ \mathcal{D}_3: & i - a \geq 0; & \mathcal{D}_4: & -i + a - 1 \geq 0; \\ & i - 2p - 1 \geq 0; & & -i + 2p \geq 0, \end{aligned}$$

где $p = i \div 2$.

3 Формы представления кусочно-квазиаффинных функций

На практике кусочно-квазиаффинные функции удобно хранить в виде бинарных деревьев, которые получаются на выходе параметризованного метода Гомори (см. документацию к [5]). Такая форма представления ККА-функций позволяет быстро вычислять значения и проводить операции над функциями. По сути, это дерево бинарного поиска нужного кусочка области определения ККА-функции.

Определение 3.1. Будем называть формой Фотрье [3] представления ККА-функции $\Phi: \mathbb{Z}^m \rightarrow \mathbb{Z}^n$, квазилинейно зависящей от вектора параметров r , бинарное дерево вложенных друг в друга операторов if, удовлетворяющее условиям:

1. Каждый узел дерева, не являющийся листом, состоит из

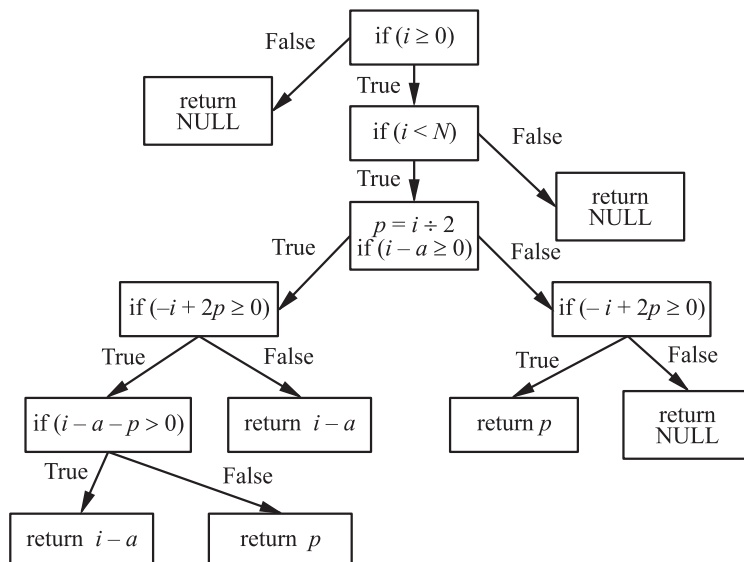


Рис. 2 Схема программы из примера 2.1

- набора определений локальных параметров p вида (7) (возможно пустого);
- оператора ветвления
 $\text{if } (U) \Phi_{\text{true}}$
 $\text{else } \Phi_{\text{false}},$
 где U — условие неотрицательности аффинной формы от аргумента функции, вектора параметров r и набора локальных параметров p данного и вышестоящих узлов до корня; Φ_{true} и Φ_{false} — поддеревья по веткам true и false.

2. Листья состоят из

- набора определений локальных параметров p вида (7) (возможно пустого);
- n -мерного вектора — значения функции, — возвращаемого оператором return (аффинная вектор-функция аргумента, параметров r и набора локальных параметров p данного и вышестоящих узлов до корня) или NULL, если на данной подобласти функция не определена.

3. Областью определения такой функции будем считать множество, на котором функция возвращает значения, отличные от NULL (NULL и возвращаемое значение 0 предполагаются различными).

В документации к [5] данному определению соответствует грамматика для Quast_group (QUAST — QUasi Affine Selection Tree).

Каждое такое дерево естественно представляется в виде программы на каком-либо языке програм-

мирования. Например, функции из примера 2.1 соответствует программа на языке Си:

```

ValueType Φ(int i) {
    if (i>=0) {
        if (i<N) {
            int p = i/2;
            if ( i-a>=0 ) {
                if ( -i+2*p>=0 ) {
                    if ( i-a-p>0 ) {return i-a;}
                    else {return p;}
                }
                else {return i-a;}
            }
            else {
                if ( -i+2*p>=0 ) {return p;}
                else {return NULL;}
            }
        }
        else {return NULL;}
    }
    else {return NULL;}
}
    
```

Она описывается двоичным деревом, показанным на рис. 2.

Лемма 3.1. Форма Фотрье корректно задает функцию. Эта функция является кусочно-квазиаффинной в смысле определения 2.6.

Одной ККА-функции может соответствовать несколько различных представлений в виде деревьев. Введем характеристики дерева.

Обозначения. Через NodeNum(Φ) и LeafNum(Φ) будем обозначать число узлов «if» и число листь-

ев дерева ККА-функции, которые не возвращают NULL.

Определение 3.2. Сложностью S_∞ представления ККА-функции в виде дерева будем называть максимальное количество вычислений неравенств в условиях операторов if, необходимое для того, чтобы посчитать значение функции в точке. Таким образом, сложность S_∞ — это максимальная глубина дерева.

Сложность приближенно показывает, насколько быстрым будет соответствующий дереву код на каком-либо языке программирования. Наряду с S_∞ для ККА-функции Φ с ограниченной областью определения можно рассматривать сложность

$$S_p = \left(\frac{1}{|\mathcal{D}(\Phi)|} \sum_{I \in \mathcal{D}(\Phi)} \mu(I)^p \right)^{1/p},$$

где $\mu(I)$ — количество неравенств, проверяемых при вычислении $\Phi(I)$, $p \in \mathbb{R}$, $p \geq 1$.

При выполнении операций с ККА-функциями удобно пользоваться расширенным представлением, которое отличается от формы Фотрье лишь тем, что в условиях операторов if разрешено использовать «short-circuit» операции логического «и», «или» и отрицания в произвольных комбинациях с неравенствами.

Определение 3.3. Будем называть такую форму представления ККА-функции расширенной формой.

Например, у представления на рис. 2 можно два верхних оператора if объединить в один:

```
if ((i>=0) && (i<N))
{ ... }
else return NULL.
```

Определение 3.4. Сложностью S_∞ расширенного представления ККА-функции также будем называть максимальное количество вычислений неравенств в условиях операторов if, необходимое для того, чтобы посчитать значения функции в точке. При этом предполагается, что логические операции производят вычисление своих операндов слева направо и вычисляют минимальное число операндов, необходимое для определения результата выражения (т. е. являются «short-circuit»).

Лемма 3.2. Для каждой расширенной формы существует форма Фотрье, задающая ту же функцию с такой же сложностью.

¹Приставку «квази» здесь опустить нельзя, так как кусочки области определения P_M являются квазивыпуклыми \mathbb{Z} -многогранниками.

Лемма 3.3. Любую ККА-функцию можно представить в расширенной форме и, в соответствии с леммой 3.2, в форме Фотрье.

Теорема 3.1. Класс функций, которые можно задать при помощи расширенного представления, совпадает с классом функций, задаваемых формой Фотрье и равен классу ККА-функций.

Доказательство теоремы следует из лемм 3.1–3.3.

Кусочно-квазиффинные множества удобно хранить в виде древовидных структур, подобных представлениям ККА-функций. Свяжем с произвольным таким множеством M характеристическую функцию P_M , которая равна единице на M и не определена (return NULL) во всех остальных точках. Функция P_M — кусочно-квазиффинная¹.

Определение 3.5. Представлением ККА-множества M в виде дерева операторов if будем называть представление соответствующей характеристической ККА-функции P_M .

Определение 3.6. Сложностью представления ККА-множества M будем называть сложность представления ККА-функции P_M .

Если M — квазивыпуклый \mathbb{Z} -многогранник, описываемый системой из m неравенств, то соответствующее функции P_M дерево линейно: оно состоит из вложенных друг в друга операторов if с возвращаемым значением NULL по ветке else. Значит, сложность S_∞ такого представления P_M равна m .

4 Операции с кусочно-квазиффинными функциями

Все ККА-множества и ККА-функции в данном разделе могут квазилинейно зависеть от внешних параметров. Заметим, что ККА-функцию $\Phi_r : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$, зависящую от набора $r = (r_1, r_2, \dots, r_s)$ целочисленных параметров, можно рассматривать как обычную ККА-функцию $m + s$ целых переменных (I, r) с областью определения $\mathcal{D}(\Phi_r) = \{(I, r) \in \mathbb{Z}^{m+s} \mid \Phi_r(I) \neq \text{NULL}\}$. Так же и ККА-множество M_r , зависящее от набора параметров r , можно рассматривать как обычное ККА-множество в расширенном пространстве переменных-параметров. Для многих операций смысловое отличие части переменных r от I не имеет значения. Случаи, где наличие параметров играет существенную роль, будут оговариваться отдельно.

Чтобы посчитать функцию Φ_r для некоторого фиксированного набора параметров $r = r_0$, нужно положить в древовидном представлении Φ_r $r = r_0$. Область определения полученной функции $\mathcal{D}(\Phi_{r_0})$ является сечением $\mathcal{D}(\Phi_r) \subset \mathbb{Z}^{m+s}$ гиперплоскостью $r = r_0$. Чтобы определить, для каких значений параметров r область определения Φ_r непуста, нужно вычислить проекцию $\mathcal{D}(\Phi_r) \subset \mathbb{Z}^{m+s}$ на пространство параметров \mathbb{Z}^s (вычисление проекции сводится к вычислению образа и рассматривается в п. 4.2.6). Точно так же вычисляется множество значений параметров, для которых непусто произвольное ККА-множество.

4.1 Базисные операции

4.1.1 Суперпозиция кусочно-квазиаффинных функций

Суперпозиция $\Phi_1 \circ \Phi_2 = \Phi_1(\Phi_2)$ строится следующим образом. Пусть функция Φ_1 описывается деревом T_1 , а функция Φ_2 — T_2 . Каждый оператор «return q_2 » дерева T_2 заменим на конструкцию T'_1 , которая получается из T_1 заменой аргумента во всех условиях и возвращаемых выражениях на q_2 . Сложность суперпозиции подчиняется неравенству $S_\infty(\Phi_1 \circ \Phi_2) \leq S_\infty(\Phi_1) + S_\infty(\Phi_2)$.

4.1.2 Составление из нескольких координатных кусочно-квазиаффинных функций одной

Пусть Φ_1, Φ_2 — ККА-функции, действующие из \mathbb{Z}^m в \mathbb{Z}^{n_1} и \mathbb{Z}^{n_2} с областями определения $\mathcal{D}(\Phi_1), \mathcal{D}(\Phi_2)$ соответственно. Составим из них функцию $\Phi = (\Phi_1, \Phi_2) : \mathbb{Z}^m \rightarrow \mathbb{Z}^{n_1+n_2}$ с областью определения $\mathcal{D}(\Phi) = \mathcal{D}(\Phi_1) \cap \mathcal{D}(\Phi_2)$. Она является ККА, так как на частях области определения — всевозможных попарных пересечениях частей $\mathcal{D}(\Phi_1)$ и $\mathcal{D}(\Phi_2)$ — функция $\Phi = (\Phi_1, \Phi_2)$ квазиаффинная. Построим соответствующее ей дерево. Пусть T_1 и T_2 — древовидные представления функций Φ_1 и Φ_2 . Каждый оператор «return q_1 » дерева T_1 заменим на конструкцию T_2 , в которой вместо операторов «return q_2 », возвращающих значение функции Φ_2 , стоят операторы «return (q_1, q_2) », возвращающие значение функции Φ . Сложность S_∞ построенного таким образом дерева не превышает суммы сложностей $S_\infty(\Phi_1)$ и $S_\infty(\Phi_2)$.

4.1.3 Удаление недостижимых веток

В процессе выполнения операций над ККА-функциями возникают древовидные представления со значениями на листьях, которые никогда не возвращаются функцией, потому что соответствующие им части области определения не содержат целых

точек. Опишем преобразование, которое удаляет все такие недостижимые ветви. Если ККА-функция зависит от параметров линейным образом, то будем считать эти параметры дополнительными аргументами. Таким образом, будут удалены ветви дерева, являющиеся недостижимыми для всех значений параметров.

Алгоритм. В цикле, пока в дереве Φ существует поддерево ϕ , удовлетворяющее одному из следующих условий, выполняем:

1. Если поддерево ϕ не является листом и все его листья возвращают NULL, заменим ϕ на лист return NULL.
2. Пусть ϕ не является листом:

$$\phi = \begin{cases} \text{if } (U) & \phi_{\text{true}}, \\ \text{else} & \phi_{\text{false}}. \end{cases}$$

Рассмотрим путь от корня дерева Φ к поддереву ϕ . Объединим в систему условия операторов if вдоль пути. При этом если после очередной вершины вдоль пути движение происходит по ветке «true», то включаем условие «как есть», а если по ветке «false» — то включаем противоположное условие. Данная система вместе с неравенствами (6), определяющими локальные параметры p , описывает некоторое ККА-множество. Обозначим его B .

Если условие U или его отрицание $\neg U$ не имеют целых решений внутри B , то в первом случае заменяем ϕ на ϕ_{false} , а во втором — на ϕ_{true} . Проверку совместимости условия внутри B можно осуществлять любым алгоритмом проверки разрешимости в целых числах системы неравенств, например тем же методом отсечений Гомори, который используется при поиске лексикографического экстремума в многограннике; можно применить Омега-тест [13].

Удаление недостижимых веток и лишних условий, выполненное в п. 2 алгоритма, приводит к эквивалентному дереву, все листья которого соответствуют непустым в \mathbb{Z}^n частям области определения ККА-функции F . Более того, благодаря п. 1 из дерева удаляются ветви, у которых все листья возвращают NULL. Таким образом, после применения данного преобразования пустые ККА-множества и только они сводятся к единственному узлу «return NULL».

Чтобы выполнить это преобразование в худшем случае потребуются проверить разрешимость в целых числах $2\text{NodeNum}(F')$ систем не более чем $S_\infty(F')$ неравенств, где F' — форма Фотрье функции F .

4.1.4 Лексикографические экстремумы множеств, зависящих от параметров

Обозначение. $S_\infty(\text{lex.max } D)$ — оценка сложности дерева ККА-функции из теоремы 2.5 П. Фотрье [3].

Теорема 4.1 (обобщение теоремы Фотрье на квазивыпуклые \mathbb{Z} -многогранники). Пусть D — квазивыпуклый \mathbb{Z} -многогранник, квазилинейно зависящий от вектора параметров $r = (r_1, \dots, r_s)^T$. Тогда функции $\Phi_{\max}(r) = \text{lex.max}_{I \in D(r)} I$ и $\Phi_{\min}(r) = \text{lex.min}_{I \in D(r)} I$ являются

ККА в смысле определения 2.6. Существуют их древовидные представления со сложностью, не больше чем $S_\infty(\text{lex.max } \tilde{D})$ и $S_\infty(\text{lex.min } \tilde{D})$, где \tilde{D} — соответствующий D выпуклый \mathbb{Z} -многогранник в пространстве переменных (I, p) (см. (7)).

Задача сводится к поиску лексикографического экстремума в обычном выпуклом \mathbb{Z} -многограннике \tilde{D} в расширенном пространстве переменных (I, p) размерности $n + k$ и взятию первых n компонент результата.

Лемма 4.1. Функция $\text{LM} : \mathbb{Z}^{2n} \rightarrow \mathbb{Z}^n$, $\text{LM}(I, J) = \text{lex.max}\{I, J\}$ является кусочно-аффинной, и для нее существует представление со сложностью $S_\infty(\text{LM}) = 2n - 1$.

Теорема 4.2 (обобщение теоремы Фотрье на ККА-множества). Пусть $M \subset \mathbb{Z}^n$ — ККА-множество, квазилинейно зависящее от вектора параметров $r = (r_1, \dots, r_s)^T$. Тогда функции $\Phi_{\max}(r) = \text{lex.max}_{I \in M(r)} I$ и $\Phi_{\min}(r) = \text{lex.min}_{I \in M(r)} I$ являются ККА в смысле определения 2.6. Сложность S_∞ представления функции $\Phi_{\max}(r)$ подчиняется неравенству:

$$S_\infty(\Phi_{\max}) \leq (\nu - 1)(2n - 1) + \sum_{i=1}^{\nu} S_\infty(\text{lex.max } D_i), \quad (8)$$

где D_i — квазивыпуклые \mathbb{Z} -многогранники, составляющие ККА-множество M , $i = 1, 2, \dots, \nu$, $\nu = \text{LeafNum}(M)$, $S_\infty(\text{lex.max } D_i)$ — сложность представления ККА-функции из теоремы 4.1. Для Φ_{\min} в неравенстве (8) нужно lex.max поменять на lex.min .

Доказательство проведем для максимума. Согласно теореме 4.1 вычислим лексикографические максимумы внутри всех многогранников D_i . Получим ν ККА-функций f_1, f_2, \dots, f_ν вектора параметров r . Искомая функция

$$\Phi_{\max}(r) = \text{lex.max}\{f_1(r), f_2(r), \dots, f_\nu(r)\}$$

является суперпозицией функций $\text{LM}(I_1, I_2, \dots, I_\nu) = \text{lex.max}\{I_1, I_2, \dots, I_\nu\}$ и $F(r) = (f_1(r), f_2(r), \dots, f_\nu(r))$. Первая является кусочно-аффинной в силу леммы 4.1, вторая — поскольку

составлена из ККА-функций. Суперпозиция ККА-функций тоже является ККА. Оценим сложность S_∞ ее представления. Сложность суперпозиции не больше суммы сложностей ее компонент. Функцию $\text{LM}(I_1, \dots, I_\nu)$ можно представить как суперпозицию $(\nu - 1)$ функций lex.max от двух аргументов, поэтому $S_\infty(\text{LM}(I_1, I_2, \dots, I_\nu)) \leq (\nu - 1)(2n - 1)$. В итоге получим

$$S_\infty(\Phi_{\max}) \leq S_\infty(\text{LM}) + S_\infty(F) \leq (\nu - 1)(2n - 1) + \sum_{i=1}^{\nu} S_\infty(\text{lex.max } D_i).$$

4.2 Небазисные операции

4.2.1 Поточечные операции над кусочно-квазиаффинными функциями

Поточечные операции над ККА-функциями, например сложение $\Phi_1 : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ и $\Phi_2 : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$, являются суперпозицией составной функции $\Phi = (\Phi_1, \Phi_2)$ и соответствующей операции: $\Phi_1 + \Phi_2 = \text{Sum} \circ \Phi$, где функция $\text{Sum} : \mathbb{Z}^{2n} \rightarrow \mathbb{Z}^n$ действует по правилу $\text{Sum}(J_1, J_2) = J_1 + J_2$. Сложность представления результата не превосходит суммы сложностей деревьев Φ_1 и Φ_2 .

4.2.2 Объединение, пересечение, разность и дополнение кусочно-квазиаффинных множеств

Характеристические функции объединения, пересечения и дополнения ККА-множеств получаются в результате поточечных операций над характеристическими функциями аргументов: $P_{M_1 \cup M_2} = P_{M_1} \vee P_{M_2}$, $P_{M_1 \cap M_2} = P_{M_1} \wedge P_{M_2}$, $P_{\bar{M}} = \neg P_M$. Разность выражается через остальные операции.

Если множество зависит от параметров, то считается, что оно пустое для тех параметров, для которых не определено. В некоторых задачах это не так, и тогда приходится еще применять операцию пересечения со множеством допустимых значений параметров.

4.2.3 Прообраз кусочно-квазиаффинного множества при кусочно-квазиаффинном отображении

Характеристическая функция прообраза ККА-множества $B \subset \mathbb{Z}^n$ при ККА отображении $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ выражается через известные операции по формуле $P_{\Phi^{-1}(B)} = P_B \circ \Phi$. Значит, $S_\infty(\Phi^{-1}(B)) \leq S_\infty(B) + S_\infty(\Phi)$.

4.2.4 Ограничение кусочно-квазиаффинной функции на множество в области значений аргумента и в области значений функции

Ограничение $\Phi|_A$ ККА-функции $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ на ККА-множество $A \subset \mathbb{Z}^m$ вычисляется по формуле $\Phi|_A = E \circ \Psi$, где $\Psi : \mathbb{Z}^m \rightarrow \mathbb{Z}^{n+1}$ — составная функция $\Psi = (\Phi, P_A)$; $E : \mathbb{Z}^{n+1} \rightarrow \mathbb{Z}^n$ — оператор проектирования на \mathbb{Z}^n : $E(J, k) = J$, $J \in \mathbb{Z}^n$, $k \in \mathbb{Z}$. Отсюда $S_\infty(\Phi|_A) \leq S_\infty(A) + S_\infty(\Phi)$.

Определение 4.1. Ограничением функции Φ на множество B в области значений будем называть функцию Φ_1 , действующую так же, как и Φ с областью определения $\mathcal{D}(\Phi_1) = \{I \in \mathcal{D}(\Phi) \mid \Phi(I) \in B\}$. Будем обозначать такое ограничение: $B|\Phi$.

Ясно, что $B|\Phi = \Phi|_{\Phi^{-1}(B)} = E \circ (\Phi, P_B \circ \Phi)$. Заметим, что функцию $(\Phi, P_B \circ \Phi)$ можно составить так, чтобы ее сложность не превышала суммы $S_\infty(B) + S_\infty(\Phi)$. Следовательно, $S_\infty(B|\Phi) \leq S_\infty(B) + S_\infty(\Phi)$.

4.2.5 Обращение кусочно-квазиаффинной функции

Произвольно взятая ККА-функция $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ может оказаться неинъективной и, следовательно, необратимой. Однако даже в этом случае у нее существуют правые обратные функции Φ^{-1} такие, что для всех значений параметров и всех точек J из области значений функции Φ (т.е. образа области определения) выполняется равенство:

$$\Phi(\Phi^{-1}(J)) = J.$$

У точки J из области значений Φ может быть несколько прообразов, следовательно, и правых обратных функций существует множество. Многие из них имеют очень большую сложность. Рассмотрим две правых обратных функции: лексикографически минимальную и лексикографически максимальную:

$$\Phi_{\min}^{-1}(J) = \text{lex.min}_{\substack{K \in \mathcal{D}(\Phi); \\ \Phi(K)=J}} K; \quad \Phi_{\max}^{-1}(J) = \text{lex.max}_{\substack{K \in \mathcal{D}(\Phi); \\ \Phi(K)=J}} K. \quad (9)$$

Предполагается, что область определения функции Φ ограничена для всех значений параметров. Таким образом, лексикографические экстремумы существуют для всех J из образа Φ .

Теорема 4.3 Функции Φ_{\min}^{-1} и Φ_{\max}^{-1} являются ККА. Сложность представления Φ_{\max}^{-1} подчиняется неравенству:

$$S_\infty(\Phi_{\max}^{-1}) \leq (\nu - 1)(2m - 1) + \sum_{i=1}^{\nu} S_\infty(\text{lex.max } D_i), \quad (10)$$

где D_i — квазивыпуклые \mathbb{Z} -многогранники, составляющие ККА-множество $M = \{K \in \mathcal{D}(\Phi) \mid \Phi(K) = J\}$, зависящее от вектора параметров J , $i = 1, 2, \dots, \nu$, $\nu = \text{LeafNum}(M) = \text{LeafNum}(\Phi)$, $S_\infty(\text{lex.max } D_i)$ — сложность представления ККА-функции из теоремы 4.1. Для Φ_{\min} в неравенстве (10) нужно lex.max поменять на lex.min .

Данное утверждение непосредственно получается, если применить теорему 4.2 к формулам (9).

Если область $\mathcal{D}(\Phi)$ не ограничена, то кусочно-квазиаффинную правую обратную функцию можно найти, сделав замену $I = R(r) = (r_1 - r_2, r_2 - r_3, \dots, r_n - r_{n+1})$ и перейдя к неотрицательным переменным r (условия неотрицательности должны быть записаны в древовидном представлении функции R). Тогда искомая правая обратная к Φ функция получается в результате суперпозиции $R \circ (\Phi \circ R)^{-1}$, где $(\Phi \circ R)^{-1}$ — лексикографически минимальная правая обратная к $\Phi \circ R$.

4.2.6 Образ кусочно-квазиаффинного множества при кусочно-квазиаффинном отображении

Определение 4.2. Образом $\Phi(B)$ множества $B \subset \mathbb{Z}^m$ при отображении $\Phi : \mathbb{Z}^m \rightarrow \mathbb{Z}^n$ будем называть множество всех целых точек $\Phi(I) \in \mathbb{Z}^n$ таких, что $I \in B$.

Заметим, что образом отрезка $[1; 5]_{\mathbb{Z}}$ при отображении $\Phi(x) = 2x$ является множество $\{2, 4, 6, 8, 10\}$, а не отрезок $[2, 10]$. Отображения, не действующие в \mathbb{Z}^n , например, $\Phi(x) = x/2$, $x \in \mathbb{Z}$, здесь не рассматриваются, хотя для многих остальных операций такое обобщение возможно.

Сведем задачу к поиску прообраза множества B при правом обратном отображении $(\Phi|_B)^{-1}$ к ограничению $\Phi|_B$. Тогда $\Phi(B) = ((\Phi|_B)^{-1})^{-1}(B)$ и $S_\infty(\Phi(B)) = S_\infty((\Phi|_B)^{-1})$. В качестве $(\Phi|_B)^{-1}$ можно взять любую ККА правую обратную функцию к $\Phi|_B$, например лексикографически максимальную (минимальную).

4.2.7 Проверка на пустоту и равенство для всех значений параметров

Кусочно-квазиаффинное множество пусто при всех значениях параметров тогда и только тогда, когда после преобразования удаления недостижимых веток его представление превращается в единственный узел «return NULL». Доказательство следует из свойств алгоритма преобразования.

Кусочно-квазиаффинные множества M_1 и M_2 равны при всех значениях параметров тогда и только тогда, когда $(M_1 \setminus M_2) \cup (M_2 \setminus M_1) = \emptyset$.

Кусочно-квазиаффинная функция Φ тождественно равна нулю на своей области определения

для всех значений параметров тогда и только тогда, когда ККА-множество $\{I \in \mathbb{Z}^m \mid \Phi(I) \neq 0\}$ пустое.

Кусочно-квазиаффинные функции Φ_1 и Φ_2 равны при всех значениях параметров тогда и только тогда, когда совпадают их области определения $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2)$ и $\Phi_1 - \Phi_2 \equiv 0$.

4.2.8 Поиск параметров, при которых выполняется проверка на пустоту и равенство

Для определения множества значений параметров, для которых ККА-множество непусто, нужно вычислить его проекцию в расширенном пространстве переменных-параметров на подпространство параметров (см. п. 4.2.6). В результате получится древовидное представление множества значений параметров, для которых существует хотя бы одно значение переменной, принадлежащее множеству.

Задача о поиске параметров, для которых совпадают два множества, сводится к предыдущей, так как $M_1 = M_2 \Leftrightarrow (M_1 \setminus M_2) \cup (M_2 \setminus M_1) = \emptyset$. При этом считается, что множество пустое для тех параметров, для которых оно не определено. Если оба множества пусты, то они совпадают.

Для поиска параметров, при которых ККА-функция Φ тождественно равна нулю на своей области определения, нужно спроектировать ККА-множество $\{I \in \mathbb{Z}^m \mid \Phi(I) \neq 0\}$ на подпространство параметров. При этом считается, что $\Phi = 0$ для тех значений параметров, для которых $\mathcal{D}(\Phi) = \emptyset$. Если в контексте задачи это не так, то нужно пересечь результат со множеством значений параметров, для которых $\mathcal{D}(\Phi) \neq \emptyset$.

Задача о поиске параметров, для которых совпадают две функции, сводится к предыдущей, так как $\Phi_1 = \Phi_2 \Leftrightarrow \mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2)$ и $\Phi_1 - \Phi_2 = 0$.

5 Заключение

Разработанная библиотека операций предоставляет унифицированный интерфейс для исследования ККА-функций и ККА-множеств. В ОРС [20] они применяются для уточнения зависимостей, при генерации кода на общую и распределенную память, отображении гнезд циклов на многоконвейерную архитектуру, проверке применимости преобразований циклов (перестановка, слияние, разрезание). Идет работа над реализацией алгоритма вычисления целых точек в произвольном ККА-множестве, зависящем от параметров, с использованием порождающих функций Барвинка (см. [10] и ссылки там). Это позволит на этапе компиляции определять число различных ячеек памяти, к которым происходит обращение в данном операторе

программы; количество итераций в цикле; число кеш-промахов; количество процессорных элементов, необходимых для запуска цикла на ПЛИС; объем динамически выделяемой памяти и т. п.

В заключение автор приносит искреннюю благодарность Аркадию Валентиновичу Климову за ряд критических замечаний и помощь в работе над статьей.

Литература

1. Воеводин В. В., Воеводин Вл. В. Параллельные вычисления. — СПб.: БХВ-Петербург, 2002.
2. Штейнберг Б. Я. Информационные зависимости и высокоуровневые распараллеливающие преобразования программ: Электронный учебник, 2007. 167 с. <http://ops.rsu.ru/works.shtml>.
3. Feautrier P. Parametric Integer Programming. — Laboratoire MASI, Institut Blaise Pascal, Université de Versailles St-Quentin, 1988. P. 25.
4. Feautrier P. Dataflow analysis of scalar and array references // Int. J. Parallel Programming, 1991. Vol. 20. No. 1. P. 23–52.
5. The Parametric Integer Programming's Home. <http://www.PipLib.org>.
6. Шульженко А. М. Расщепление многомерных циклов для эффективного распараллеливания // Параллельные вычисления в задачах математической физики: Труды Всеросс. науч.-технич. конф. — Ростов-на-Дону, 2004. С. 186–194.
7. Шульженко А. М. Автоматическое определение циклов `ParDo` в программе // Известия вузов. Северокавказский регион. Естественные науки. Приложение, 2011. Вып. 5. С. 77–88.
8. Штейнберг Б. Я. Оптимизация размещения данных в параллельной памяти. — Ростов-на-Дону: ЮФУ, 2010. 255 с.
9. Klimov A. V. Transforming affine nested loop programs to dataflow computation model // PSI'11: Ershof Informatic Conference. — Novosibirsk, 2011. P. 274–285.
10. Seghir R., Verdoolaege S., Beyls K., Loechner V. Analytical computation of Ehrhart polynomials and its application in compile-time generated cache hints. Research Report of the Université Louis Pasteur icps-2004-118, 2004.
11. Штейнберг Р. Б. Использование решетчатых графов для исследования многоконвейерной модели вычислений // Известия вузов. Северокавказский регион. Естественные науки, 2009. Вып. 2. С. 16–18.
12. Гуда С. А. Оценки длины критического пути в решетчатом графе // Параллельные вычисления и задачи управления: Труды IV Междунар. конф. — М., 2008. С. 1253–1267.
13. Pugh W. The Omega test: A fast and practical integer programming algorithm for dependence analysis // 1991 ACM/IEEE Conference on Supercomputing Proceedings. — New York, NY, USA: ACM, 1991.

14. *Maslov V.* Lazy array data-flow dependence analysis // 21st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages Proceedings, 1994. P. 311–325.
15. PolyLib — a library of polyhedral functions. <http://icps.u-strasbg.fr/polylib>.
16. Graphite: Gimple Represented as Polyhedra. <http://gcc.gnu.org/wiki/Graphite>.
17. Parma Polyhedra Library. <http://bugseng.com/products/ppl>.
18. Integer Set Library. <http://freecode.com/projects/isl>; <http://www.kotnet.org/~skimo/isl>.
19. *Штейнберг Б. Я., Нис З. Я., Петренко В. В., Черданцев Д. Н., Штейнберг Р. Б., Шульженко А. М.* Состояние и возможности Открытой распараллеливающей системы (лето 2006 г.) // Перспективы систем информатики: Труды VI Междунар. конф., рабочего семинара «Наукоемкое программное обеспечение». — Новосибирск, 2006. С. 122–125.
20. Открытая распараллеливающая система. <http://ops.rsu.ru>.
21. *Штейнберг Б. Я., Алымова Е. В., Баглий А. П. и др.* Особенности реализации распараллеливающих преобразований программ в ДВОР // Параллельные вычисления и задачи управления: Труды Междунар. конф. — М.: ИПУ РАН, 2010. С. 787–854.
22. *Штейнберг Б. Я., Абрамов А. А., Баглий А. П. и др.* Уточнение зависимостей программы в ДВОР // Параллельные вычисления и задачи управления: Труды Междунар. конф. — М.: ИПУ РАН, 2010. С. 855–864.
23. *Штейнберг Б. Я., Абрамов А. А., Алымова Е. В. и др.* Диалоговый высокоуровневый оптимизирующий распараллеливатель (ДВОР) // Научный сервис в сети Интернет: суперкомпьютерные центры и задачи: Труды Междунар. суперкомпьютерной конф. — Новороссийск, 2010. — М.: МГУ, 2010. С. 71–75.
24. *Noakala S., Risset T.* A Library for Z-polyhedral operations: Technical report PI-1330. — Rennes, France: IRISA, 2000.

МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ СОЗДАНИЯ ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ВЫЧИСЛЕНИЯ ИНДИКАТОРОВ ТЕМАТИЧЕСКИХ ВЗАИМОСВЯЗЕЙ НАУКИ И ТЕХНОЛОГИЙ*

В. А. Минин¹, И. М. Зацман², М. Г. Кружков³, Т. П. Норемян⁴

Аннотация: Анализируется зарубежный опыт вычисления индикаторов тематических взаимосвязей науки и технологий. Цель анализа заключается в разработке принципов создания отечественных информационных систем для вычисления индикаторов взаимосвязей с учетом исторически сложившейся в нашей стране структуры наследуемых научных и патентных информационных ресурсов. Этот вид информационных систем является новым для российской научно-технической сферы. Их создание необходимо для мониторинга и оценивания программ научных исследований и принятия решений на всех этапах программной деятельности. В статье предлагается методология определения индикаторов тематических взаимосвязей в отечественной научно-технической сфере как основа создания информационных систем, предназначенных для вычисления их значений.

Ключевые слова: взаимосвязи науки и технологий; классификация научных направлений; международная патентная классификация; рубрицирование научных документов

1 Введение

1 ноября 2012 г. на заседании Правительства РФ был рассмотрен проект «Государственной программы РФ «Развитие науки и технологий» на 2013–2020 годы», включающей шесть подпрограмм. Для подпрограммы номер 2 «Прикладные проблемно-ориентированные исследования и развитие научно-технического задела в области перспективных технологий» предлагается установить два следующих индикатора⁵, характеризующих достижение цели Государственной программы⁶ в части проблемно-ориентированных исследований:

- (1) коэффициент изобретательской активности (число отечественных патентных заявок на изобретения, поданных в России в расчете на 10 тыс. чел. населения);
- (2) число патентных заявок на изобретения, поданных отечественными заявителями в России

из организаций — участников Государственной программы [1, с. 48].

Из названий этих двух индикаторов следует, что в процессе оценивания научно-технического задела в области перспективных технологий в этой программе учитывается только изобретательская активность. Второй, но не менее важный аспект, а именно: цитируемость в описаниях изобретений публикаций, являющихся непосредственными результатами научных исследований и разработок, — не нашел своего отражения в списке индикаторов этой программы. Здесь важно отметить, что современные методологии, реализуемые в информационных системах индикаторного оценивания процессов трансформации знаний в новые технологии, состоят из двух основных компонентов [2]:

- (1) индикаторное оценивание тематических взаимосвязей результатов исследований и разработок с технологической модернизацией (что да-

* Работа выполнена при поддержке РГНФ, грант № 12-02-12019в.

¹ Российский фонд фундаментальных исследований, minin@rfr.ru

² Институт проблем информатики Российской академии наук, iz_ipi@al70.ipi.ac.ru

³ Институт проблем информатики Российской академии наук, magnit75@yandex.ru

⁴ Институт проблем информатики Российской академии наук, izzittami@gmail.com

⁵ Финансовые индикаторы программы, например удельный вес внебюджетных средств во внутренних затратах на исследования и разработки, в статье не рассматриваются.

⁶ Целью всей программы является «формирование конкурентоспособного и эффективно функционирующего сектора исследований и разработок и обеспечение его ведущей роли в процессах технологической модернизации российской экономики» [1, с. 9].

ет возможность получать экспертные оценки инновационно-технологического потенциала направлений научных исследований);

- (2) индикаторное оценивание изобретательской активности в процессе исследований и разработок (как правило, оценивание ведется по направлениям технологического развития).

Таким образом, в Государственной программе РФ «Развитие науки и технологий» на 2013–2020 гг. используется только второй компонент, но полностью отсутствует первый. Одна из причин заключается в том, что в нашей стране отсутствуют те информационные системы, которые могли бы вычислять индикаторы взаимосвязей научных направлений и технологий, сопоставляя накопленные научные и патентные информационные ресурсы [3, 4].

Возможно, по этой причине разработчики программы и ограничились традиционным индикаторным оцениванием изобретательской активности в процессе исследований и разработок в рамках этой программы, хотя имеющийся зарубежный опыт, описанию которого посвящен третий раздел статьи, свидетельствует о ключевой роли количественных индикаторов взаимосвязей для многоаспектного оценивания научно-технического задела в области перспективных технологий в интересах принятия решений на всех этапах программной деятельности.

Для изменения сложившейся ситуации Российский гуманитарный научный фонд (РГНФ) начал в 2012 г. финансирование инициативного проекта по гранту № 12-02-12019в с целью создания экспериментального образца информационной системы, предназначенного для индикаторного оценивания взаимосвязей результатов научных исследований с развитием информационных технологий (ИТ). Основной целью статьи является описание имеющегося задела и полученных в 2012 г. результатов этого проекта.

2 Проблемы программно-целевого планирования в научно-технической сфере

В течение 2005–2011 гг. сформировалось новое направление исследований «Методы и технологии

информационного мониторинга в научно-технической сфере». Результаты, полученные специалистами ЦЭМИ РАН и ИПИ РАН, опубликованы в подразд. 5.4 монографии «Мезоэкономика развития» [5], двух книгах [6, 7] и 19 статьях [2–4, 8–23], в том числе в трех зарубежных публикациях. В этих работах была сформулирована проблема информационного мониторинга, анализа и оценивания научной деятельности, в том числе программ научных исследований¹ (далее по тексту — проблема мониторинга), а также оценивания влияния их результатов на социально значимые сферы деятельности, в том числе на создание новых технологий.

В процессе постановки этой проблемы учитывались методические документы и нормативно-правовые акты, предписывающие переход к программно-целевым методам бюджетного планирования, ориентированным на повышение результативности использования бюджетных расходов. В частности, постановление Правительства РФ от 22 мая 2004 г. № 249 «О мерах по повышению результативности бюджетных расходов» (далее — постановление № 249) предусматривает переход на среднесрочное бюджетирование, ориентированное на результаты (СБОР), во всех социально значимых сферах деятельности, в том числе в научно-технической сфере. Основные принципы СБОР сформулированы в Концепции реформирования бюджетного процесса в РФ в 2004–2006 гг., одобренной постановлением № 249.

С точки зрения разработки информационных систем мониторинга ключевым методическим документом служат рекомендации по подготовке Докладов о результатах и основных направлениях деятельности субъектов бюджетного планирования на 2006–2008 гг. [24]. Согласно этим рекомендациям используемые системы индикаторов и показателей должны соответствовать следующим 9 требованиям, которые были уточнены в процессе постановки проблемы мониторинга следующим образом:

- (1) *адекватность*: показатель (индикатор²) должен характеризовать прогресс в достижении цели или решении задачи; используемые комплексы показателей (индикаторов) должны охватывать *все существенные аспекты достижения цели и/или решения задач ПД*;
- (2) *точность*: погрешности измерения не должны приводить к искаженному представлению о результатах, эффективности и результативности ПД;

¹Далее наряду со словосочетанием «программа научных исследований» будет использоваться термин «программная деятельность» (кратко — ПД). Отличие ПД от программ заключается в том, что кроме периода действия программы рассматривается стадия формирования и стадия оценивания результатов программы после ее завершения.

²Термин «показатель» используется в этой статье как родовой по отношению к термину «индикатор». Более подробно вопрос о соотношении этих терминов, включая их дефиниции, рассмотрен в работе [10].

- (3) *объективность*: не допускается использование показателей, улучшение отчетных значений которых возможно при ухудшении реального положения дел; используемые показатели должны в наименьшей степени создавать стимулы к искажению результатов ПД;
- (4) *достоверность*: способ сбора и обработки исходной информации должен допускать возможность проверки точности как собранных, так и обработанных данных в процессе независимого мониторинга и оценивания ПД;
- (5) *однозначность*: определение показателя (индикатора) должно обеспечивать *одинаковое понимание существа измеряемой характеристики*;
- (6) *экономичность*: получение отчетных данных должно производиться с минимально возможными затратами, применяемые показатели должны в максимальной степени основываться на уже существующих процессах сбора информации о ПД;
- (7) *сопоставимость*: выбор показателей следует осуществлять исходя из необходимости непрерывного накопления данных и обеспечения их сопоставимости за отдельные периоды с показателями, используемыми для оценки прогресса в решении сходных (смежных) задач, а также с показателями, используемыми в международной практике;
- (8) *своевременность и регулярность*: отчетные данные о ПД должны поступать со строго определенной периодичностью и с незначительным временным лагом между моментом сбора информации и моментом использования результатов ее обработки;
- (9) *уникальность*: показатели достижения цели ПД не должны представлять собой объединение нескольких показателей, характеризующих решение отдельных относящихся к этой цели задач ПД.

Перечисленные требования представляют собой в совокупности методический фактор, который приобретает особую актуальность для оценивания ПД в тех случаях, когда на основе значений индикаторов планируется формировать и реализовывать стратегию в сфере науки, в том числе распределять бюджетные средства по фундаментальным и технологически ориентированным научным направлениям. С позиции разработчиков информационных систем и технологий мониторинга, с помощью которых должны вычисляться значения индикаторов

ПД, отсутствуют необходимые теоретические основы их разработки, учитывающие *все 9 требований, начиная с требования адекватности*.

Лакуны в современной системе знаний о мониторинге ПД анализировались в отечественных и зарубежных исследованиях неоднократно [22, 23, 25, 26], а результаты этих исследований обсуждались на семинарах.

Например, в материалах семинара по методическим вопросам оценивания федеральных научно-исследовательских программ США, состоявшегося 4–5 декабря 2003 г., приведена сводная таблица из 24 нерешенных задач [25]. Решение некоторых из них требует проведения фундаментальных и прикладных научных исследований.

В эту таблицу включены следующие две проблемы, иллюстрирующие необходимость дальнейшего развития теоретических основ мониторинга:

- (1) отсутствуют модели и методы мониторинга, анализа и оценивания научно-исследовательских программ, обеспечивающие проверку точности данных, используемых для оценки результативности этих программ, т. е. не выполняется требование **достоверности**;
- (2) между экспертами Административно-бюджетного управления и федеральных агентств США нередко возникают конфликты из-за различного понимания смысла индикаторов результатов, эффективности реализации и результативности научно-исследовательских программ (не выполняется требование **однозначности**).

Ряд актуальных задач мониторинга и оценивания ПД был рассмотрен в процессе международной экспертизы итогов мониторинга и оценивания рамочных программ ЕС [26]. Лакуны в системе знаний о мониторинге ПД и, как следствие, **неадекватность** комплексов индикаторов, используемых для оценивания конкретных программ, стали предметом анализа и в отечественных исследованиях. Был проведен анализ комплекса индикаторов одной из отечественных научных программ и показана неполнота используемого в этой программе набора индикаторов результатов, эффективности и результативности [23]. В процессе анализа рассматривались три основные категории результатов ПД [22, 24, 26]¹:

- (1) *непосредственные результаты* исследовательских программ (outputs), например публикации и доклады по итогам выполнения проек-

¹См. также Акт Конгресса США о результатах и результативности государственного управления (Government Performance and Results Act of 1993), принятый 5 января 1993 г.

тов программ (к непосредственным результатам относится только факт публикации или выступления с докладом, но не их содержание);

- (2) *целевые ожидаемые* и фактически полученные *конечные результаты* (outcomes), которые могут быть заданы как для программы в целом, так и для отдельных ее тематических направлений исследований и проектов (к целевым результатам относятся запланированные итоги программ в целом, например увеличение доли молодых ученых среди исследователей, и/или содержательные итоги работ тематических направлений и проектов программ, например выявление закономерностей распространения некоторого вида организмов или синтез нового вещества);
- (3) социально-экономические, технологические и иные изменения, связанные с итогами реализации проектов программ (impacts & linkages), которые представляют собой эффект от применения (влияния) непосредственных или целевых результатов на развитие как самой сферы науки, так и других социально значимых сфер деятельности общества — развитие сфер образования и здравоохранения, технологического и экономического развития, модернизации правовой сферы и т. д. (далее по тексту — *результаты влияния*).

Было показано, что анализируемый комплекс индикаторов оценивания программы не содержит индикаторов результатов влияния [23]. В результате проведенного анализа были сформулированы две новые задачи:

- разработка методов и моделей проектирования новых индикаторов ПД с целью повышения степени адекватности комплексов индикаторов, используемых для оценивания конкретных программ;
- создание информационных систем для вычисления уже спроектированных индикаторов взаимосвязей науки и технологий.

Что касается первой задачи, то методы проектирования новых индикаторов были рассмотрены в работах [27, 28]. Эти методы были разработаны на основе семиотических моделей [29–34]. В данной статье предлагается подход к решению второй задачи. При этом предполагается, что некоторые индикаторы взаимосвязей науки и технологий уже спроектированы и разработчики информационных систем мониторинга могут использовать описания этих индикаторов в процессе разработки технологий вычисления их значений. Это предположение

обосновано тем, что за рубежом исследования взаимосвязей науки и технологий ведутся уже не один десяток лет (см. следующий раздел статьи) и ряд индикаторов спроектирован и опробован на практике. Однако это не означает, что проектирование новых индикаторов оценивания взаимосвязей перестало быть актуальной задачей.

3 Взаимосвязи науки и технологий

Один из наиболее сложных аспектов в исследовании взаимосвязей науки и технологий состоит в индикаторном оценивании процессов переноса знаний, в том числе отраженных в научных публикациях, из разных областей исследований в сферу технологического развития. Причина пристального внимания к индикаторному оцениванию процессов переноса знаний заключается в том, что финансирование научных исследований, ориентированных на развитие научно-технического задела в области перспективных технологий, связано, с одной стороны, с большим риском. С другой стороны, есть риск упустить новые прорывные решения и потерять конкурентоспособность [35]. Это касается конкурентоспособности как отдельных предприятий в конкретной технологической сфере, так и государства в целом [36].

Поэтому и возникла потребность в решении задачи оценивания взаимосвязей науки и технологий в процессе создания информационных систем мониторинга как инструментов обеспечения стратегического планирования и управления финансированием исследований, ориентированных на технологическое развитие. При решении этой задачи один из самых сложных вопросов заключается в том, как зафиксировать факт передачи и использования в технологической сфере результатов научных исследований [35].

В силу указанных причин и стали разрабатываться индикаторы оценивания взаимосвязей, а также методы и алгоритмы определения их значений. Один из предложенных подходов заключается в том, что процессы передачи знаний от науки к технологиям отслеживаются с помощью научных публикаций, цитируемых экспертами в отчетах о патентном поиске и/или авторами изобретений в их описаниях, что конечно не отражает весь спектр взаимосвязей науки и технологий.

По мнению Тийссена с соавторами, основная сложность в отслеживании всего спектра заключается в том, что явные процессы переноса знаний являются многочисленными, а неявные — труд-

но идентифицируемыми [37]. Тем не менее в процессе исследования этой проблемы за последние 30 лет наблюдается явный прогресс, так как был предложен ряд индикаторов, характеризующих взаимосвязи науки и технологий на макроуровне (определяется интегральная интенсивность цитирования результатов всех тех научных исследований, которые связаны с развитием технологий, см. табл. 1) [38] и более детальные взаимосвязи отдельных научных дисциплин и видов технологий [39].

В 1985 г. в процессе сопоставительного анализа статей по биологии и библиографических ссылок на статьи в массивах описаний изобретений по биотехнологиям были экспериментально зафиксированы взаимосвязи между научными публикациями и развитием биотехнологий. В процессе анализа различались патентные ссылки, т. е. ссылки на ранее выданные патенты, и непатентные ссылки, среди которых выделялись ссылки на статьи из журналов, включенных в Указатель научного цитирования (Science Citation Index или SCI), и ссылки на статьи из других журналов. Было обработано более 6500 патентных и непатентных ссылок, извлеченных из описаний 399 патентов [39].

Позже, в конце прошлого века, было проведено широкомасштабное исследование взаимосвязей науки и технологий с использованием более одного миллиона непатентных ссылок, извлеченных из десятилетнего массива описаний изобретений США и Европейского патентного ведомства (ЕПВ). Было показано экспериментально, что *75% научных статей*, цитируемых в этих массивах описаний изобретений по широкому спектру технологий, были подготовлены по результатам, полученным в *некоммерческом секторе научной сферы*. Это дало возможность авторам исследования сделать вывод о сильной зависимости технологического развития от степени государственной поддержки науки в США и европейских странах [40].

В конце прошлого века Мэнсфилд провел серию экспериментальных исследований, взяв в качестве исходных данных сведения об итогах работы ряда американских компаний. Его работы содержат эмпирические данные о другом аспекте взаимосвязей результатов научных исследований и технологических инноваций. На основе анализа этих данных он пришел к выводу, что *10% технологических новшеств* не были бы изобретены или же были бы созданы с большой задержкой, если бы они были сделаны без использования результатов соответствующих академических исследований [41–43].

Один из основных вызовов в современных исследованиях взаимосвязей науки и технологий состоит в том, чтобы разработать методологию количественного и качественного оценивания этих

взаимосвязей как для целых областей знаний, так и для отдельных направлений исследований. Даже при наличии качественных экспертных оценок вклада тех или иных научных направлений в технологическое развитие все равно необходимо вычислять количественные индикаторы взаимосвязей с целью верификации качественных оценок.

Проведенный анализ зарубежного опыта показывает, что вычисление индикаторов взаимосвязей науки и технологий требует автоматизированной обработки больших объемов слабоструктурированных полнотекстовых описаний изобретений и сопоставления результатов обработки патентной информации с названиями источников научных публикаций (журналов или материалов конференций) и с названиями самих статей, хранящихся в научных электронных библиотеках. Это сопоставление дает возможность определить тематику того научного направления, к которому относятся научные публикации, цитируемые в описаниях изобретений. В процессе обработки тысяч и миллионов описаний изобретений используются сложные алгоритмы парсинга, с помощью которых анализируются патентные тексты, в которых выделяются и структурируются библиографические ссылки на научные публикации [38].

4 Информационные ресурсы для определения взаимосвязей

Для сопоставления результатов обработки описаний изобретений с научной информацией используются те патентные информационные ресурсы, в которых цитируются научные публикации. После идентификации каждой научной публикации в описании изобретения определяется источник публикации и ее тематика, что и дает возможность устанавливать тематические взаимосвязи между научными направлениями и технологиями.

Тематика каждого научного направления задается в виде одной или нескольких рубрик выбранной системы классификации областей знаний. Тематика каждой анализируемой технологии задается в виде списка рубрик Международной патентной классификации (МПК). Наиболее часто используются списки рубрик МПК из номенклатуры, разработанной Фраунгоферовским институтом системотехники и инновационных исследований (Fraunhofer Gesellschaft-Institute für Systemtechnik und Innovationsforschung — FhG-ISI). Примеры списков рубрик МПК из номенклатуры FhG-ISI приведены в табл. 1 для пяти видов технологий. Последний столбец содержит данные интегральной

Таблица 1 Коды МПК и интегральные интенсивности цитирования для различных технологий

Название технологий	Коды МПК	Интегральная интенсивность цитирования
Биотехнологии	C07G; C12M, N, P, Q, R, S	138,43
Фармацевтические	A61K	83,71
Полупроводниковые	H01L	56,44
Оптические	G02; G03B, C, D, F, G, H; H01S	21,89
Информационные	G06; G11C; G10L	20,39

интенсивности цитирования результатов тех научных исследований, которые связаны с развитием технологий, указанных в первом столбце. Интегральная интенсивность цитирования определена в табл. 1 как число цитируемых научных публикаций на 100 описаний изобретений [38].

В номенклатуре FhG-ISI предметная область ИТ, для которой планируется разработать экспериментальный образец информационной системы, предназначенной для индикаторного оценивания тематических взаимосвязей, описывается следующими тремя рубриками МПК (см. табл. 1):

G06 — «Вычисление; счет» (эта рубрика МПК включает оптические вычислительные устройства, обработку цифровых данных с помощью компьютеров, аналоговые и гибридные компьютеры);

G11C — «Запоминающие устройства статического типа»;

G10L — «Анализирование или синтезирование речи; распознавание речи».

Таким образом, экспериментальный образец информационной системы должен содержать патентные информационные ресурсы по трем рубрикам МПК: G06, G11C и G10L. Используя эти ресурсы, система должна будет по запросам экспертов определять индикаторы тематических взаимосвязей для предметной области ИТ, т. е. для выбранного экспертом научного направления (дисциплины) будет вычисляться частота цитируемости публикаций этого направления в описаниях изобретений по ИТ, имеющих коды рубрик G06, G11C и G10L.

Используемые за рубежом подходы к определению количественных индикаторов тематических взаимосвязей науки и технологий основаны, как правило, на следующих исходных положениях.

Во-первых, эти индикаторы трактуются как частотные оценки, пропорциональные числу публикаций по тем научным направлениям, которые цитируются в описаниях изобретений, относящихся к рассматриваемой технологии. Так как научные публикации из одной области знаний или одного научного направления могут цитироваться в опи-

саниях изобретений из разных рубрик МПК, то в этом случае результат количественного оценивания является векторной величиной, а для нескольких областей знаний или научных направлений — матрицей (табл. 2) [44, с. 420]. При получении этих оценок учитывается тот факт, что цитирование научных справочников и классических научных трудов в описаниях изобретений, скорее всего, будет указывать на хрестоматийные, а не на новые научные результаты. Кроме того, достоверность частотных оценок во многом будет зависеть от степени представительности используемых массивов патентных и научных информационных ресурсов.

Во-вторых, рассматриваемые частотные оценки отражают тематические взаимосвязи научных результатов с технологиями, но не отражают обратных связей (в частности, влияния результатов технологических разработок на инициирование фундаментальных и прикладных исследований и получение в итоге новых научных результатов).

В-третьих, если в отдельно взятом описании изобретения нет ссылок на научные публикации, то это не должно интерпретироваться как отсутствие связей этого отдельно взятого изобретения с научными результатами, так как не все случаи передачи знаний обязательно эксплицируются в виде цитируемых научных публикаций [45].

Авторы «Третьего европейского отчета по научно-технологическим индикаторам» отмечают особое положение такой области знаний, как математика [44, с. 421]. С одной стороны, в описаниях изобретений к патентам (далее по тексту — в патентах) редко встречаются ссылки на математические публикации (см. табл. 2), с другой стороны, очевидно, что математические методы и модели являются необходимыми во многих сферах технологий (производства).

Остановимся на этом примере процентных индикаторов, определенных с использованием научных и патентных информационных ресурсов, более подробно. Набор индикаторов представляет собой матрицу чисел, характеризующих взаимосвязи

Таблица 2 Индикаторы взаимосвязей сфер технологий (производства) и областей знаний

Сферы технологий (производства)	Науки о Земле и других планетах	Сельскохозяйственные науки	Химия	Медицинские науки	Технические науки	Науки о жизни	Физика	Математика	Междисциплинарные проблемы
Биотехнологии	0,1	4,4	2,0	21,3	0,2	52,8	0,1	0,0	19,1
Фармацевтические и косметические средства	0,0	2,5	5,7	42,4	0,2	34,0	0,1	0,0	14,9
Технологии тонкого органического синтеза	0,0	2,8	10,8	28,8	0,2	40,6	0,1	0,0	16,5
Сельское хозяйство и химическое производство пищевых продуктов	0,1	33,4	2,1	4,2	0,8	48,1	0,0	0,0	11,2
Контрольно-измерительные технологии	0,4	1,5	6,3	29,0	6,6	32,4	10,3	0,0	13,3
Нефтехимическая промышленность и химическая переработка сырья	0,2	8,1	10,6	32,0	1,1	33,3	1,1	0,0	13,6
Технологии производства пищевых продуктов	0,0	15,6	3,9	15,6	3,9	41,7	0,6	0,0	18,7
Полупроводниковая промышленность	0,5	0,5	13,1	1,9	23,8	0,3	58,7	0,0	1,1
Телекоммуникационные технологии	0,6	1,9	1,0	2,5	77,0	0,7	15,7	0,2	0,3
Ядерная техника и технологии	1,6	0,0	8,1	17,7	37,1	4,8	24,2	0,0	6,5
Информационные технологии	1,2	1,0	1,2	6,8	71,1	5,4	11,4	0,2	1,7
Космическая техника	0,0	0,0	20,0	0,0	50,0	0,0	30,0	0,0	0,0
Оптическая промышленность	0,2	0,0	12,0	0,7	22,5	1,7	61,2	0,0	1,4
Медицинские технологии	0,0	1,1	2,8	51,7	4,0	23,4	6,8	0,0	9,9
Технологии обработки поверхностей и лакокрасочные технологии	1,7	0,6	32,8	1,7	16,9	2,8	39,0	0,0	4,5
Технологии химии полимеров	0,2	4,3	42,6	13,3	3,1	26,0	1,0	0,0	9,3
Аудиовизуальные технологии	0,0	0,0	0,0	2,9	63,8	0,0	32,6	0,0	0,7
Металлургическая промышленность и производство материалов	3,4	0,0	29,9	3,4	34,0	2,7	19,7	0,0	6,8
Электротехническая промышленность	0,0	0,4	23,3	3,6	25,7	0,4	42,2	0,0	4,4
Химическое машиностроение	2,9	2,9	42,7	9,7	18,4	9,7	4,9	0,0	8,7

Примечание 1. Междисциплинарные проблемы позиционируются как самостоятельная область знаний.

Примечание 2. Данные таблицы получены в результате обработки европейских патентных заявок, поданных в период времени 1992–1996 гг.

Примечание 3. Сумма значений процентных индикаторов по строкам может быть меньше 100%, так как в первой строке таблицы перечислены не все области знаний.

сфер технологий (производства) и областей знаний, определенную на основе информационных ресурсов ЕПВ. Эта матрица получена в результате выборки и редуцирования данных из упомянутого европейского отчета [44].

В первой строке табл. 2 приведены области знаний. В первой колонке перечислены 20 сфер тех-

нологий (производства). Каждый столбец табл. 2, кроме первого, содержит вектор частотного распределения научных публикаций одной из 9 перечисленных областей знаний по 20 сферам технологий. Для вычисления каждого значения индикатора считалось число тех научных публикаций этой области знаний, которые цитировались в патентах, относя-

щихся к указанной в начале строки сфере технологий.

Например, строка «Биотехнологии» содержит число 52,8. Это число говорит о том, что в обработанном массиве патентов ЕПВ по биотехнологиям 52,8% всех научных цитат являются ссылками на научные публикации по наукам о жизни. Строка «Телекоммуникационные технологии» содержит числа 0,6 и 0,2, которые говорят о том, что в патентах этого вида технологий 0,6% всех цитат являются ссылками на научные публикации по наукам о Земле и 0,2% — ссылками на математические публикации.

Очевидно, что каждая из областей знаний может быть структурирована более детально и для нее может быть получена своя матрица индикаторов, определенных с использованием научных и патентных информационных ресурсов. Пример более детальной структуризации с индикаторами взаимосвязей теоретических и прикладных научных дисциплин с ИТ, вычисленными с использованием патентов США и ЕПВ, был рассмотрен в работе [46].

Отметим, что данные диаграмм, приведенных в этой работе, и табл. 2 иллюстрируют существенные отличия в цитировании научных публикаций для разных теоретических и прикладных дисциплин. Существуют отличия и в региональном разрезе, например доля научных публикаций по компьютерной науке, на которые есть ссылки в патентах США по ИТ, равна 2,35%, что на 2,2% меньше, чем доля научных публикаций по компьютерной науке, на которые есть ссылки в европейских патентах по ИТ [46].

Таким образом, ретроспективное сопоставление научных и патентных информационных ресурсов является основой для индикаторного оценивания взаимосвязей результатов науки и развития технологий. При этом вычисление значений индикаторов не является итоговым этапом методологии оценивания этих взаимосвязей. Заключительным этапом методологии является экспертная оценка вычисленных значений индикаторов. При этом мнения экспертов далеко не всегда будут соответствовать вычисленным значениям индикаторов.

Естественно, возникает вопрос о том, что если используемая методология индикаторного оценивания взаимосвязей имеет объективную основу в виде ретроспективных научных и патентных информационных ресурсов, то зачем нужен завершающий этап — экспертиза вычисленных значений индикаторов. Причина кроется в том, что значения полученных индикаторов являются объективными, но только *косвенными мерами* взаимосвязей науки и технологий. Поэтому для их верификации и используются субъективные, но прямые методы

экспертизы, включающие содержательный анализ сопоставляемых научных и патентных информационных ресурсов. Если косвенные объективные оценки, вычисленные на основе накопленных информационных ресурсов, совпадают с субъективными и прямыми согласованными экспертными оценками, то только в этом случае они могут служить надежной основой для выбора приоритетных направлений ориентированных научных исследований, являющихся ключевыми для создания перспективных технологий.

Таким образом, сопоставление научных и патентных информационных ресурсов является ключевым этапом методологии индикаторного оценивания процессов трансформации знаний в новые технологии.

5 Предлагаемая методология индикаторного оценивания

Как отмечалось ранее, современные методологии оценивания процессов трансформации знаний в новые технологии состоят из двух основных компонентов: (1) индикаторного оценивания взаимосвязей результатов исследований и разработок с технологической модернизацией и (2) индикаторного оценивания изобретательской активности в процессе исследований и разработок. В России до настоящего времени используется только второй компонент, но полностью отсутствует первый, так как в нашей стране нет информационных систем для индикаторного оценивания взаимосвязей.

Важно отметить, что отечественные патентные информационные ресурсы не удовлетворяют требованиям зарубежных вариантов методологии оценивания процессов трансформации знаний в новые технологии (по полноте и структуре ресурсов), что исключает возможность копирования зарубежных вариантов [3, 4]. Поэтому актуальной является задача разработки нового варианта методологии, который учитывал бы исторически сложившуюся структуру и наполнение отечественных патентных информационных ресурсов.

Приведем описание предлагаемого варианта методологии определения индикаторов тематических взаимосвязей науки и технологий в интересах выбора направлений ориентированных научных исследований.

Предлагаемый вариант методологии включает следующие 8 основных этапов.

1. Определение временного периода для вычисления ретроспективного тренда количественных индикаторов, выбор отечественных систем

классификации областей знаний¹, составление списков анализируемых технологических областей, формирование перечня индикаторов, которые необходимы в интересах выбора направлений ориентированных научных исследований.

2. Формирование массивов патентных информационных ресурсов по анализируемым технологическим областям (с использованием списков рубрик МПК из номенклатуры FhG-ISI) для определенного временного отрезка вычисления ретроспективного тренда количественных индикаторов.
3. Формирование базы данных научных публикаций, цитируемых в описаниях изобретений, в виде их библиографических описаний (далее по тексту — указатель цитирования), которая создается на основе сформированных массивов патентных информационных ресурсов по анализируемым технологическим областям. Для каждой публикации указывается идентификатор изобретения в виде номера патента и/или заявки на выдачу патента и все индексы МПК этого изобретения.
4. Автоматизированная структуризация библиографических описаний с одновременной нормализацией и пополнением нормативных списков отечественных и иностранных журналов (для статей), названий конференций (для научных докладов и сообщений) и названий издательств (для книг и трудов конференций с учетом того, что труды ряда конференций публикуются в периодических изданиях) с целью обеспечения автоматизированного рубрицирования.
5. Автоматизированное рубрицирование библиографических описаний цитируемых научных публикаций с использованием названий журналов, издательств, конференций и названий статей, а также с использованием выбранной системы классификации областей знаний.
6. Вычисление индикаторов тематических взаимосвязей как частотных оценок, полученных в процессе сопоставления рубрик МПК описаний изобретений, в которых встретились ссылки на научные публикации, и рубрик системы классификации областей знаний, к которым относятся цитируемые научные публикации, а также визуализация вычисленных значений индикаторов.
7. Получение экспертных оценок для вычисленных значений индикаторов взаимосвязей на

основе аналитических исследований и содержательного анализа сопоставляемых научных и патентных информационных ресурсов.

8. Проверка согласованности экспертных оценок и вычисленных значений индикаторов, фиксация тех пар рубрик МПК и системы классификации областей знаний, в которых мнения экспертов не согласованы между собой (и/или не согласуются с вычисленными значениями индикаторов), и для всех случаев несогласованности — дополнение базы данных цитируемых научных публикаций аннотациями и другими дополнительными полями (с последующим переходом на этап 5).

Сопоставим предлагаемый вариант методологии с европейским вариантом [47]. Для первого этапа отметим одно принципиальное отличие. В предлагаемом варианте имеется возможность выбора систем классификации областей знаний, которая отсутствует в европейском варианте. С точки зрения принятия решений по финансированию направлений ориентированных исследований эта возможность позволяет применять ту систему классификации областей знаний, которая используется при принятии решений (например, ГРНТИ, классификатор направлений фундаментальных научных исследований РАН, классификаторы РФФИ, РГНФ и т. д.).

В европейском варианте применяется только одна система классификации областей знаний (SCI-ISI Journal Classification System), по рубрикам которой распределены журналы Указателя научного цитирования SCI. Эта система классификации используется в процессе рубрицирования цитируемых научных публикаций, которые идентифицируются с помощью пяти полей: (1) фамилия первого автора; (2) источник публикации; (3) год публикации; (4) том; (5) номер первой страницы публикации. Используя название источника публикации, с помощью этой системы классификации определяется тематическая рубрика публикации [47, с. 407].

На третьем этапе отметим второе принципиальное отличие. В предлагаемом варианте методологии предусмотрено формирование указателя цитирования. При этом в указатель включаются и затем используются все научные публикации, цитируемые в описаниях отобранных изобретений (в европейском варианте рассматриваются только так называемые «front-page публикации», которые цитируются на первых страницах описаний отобранных изобретений).

¹Основным критерием для выбора системы классификации областей знаний является ее использование в процессе принятия решений по финансированию направлений ориентированных научных исследований.

На пятом этапе отметим еще одно важное отличие. В европейском варианте для определения тематической рубрики публикации используется только название источника публикации, а в предлагаемом варианте используются и названия источников, и названия цитируемых научных публикаций. Ключевые слова из названий публикаций используются для уточнения рубрики публикации в тех случаях, когда одному источнику соответствует несколько рубрик. С этой целью для каждой рубрики системы классификации областей знаний, которая используется при принятии решений, планируется формировать ее терминологический портрет, используемый в процессе уточнения рубрики публикации.

Еще одно принципиальное отличие касается восьмого этапа методологии, где появляется возможность существенно снизить область несогласованных экспертных оценок для индикаторов за счет использования аннотаций научных публикаций. Ключевые слова аннотаций планируется использовать в тех случаях, когда одни названия источников и цитируемых научных публикаций дают несогласованные экспертные оценки и/или эти оценки не согласованы с полученными значениями индикаторов.

6 Заключение

Предлагаемый вариант методологии позволит вычислять значения количественных индикаторов взаимосвязей науки и технологий как для целых областей знаний, так и для отдельных направлений исследований в целях идентификации технологически ориентированных научных направлений.

Впервые в нашей стране появляется возможность в процессе идентификации ориентированных научных направлений использовать *объективную основу*, а именно: накопленные научные и патентные информационные ресурсы.

Предлагаемый вариант методологии обладает рядом следующих принципиальных отличий от зарубежных аналогов:

- выбор именно тех систем классификации областей знаний, которые используются в процессе принятия решений;
- использование всех научных публикаций, цитируемых в описаниях отобранных изобретений, для определения значений индикаторов взаимосвязей науки и технологий (одновременно планируется вычислять и варианты индикаторов с использованием только «front-page публикаций», что даст возможность сопоставить

полученные отечественные результаты с зарубежными);

- формирование и применение терминологических портретов рубрик отобранных систем классификации областей знаний;
- использование ключевых слов из названий публикаций для уточнения рубрики публикации в тех случаях, когда одному источнику публикации приписано несколько рубрик систем классификации областей знаний;
- использование ключевых слов из аннотаций публикаций в тех случаях, когда названия источников и цитируемых научных публикаций дают несогласованные экспертные оценки и/или эти оценки не согласованы с полученными значениями индикаторов.

В соответствии с предлагаемым вариантом методологии разработаны: концепция и архитектура информационной системы, предназначенной для индикаторного оценивания взаимосвязей науки и технологий; архитектуры функциональных подсистем, обеспечивающих реализацию методологии; основные проектные решения для функциональных подсистем и видов обеспечения этой системы.

Разрабатываемая информационная система индикаторного оценивания взаимосвязей науки и технологий не имеет аналогов в российской научно-технической сфере. Ее создание необходимо для проведения мониторинга, многоаспектного оценивания программ научных исследований и прогнозирования научно-технологического развития страны.

Литература

1. Проект государственной программы РФ «Развитие науки и технологий» на 2013–2020 годы. <http://минобрнауки.рф/документы/2475>.
2. *Архипова М. Ю., Зацман И. М., Шульга С. Ю.* Индикаторы патентной активности в сфере информационно-коммуникационных технологий и методика их вычисления // Экономика, статистика и информатика. Вестник УМО, 2010. № 4. С. 93–104.
3. *Зацман И. М., Шубников С. К.* Принципы обработки информационных ресурсов для оценки инновационного потенциала направлений научных исследований // Электронные библиотеки: перспективные методы и технологии, электронные коллекции — RCDL'2007: Труды IX Всеросс. научной конф. — Переславль: Ун-т города Переславля, 2007. С. 35–44.
4. *Зацман И. М., Курчавова О. А., Галина И. В.* Информационные ресурсы и индикаторы для оценки инновационного потенциала направлений научных исследований // Системы и средства информатики. Доп. вып. — М.: Наука, 2008. С. 159–175.

5. Мезоэкономика развития / Под ред. чл.-корр. РАН Г. Б. Клейнера. — М.: Наука, 2011. 805 с.
6. Клейнер Г. Б., Голиченко О. Г., Зацман И. М. Основные принципы разработки системы мониторинга функционирования исследовательских организаций. — М.: ЦЭМИ РАН, 2007. 62 с.
7. Зацман И. М., Вережкин Г. Ф., Шубников С. К. Моделирование систем мониторинга. — М.: ИПИ РАН, 2008. 115 с.
8. Шубников С. К. Формы документов в системах информационного обеспечения оценки результативности научной деятельности // Системы и средства информатики. — М.: Наука, 2005. Вып. 15. С. 59–76.
9. Зацман И. М. Информационные ресурсы для систем мониторинга в сфере науки // Системы и средства информатики. — М.: Наука, 2005. Вып. 15. С. 288–318.
10. Зацман И. М. Терминологический анализ нормативно-правового обеспечения создания систем мониторинга и оценки результативности в сфере науки // Экономическая наука современной России, 2005. № 4. С. 114–129.
11. Зацман И. М., Вережкин Г. Ф. Информационный мониторинг сферы науки в задачах программно-целевого управления // Системы и средства информатики. — М.: Наука, 2006. Вып. 16. С. 164–189.
12. Шубников С. К., Ложилова Е. Ю., Косарик В. В. Принципы систематизации и стандартизации описания структур информационных ресурсов в сфере науки // Системы и средства информатики. — М.: Наука, 2006. Вып. 16. С. 190–213.
13. Зацман И. М. Полидоменные модели в системах оценки инновационного потенциала и результативности научных исследований // Компьютерная лингвистика и интеллектуальные технологии: Труды Междунар. конф. Диалог-2006. — М.: РГГУ, 2006. С. 178–183.
14. Зацман И. М. Полидоменные модели электронных библиотек систем мониторинга сферы науки // Электронные библиотеки: перспективные методы и технологии, электронные коллекции — RCDL'2006: Труды VIII Всеросс. науч. конф. — Ярославль: ЯрГУ, 2006. С. 75–81.
15. Зацман И. М., Вережкин Г. Ф., Дрынова И. В., Курчавова О. А., Ларин Н. В., Норежян Т. П. Моделирование систем информационного мониторинга как проблема информатики // Системы и средства информатики. Спец. вып. Научно-методологические проблемы информатики. — М.: ИПИ РАН, 2006. С. 112–139.
16. Зацман И. М., Кожунова О. С. Семантический словарь системы информационного мониторинга в сфере науки: задачи и функции // Системы и средства информатики. — М.: Наука, 2007. Вып. 17. С. 124–141.
17. Zatsman I., Kozhunova O. Evaluating for institutional academic activities: Classification scheme for R&D indicators // 10th Conference (International) on Science and Technology Indicators (STI'2008): Book of Abstracts. — Vienna: ARC GmbH, 2008. P. 428–431.
18. Кожунова О. С. Семантический словарь системы информационного мониторинга в сфере науки и ресурс Eurowordnet: структура, задачи и функции // Системы и средства информатики. — М.: Наука, 2008. Вып. 18. С. 156–170.
19. Архипова М. Ю., Зацман И. М. Основные тенденции патентной активности в сфере информационных и телекоммуникационных технологий // Институциональные основы инновационных процессов: Мат-лы 4-х Друкеровских чтений. — М.: Доброе слово, 2008. С. 201–206.
20. Архипова М. Ю., Зацман И. М., Хавансков В. А. Индикаторы патентной активности РАН // Институциональные концепции менеджмента: Мат-лы 6-х Друкеровских чтений. — Екатеринбург: УрГУ, 2009. Т. 1. С. 141–150.
21. Zatsman I., Kozhunova O. Evaluation system for the Russian Academy of Sciences: Objectives—Resources—Results approach and R&D indicators // 2009 Atlanta Conference on Science and Innovation Policy Proceedings / Eds. S. E. Cozzens and P. Catalan. <http://smartech.gatech.edu/bitstream/1853/32300/1/104-674-1-PB.pdf>.
22. Зацман И. М. Категоризация результатов и индикаторов программ научных исследований в информационных системах мониторинга // Системы и средства информатики. Доп. вып. — М.: ИПИ РАН, 2009. С. 200–219.
23. Zatsman I., Durnovo A. Incompleteness problem of indicators system of research programme // 11th Conference (International) on Science and Technology Indicators (STI'2010): Book of Abstracts. — Leiden: CWTS, 2010. P. 309–311.
24. Методические рекомендации по подготовке Докладов о результатах и основных направлениях деятельности субъектов бюджетного планирования на 2006–2008 годы. http://www.minfin.ru/common/img/uploaded/library/2005/07/metod_270705.doc.
25. Planning for performance and evaluating results of public R&D programs // Meeting the OMB PART Challenge: Workshop Report. — Washington: The Washington Research Evaluation Network, 2004.
26. Special Report No.9/2007 concerning “Evaluating the EU Research and Technological Development (RTD) framework programmes — could the Commission’s approach be improved?” // Official J. Eur. Union C26, 30.01.2008. P. 1–38.
27. Zatsman I., Durnovo A. Program-oriented indicators: Production and application in science // Системы и средства информатики, 2012. Т. 22. № 1. С. 110–120.
28. Zatsman I., Durnovo A. Proactive dictionary of evaluation system as a tool for science and technology indicator development // 17th Conference (International) on Science and Technology Indicators Proceedings. — Montréal: Science-Metrix and OST, 2012. Vol. 2. P. 905–906.

29. *Зацман И. М., Косарик В. В., Курчавова О. А.* Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 54–69.
30. *Зацман И. М.* Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // Информатика и её применения, 2009. Т. 3. Вып. 2. С. 65–81.
31. *Зацман И. М.* Нестационарная семиотическая модель компьютерного кодирования концептов, информационных объектов и денотатов // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 87–101.
32. *Зацман И. М., Дурново А. А.* Моделирование процессов формирования экспертных знаний для мониторинга программно-целевой деятельности // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 84–98.
33. *Zatsman I.* Tracing emerging meanings by computer: Semiotic framework // 13th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Limited, 2012. Vol. 2. P. 1298–1307.
34. *Zatsman I.* Denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation // 11th IEEE Conference (International) on Cognitive Informatics & Cognitive Computing Proceedings. — Los Alamitos, CA: IEEE Computer Society Press, 2012. P. 27–34.
35. *Schmoch U.* Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. P. 193–211.
36. *Computational science: Ensuring America's competitiveness. Report to the President.* — Arlington, VA: National Coordination Office for Information Technology Research and Development, 2005. 104 p.
37. *Tijssen R. J. W., Buter R. K., Van Leeuwen Th. N.* Technological relevance of science: An assessment of citation linkages between patents and research papers // *Scientometrics*, 2000. Vol. 47. No. 2. P. 389–412.
38. *Van Looy B., Zimmermann E., Veugelers R., Verbeek A., Mello J., Debackere K.* Do science-technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains // *Scientometrics*, 2003. Vol. 57. No. 3. P. 355–367.
39. *Narin F., Noma E.* Is technology becoming science? // *Scientometrics*, 1985. Vol. 7. No. 3–6. P. 369–381.
40. *Narin F., Olivastro D.* Linkage between patents and papers: An interim EPO/US comparison // *Scientometrics*, 1998. Vol. 41. No. 1–2. P. 51–59.
41. *Mansfield E.* Academic research and innovation // *Research Policy*, 1991. Vol. 20. Is. 1. P. 1–12.
42. *Mansfield E.* Academic research underlying industrial innovations: Sources, characteristics and financing // *Review of Economic and Statistics*, 1995. Vol. 77. No. 1. P. 55–62.
43. *Mansfield E.* Academic research and industrial innovation: An update of empirical findings // *Research Policy*, 1998. Vol. 26. Is. 7–8. P. 773–776.
44. *Third European Report on Science & Technology Indicators.* — Luxembourg: Office for Official Publications of the European Communities, 2003. 451 p.
45. *Nonaka I., Takeuchi H.* The knowledge-creating company. — N.Y.: Oxford University Press, 1995 (пер. на русский язык: *Нонака И., Такеучи Х.* Компания — создатель знания. — М.: Олимп-бизнес, 2003).
46. *Зацман И. М., Кожунова О. С.* Предпосылки и факторы конвергенции информационной и компьютерной наук // Информатика и её применения, 2008. Т. 2. Вып. 1. С. 77–98.
47. *Verbeek A., Debackere K., Luwel M., Andries P., Zimmermann E., Deleus D.* Linking science to technology: Using bibliographic references in patents to build linkage schemes // *Scientometrics*, 2002. Vol. 54. No. 3. P. 399–420.

СТРАТЕГИИ ВЫРАВНИВАНИЯ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ: СЕМАНТИЧЕСКИЕ АСПЕКТЫ*

Е. Б. Козеренко¹

Аннотация: Данная статья посвящена проблемам проектирования и разработки лингвистически мотивированных механизмов выравнивания параллельных текстов и выявления грамматических (функционально-семантических) соответствий для формирования статистических портретов языковых употреблений, которые в дальнейшем будут встроены в гибридные модели машинного перевода. Гибридными называются такие модели, в которых для обработки естественного языка применяются как статистические механизмы, так и механизмы, основанные на правилах. Представленный в данной работе подход заключается в использовании исходной расширяемой грамматики, которая в процессе развития дополняется соответствиями, извлеченными из параллельных текстов. В качестве исходной грамматики используется *когнитивная трансферная грамматика* (КТГ), основанная на *трансферах* (двухязычных фразовых структурах), в которой представлены когнитивные и функциональные характеристики фразовых структур.

Ключевые слова: выравнивание; параллельные тексты; синтаксис; семантика; фразовые структуры; гибридные модели; машинный перевод

1 Введение

В данной работе проблема выравнивания параллельных текстов рассматривается в связи с задачей создания лингвистически достоверных механизмов для построения модулей обучения в системах машинного перевода и извлечения знаний из текстов. Современный этап исследований в области обработки естественного языка характеризуется стремительной «гибридизацией» подходов и моделей. Включение статистических характеристик в системы, основанные на правилах, позволяет отразить динамику и разнообразие языковых форм и значений, порождаемых в процессе речевой деятельности.

Основная цель исследований, проводимых автором данной работы, — извлечение из параллельных текстов на разных языках таких фразовых структур, которые выражают одинаковые значения, и включение их в систему правил расширяемой грамматики для решения задач машинного перевода и извлечения знаний из многоязычных предметно-ориентированных текстов. Расширяемая грамматика, которая при этом используется, содержит когнитивные и функциональные характеристики фразовых структур и базируется на формализме КТГ [1]. Основным конструктивным элементом КТГ являются *трансферы* (двухязычные фразовые структуры), которые задаются лексиконом нетерминальных символов. Этот лексикон динамиче-

ски пополняется новыми структурами в процессе выравнивания параллельных текстов. Формализм КТГ отражает парадигматические свойства рассматриваемых языков, т. е. язык как систему, тогда как выравниваемые параллельные тексты являются реализацией речи, т. е. языка как деятельности. Ключевыми аспектами этой деятельности являются динамика и синтагматика. Основной единицей анализа синтагматических свойств параллельных текстов является предложение.

Выравнивание по предложениям и словам проводится с использованием статистических методов. На первом этапе получаем параллельные корпуса, выравненные по предложениям, на следующем этапе производится пословное выравнивание. Далее проводится выравнивание по *фразам*.

В статистической исследовательской парадигме термин *фраза* означает произвольный сегмент предложения, выделенный статистическим инструментом и вовсе не являющийся нетерминальным символом (синтаксической единицей) какой-либо формальной грамматики. В отличие от такого подхода, в представленных исследованиях *фраза* — это синтаксически значимая единица в составе предложения, которая рассматривается в парадигматическом и синтагматическом аспектах. При этом сочетаются подходы КТГ и категориальной грамматики SUG (*Semiotic Universal Grammar*) [2, 3].

* Работа выполнена при частичной поддержке РФФИ (проект 11-06-00476-а).

¹ Институт проблем информатики Российской академии наук, kozerenko@mail.ru

Основные проблемы выравнивания параллельных текстов и, соответственно, обучения статистических процессоров естественного языка обусловлены наличием значительных трансформаций предложений исходного и целевого текстов, поскольку каждый язык использует свои специфические механизмы описания референтной ситуации. Переводческие трансформации исследуются в теории и практике перевода, структурной и прикладной лингвистике. Усилия исследователей направлены на выявление наиболее типичных (частотных) и значимых трансформаций в изучаемых параллельных текстах и развитии типологии трансформаций, представленных в исходном формализме КТГ. Таким образом, необходимо разработать и применить такие стратегии и инструменты, которые обеспечивают наиболее адекватные средства для описания и выявления сопоставимых языковых структур.

Важным проектным решением, которое вытекает из опыта исследований и разработок, является различие между задачами машинного перевода методом трансфера, исполняемого на уровне *трансфем* (синтаксических структур с функционально-семантической мотивацией) [1], и задачей извлечения знаний из текстов на разных языках. Для извлечения знаний создается метод многоязычной обработки, который основан на принципе *интерлингвы* (языка-посредника). В рассматриваемых проектах в роли интерлингвы используется язык расширенных семантических сетей (РСС) [4]. Соответственно, семантически-ориентированное выравнивание параллельных текстов проводится в двух режимах: (1) сопоставления на уровне *трансфем* и (2) сопоставления на уровне *концептов* и *отношений*. Обе стратегии рассматриваются в разд. 2 и 3. В разд. 4 обсуждаются современные подходы к проблеме выравнивания и анализа параллельных текстов. Функционально-семантические методы выравнивания с примерами и статистическими данными приводятся в разд. 5. В заключительном разделе представлены направления дальнейших исследований.

2 Выравнивание на основе трансфем

Трансфема — это единица когнитивного переноса, устанавливающая функционально-семантическое соответствие между структурами исходного языка L_s и структурами целевого языка L_T . Для выравнивания параллельных текстов трансфемы задаются как правила переписывания, в которых

в левой части стоит нетерминальный символ, а в правой — выравненные пары цепочек терминальных и нетерминальных символов, принадлежащих исходному и целевому языкам (например, русскому и английскому):

$$T \rightarrow \langle \rho, \alpha \sim \rangle,$$

где T — нетерминальный символ, ρ и α — цепочки терминальных и нетерминальных символов, принадлежащих русскому и английскому языкам, \sim — символ соответствия между нетерминальными символами, входящими в ρ , и нетерминальными символами, входящими в α . При выравнивании параллельных текстов на основе когнитивной трансферной грамматики процесс деривации начинается с пары связанных исходных символов S_ρ и S_α , далее на каждом шаге связанные нетерминальные символы попарно переписываются с использованием двух компонентов единого правила.

Для автоматического извлечения правил из параллельных текстов на основе когнитивной трансферной грамматики необходимо предварительно выравнивать тексты по предложениям и словам. Извлекаемые правила опираются на пословные выравнивания таким образом, что вначале идентифицируются исходные фразовые пары с использованием такого же критерия, как и большинство статистических моделей перевода, основанных на фразовом подходе [5], а именно: должно быть хотя бы одно слово внутри фразы на одном языке, выравненное с некоторым словом внутри фразы на другом языке, но никакое слово внутри фразы на одном языке не может быть выравнено ни с каким словом за пределами парной ей фразы на другом языке.

При выравнивании параллельных текстов ставятся следующие цели:

- извлечь фразовые структуры, выражающие одинаковые значения в разных языках — *трансфемы*;
- усилить и расширить исходную грамматику;
- разработать типологию *соответствий* M (от английского слова *match* — *пара, соответствие*);
- исследовать контекстные характеристики *трансфем* T и межязыковых *соответствий* M .

Первоочередной задачей при выявлении переводческих трансформаций в параллельных текстах является выявление наиболее предпочтительных для каждого языка (русского, английского, французского) способов конфигурирования фраз (последовательностей категориальных вершин). Эти предпочтения задаются в виде частотных характеристик для каждого типа соответствий, которые

затем в виде гибридных правил включаются в расширяемую грамматику. Особое внимание уделяется ситуациям категориального сдвига (конверсиям) при переводе фраз с одного языка на другой. Категория фразовой структуры определяется категорией ее головной вершины. Таким образом, когда в процессе трансфера происходит конверсия и категория головной вершины изменяется, новое категориальное значение присваивается всей фразовой структуре. Так, имя существительное, модифицирующее другое существительное в английском языке, становится в целом ряде случаев прилагательным при переводе на русский язык (например, *stone wall* — *каменная стена*); неличная форма глагола, играющая роль глагольного модификатора, становится личной глагольной формой в придаточном обстоятельстве предложении и т. д.

В ходе исследований было выявлено 34 типа основных трансформаций, наиболее статистически значимыми являются глагольно-именные трансформации, пассивизация (при этом в русском, английском и французском языках существует несколько способов оформления пассива), топикализация, падежный сдвиг (например, *именительный* — *дательный* и т. п.), преобразования личных форм глагола в неличные и наоборот. Были исследованы синонимичные способы выражения одинаковых функционально-семантических значений — это явление в литературе часто обозначают термином *перифраза* (*paraphrasing*) — как в пределах одного языка, так и в языковых парах.

Динамика выравнивания реализуется в *соответствиях* M , которые фиксируются посредством механизма категориальной грамматики SUG [2, 3].

В целом процесс выравнивания организуется в два этапа: предварительный и семантически-ориентированный. На этапе предварительного выравнивания выполняются следующие процедуры с использованием специального набора инструментальных систем: выравнивание по предложениям — ABVYU aligner [6]; выравнивание по словам — Verbalizator (tokenizer), Berkeley aligner [7]; выравнивание поддеревьев — Cognitive Dwarf [8]; выравнивание по структурам и формирование конкордансов: Sketch Engine [9].

Семантически-ориентированный этап выравнивания параллельных текстов проводится в двух режимах: (1) сопоставления на уровне *трансфем*, при этом выявляются языковые структуры, выражающие сходные функциональные значения в параллельных текстах, и (2) сопоставления на уровне *концептов и отношений*.

Режим выравнивания первого типа базируется на понятиях *трансфемы* T (*transfeme*) и *соответствия* M (*match*).

Определение. Трансфема T — это единица парадигматического плана, относящаяся к *языку* как системе, соответствие M — единица синтагматического плана, относящаяся к *речи* (*дискурсу*); таким образом, трансфемы T реализуются в соответствиях M .

Соответствие M может быть шире, чем трансфема T , и часто включает контекст.

Функционально-семантическое выравнивание на основе трансфем включает:

- выявление фразовых структур, управляемых головными вершинами;
- выравнивание головных вершин с учетом регулярных соответствий и трансформаций.

Функционально-семантическое выравнивание осуществляется на основе формализма КТГ [1]:

$$G_{CT} = \{T_{L_1}, T_{L_2}, N_{L_1}, N_{L_2}, P_{CA}, P_{CT}, S_{L_1}, S_{L_2}, M, D\},$$

где T_{L_1} и T_{L_2} — множества терминальных символов языков L_1 и L_2 ; N_{L_1} и N_{L_2} — множества нетерминальных символов языков L_1 и L_2 ; P_{CA} и P_{CT} — правила анализа и синтеза на основе когнитивного трансфера; S_{L_1} и S_{L_2} — пара исходных символов языков L_1 и L_2 , с которых начинается процесс анализа и выравнивания предложений; M — функция установления соответствия между языковыми структурами L_1 и L_2 ; D — функция, приписывающая значения вероятности каждому правилу из множеств P_{CA} , P_{CT} . Ядром формализма КТГ является лексикон трансфем, который относится к множеству нетерминальных символов.

3 Выравнивание на основе концептов и отношений

Выравнивание на основе концептов (сущностей) и отношений (связей) параллельных и концептуально-сопоставимых текстов на различных языках направлено на выявление языковых реализаций структур знаний и формирования многоязычных баз знаний, которые затем будут применяться в интеллектуальных аналитических системах. Такой режим выравнивания будем называть концептуально-ориентированным, инструментом для него служит лингвистический процессор Semantix. При этом выравнивание производится по языковым объектам, выражающим сущности (концепты), действия (предикаты) и связи. Для выравнивания по сущностям сопоставляются фразы с именными головными вершинами, для выравнивания по действиям — глагольные вершины (при

этом рассматриваются как личные, так и неличные формы глаголов), для выравнивания по связям сопоставляются генитивные предложные и беспредложные фразы.

Механизмы концептуально-ориентированного выравнивания основаны на аппарате РСС [4], обладающих достаточной выразительной мощностью для представления естественно-языковых структур с высокой степенью вложенности и выполняющих роль *интерлингвы* (языка-посредника). Базовым структурным элементом РСС является именованный N -местный предикат, называемый *фрагментом*. В основе РСС лежит множество вершин (V), из которых состоят элементарные фрагменты: $V_0(V_1, V_2, \dots, V_k/V_{k+1})$, где $V_0, V_1, V_2, \dots, V_k, V_{k+1}, V, k > 0$. Данный фрагмент представляет k -арное отношение. Фрагментам приписываются роли, вершина V_0 соответствует имени отношения, вершины V_1, V_2, \dots, V_k соответствуют объектам, которые объединены в отношение, а вершина V_{k+1} , отделенная косой чертой (/) от всей структуры в целом, соответствует вершине связи. V_{k+1} называется C -вершиной, и все эти элементы образуют РСС [4]

Все множество языковых объектов задается в виде предикатно-аргументных структур. Анализ предложения производится на основе унификационной грамматики. Слова и конструкции, играющие роль предикатов, в предложении служат «опорными» элементами, и результат анализа предложения образует один «расширенный» предикат, соответствующий главному предикату данного предложения. Модели управления и трансформационные свойства задаются в словаре в рамках словарных статей глаголов. В результате трансформаций происходит сдвиг моделей управления.

Особое внимание в исследованиях уделяется номинализации и изменению предложного управления на беспредложное, например: *стрелять по уткам — стрелять уток*.

Таким образом, концептуально-ориентированное выравнивание — это процесс извлечения знаний в многоязычном режиме и наполнения базы знаний для создания предметно-ориентированных интеллектуальных экспертных систем.

4 Современные подходы к выравниванию параллельных текстов: статистические и лингвистические аспекты

Статистические подходы к выравниванию параллельных текстов имеют целью установление

наиболее вероятного выравнивания A для двух заданных параллельных текстов S и T :

$$\arg \max_A P(A|S, T) = \arg \max_A P(A, S, T).$$

В статистическом машинном переводе делается попытка построить модель вероятности перевода $P(r_1^J|e_1^I)$, которая описывает соотношения между некоторой цепочкой r_1^J на исходном языке и цепочкой e_1^I на целевом языке. В статистической модели выравнивания текстов $P(r_1^J, a_1^J|e_1^I)$ вводится «скрытое» выравнивание a_1^J , задающее отображение из исходной позиции j в целевую позицию a_j . Соответствие между моделью перевода и моделью выравнивания задается следующим образом:

$$P(r_1^J|e_1^I) = \sum_{a_1^J} P(r_1^J, a_1^J|e_1^I).$$

Выравнивание a_1^J может содержать выравнивания $a_j = 0$ с пустым словом e_0 для тех слов исходного языка, которые не нашли ни одного соответствия среди слов целевого языка. В целом статистическая модель зависит от множества неизвестных параметров θ , которые извлекаются из обучающей выборки. Следующее выражение представляет зависимость модели от набора параметров:

$$P(r_1^J, a_1^J|e_1^I) = p_\theta(r_1^J, a_1^J|e_1^I).$$

Для выявления неизвестных параметров θ задается обучающий корпус параллельных текстов, содержащий S пар предложений $\{(r_s, e_s) : s = 1, \dots, S\}$. Для каждой пары (r_s, e_s) переменная выравнивания обозначается как $a = a_1^J$. Неизвестные параметры устанавливаются посредством максимизации сходства параллельных текстов в корпусе:

$$\hat{\theta} = \arg \max_{\theta} \prod_{s=1}^S \sum_a p_\theta(r_s, a|e_s).$$

Как правило, максимизация для таких моделей выполняется на основе алгоритма максимизации ожидания [19] или ему подобных. Такой алгоритм полезен для решения проблемы оценки параметров, но не является необходимым для статистического подхода. Следовательно, несмотря на то что существует большое число выравниваний для данной пары предложений, всегда можно найти наилучшее выравнивание

$$\hat{a}_1^J = \arg \max_{a_1^J} p_{\hat{\theta}}(r_1^J, a_1^J|e_1^I).$$

Выравнивание \hat{a}_1^J также называется выравниванием Витерби для пары предложений (r_1^J, e_1^I) .

Оценка качества выравнивания Витерби по сравнению с некоторым эталонным выравниванием осуществляется вручную. Параметры моделей статистического выравнивания оптимизируются с учетом критерия максимального сходства, который далеко не всегда отражает качество выравнивания.

В этой связи чрезвычайную важность приобретает проблема перифразирования, и ей уделяется все больше внимания в работах ведущих исследовательских групп в области компьютерной лингвистики [11–21]. Модель перифразы, описанная в работе [19], обучалась на основе корпуса European1. Авторы использовали десять параллельных корпусов между английским и каждым из следующих языков: датским, голландским, финским, французским, немецким, греческим, итальянским, португальским, испанским и шведским, приблизительно с 30 млн слов для каждого из этих языков и 315 млн английских слов. Автоматические словные выравнивания были получены с помощью Giza++ [20]. Английская часть каждого параллельного корпуса анализировалась с применением парсера Vikel [11]. Были разобраны в общей сложности 1,6 млн уникальных предложений. Языковая модель на основе триграмм была обучена на этих английских предложениях с использованием инструментария языкового моделирования SRI [12]. Этот инструментарий поддерживает создание и оценку разнообразных типов языковых моделей на основе статистики N -грамм, а также ряд сопутствующих задач, таких как статистическая разметка (тегирование) и манипуляции списками N -лучших и решетками слов.

Извлечение перифраз из двуязычных параллельных корпусов было описано в работах Bannard и Callison-Burch [17], которые выводили перифразы с использованием методов статистического машинного перевода на основе фраз [14]. Затем путем введения комплексных синтаксических меток вместо использования только нетерминальных символов из деревьев разбора авторы смогли добиться существенного улучшения по сравнению с основным методом. Синтаксические ограничения значительно улучшают качество этого метода перифразы. В работе [18] представлен новый подход к перифразе на основе двуязычного корпуса. Автор демонстрирует, что более высокое качество может быть достигнуто, если ввести такое ограничение: перифраза должна иметь тот же синтаксический тип, что и исходная фраза. В работе предложены ограничения на перифразы на двух этапах: когда они выводятся на основе разобранных параллельных корпусов и когда они подставляются в разобранные тестовые предложения. Автор ввел синтаксические ограничения, пометив все фразы и перифразы (даже

не входящие в число составляющих) с помощью слеш-категорий, используемых в категориальной грамматике CCG (комбинаторной категориальной грамматике). Однако автор не дает ни формального определения некоторой синхронной грамматики, ни предлагает систему декодирования, поскольку его система представлений не содержит иерархических правил (деревьев разбора). Правила синхронной грамматики для перевода извлекаются из пар предложений, автоматически разобранных и выровненных по словам. Методы извлечения варьируются в зависимости от того, извлекают ли они только минимальные правила для фраз, у которых есть доминирующие узлы в дереве разбора, или более сложные правила, включающие фразы, не входящие в число составляющих [21].

Главной мотивацией для использования синтаксических перифраз наряду с их исходными фразовыми соответствиями является их потенциальная возможность в более общем виде отразить лингвистические трансформации, сохраняющие значение. Система, обладающая механизмом синтаксических перифраз, должна быть способна обучаться хорошо оформленным обобщенным фразовым структурам, которые можно будет применять для анализа невидимых данных.

5 Типология наиболее частотных языковых трансформаций в параллельных текстах

Для создания *типологии трансформаций* были рассмотрены следующие приемы синхронного перевода:

- полный перевод лексико-грамматических форм (ЛГФ), когда имеет место полное соответствие между структурами исходного и целевого языков по форме, функции и значению;
- нулевой перевод (ЛГФ используются по-разному);
- частичный перевод (совпадают несколько содержательных функций ЛГФ);
- функциональная замена (имеет место функциональная идентичность ЛГФ);
- конверсия (замена одной грамматической категории на другую).

Фокусная выборка трансформаций формируется на основе их частотности в параллельных текстах:

- *номинализация* (35% в англо-русской языковой паре);

- пассивизация (18%–24% в русско-английском направлении перевода);
- адъективно-адвербиальные трансформации (12%–14% в обоих направлениях англо-русской языковой пары);
- субъектно-объектные трансформации (28% в обоих направлениях англо-русской языковой пары).

Рассмотрим некоторые примеры:

Косвенный объект → Субъект

Серьезными разногласиями была отмечена встреча сторон. — *Serious disagreements arose during the meeting of the sides.*

Прямой объект → Субъект

Новую планету назвали в честь ее открывателя. — *The new planet was named after its discoverer.*

Предложная фраза → Субъект

На встрече договорились. — *The meeting reached the conclusion.*

Функционально-семантическая мотивация фразовых структур представлена множеством нетерминальных символов с их частотными характеристиками, например:

(“Объектный инфинитив”, 400,0)

(“Объектный *be*-инфинитив”, 600,0)

(“Субъектный инфинитив с причастием прошедшим”, 580,0)

(“Субъектный инфинитив с прилагательным”, 490,0)

(“Субъект с причастием”, 570,0)

Будем использовать типизацию соответствий и статистические портреты трансформаций. Так, русский язык примерно на 35% более номинативен, чем английский:

In vacuum molecules have large space in which to move (V);

В вакууме молекулы имеют большое пространство для движения (N);

in which to move (V) → *для движения* (N).

Наиболее продуктивные типы глагольно-номинативных трансформаций коррелируют со следующими функциональными значениями:

- обстоятельство цели и следствия, выраженное инфинитивом (58%);
- сложный предикат с инфинитивом (*be* + infinitive) (51%).

Эта статистика учитывается в многовариантных правилах когнитивного трансфера. Соответствия, или *события*, составляют множество хорошо оформленных нетерминалов, а динамически выявляемые контекстные структуры представляются в нотации категориальной грамматики (SUG) [2, 3].

Для разрешения неоднозначности языковых объектов и структур используются векторные пространства. Основным прием выравнивания параллельных тестов по трансфемам — установление соответствий между головными вершинами фразовых структур, в настоящее время осуществляется в полуавтоматическом режиме, автоматическое сопоставление находится в стадии разработки: выравнивание по головным элементам фраз и сравнение фразовых структур.

Трансфемы T реализуются в соответствиях M . Будем различать два типа соответствий: прямые соответствия без трансформаций M^d и соответствия, в которых имеют место трансформации M^t . Соответствия M^d демонстрируют прямое соответствие категориальных значений головных элементов сопоставляемых структур (S_s and S_o), например имя существительное в исходном тексте будет соответствовать имени существительному в целевом тексте, и фразовые структуры будут подобными: NP (именная фраза) в исходном и целевом текстах. M — это соответствия, содержащие трансформации, например имя существительное в исходном тексте будет соответствовать глагольному элементу в целевом тексте, и фразовые структуры будут различаться, например: NP → VP.

Стратегии выравнивания параллельных текстов и установления соответствий, представленные в данной работе, учитывают возможные трансформации $M_{s \rightarrow o}^t$, где M — это соответствие, t — трансформация, s — исходный (*source*) язык, o — целевой (*objective*) язык; $S_M^{CT} \sum M_{s \rightarrow o}^t$ образует пространство соответствий когнитивного трансфера (S_M^{CT}).

Таким образом, основная цель формирования пространства двуязычных соответствий — обеспечение обучающей выборки для машинного обучения систем автоматического перевода, основанных на символично-статистических (гибридных) подходах.

Рассмотрим примеры представления англо-русских выравниваний параллельных текстов.

The second step of the above described process is conducted in the presence of a metal catalyst.

NP (Subj) PP VP (Passive) PP (English).

Вторую (Adj, F, Sg, Acc) стадию (N, F, Sg, Acc) вышеописанного (N, F, Sg, Acc) способа (N, M, Sg, Gen) проводят (V, Pl, 3-d Person, Pres) в (P) присутствии (N, Neutr, Sg, Prep) металлического (Adj, M, Sg, Gen) катализатора (N, M, Sg, Gen).

NP (Accusative) NP (Gen) VP (Person Undefined — Plural) PP (Russian).

Формальное представление трансформации выглядит следующим образом:

NP (Subj) PP VP (Passive) PP →
 → NP (Accusative) NP (Gen) VP
 (Person Undefined — Plural) PP

Было выявлено 134 типа соответствий *M*, которые принадлежат пространствам когнитивного трансфера.

Методы машинного обучения [22, 23] обеспечивают системе возможность извлекать значения, которые ожидаются в рамках определенных контекстов, следовательно, вполне естественно использовать статистические данные, извлеченные из больших массивов текстовой информации и использовать их для вычисления вероятности реализации определенного значения в соответствующем контексте.

Успех вероятностной модели зависит, помимо прочего, от адекватного определения *события*. Обычный тип события в вероятностной обработке естественного языка — совместное появление одного или нескольких слов в определенном контексте. В данном случае *слово* означает не словоформу (т.е. цепочку символов в том виде, как она появилась в тексте), а его основу, или *лемму*. Контекст, в свою очередь, определяется как упорядоченное множество прилегающих слов с обеих сторон каждого словоупотребления, при этом ширина такого контекстного окна задается произвольным целым числом. В то время как основа лексической единицы (лексемы) может быть разумным (если не идеальным) способом представления значений, определение контекста, обычно употребляемое в статистической обработке естественного языка, — это явное упрощение, которое затемняет важные лингвистические отношения. Вполне разумно предположить, что вероятность слова, имеющего определенное значение, зависит от других слов, расположенных поблизости. Однако этот «мешок слов» неизбежно скрывает сложные аспекты, для которых требуются более точные процедуры выявления.

6 Заключение

В данной статье рассматривается языковая пара русский—английский, однако проведенные эксперименты показывают, что основные выводы справедливы для языков с близкой структурой, например белорусского и украинского. Процессы синтаксических трансформаций также очень сходны в ряде других европейских языков. Включение

статистических данных в системы, основанные на правилах, позволяет отразить динамику и разнообразие языковых форм и значений, порождаемых в процессе речевой деятельности. Функциональная и когнитивная мотивация правил исходной грамматики позволяет увеличить точность соответствий на 23%—42% (по оценке экспертов) в зависимости от типа сопоставляемых текстов. Дальнейшие исследования и разработки направлены на расширение числа сопоставляемых языковых пар и формирование лингвистического ресурса размеченных параллельных текстов по научной и экономической тематике для обучения систем машинного перевода и обработки знаний.

Литература

1. 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.
2. Shaumyan S. Categorical grammar and semiotic universal grammar // IC-AI'03: Conference (International) on Artificial Intelligence Proceedings. — Las Vegas, USA: CSREA Press, 2003. P. 623—629.
3. Kozerenko E. B., Shaumyan S. Discourse projections of semiotic universal grammar // Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA: CSREA Press, 2005. P. 3—9.
4. Kuznetsov I. P., Kozerenko E. B., Matskevich A. G. Intelligent extraction of knowledge structures from natural language texts // 2011 IEEE/WIC/ACM Joint Conferences (International) on Web Intelligence and Intelligent Agent Technology — Workshops WI-IAT 2011: Proceedings. P. 269—272.
5. Och F. J., Ney H. A systematic comparison of various statistical alignment models // Computational Linguistics, 2003. Vol. 29. No. 1. P. 19—51.
6. The web site for ABBYY Aligner. <http://aligner.abbyyonline.com/ru>.
7. The web site for Berkeley Aligner. <http://snap.cs.berkeley.edu>.
8. The description of Cognitive Dwarf. <http://www.isa.ru/proceedings/images/documents/2008-38/91-109.pdf>.
9. The web site for Sketch Engine. <http://www.sketchengine.co.uk>.
10. Dempster A. P., Laird N. M., Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm // J. R. Stat. Soc. Ser. B, 1977. Vol. 39. No. 1. P. 1—22.
11. Bikel D. Design of a multi-lingual, parallel processing statistical parsing engine // HLT-2002 Proceedings, 2002. P. 178—182.

12. *Stolcke A.* SRILM — an extensible language modeling toolkit // Conference (International) on Spoken Language Processing Proceedings. — Denver, Colorado, 2002. P. 901–904.
13. *Pang B., Knight K., Marcu D.* Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences // NAACL'03: Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Proceedings, 2003. Vol. 1. P. 102–109.
14. *Koehn P., Och F.J., Marcu D.* Statistical phrase-based translation // NAACL'03: Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology Proceedings, 2003. Vol. 1. P. 48–54.
15. *Galley M., Hopkins M., Knight K., Marcu D.* What's in a translation rule? // HLT/NAACL Proceedings, 2004. P. 273–280.
16. *Koehn P.* A parallel corpus for statistical machine translation // MT-Summit Proceedings. — Phuket, Thailand, 2005. P. 79–86.
17. *Bannard C., Callison-Burch C.* Paraphrasing with bilingual parallel corpora // ACL Proceedings, 2005. P. 597–604.
18. *Callison-Burch C.* Syntactic constraints on paraphrases extracted from parallel corpora // EMNLP-2008 Proceedings, 2008. P. 196–205.
19. *Koehn P.* Statistical machine translation. — Cambridge: University Press, 2009.
20. The web site for GIZA++. <http://www.statmt.org/moses/giza/GIZA++.html>.
21. *Ganitkevitch Ju., Callison-Burch C., Napoles C., Van Durme B.* Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation // Conference on Empirical Methods in Natural Language Processing Proceedings, 2011. P. 1168–1179.
22. *Malkov K.V. Tunitsky D.V.* On extreme principles of machine learning in anomaly and vulnerability assessment // MLMTA'06: Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA, 2006. P. 24–29.
23. *Bogatyrev K.* In defense of symbolic NLP // MLMTA'06: Conference (International) on Machine Learning, Models, Technologies and Applications Proceedings. — Las Vegas, USA, 2006. P. 63–68.

ПОСТРОЕНИЕ СЕМАНТИЧЕСКИХ ВЕКТОРНЫХ ПРОСТРАНСТВ РАЗЛИЧНЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ*

Ю. И. Морозова¹

Аннотация: Данная работа посвящена актуальным проблемам исследования семантики лингвистических единиц с использованием корпусных методов. В работе дается описание нового направления лингвистических исследований — дистрибутивной семантики. Предлагается расширение существующих моделей дистрибутивной семантики за счет перехода от описания лексем к описанию значимых словосочетаний. Описывается методика построения семантических векторных пространств (СВП) для различных предметных областей.

Ключевые слова: дистрибутивная семантика; векторные пространства; значимые словосочетания; колокации

1 Обзор моделей дистрибутивной семантики

Дистрибутивная семантика — область научных исследований, занимающаяся вычислением степени семантической близости между лингвистическими единицами на основании их дистрибутивных (контекстных) признаков в больших массивах лингвистических данных. Модели векторных пространств находят все более широкое применение в исследованиях, связанных с семантическими моделями естественного языка, и имеют разнообразный спектр потенциальных и действующих приложений. Основными сферами применения дистрибутивных моделей являются: разрешение лексической неоднозначности, информационный поиск, кластеризация документов, автоматическое формирование словарей (словарей семантических отношений, двуязычных словарей), создание семантических карт, моделирование перифраз, определение тематики документа, определение тональности высказывания, биоинформатика.

Теоретические основы данного направления восходят к дистрибутивной методологии З. Харриса [1, 2]. Близкие идеи выдвигали основоположники структурной лингвистики Ф. де Соссюр и Л. Витгенштейн. Дистрибутивная семантика основывается на дистрибутивной гипотезе о том, что лингвистические элементы со схожей дистрибуцией имеют близкие значения [3, 4].

В качестве вычислительного инструмента и способа представления моделей используется линейная алгебра. Информация о дистрибуции лингвистических единиц представляется в виде

многомерных векторов, а семантическая близость между лингвистическими единицами вычисляется как расстояние между векторами. Многомерные векторы образуют матрицу, где каждый вектор соответствует лингвистической единице (слово или словосочетание), а каждое измерение вектора соответствует контексту (документ, параграф, предложение, словосочетание, слово).

Для вычисления меры близости между векторами могут использоваться различные формулы: расстояние Минковского, расстояние Манхэттена, евклидово расстояние, расстояние Чебышёва, скалярное произведение, косинусная мера. Наиболее популярной является косинусная мера:

$$\frac{x \bullet y}{|x| \bullet |y|} = \frac{\sum_{i=1}^n x_i \bullet y_i}{\sqrt{\sum_{i=1}^n x_i^2} \bullet \sqrt{\sum_{i=1}^n y_i^2}}.$$

Существует множество разновидностей моделей дистрибутивной семантики, которые различаются по следующим параметрам:

- тип контекста (размер контекста, правый или левый контекст, ранжирование);
- количественная оценка частоты встречаемости слова в данном контексте (абсолютная частота, энтропия, совместная информация и пр.);
- метод вычисления расстояния между векторами (косинус, скалярное произведение, расстояние Минковского и пр.);

* Работа выполнена при частичной поддержке РФФИ (проект 11-06-00476-а).

¹ Институт проблем информатики Российской академии наук, yulia-ipi@yandex.ru

– метод уменьшения размерности матрицы (случайная проекция, сингулярное разложение и пр.).

Наиболее известными моделями дистрибутивной семантики являются латентный семантический анализ, разработанный для решения проблемы синонимии при информационном поиске [5], и модель языка как гиперпространства, разработанная как модель семантической памяти человека [6].

Концепция СВП впервые была реализована в информационно-поисковой системе SMART [7]. 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.

Впоследствии концепция СВП была успешно применена для других семантических задач. Например, в работе [8] контекстное векторное пространство было использовано для оценки семантической близости слов. Данная система достигла результата 92,5% на тесте по выбору наиболее подходящего синонима из стандартного теста английского языка TOEFL, в то время как средний результат при прохождении теста человеком 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 [9].

2 Выделение значимых словосочетаний

Целью данной работы является применение модели СВП для построения концептуальных моделей различных предметных областей. Развитие существующих подходов к построению СВП заключается в использовании значимых словосочетаний (ЗС) вместо отдельных лексем. Под ЗС понимаются лексические последовательности, имеющие тенденцию к совместной встречаемости.

В лингвистике для обозначения ЗС используется также термин «коллокация». Этот термин был впервые введен в «Словаре лингвистических терминов» О. С. Ахмановой [10]. Исследованиям коллокаций русского языка посвящено большое количество литературы, например монография Е. Г. Бо-

рисовой [11]. В теоретической лингвистике под коллокациями понимают словосочетания из двух или более слов, которые обуславливают друг друга семантически и грамматически [12]. В корпусной лингвистике коллокациями называют статистически устойчивые словосочетания, причем они могут быть как фразеологизированными, так и свободными.

Для выделения значимых словосочетаний в компьютерной лингвистике используются различные статистические меры (меры ассоциации, меры ассоциативной связанности, *англ.* association measures), вычисляющие силу связи между элементами в составе коллокации. В литературе упоминается несколько десятков мер ассоциативной связанности. Чаще других используются MI, t-score и log-likelihood [13].

Мера MI (mutual information), введенная в работе [14], сравнивает зависимые контекстно-связанные частоты с независимыми частотами слов в тексте. Если значение MI превосходит определенное пороговое значение, то словосочетание считают статистически значимым. Мера MI вычисляется по следующей формуле:

$$MI = \log_2 \frac{f(n, c) \times N}{f(n) \times f(c)},$$

где n — первое слово словосочетания; c — второе слово словосочетания; $f(n, c)$ — частота совместной встречаемости двух слов; $f(n)$, $f(c)$ — абсолютные частоты встречаемости каждого слова по отдельности; N — общее число словоупотреблений в корпусе.

Мера t-score также используется при ответе на вопрос, насколько не случайным является сочетание двух или более слов в тексте. Для вычисления t-score используется следующая формула:

$$t\text{-score} = \frac{f(n, c) - f(n) \times f(c)/N}{\sqrt{f(n, c)}}.$$

Также достаточно часто применяется мера, известная под названием log-likelihood, или логарифмическая функция правдоподобия, введенная в работе [15]. Для вычисления log-likelihood применяется следующая формула:

$$\log\text{-likelihood} = 2 \sum f(n, c) \times \log_2 \frac{f(n, c) \times N}{f(n) \times f(c)}.$$

Применив различные меры ассоциативной связанности слов к материалам научных патентов, авторы составили частотный словарь значимых словосочетаний для предметной области научных патентов. Примеры выделенных значимых словосочетаний: *благородный металл, вспомогательное*

устройство, жесткий элемент, измерительная ячейка, опорный карниз, оптический луч, система охлаждения, тяжелая фракция.

3 Построение семантического векторного пространства

Модель семантического векторного пространства, которую планируется построить в ходе данного исследования, обладает следующими характеристиками:

- тип изучаемых единиц: значимые словосочетания;
- тип контекста: лексемы и словосочетания, размер контекста — предложение, ранжирование контекста — нет;
- количественная оценка частоты встречаемости изучаемой единицы в данном контексте: абсолютная частота;
- метод вычисления расстояния между векторами: косинусная мера.

Приведем 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. В таблице используются сокращенные обозначения: c_1 — искусственный интеллект; c_2 — компьютерная лингвистика; c_3 — дискретная математика; c_4 — конструктивная математика; c_5 — интеллект, интеллектуальный; c_6 — математика, математический; c_7 — изучать, изучение.

Применив формулу вычисления косинусной меры между контекстными векторами, получим

	c_1	c_2	c_3	c_4	...
c_1	0	1	0	0	...
c_2	1	0	0	0	...
c_3	0	0	0	0	...
c_4	0	0	0	0	...
...

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.

4 Заключение

В работе были рассмотрены основные направления и модели нового направления исследований в компьютерной лингвистике — дистрибутивной семантики. На основании автоматической обработки больших массивов лингвистических данных возможно создавать различные лингвистические ресурсы: семантические словари, многоязычные словари, семантические карты предметных областей. В качестве математической модели используются многомерные векторы и матрицы линейной алгебры, что представляет собой удобный формализм для компьютерной реализации. В рамках данного направления предлагается разработать методику построения семантических векторных пространств для различных предметных областей, где в качестве изучаемых единиц будут выступать значимые словосочетания, выделенные из текстов с использованием мер ассоциативной связанности слов.

Литература

1. Harris Z. S. Papers in structural and transformational linguistics. — Dordrecht: Reidel, 1954.
2. Harris Z. S. Mathematical structures of language. — New York: John Wiley & Sons, 1968.
3. Sahlgren M. The distributional hypothesis // From context to meaning: Distributional models of the lexicon in

- linguistics and cognitive science (Special issue of the Italian Journal of Linguistics). — Pisa: Pacini Editore, 2008. Vol. 20. No. 1. P. 33–53.
4. *Turney P. D., Pantel P.* From frequency to meaning: Vector space models of semantics // *J. Artificial Intelligence Research*. — Menlo Park, California: AAAI Press, 2010. No. 37. P. 141–188.
 5. *Landauer Th. K., McNamara D. S., Dennis S., Kintsch W.* Handbook of Latent Semantic Analysis. — Mahwah, NJ: Lawrence Erlbaum, 2007.
 6. *Lund K., Burgess C.* Producing high-dimensional semantic spaces from lexical co-occurrence // *Behavior Research Methods, Instruments & Computers*. — New York: Psychonomic Society, 1996. Vol. 28. No. 2. P. 203–208.
 7. *Salton G. M.* The SMART retrieval system: Experiments in automatic document processing. — Eaglewood Cliffs, NJ: Prentice-Hall, 1971.
 8. *Rapp R.* Word sense discovery based on sense descriptor dissimilarity // 9th MT Summit Proceedings. — New Orleans, LA, 2003. P. 315–322. <http://www.amtaweb.org/summit/MTSummit/FinalPapers/19-Rapp-final.pdf>.
 9. *Turney P.* A uniform approach to analogies, synonyms, antonyms and associations // 22nd Conference (International) on Computational Linguistics (COLING) Proceedings, 2008. P. 905–912. <http://www.aclweb.org/anthology-new/C/C08/C08-1114.pdf>.
 10. *Ахманова О. С.* Словарь лингвистических терминов. — М.: Советская энциклопедия, 1966.
 11. *Борисова Е. Г.* Коллокации. Что это такое и как их изучать. — 2-е изд., стер. — М.: Филология, 1995.
 12. *Иорданская Л. Н., Мельчук И. А.* Смысл и сочетаемость в словаре. — М.: Языки славянских культур, 2007.
 13. *Захаров В. П., Хохлова М. В.* Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: Труды Международ. конф. Диалог'2010. — М.: РГГУ, 2010.
 14. *Church K., Hanks P.* Word association norms, mutual information, and lexicography // *Computational Linguistics*, 1996. Vol. 16. No. 1. P. 22–29.
 15. *Dunning T.* Accurate methods for the statistics of surprise and coincidence // *Computational Linguistics*, 1993. Vol. 19. No. 1. P. 61–74.

ОЦЕНКА СЕМАНТИЧЕСКОЙ АДЕКВАТНОСТИ ТЕКСТОВ ИНФОРМАЦИОННЫМ МЕТОДОМ

Л. А. Кузнецов¹, В. Ф. Кузнецова²

Аннотация: Рассматривается проблема автоматизации проверки знаний обучаемых сравнением ответов учащихся с эталонными ответами, хранящимися в базе данных. Для оценки степени соответствия ответа эталону разрабатывается оригинальная методология, опирающаяся на теорию информации. Приведены некоторые результаты, иллюстрирующие применение методологии для автоматической оценки близости англоязычных текстов — изложений, написанных студентами.

Ключевые слова: семантическое подобие текстов; вероятностная модель текста; теория информации; энтропия; взаимная информация текстов; автоматизация оценки знаний

1 Введение

Актуальной задачей внедрения современных информационных технологий в сферу образования является разработка средств автоматизации оценки знаний обучаемых. Разработка инструментов автоматической оценки уровня знаний позволила бы решить многочисленные и хорошо известные проблемы аттестации обучаемых на всех этапах изменения их статуса.

Базой для разработки таких инструментов являются методы интеллектуальной обработки информации, содержащейся в информационных объектах, представленных на естественном языке. Эти методы позволяют автоматически оценивать семантическую близость информационных объектов. В настоящее время в информационно-поисковых системах при классификации текстов, при проверке текстов на плагиат [1, 2] применяются статистические подходы на основе векторно-пространственной модели текста, предложенной Г. Солтоном с соавторами в 1975 г. [3]. В ней текст представляется вектором частот входящих в него слов, а оценка близости текстов равна косинусу угла между векторами текстов.

Такой подход не позволяет использовать для оценки степени близости содержательные морфологические и синтаксические характеристики естественного языка, посредством которых и отражается семантика объектов. В нем сравнение базируется на количественной оценке частот слов в текстах и поэтому фактическая близость текстов устанавливается последующим субъективным анализом. На его основе не представляется возможным автоматизировать процесс оценки степени семантического

подобия информационных объектов (текстов), в частности он не может быть применен при оценке знаний.

В данной работе рассматривается задача оценки семантической близости текстов, которая при положительном решении могла бы использоваться в качестве основы систем автоматизированной проверки уровня знаний, усвоенных обучаемыми. Сейчас распространена система проверки знаний на основе тестов, что объясняется простотой ее автоматизации. Пятьдесят лет назад тестирующие установки такого типа изготовлялись на релейных схемах, а в настоящее время та же примитивная идеология перенесена на компьютеры. Очевидно, что традиционный письменный экзамен, при котором обучающийся дает полноценный ответ, обеспечивает более качественную проверку уровня знаний по сравнению с тестовой системой, но в настоящее время он не применяется, например при проведении ЕГЭ, из-за значительных затрат квалифицированных проверяющих для «ручной» проверки.

Автоматизация приема полноценных ответов обучаемых упирается в проблему разработки методологии автоматической оценки семантической близости информационных объектов, представленных на естественном языке. Настоящая статья посвящена оригинальной методологии [4] автоматической оценки семантической близости текстов, опирающейся на представление текста в виде вероятностно-статистической модели, в рамках которой на основе морфологической и синтаксической детализации текстов для отражения содержащейся в них информации вводится система семантических компонентов.

¹Липецкий государственный технический университет, kuznetsov.leonid48@gmail.com

²Липецкий государственный технический университет, kuznetsov@stu.lipetsk.ru

Ниже иллюстрируется методология формализации принципиальной базовой идеи сопоставления семантического содержания информационных объектов, представленных на естественном языке. Максимально простая трактовка ее содержательного существа состоит в следующем. При субъективной (интеллектуальной) оценке подобия содержания текстов не считаются слова и вообще не обращается внимание на конкретные слова, а интуитивно сопоставляется: о ком (чем) идет речь, что он делает (с ним делается), когда, где, как, при каких условиях и т. п. Ответ на любой из подобных вопросов, извлекаемый из текста, назовем для определенности семантическим компонентом. При субъективном анализе фактически сопоставляются семантические компоненты текста, которые представляются в нем случайными наборами слов. Предлагаемая методология автоматической оценки адекватности текстов базируется на формальном представлении такой же схемы.

Принципиальным является вероятностный подход к наполнению сопоставляемых семантических компонентов текста, отражающих: кто, что, где, когда, как и т. д., которые могут быть представлены случайными наборами конструкций языка. В большинстве структурированных языков могут быть введены формальные правила соотнесения семантических компонентов текста с морфологическими (существительными, глаголами, прилагательными и т. д.) и синтаксическими (подлежащими, сказуемыми, обстоятельствами и т. д.) компонентами языка. Правила морфологической и синтаксической структуризации текстов, т. е. отнесения слов к конкретным морфологическим и синтаксическим компонентам, представлены в грамматиках языков, которые во многих случаях уже имеют автоматизированные версии.

Предлагаемая методология базируется на формировании множества семантических компонентов на основании одного из текстов, принимаемого (условно) за эталонный. При формализации семантические компоненты классифицируются случайными событиями. Содержание семантических компонентов формируется из отдельных элементов исследуемого текста — слов или составных конструкций, которые трактуются элементарными случайными событиями (исходами). Система семантических компонентов (множество случайных событий) определяет семантическую структуру текста (алгебру вероятностной модели). Множество всех слов (исходов) текста, алгебра и вероятности случайных событий образуют вероятностную модель информационного объекта, представленного на естественном языке. Для конкретного текста модель отражает его вероятностно-статистиче-

ский образ. Другие сравниваемые тексты (копии) структурируются по введенной системе семантических компонентов (алгебре) и определяются их вероятностно-статистические образы. Количество информации в вероятностных образах объектов оценивается энтропией, а количество совместной (совпадающей) информации в образах двух объектов, оценивается взаимной информацией. Методология базируется на сопоставлении содержания семантических компонентов текстов, которые представляются случайными событиями, и поэтому степень их подобия (количество взаимной информации) является вероятностной мерой, изменяющейся от нуля при полном несовпадении текстов до количества информации в каждом из них при полном совпадении.

Представление сравниваемых текстов в виде вероятностно-статистических образов [4] позволяет по единой шкале оценить количество информации в текстах, количество общей информации в текстах и количество информации, отличающей тексты, т. е. содержащейся в одном и не содержащейся в другом. Количественная мера информации, содержащейся в объектах, и степени ее подобия вводятся с использованием аппарата теории информации.

При использовании этой методологии для оценки знаний абстрактное количество взаимной информации требуется градуировать в баллах по принятой N -балльной шкале оценки знаний. Техника градуировки кратко поясняется ниже на примере 100-балльной шкалы, используемой в ЛГТУ.

В примерах, иллюстрирующих адекватность меры количества взаимной информации для оценки уровня подобия объектов и некоторые возможности ее адаптации к особенностям конкретных информационных объектов, используется англоязычный материал. В качестве фактической эмпирической основы для исследования используются результаты изложения, написанного студентами на английском языке. Изложения были проверены и оценены преподавателем английского языка. Тексты изложений и эталон, прочитанный студентам преподавателем, были введены в систему автоматической оценки. Методика формально-математической оценки изложений и полученные результаты представлены в статье.

2 Формирование вероятностно-статистической модели текста

В теории вероятностей [5] вводится вероятностная модель, в рамках которой отражается вся име-

ющаяся информация об объекте и предоставляется возможность структуризации этой информации для отражения и исследования его содержательных особенностей, представляющих интерес. Наиболее полная информация о случайном объекте представляется распределением вероятностей, т. е. множеством элементарных исходов $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ случайной величины и их вероятностями $p(\omega_i)$. Содержательное существо объекта может быть введено в модель алгеброй, т. е. структуризацией множества реализаций Ω в виде системы случайных событий $A_j \in \Omega$, отражающих состояние или поведение вероятностного объекта в представляющем интерес смысле.

Вероятностная модель в общем случае представляется в виде:

$$M_\omega = \{\Omega, \aleph, P(A_j)\}, \quad (1)$$

где $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ — пространство элементарных исходов; $\aleph = \{A_1, A_2, \dots, A_j, \emptyset, \Omega\}$ — алгебра событий A_j , составленных с помощью операций логического сложения, умножения и отрицания из элементарных событий ω_i и дополненная невозможным \emptyset и достоверным Ω событиями; $P(A_j)$ — вероятности событий, составляющих алгебру, которые рассчитываются по вероятностям элементарных исходов $p(\omega_i)$, $i = 1, 2, \dots, n_\omega$, составляющих это событие:

$$P(A_j) = \sum_{\omega \in A_j} p(\omega_i). \quad (2)$$

Общая вероятностная модель случайного объекта трансформируется в вероятностно-статистическую модель информационного объекта приданием соответствующего содержания ее компонентам. Так, множеством элементарных исходов $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ текстового документа являются слова, составляющие текст, выбор которых для отражения некоторого содержания достаточно случаен. Важнейшим компонентом вероятностной модели является алгебра событий \aleph , которая в контексте статьи представляет систему семантических компонентов, по которым «раскладывается» содержание текста. Система семантических компонентов формируется с использованием частей речи, членов предложения, иных конструируемых из них структур, которые отражают семантическое содержание представляемого текста. В приводимых ниже примерах используется вероятностно-статистическая морфологическая модель (ВСММ), в которой семантические компоненты (случайные события) A_j , определяющие структуризацию текста, отождествляются с частями речи [4]. При этом артикли, частицы, междометия и другие служебные части

речи игнорировались и рассматривался следующий набор случайных событий:

$$\begin{aligned} \aleph = (A_1 = \text{Существительное}, A_2 = \text{Глагол}, \\ A_3 = \text{Прилагательное}, A_4 = \text{Наречие}, \\ A_5 = \text{Числительное}, \\ A_6 = \text{Неопределенное слово}). \quad (3) \end{aligned}$$

Алгебры образов эталона и ответа имеют одну и ту же систему событий, но их вероятности могут быть различными. Вероятности событий A_j , составляющих алгебру \aleph , вычисляются по вероятностям $p(\omega_i)$ отдельных слов, входящих в случайные события, в соответствии с (2). В результате получаются вероятностно-статистические морфологические образы эталона и ответов в виде модели (1):

$$\begin{aligned} M_\Theta = \{\Omega_\Theta, \aleph, P(A_j^\Theta)\}; \\ M_O = \{\Omega_O, \aleph, P(A_j^O)\}. \end{aligned} \quad (4)$$

Образы эталона M_Θ и ответа M_O содержат всю информацию о сравниваемых текстах как случайных объектах. В них сравниваемые тексты структурируются в виде систем семантических компонентов, отражающих то, о чем (или о ком) идет в них речь, что они делают (или с ними делается), как, каким образом, при каких условиях и т. д., и т. п. Система семантических компонентов для эталона и для ответа одна и та же, она может автоматически синтезироваться при анализе эталона. Вес, уровень значимости отдельных событий в контексте определяются их вероятностями (2).

Структурированное представление образов (4) вероятностной модели (1) может быть оформлено в виде таблицы, подобной табл. 1 для алгебры (3). В ней шапка отражает \aleph — систему семантических компонентов (3), общую для эталона и ответа. Содержание таблицы представляет наполнение семантических компонентов (в примере — множество слов Ω , составляющих конкретный информационный объект, ИО). Частоты отдельных слов $p(\omega_i)$ представлены в таблице, по ним определяются показанные в последней строке таблицы вероятности случайных событий $P(A_j)$. Таким образом, табл. 1 представляет в полном объеме вероятностно-статистический морфологический образ (ВСМО) модели (1) соответствующего ИО. В общем случае алгебра может быть расширена включением всех существенных для отражения содержания текста семантических компонентов. Процесс расширения может выполняться автоматически, но это вопрос отдельного рассмотрения.

При оценке близости текстов используется и совместная энтропия текстов, которая определяется

Таблица 1 Табличное представление вероятностной модели ИО с алгеброй (3)

Семантические компоненты								
Существительное		Глагол		Прилагательное		...	Неопределенное	
Слово	Частота	Слово	Частота	Слово	Частота	...	Слово	Частота
Advice	$p(\text{advice})$	Allow	$p(\text{allow})$	American	$p(\text{American})$...	Of	$p(\text{of})$
Bang	$p(\text{bang})$	Appear	$p(\text{appear})$	British	$p(\text{British})$...	For	$p(\text{for})$
...
$P(A_1)$		$P(A_2)$		$P(A_3)$...	$P(A_6)$	

по общей модели вероятностного объекта «эталон–ответ», имеющей вид:

$$M_{\Theta O} = \{ \Omega_{\Theta O}, \aleph, P(A_j) \}, \quad (5)$$

где $\Omega_{\Theta O} = \{ \Omega_{\Theta} + \Omega_O \}$ — множество элементарных событий: объединение всех слов, принадлежащих и эталону, и ответу; $\aleph_{\Theta O}$ — система событий, получающаяся разделением каждого события A_j из (3) на события:

$$A_j^{\Theta O} = \{ \omega_{ik} \in \Omega_O + \Omega_{\Theta} | \omega_{ik} \in A_j^{\Theta} \cap \omega_{ik} \in A_j^O \}; \quad (6a)$$

$$A_{\Theta j}^{\Theta} = \{ \omega_{ik} \in \Omega_O + \Omega_{\Theta} | \omega_{ik} \in A_j^{\Theta} \cap \omega_{ik} \notin A_j^O \}; \quad (6б)$$

$$A_{\Theta j}^O = \{ \omega_{ik} \in \Omega_O + \Omega_{\Theta} | \omega_{ik} \in A_j^{\Theta} \cap \omega_{ik} \in A_j^O \}, \quad (6в)$$

где верхним индексом отмечен образ, в который входит рассматриваемое слово, а нижним — образ, в который слово не входит.

Событие (6a) составляется из слов, входящих в событие A_j эталона и ответа, (6б) объединяет слова, входящие в событие A_j эталона, но отсутствующие в ответе, (6в) объединяет слова, не входящие в событие A_j в эталоне, но имеющиеся в ответе. Других вариантов для слов, присутствующих в двух сопоставляемых текстах, нет.

Таким образом, в модели (5) используется детализация текста, подобная (3), но с определением принадлежности слова к каждому тексту (эталону и ответу):

$$\aleph_{\Theta O} = \{ A_j = [A_j^{\Theta O}, A_{\Theta j}^{\Theta}, A_{\Theta j}^O], j = 1, 2, \dots, J \}.$$

Для получения количественных оценок уровня подобия ответа эталону, тексты формализуются в виде M_{Θ} , M_O и $M_{\Theta O}$. Вся информация, содержащаяся в образах вероятностных моделей текстов, представляется в виде таблиц формата, подобно табл. 1. Естественно, алгебра моделей определяется адекватно содержанию ИО. Формирование алгебр базируется на автоматическом анализе текста и установлении принадлежности каждого слова

к определенному событию алгебры (3). В настоящее время имеются коммерческие и бесплатные программные продукты для автоматического морфологического и синтаксического анализа текстов. Для построения ВСММ использовался¹ свободно распространяемый продукт российской компании Cognitive Technologies — система синтаксического анализа и машинного перевода Cognitive Dwarf [6], ориентированная на русский и английский языки. Система позволяет выполнять морфологический и синтаксический разбор текста со скоростью порядка нескольких килобайт в секунду.

3 Мера семантической близости информационных объектов

Оценка семантической близости информационных объектов — эталона и его искаженной копии, формально представленных моделями вида (4), может быть выполнена на основе представлений теории информации. В теории информации К. Шенноном [7] для оценки близости сообщения, полученного приемником, сообщению, посланному передатчиком, введена количественная мера информации и потерь информации при ее передаче. Задача оценки уровня знаний даже в содержательном смысле подобна передаче сообщений: эталон — это переданное обучаемому сообщению (знание), а его ответ может трактоваться как принятое сообщение (усвоенное знание). Поэтому естественно использовать для оценки уровня адекватности (семантического подобия) ответа эталону количество их совместной (взаимной) информации, которое в теории информации введено для оценки близости переданного и принятого сообщений.

Количество информации, содержащееся в случайном объекте M_{ω} (1), оценивается энтропией:

$$H_{\omega} = - \sum_{A_j \in \aleph} P(A_j) \ln P(A_j), \quad (7)$$

¹Некоторые из используемых ниже количественных данных взяты из магистерской диссертации А. С. Кондаурова, выполненной под руководством Л. А. Кузнецова.

где обозначения совпадают с использованными в (1).

Из (7) следует, что энтропия — мера количества информации случайной величины A — определяется распределением ее вероятностей. Вследствие того, что $P(A) \leq 1$ и $\ln P(A) \leq 0$, для получения положительной величины в (7) используется минус.

Образы эталона и ответа (4) позволяют оценить количество информации в каждом из них значениями энтропии, а именно:

$$\left. \begin{aligned} H_{\Xi} &= - \sum_{A_j^{\Xi} \in \mathbb{N}} P(A_j^{\Xi}) \ln P(A_j^{\Xi}); \\ H_O &= - \sum_{A_j^O \in \mathbb{N}} P(A_j^O) \ln P(A_j^O). \end{aligned} \right\} \quad (8)$$

Образ общего объекта «эталон–ответ» позволяет вычислить совместную энтропию двух текстов в виде:

$$\begin{aligned} H_{\Xi O} &= - \sum_{A_j \in \mathbb{N}} [P(A_{Oj}^{\Xi}) \ln P(A_{Oj}^{\Xi}) + \\ &+ P(A_{Oj}^O) \ln P(A_{Oj}^O) - P(A_j^{\Xi O}) \ln P(A_j^{\Xi O})]. \quad (9) \end{aligned}$$

На основании значений энтропии (8), (9) определяется совместная или взаимная информация:

$$I_{\Xi O} = H_{\Xi} + H_O - H_{\Xi O}. \quad (10)$$

В данном контексте совместная информация может трактоваться как количество информации из эталона, содержащееся в ответе. Вследствие симметрии (10) можно говорить и о количестве информации из ответа, содержащемся в эталоне.

Нетрудно проверить, что получаемое в соответствии с (10) количество совместной информации согласуется с интуитивными представлениями. Для этого можно рассмотреть два предельных варианта: (1) полное совпадение ответа и эталона и (2) полное их несовпадение.

Если эталон и ответ полностью совпадают, то в (8) $A_j^{\Xi} = A_j^O$ для всех j и, следовательно, $H_{\Xi} = H_O$. Далее, из (6) видно, что в этом случае события $A_j^{\Xi O}$ будут совпадать с событиями $A_j^{\Xi} = A_j^O$ и $P(A_j^{\Xi O}) = P(A_j^{\Xi}) + P(A_j^O)$ для всех $A_j^{\Xi O}$, следовательно, $H_{\Xi O} = 0$. Поэтому при полном совпадении эталона и ответа количество взаимной информации (10) будет $I_{\Xi O} = H_{\Xi} + H_O$, т.е. оно равно ее содержанию в эталоне и ответе, который совпадает с эталоном.

Другой крайний случай получается при полном несовпадении. При этом, очевидно, не будет слов, принадлежащих одновременно эталону и ответу. Поэтому все события (6а) будут иметь нулевую вероятность ($P(A_j^{\Xi O}) = 0$ для всех j). В теории информации принято считать $0 \ln 0 = 0$, так что

вычитаемое в правой части (9) будет равно нулю. В этом случае будет иметь место совпадение случайных событий $A_{\Xi j}^O = A_j^O$, $A_{Oj}^{\Xi} = A_j^{\Xi}$ и будут равны вероятности $P(A_{\Xi j}^O) = P(A_j^O)$, $P(A_{Oj}^{\Xi}) = P(A_j^{\Xi})$. Поэтому из (9) следует $H_{\Xi O} = H_{\Xi} + H_O$. Подстановка этого результата в (10) показывает, что количество совместной информации в этом случае получается равным нулю.

Таким образом, количество взаимной информации по (10) при описанном способе определения совместных событий изменяется от нуля до количества информации, содержащегося в эталоне и ответе, что соответствует содержательному смыслу оценки уровня адекватности ответа эталону. Это является достаточным обоснованием использования количества взаимной информации в качестве меры соответствия информационных объектов, представленных на естественном языке.

Заданемся вопросом: а какова будет мера близости эталона и ответа, если они отличаются только порядком слов? Ответ на этот вопрос может быть извлечен из алгебры событий (3) и табличной иллюстрации (см. табл. 1) образа объекта, представленного с ее использованием. По мнению авторов, широкие возможности предлагаемой методики обеспечиваются именно вариацией алгебры, которая позволяет для разложения текста использовать фактически любую систему морфологических и синтаксических конструкций языка, которые, в данном применении вероятностной модели, являются случайными событиями. Используемая в статье алгебра (3) — частный случай, ориентированный на наглядность представления.

Алгебра (3) не учитывает порядок следования слов. Более того, отмечено, что предлоги и частицы для простоты игнорировались. Поэтому по текстам, отличающимся только порядком слов, будут синтезированы совпадающие образы в виде табл. 1, в которых по морфологической принадлежности будут собраны совпадающие столбцы существительных, прилагательных и т.д. Вероятности, вычисленные по совпадающим таблицам, также совпадут, так что при используемой алгебре, не учитывающей порядок, количество информации от него не зависит.

Однако ничто не мешает в алгебру ввести события вида «существительное → глагол», «прилагательное → существительное», «предлог → существительное» и т.д., включая многословные конструкции, которые могут оказаться полезными в конкретном варианте последующего использования системы.

Одной из наиболее важных отличительных особенностей предлагаемой методологии является ее формально-математическая основа, которая позво-

ляет формализовать сам процесс синтеза автоматизированной системы оценки близости текстов. В контексте сопоставления текстов алгебра представляет собой модель текста.

Известно, что модель объекта не может с ним совпадать, а является лишь некоторым его отражением. Модели используются для оптимизации управления объектами, и при этом они должны адекватно отражать свойства объекта, учет которых необходим для эффективного управления. Чтобы приблизить модель к объекту, используется ее идентификация, призванная повышать степень адекватности модели объекту в разрезе решаемых с помощью модели задач.

Формализация текста позволяет видеть возможность структурной идентификации модели представления текста, т. е. алгебры (3). Формализация в виде структуры, приведенной в табл. 1, позволяет оценить вероятности отдельных морфологических и синтаксических компонентов, по вероятностям могут быть сконструированы дополнительные составные компоненты, которые добавляются в алгебру. Этот процесс может быть организован в виде итерационной процедуры последовательных приближений. Критерием оптимальности алгебры и целесообразности продолжения процесса идентификации алгебры (или, что одно и то же, обучения системы) может быть количество взаимной информации между эталоном и образом (ответом). При повышении уровня детализации текста эталона будет увеличиваться количество событий (столбцов) в табл. 1 и возрастать энтропия. Ответ будет структурироваться в виде такой же таблицы, с тем же набором случайных событий. Процесс детализации может быть прекращен, когда величина изменения взаимной информации в следующих одна за дру-

гой итерациях станет меньше заданного порогового значения.

Ниже на достаточно наглядных примерах иллюстрируются некоторые аспекты количества взаимной информации как меры близости, или подобия, информационных объектов, представленных на естественном, английском, языке.

4 Пример содержательной интерпретации формальных конструкций

Формально-математические конструкции, используемые при оценке адекватности текстов, возможно несколько непривычные, будут проиллюстрированы на очень простом примере. В качестве эталонного текста используется предшествующий данному разделу и выделенный полужирным курсивом абзац: «Ниже на ~~достаточно наглядных~~ примерах иллюстрируются ~~некоторые аспекты~~ количества взаимной информации ~~как меры близости~~, или подобия, информационных объектов, представленных на естественном, ~~английском~~, языке».

Ответом будет следующий, несколько измененный вариант этого же текста: «Ниже на ~~ярких~~ примерах иллюстрируются ~~главные характеристики~~ количества взаимной информации, ~~являющейся~~ мерой подобия информационных объектов, представленных на естественном языке». Для наглядности слова, убранные из эталона, зачеркнуты одной линией, а вставленные в ответ — двойной линией.

В примере используется алгебра, содержащая шесть случайных событий A_1 – A_6 , которые показаны в шапке табл. 2, представляющей конкрети-

Таблица 2 Распределение реализаций (слов) по системе событий алгебры морфологической модели текста

Существительное, A_1	Глагол ¹ , A_2	Прилагательное ² , A_3	Наречие, A_4	Предлог, A_5	Союз, A_6
примерах аспекты количества информации меры близости подобия объектов языке характеристики	иллюстрируются представленных являющейся	наглядных некоторые взаимной информационные естественном английском ярких главные	ниже достаточно	на на	Или как

Примечания:

¹К глаголам отнесены причастия, являющиеся атрибутивной глагольной формой.

²К прилагательным отнесено местоимение для сокращения таблицы, имеющей иллюстративное значение для демонстрации способа вычисления количественной оценки.

Таблица 3 Вероятности и энтропии случайных событий (3), (6) в текстах примера

Объект	Характеристика	Алгебра — система случайных событий					
		A_1	A_2	A_3	A_4	A_5	A_6
Эталон	Вероятность	0,210	0,048	0,143	0,048	0,048	0,048
	Энтропия	0,330	0,145	0,278	0,145	0,145	0,145
Ответ	Вероятность	0,190	0,071	0,119	0,024	0,048	0
	Энтропия	0,316	0,188	0,253	0,089	0,145	0
$A^{\ominus\circ}$	Вероятность	0,333	0,095	0,143	0,048	0,095	0
	Энтропия	0,366	0,224	0,278	0,145	0,224	0
A_{\circ}^{\ominus}	Вероятность	0,048	0	0,071	0,024	0	0,048
	Энтропия	0,145	0	0,188	0,080	0	0,145
A_{\ominus}°	Вероятность	0,028	0,024	0,048	0	0	0
	Энтропия	0,089	0,089	0,145	0	0	0

зацию табл. 1. В табл. 2 представлены реализации и их распределение по всем случайным событиям моделей (4)–(6):

- эталон образуют слова, не зачеркнутые и зачеркнутые одной чертой;
- ответ — слова, не зачеркнутые и зачеркнутые двойной чертой;
- незачеркнутые слова входят в оба текста;
- слова, зачеркнутые одной чертой, входят только в эталонный текст;
- слова, зачеркнутые двойной чертой, входят только в ответ.

В табл. 3 приведены количественные значения вероятностей и энтропий, вычисленных по данным табл. 2. Вероятности реализаций (слов) вычислялись в виде относительных частот: количество вхождений слова / общее число слов.

Построчным суммированием энтропий в табл. 3 получается:

- энтропия эталона — 1,188,
- энтропия ответа — 0,992,
- энтропия $A^{\ominus\circ}$ — 1,237,
- энтропия A_{\circ}^{\ominus} — 0,827,
- энтропия A_{\ominus}° — 0,323.

Количество взаимной информации в текстах примера определяется по энтропиям в виде: $I = H_{\ominus} + H_{\circ} - H_{\ominus\circ} = H_{\ominus} + H_{\circ} - H_{\circ}^{\ominus} - H_{\ominus}^{\circ} + H^{\ominus\circ} = 2,266$. Полученный результат приближается к варианту совпадения текстов, что согласуется с незначительным уровнем искажения ответа по отношению к эталону.

Завершая пример, можно отметить, что в зависимости от задачи вероятности могут исчисляться различными способами, что может привести к некоторым масштабным преобразованиям количественных характеристик. Может быть использован

аппарат условных вероятностей, т. е. введено отличие отдельных слов по их принадлежности к отдельным образцам сопоставляемых текстов. В этом случае также работает предлагаемая методология через условные энтропии и информацию. Все многообразие возможных модификаций методологии не представляется возможным изложить в пределах одной статьи, но практически готовы работы, в которых достаточно наглядно будут продемонстрированы возможности адаптации алгебры, т. е. модели представления текста, и оценки семантической близости слов русского языка, которые следуют из предлагаемого подхода.

5 Проверка монотонности информационной меры оценки близости текстов

Важным требованием к мере оценки близости текстов является монотонность ее изменения при монотонном изменении уровня адекватности текстов. Параметры меры фактического соответствия реальной близости текстов могут варьироваться, однако для эффективной практической реализации автоматического сравнения текстов монотонность меры должна иметь место.

Для исследования характера изменения совместной информации при задаваемом монотонном изменении уровня совпадения текстов выполнен эксперимент планированного искажения текста. На основе одного эталонного текста на английском языке, состоящего из 174 слов, который обозначается T_{\ominus} , синтезирован массив его искаженных копий T_j , $j = 1, 2, \dots, 20$. Копии формировались следующим образом. Эталонный текст был разбит на 20 частей. Первая искаженная копия T_1 получена из эталонного текста путем замены 1/20

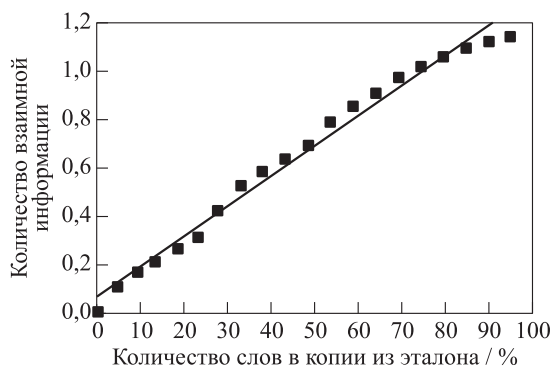


Рис. 1 Зависимость количества совместной информации от доли слов из эталона, сохраненных в копии

(по одному слову в каждой части) его слов на слова из текста с другой тематикой, вторая копия T_2 получается заменой в эталоне 2/20 слов и т. д. Последняя искаженная копия не содержит ни одного слова эталонного текста (заменено 20/20). Для каждой искаженной копии T_j определена прямая оценка ее соответствия эталону в виде доли слов из эталонного текста, сохраненных в копии, по 100-балльной шкале:

$$Q(T_{\Xi}, T_j) = \frac{m_j}{n} \cdot 100, \quad (11)$$

где m_j — число слов из эталонного текста T_{Ξ} в копии T_j ; n — число слов в эталоне.

Для полученного массива из 20 английских текстов с известными оценками (11) рассчитаны значения количества совместной информации $I_{\Xi O_j}$ между копиями T_j и эталонным текстом T_{Ξ} по (10). На рис. 1 показано соответствие между количеством совместной информации, определенным по (10), и оценкой (11) степени искажения эталона при формировании копии.

На рис. 1 видно, что значение совместной информации монотонно изменяется при монотонном изменении реального уровня искажения текстов, причем между ними существует линейная связь с коэффициентом корреляции 0,992.

6 Сопоставление информационной меры с реальными оценками преподавателя и градуировка информационной меры по N -балльной шкале оценок

Возможность практического использования методологии была оценена сопоставлением реальных

оценок, выставленных преподавателем, с оценками, сформированными автоматически на основании количества взаимной информации. Исходными данными для такого исследования являлись результаты контрольной проверки знаний студентов по английскому языку. Контроль осуществлялся написанием изложения по обычной технологии — преподаватель прочитал эталонный текст, который был по памяти воспроизведен студентами. Целью написания изложения в соответствии с методикой обучения языку была проверка знания орфографии и умения передачи содержания. Изложения были проверены и оценены преподавателем английского языка стандартным образом на предмет орфографии (первая оценка), полноты и правильности передачи содержания (вторая оценка) по 100-балльной шкале. В настоящем исследовании использовалась вторая оценка за полноту и правильность изложения эталонного текста, отражающая семантическую близость ответа эталону, орфографические ошибки игнорировались (автоматически исправлялись).

Оценка по количеству взаимной информации осуществлялась автоматизированной системой, реализующей изложенную выше методологию, использующую для классификации слов Cognitive Dwarf [6]. Эталонный текст и 22 варианта его изложения студентами были введены в систему. Для текстов формировались вероятностно-статистические морфологические образы с алгеброй событий (3). По ним определялись энтропии образов (8), (9) и количество взаимной информации (10).

Количество взаимной информации (10) является абстрактной величиной, поэтому для сопоставления необходимо произвести его градуировку в баллах принятой в образовательном учреждении системы оценок. Фактически градуировка сводится к установлению однозначного соответствия значений информационной меры (10) оценкам по выбранной шкале. В простейшем случае это соответствие может быть получено в виде линейной регрессии $y = a + bx$, где x — информационная мера соответствия ответа эталону, вычисляемая автоматически, а y — оценка по 100-балльной шкале, принятой в образовательном учреждении. Для градуировки, т. е. преобразования количества взаимной информации в реальные оценки могут быть использованы более сложные модели, позволяющие учесть конкретные содержательно-методические особенности дисциплин и регламентов. Шкала оценок может быть произвольной N -балльной.

Для иллюстрации приведем два варианта модели градуировки автоматически формируемой оценки по принятой в университете 100-балльной шка-

Таблица 4 Результаты расчета оценок изложений студентов

№ ответа студента	Оценка преподавателя	Количество информации, $I_{ЭО}$	Оценка по количеству информации		Векторно-пространственная модель текста Г. Солтона	
			Без взвешивания	С весами частей речи	$\cos(T_0, T_j)$	Оценка по $\cos(T_0, T_j)$
1	25	0,2413	41,83	35,35	0,444	32,74
2	30	0,2485	42,28	30,24	0,532	43,01
3	35	0,3365	47,68	32,31	0,589	49,54
4	45	0,3348	47,58	47,18	0,638	55,34
5	45	0,5127	58,50	50,15	0,694	61,82
6	50	0,3681	49,62	44,24	0,495	38,64
7	50	0,4157	52,55	60,10	0,730	65,94
8	53	0,5318	59,67	54,50	0,788	72,77
9	53	0,2753	43,92	43,05	0,606	51,57
10	55	0,4472	54,48	52,92	0,713	63,98
11	60	0,4954	57,44	60,40	0,617	52,83
12	60	0,4490	54,59	65,80	0,641	55,61
13	65	0,5648	61,70	70,57	0,735	66,52
14	78	0,8275	77,83	74,87	0,829	77,48
15	80	0,8549	79,52	88,53	0,846	79,43
16	80	0,4352	53,74	65,45	0,601	51,03
17	85	1,0268	90,07	90,43	0,868	82,09
18	85	1,0163	89,42	80,51	0,883	83,73
19	88	0,9875	87,66	88,62	0,867	81,94
20	88	1,0240	89,90	81,74	0,889	84,48
21	90	0,5166	58,74	83,32	0,641	55,69
22	97	1,1603	98,27	96,74	0,944	90,82

ле, результаты применения которых представлены в табл. 4. Первый вариант имел вид регрессии $y = a + bx$, для которой параметры a и b определялись методом наименьших квадратов по рядам значений оценок преподавателя (см. столбец 2 в табл. 4) и количества взаимной информации (см. столбец 3 в табл. 4). По этим данным получена линейная регрессия: $y = 27,02 + 61,41x_j$, где $x_j = I_{ЭОj}$ — количество взаимной информации в эталонном тексте и j -м изложении, а y — автоматически формируемая по нему оценка.

Второй вариант модели пересчета информационной оценки в 100-балльную отличался от первого введением учета различной значимости в текстах семантических компонентов. При использовании алгебры (3) это сводится к учету различной значимости частей речи. Для иллюстрации возможностей такого варианта уточнения перед слагаемыми A_j , соответствующими различным частям речи (см. табл. 1), в (8), (9) были введены параметры β_j , значения которых определялись методом наименьших квадратов в процессе получения уравнения регрессии. Фактически эти параметры отражают различный «вклад» семантических компонентов в соответствие между эталоном и ответом. В ре-

зультате было получено следующее уравнение множественной регрессии: $y_1 = 0,8436 - 242,23A_1 + 451,27A_2 + 131,60A_3 - 262,16A_4 - 110,75A_5$, где y_1 — оценка по 100-балльной шкале, а A_j , $j = 1, 2, \dots, 5$, — части речи (семантические компоненты — случайные события алгебры (3)).

Значения оценок y и y_1 приведены соответственно в четвертом и пятом столбцах табл. 4. Среднеквадратичное отклонение оценок y от оценок преподавателя составляет 11,04 балла, коэффициент корреляции равен 0,847. Некоторое усложнение формулы перевода количества информации в 100-балльную оценку введением весовых коэффициентов позволило существенно снизить отклонение оценок, сформированных автоматически, от оценок, выставленных преподавателем. Значения 100-балльных оценок, вычисленных с использованием множественной регрессии с весовыми коэффициентами частей речи показаны в пятом столбце табл. 4. Их среднеквадратичное отклонение от оценок преподавателя составило 6,324 балла, а множественный коэффициент корреляции достиг 0,955. Видно, что среднее отклонение оценки, формируемой автоматически по количеству совместной информации, от оценки преподавателя уменьшилось

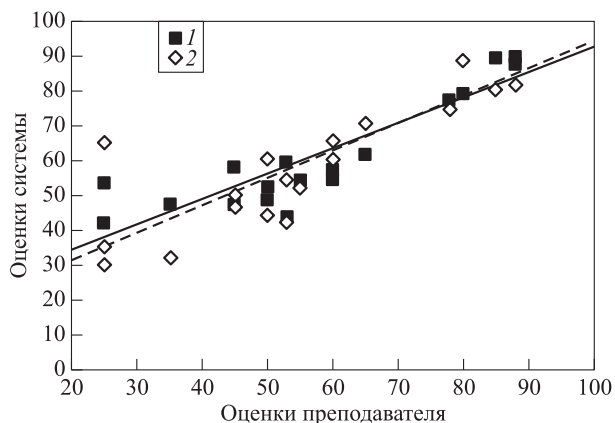


Рис. 2 Сопоставление оценок преподавателя и оценок системы: 1 — при градуировке без взвешивания частей речи; 2 — при взвешивании частей речи

почти в два раза. Этот пример показывает возможности повышения адекватности автоматически формируемых оценок идентификацией формул перевода количества информации в баллы реальной шкалы оценок. Кроме того, для повышения адекватности автоматической оценки, следуя методике [4], может быть использован механизм вариации алгебры, т. е. системы семантических компонентов — случайных событий, отраженной в табл. 1.

На рис. 2 представлено наглядное сопоставление оценок, выставленных преподавателем, и оценок y (без взвешивания) и y_1 (с взвешиванием частей речи), сформированных автоматически по количеству взаимной информации. И на рис. 2 видны две точки, которые вносят наибольший вклад в ошибку. В табл. 4 они выделены полужирным шрифтом (16-я и 21-я). При детальном анализе выяснилось, что оценки, выставленные преподавателем, были не в полной мере адекватны содержанию изложений. Интересно отметить, что исключение этих точек и расчет по оставшимся 20 точкам принципиально меняет результат: среднеквадратичная ошибка оценки y становится равной 1,593 балла, а оценки y_1 — 1,248 балла.

Выше отмечалось, что в настоящее время многие информационно-поисковые системы базируются на векторно-пространственной модели текста, предложенной Г. Солтоном с соавторами [3], в которой мерой близости текстов служит косинус угла между векторами, отражающими надлежащим образом преобразованные тексты. В принятых ранее обозначениях это

$$\cos(T_{\Theta}, T_j) = \frac{(T_{\Theta}, T_j)}{|T_{\Theta}| |T_j|}.$$

Такая мера также была вычислена для всех 22 изложений и эталонного текста. Эта мера была сопоставлена с оценками преподавателя и получена

соответствующая регрессия для перевода ее значений в 100-балльные оценки:

$$Q_{100} = -18,97 + 116,3 \cos(T_{\Theta}, T_j).$$

Результаты вычислений представлены в 6-м и 7-м столбцах табл. 4. Среднеквадратичное отклонение прогноза оценки по этой регрессии от оценки преподавателя составило 13,27 балла, а коэффициент корреляции равен 0,769. После удаления из массива 16-й и 21-й точек ошибка составила 2,155 балла. Уклонение оценки, автоматически полученной по разработанной методике, от оценки преподавателя оказалось примерно в два раза меньше. Обычно предполагается нормальное распределение для ошибки и величина ее оценивается 3σ -интервалом. В этом случае возможное максимальное уклонение оценки, формируемой автоматически по разработанной методике, от оценки преподавателя оценивается в 3,744 балла, в то время как для векторно-пространственной модели оценка ошибки дает 6,466 балла. Без удаления проблемных 16-й и 21-й точек соотношение ошибок предсказания (примерно в два раза) сохраняется.

Таким образом, из приведенного примера видно, что автоматизированная проверка знаний (оценка семантической близости текстов) не утопия. Реальный путь повышения надежности и уровня адекватности сопоставления текстов представляется в исследовании возможностей детализации системы семантических компонентов, адаптации градуировки количества информации в принятой шкале оценок, учета синонимов. Но все эти задачи укладываются в предложенную методологию, и варианты их решения достаточно хорошо просматриваются. Важно отметить, что вероятностная модель текста и информационная мера близости текстов содержательно соответствуют существу задачи семантического сопоставления текстов и имеют практически неограниченные возможности детализации и повышения на этой основе достоверности оценки семантической адекватности ответа эталону.

7 Заключение

Изложена оригинальная методологическая основа автоматизации процедуры оценки уровня семантического подобия текстов, которая может быть использована в автоматических системах обработки вербальной информации, в частности в системах оценки уровня знаний учащихся. Системы оценки знаний, разработанные на ее основе, смогут воспринимать реальные полноценные ответы учащихся на поставленные вопросы и обеспечивать автоматическую оценку уровня адекватности

ответа содержанию эталонного ответа, хранящегося в базе данных системы. Представлены примеры, иллюстрирующие содержательную интерпретацию формально-математического аппарата и возможности настройки перевода количества взаимной информации в оценку по принятой в учебном заведении шкале оценки знаний. Разработанная математическая основа может адаптироваться к задачам синтеза различных моделей представления текстов, позволяющих отразить специфику дисциплин и методических приемов.

Литература

1. *Безсуднов И. В.* Интернетика. Навигация в сложных сетях: модели и алгоритмы. — М.: Либроком, 2009. 264 с.
2. *Manning Ch. D., Raghavan P., Schütz H.* An introduction to information retrieval. — Cambridge: University Press, 2009. 569 p.
3. *Salton G., Wong A., Yang C. S.* A vector space model for automatic indexing // *Communications of the ACM*, 1975. Vol. 18. No. 11. P. 613–620.
4. *Кузнецов Л. А.* Теоретические основы автоматизированной оценки знаний // *Качество. Инновации. Образование*, 2010. № 11. С. 8–19.
5. *Гнеденко Б. В.* Курс теории вероятностей: Учебник. — 9-е изд., испр. — М.: ЛКИ, 2007. 448 с.
6. Программный пакет синтаксического разбора и машинного перевода Cognitive Dwarf. <http://cs.isa.ru:10000/dwarf/>.
7. *Шеннон К.* Работы по теории информации и кибернетике / Пер. с англ. — М.: ИЛ, 1963.

ОБОБЩЕННЫЕ ДИСПЕРСИОННЫЕ ГАММА-РАСПРЕДЕЛЕНИЯ КАК ПРЕДЕЛЬНЫЕ ДЛЯ СЛУЧАЙНЫХ СУММ*

Л. М. Закс¹, В. Ю. Королев²

Аннотация: Доказана общая теорема о необходимых и достаточных условиях сходимости распределений сумм случайного числа независимых одинаково распределенных случайных величин к однопараметрическим сдвиг-масштабным смесям нормальных законов. В качестве следствия получены необходимые и достаточные условия сходимости распределений случайных сумм независимых одинаково распределенных случайных величин к обобщенным дисперсионным гамма-распределениям. Для частного случая — специальных случайных блужданий с непрерывным временем, порожденных обобщенными дважды стохастическими пуассоновскими процессами, — приведены оценки скорости этой сходимости.

Ключевые слова: случайная сумма; обобщенное гиперболическое распределение; обобщенное обратное гауссовское распределение; обобщенное гамма-распределение; обобщенное дисперсионное гамма-распределение; смесь распределений вероятностей; идентифицируемые смеси; аддитивно замкнутое семейство; оценка скорости сходимости

1 Введение

В данной работе в качестве более гибкой альтернативы часто (и успешно) применяемым в практических исследованиях обобщенным гиперболическим распределениям рассматривается класс обобщенных дисперсионных гамма-распределений и дается теоретическое обоснование использованию его представителей в качестве асимптотических аппроксимаций при решении практических задач.

1.1 Обобщенные гиперболические распределения

Плотность *обобщенного обратного гауссовского распределения* обозначим $p_{\text{GIG}}(x; \nu, \mu, \lambda)$:

$$p_{\text{GIG}}(x; \nu, \mu, \lambda) = \frac{\lambda^{\nu/2}}{2\mu^{\nu/2} K_{\nu}(\sqrt{\mu\lambda})} x^{\nu-1} \exp\left\{-\frac{1}{2}\left(\frac{\mu}{x} + \lambda x\right)\right\}, \quad x > 0.$$

Здесь $\nu \in \mathbb{R}$,

$$\mu > 0, \lambda \geq 0, \text{ если } \nu < 0;$$

$$\mu > 0, \lambda > 0, \text{ если } \nu = 0;$$

$$\mu \geq 0, \lambda > 0, \text{ если } \nu > 0,$$

$K_{\nu}(z)$ — модифицированная бесселева функция третьего рода порядка ν :

$$K_{\nu}(z) = \frac{1}{2} \int_0^{\infty} y^{\nu-1} \exp\left\{-\frac{z}{2}\left(y + \frac{1}{y}\right)\right\} dy, \quad z \in \mathbb{C}, \quad \operatorname{Re} z > 0.$$

Соответствующую функцию распределения обозначим $P_{\text{GIG}}(x; \nu, \mu, \lambda)$:

$$P_{\text{GIG}}(x; \nu, \mu, \lambda) = \begin{cases} 0, & x < 0; \\ \int_0^x p_{\text{GIG}}(z; \nu, \mu, \lambda) dz, & x \geq 0. \end{cases}$$

Как отмечено в статье [1], обобщенное обратное гауссовское распределение введено в 1946 г. Этьеном Альфеном (Étienne Halphen), который использовал его для описания объема воды, проходящего ежемесячно через гидроэлектростанции. В [1] обобщенное обратное гауссовское распределение названо *распределением Альфена*. В 1973 г. это распределение было переоткрыто Гербертом Зихелем [2], который использовал его в качестве смешивающего закона при рассмотрении специальных смешанных пуассоновских распределений (*распределений Зихеля*, см., например, [3]) как дискретных распределений с тяжелыми хвостами. В 1977 г. эти распреде-

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м и 12-07-00115а), Министерством образования и науки (госконтракт 16.740.11.0133).

¹Альфа-банк, отдел моделирования и математической статистики, lily.zaks@gmail.com

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук, vkorolev@cs.msu.su

ления снова переоткрыл О.-Э. Барндорфф-Нильсен [4, 5], который, в частности, использовал их для описания распределения размеров частиц.

Класс обобщенных обратных гауссовских распределений довольно обширен и содержит, в частности, как распределения с экспоненциально убывающими хвостами (гамма-распределение ($\mu = 0, \nu > 0$)), так и распределения с хвостами, убывающими степенным образом (обратное гамма-распределение ($\lambda = 0, \nu < 0$), обратное гауссовское распределение ($\nu = -1/2$) и его предельный при $\lambda \rightarrow 0$ случай — распределение Леви (устойчивое распределение с характеристическим показателем, равным $1/2$, сосредоточенное на неотрицательной полуоси — распределение времени достижения стандартным винеровским процессом единичного уровня)).

Стандартную нормальную функцию распределения будем обозначать $\Phi(x)$:

$$\Phi(x) = \int_{-\infty}^x \varphi(z) dz, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\},$$

$$x \in \mathbb{R}.$$

В 1977–1978 гг. О.-Э. Барндорфф-Нильсен [4, 5] ввел класс *обобщенных гиперболических распределений* как класс специальных сдвиг-масштабных смесей нормальных законов. Пусть $\alpha \in \mathbb{R}, \beta \in \mathbb{R}$. Если функцию обобщенного гиперболического распределения с параметрами $\alpha, \beta, \nu, \mu, \lambda$ обозначить $P_{GH}(x; \alpha, \beta, \nu, \mu, \lambda)$, то по определению

$$P_{GH}(x; \alpha, \beta, \nu, \mu, \lambda) = \int_0^{\infty} \Phi\left(\frac{x - \beta - \alpha z}{\sqrt{z}}\right) p_{GIG}(z; \nu, \mu, \lambda) dz, \quad x \in \mathbb{R}. \quad (1)$$

Обратим внимание, что в (1) смешивание происходит одновременно и по параметру сдвига, и по параметру масштаба, но так как эти параметры в (1) связаны жесткой зависимостью, то фактически смесь (1) является *однопараметрической*.

Несложно убедиться, что плотность $p_{GH}(x; \alpha, \beta, \nu, \mu, \lambda)$ обобщенного гиперболического распределения имеет вид:

$$p_{GH}(x; \alpha, \beta, \nu, \mu, \lambda) = \int_0^{\infty} \frac{1}{\sqrt{z}} \varphi\left(\frac{x - \beta - \alpha z}{\sqrt{z}}\right) p_{GIG}(z; \nu, \mu, \lambda) dz = \int_0^{\infty} \frac{1}{\sqrt{2\pi z}} \exp\left\{-\frac{(x - \beta - \alpha z)^2}{2z}\right\} \times$$

$$\times \frac{\lambda^{\nu/2} z^{\nu-1}}{2\mu^{\nu/2} K_{\nu}(\sqrt{\mu\lambda})} \exp\left\{-\frac{1}{2}\left(\frac{\mu}{z} + \lambda z\right)\right\} dz = \frac{\lambda^{\nu/2} (\lambda + \alpha^2)^{\nu/2-1/4}}{2\sqrt{2\pi} \mu^{\nu/2} K_{\nu}(\sqrt{\mu\lambda})} (\mu + (x - \beta)^2)^{1/4-\nu/2} \times \exp\{\alpha(x-\beta)\} K_{\nu-1/2}\left(\sqrt{(\mu + (x - \beta)^2)(\lambda + \alpha^2)}\right).$$

Класс обобщенных гиперболических распределений очень широк и содержит, в частности, (а) симметричные и скошенные (skew) распределения Стьюдента (в том числе распределение Коши), которым в представлении (1) соответствуют смешивающие обратные гамма-распределения; (б) дисперсионные гамма-распределения (Variance Gamma (VG) distributions) (в том числе симметричные и несимметричные распределения Лапласа), которым в представлении (1) соответствуют смешивающие гамма-распределения; (в) нормальные обратные нормальные (NIG) распределения, которым в представлении (1) соответствуют смешивающие обратные нормальные распределения, и многие другие типы.

Обобщенные гиперболические распределения продемонстрировали высочайшую адекватность при их применении для описания статистических закономерностей поведения различных характеристик сложных открытых систем, в частности турбулентных систем и финансовых рынков. Публикации, посвященные моделям, основанным на обобщенных гиперболических распределениях, исчисляются сотнями. Достаточно упомянуть лишь канонические работы [5–16]. Согласно расхожему мнению, столь высокая адекватность обобщенных гиперболических моделей может быть формально объяснена большим числом настраиваемых параметров, позволяющим подогнать какую угодно модель к каким угодно данным.

Среди статистиков хорошо известно высказывание Ж. Бертрана «Give me four parameters and I shall describe an elephant; with five, it will wave its trunk» (цитируется по статье Л. ЛеКама [17]). Это обстоятельство, конечно же, играет свою роль, однако на самом деле модели типа (1) в большинстве случаев адекватны по гораздо более естественным глубоким причинам.

В прикладной теории вероятностей принято считать, что ту или иную модель можно считать в достаточной мере обоснованной (адекватной) только тогда, когда она является *асимптотической аппроксимацией*, т. е. когда существует довольно простая предельная схема (например, схема суммирования) и соответствующая предельная теорема, в которой

рассматриваемая модель выступает в качестве предельного распределения [18]. В первоисточниках упомянутые выше обобщенные гиперболические модели вводились чисто умозрительно как распределения процесса броуновского движения со случайным временем, в каждый момент имеющим то или иное обобщенное обратное гауссовское распределение. Лишь в статье [19] со ссылкой на работу А. Реньи [20] имеется довольно расплывчатый намек на то, что смеси нормальных законов могут быть предельными для сумм случайного числа случайных величин.

Однако, как ни удивительно, несмотря на то что свойства обобщенных гиперболических распределений изучены довольно полно, до недавнего времени не было дано корректного доказательства того факта, что обобщенные гиперболические распределения выступают в качестве предельных в простейшей схеме случайно остановленных случайных блужданий. И, стало быть, приводимая в некоторых работах аргументация, связывающая смешивание в модели (1) со случайным характером волатильности при применении обобщенных гиперболических распределений в финансовой математике, не имела строгого формального обоснования. Возможно, причина этого лежит в том, что в схеме «нарастающих» сумм, рассматривавшейся в [20], полное решение указанной задачи невозможно. Его можно получить, лишь рассматривая случайные суммы в рамках асимптотической схемы серий. основополагающей работой в этом направлении стала работа Б. В. Гнеденко и Х. Фахима [21].

«Асимптотическое» обоснование некоторых из упомянутых выше моделей было дано лишь недавно в статьях [22, 23], где показано, что скошенные распределения Стьюдента и дисперсионные гамма-распределения могут выступать в качестве предельных в довольно простых предельных теоремах для регулярных статистик, построенных по выборкам случайного объема, в частности в схеме случайного суммирования случайных величин, и, следовательно, могут считаться *естественными* асимптотическими аппроксимациями для распределений многих процессов, например, сходных с неоднородными случайными блужданиями.

В статье [24] приведена общая теорема о необходимых и достаточных условиях сходимости распределений сумм случайного числа независимых одинаково распределенных случайных величин к однопараметрическому сдвиг-масштабным смесям нормальных законов и в качестве следствия из нее получены необходимые и достаточные условия сходимости распределений случайных сумм независимых одинаково распределенных случайных величин к обобщенным гиперболическим распре-

делениям. На примере довольно общего и просто интерпретируемого частного случая — специальных случайных блужданий с непрерывным временем, порожденных обобщенными дважды стохастическими пуассоновскими процессами, — там же приведены оценки скорости этой сходимости. В данной работе результаты статьи [24] переносятся на обобщенные дисперсионные гамма-распределения.

1.2 Обобщенные дисперсионные гамма-распределения

Гамма-распределение и обратное гамма-распределение являются частными представителями класса обобщенных гамма-распределений (ОГ-распределения), важная роль которых в моделировании и анализе стохастической структуры информационных потоков описана в книге [25]. Обобщенные гамма-распределения были впервые описаны как единое семейство в 1962 г. в работе [26] в качестве семейства вероятностных моделей, включающего в себя одновременно гамма-распределения и распределения Вейбулла.

Обобщенным гамма-распределением называется распределение, определяемое плотностью вероятностей вида

$$f(x; \nu, \kappa, \delta) = \begin{cases} \frac{|\nu|}{\delta \Gamma(\kappa)} \left(\frac{x}{\delta}\right)^{\kappa\nu-1} \exp\left\{-\left(\frac{x}{\delta}\right)^\nu\right\}, & x \geq 0; \\ 0, & x < 0, \end{cases} \quad (2)$$

с параметрами $\nu \in \mathbb{R}$, $\kappa, \delta \in \mathbb{R}^+$, отвечающими соответственно за *степень, форму и масштаб*, где $\Gamma(\kappa)$ — эйлерова гамма-функция:

$$\Gamma(\kappa) = \int_0^{\infty} x^{\kappa-1} e^{-x} dx.$$

Семейство ОГ-распределений включает в себя практически все наиболее популярные абсолютно непрерывные распределения. В частности, семейство ОГ-распределений содержит следующие распределения.

1. Гамма-распределение ($\nu = 1$):

$$f(x; \kappa, \theta) = \frac{1}{\Gamma(\kappa)} \theta^\kappa x^{\kappa-1} e^{-\theta x}, \quad x \geq 0, \kappa > 0, \theta > 0.$$

1.1. Показательное (экспоненциальное) распределение ($\nu = 1, \kappa = 1$):

$$f(x; \theta) = \theta e^{-\theta x}, \quad x \geq 0, \theta > 0.$$

1.2. Распределение Эрланга ($\nu = 1, \kappa \in \mathbb{N}$):

$$f(x; \kappa, \theta) = \frac{1}{\Gamma(\kappa)} \theta^\kappa x^{\kappa-1} e^{-\theta x},$$

$$x \geq 0, \kappa > 0, \theta > 0.$$

1.3. Распределение хи-квадрат ($\nu = 1, \delta = 2$):

$$f(x; n) = \frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{n/2-1} e^{-x/2}, \quad x \geq 0, n \in \mathbb{N}.$$

2. Распределение Накагами ($\nu = 2$):

$$f(x; \mu, \lambda) = \frac{2(\lambda\mu)^\mu}{\Gamma(\mu)} x^{2\mu-1} e^{-\lambda\mu x^2},$$

$$x \geq 0, \mu > 0, \lambda > 0.$$

2.1. Полунормальное распределение (распределение максимума винеровского процесса на отрезке $[0, 1]$) ($\nu = 2, \kappa = 1/2$):

$$f(x; \delta) = \sqrt{\frac{2}{\pi\delta}} \exp\left\{-\frac{x^2}{2\delta^2}\right\}, \quad x \geq 0, \delta > 0.$$

2.2. Распределение Рэлея ($\nu = 2, \kappa = 1$):

$$f(x; \delta) = \frac{x}{\delta^2} \exp\left\{-\frac{x^2}{2\delta^2}\right\}, \quad x \geq 0, \delta > 0.$$

2.3. Хи-распределение ($\nu = 2, \delta = \sqrt{2}$):

$$f(x; n) = \frac{1}{2^{n/2-1} \Gamma(n/2)} x^{n-1} \exp\left\{-\frac{x^2}{2}\right\},$$

$$x \geq 0, n \in \mathbb{N}.$$

2.4. Распределение Максвелла (распределение модулей скоростей движения молекул в разреженном газе) ($\nu = 2, \kappa = 3/2$):

$$f(x; \delta) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\delta^3} \exp\left\{-\frac{x^2}{2\delta^2}\right\}, \quad x \geq 0, \delta > 0.$$

3. Распределение Вейбулла–Гнеденко ($\kappa = 1$):

$$f(x; \eta, \mu) = \frac{\eta x^{\eta-1}}{\mu^\eta} \exp\left\{-\left(\frac{x}{\mu}\right)^\eta\right\},$$

$$x \geq 0, \eta > 0, \mu > 0.$$

4. Обратное гамма-распределение ($\nu = -1$):

$$f(x; \mu, \lambda) = \frac{1}{\mu\lambda\Gamma(\lambda)} \left(\frac{\mu\lambda}{x}\right)^{\lambda+1} \exp\left\{-\frac{\mu\lambda}{x}\right\},$$

$$x \geq 0, \lambda > 0, \mu > 0.$$

4.1. Распределение Леви ($\nu = -1, \kappa = 1/2$):

$$f(x; \mu) = \sqrt{\frac{\mu}{2\pi}} \frac{1}{x^{3/2}} \exp\left\{-\frac{\mu}{2x}\right\}, \quad x \geq 0, \mu > 0.$$

5. Логнормальное распределение ($\kappa \rightarrow \infty$):

$$f(x; \mu, \delta) = \frac{1}{\delta x \sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\delta^2}\right\},$$

$$x \geq 0, \mu \in \mathbb{R}, \delta > 0.$$

Широкая применимость ОГ-распределений обусловлена возможностью их использования в качестве адекватных асимптотических аппроксимаций, поскольку практически все они выступают в качестве предельных в различных предельных теоремах теории вероятностей, а именно:

- показательное распределение выступает в качестве предельного как в схеме максимума (минимума) (см., например, [27]), так и в схеме геометрического суммирования, описывая распределение времени восстановления в прореженных процессах восстановления, выступающих моделями потоков редких событий (см., например, [28]);
- гамма-распределение является безгранично делимым и потому выступает в качестве предельного для распределений сумм независимых равномерно предельно малых случайных величин. При этом распределение Эрланга возникает как допредельное распределение суммы независимых экспоненциально распределенных случайных величин, что в терминах случайной интенсивности может означать, что если случайная интенсивность потока поступления запросов имеет гамма-распределение со значимым параметром формы, то при обработке этих запросов в основном задействованы механизмы последовательной обработки информации;
- распределение Вейбулла–Гнеденко принадлежит к так называемому первому типу предельных распределений экстремальных порядковых статистик (минимума или максимума) (см., например, [27]), что в терминах случайной интенсивности может означать, что если случайная интенсивность потока поступления запросов имеет распределение Вейбулла–Гнеденко со значимым параметром степени, то при обработке этих запросов в основном задействованы механизмы параллельной обработки информации;

- полунормальное распределение (распределение модуля стандартной нормальной случайной величины) возникает как предельное для максимальных частичных сумм независимых случайных величин (см., например, [29]);
- распределение Леви принадлежит к классу устойчивых законов и потому является предельным для сумм независимых одинаково распределенных случайных величин. Оно также является распределением времени достижения стандартным винеровским процессом (процессом броуновского движения) фиксированного уровня;
- логнормальное распределение выступает в качестве предельного для распределения размера частиц при дроблении (см., например, [30]).

Эти свойства ОГ-распределений обосновывают, в частности, целесообразность моделирования с их помощью распределения случайной интенсивности потока запросов в информационных системах. Это семейство также широко используется в других прикладных задачах в самых разных областях (см., например, [25]).

Упомянутые выше четырехпараметрические семейства скошенных распределений Стюдента и дисперсионных гамма-распределений являются подклассами введенного в работе [22] пятипараметрического семейства распределений

$$W(x; a, \sigma, \nu, \kappa, \delta) = \int_0^{\infty} \Phi\left(\frac{x - au}{\sigma\sqrt{u}}\right) f(u; \nu, \kappa, \delta) du, \quad (3)$$

где $f(u; \nu, \kappa, \delta)$ — плотность ОГ-распределения (2). В статье [22] распределения вида (3) названы *обобщенными дисперсионными гамма-распределениями*.

Задача поиска универсальной модели статистических закономерностей во многих областях, в частности в финансовой математике или в физике плазмы, подобна задаче отыскания «философского камня» в алхимии и поэтому не имеет точного решения. Однако, основываясь на вышеперечисленных аналитических и асимптотических свойствах представителей семейства ОГ-распределений и предельных теоремах для сумм независимых случайных величин как теоретико-вероятностной формализации принципа неубывания неопределенности в сложных системах [31], можно утверждать, что семейство обобщенных дисперсионных гамма-распределений является *практически* универсальным для многих задач. Оно представляется еще более гибкой моделью, нежели обобщенные гиперболические распределения, так как класс обобщенных гамма-распределений в определенном смысле ши-

ре класса обобщенных обратных гауссовских распределений, поскольку, в отличие от последнего, он содержит распределения вейбулловского (экспоненциально-степенного) типа с произвольным показателем степени в экспоненте.

2 Критерий сходимости распределений случайных сумм к обобщенным дисперсионным гамма-распределениям

Будем считать, что все случайные величины, о которых пойдет речь ниже, заданы на одном вероятностном пространстве $(\Omega, \mathfrak{A}, P)$. Пусть $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$, — семейство последовательностей одинаково распределенных в каждой последовательности (при каждом фиксированном n) случайных величин. Пусть $\{N_n\}_{n \geq 1}$ — последовательность целочисленных неотрицательных случайных величин таких, что при каждом $n \geq 1$ случайные величины $N_n, X_{n,1}, X_{n,2}, \dots$ независимы. Положим

$$S_{n,k} = X_{n,1} + \dots + X_{n,k}.$$

Во избежание недоразумений полагаем $\sum_{j=1}^0 = 0$.

Символ \implies будет обозначать слабую сходимость (сходимость по распределению).

Расстояние Леви, которое, как известно, метризует слабую сходимость в пространстве функций распределения, будем обозначать $L(\cdot, \cdot)$,

$$L(F, G) = \inf\{\epsilon : G(x - \epsilon) - \epsilon \leq F(x) \leq G(x + \epsilon) + \epsilon \forall x \in \mathbb{R}\}.$$

Каждой паре функций распределения (F, H) поставим в соответствие множество $\mathcal{M}(F|H)$, содержащее все функции распределения $Q(x)$ с $Q(0) = 0$, обеспечивающие представление характеристической функции, соответствующей функции распределения F , в виде степенной смеси характеристической функции, соответствующей функции распределения H :

$$\int_{-\infty}^{\infty} e^{itx} dF(x) = \int_0^{\infty} h^x(t) dQ(x), \quad t \in \mathbb{R},$$

где

$$h(t) = \int_{-\infty}^{\infty} e^{itx} dH(x), \quad t \in \mathbb{R}.$$

Везде далее сходимость будет подразумеваться при $n \rightarrow \infty$.

Лемма 1. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и функция распределения $H(x)$ такие, что*

$$P(S_{n,k_n} < x) \implies H(x).$$

Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость

$$P(S_{n,N_n} < x) \implies F(x)$$

распределений случайных сумм к некоторой функции распределения $F(x)$, необходимо и достаточно, чтобы существовала слабо компактная последовательность функций распределения $\{Q_n^(x)\}_{n \geq 1}$ такая, что выполняются условия:*

- (i) $Q_n^*(x) \in \mathcal{M}(F|H), n = 1, 2, \dots;$
- (ii) $L(Q_n^*, Q_n) \rightarrow 0,$

где $Q_n(x) = P(N_n < k_n x), x \in \mathbb{R}.$

Доказательство. Данное утверждение является частным случаем теоремы 4.2.1 в [31].

Напомним определение идентифицируемости смесей распределений вероятностей, предложенное в работе [32]. Для целей данной статьи достаточно рассмотреть смеси распределений из однопараметрических семейств. Пусть функция $H(x; y)$ определена на плоскости $\mathbb{R} \times \mathbb{R}$. Предположим, что функция $H(x; y)$ измерима по y при каждом фиксированном $x \in \mathbb{R}$ и является функцией распределения как функция аргумента x при каждом фиксированном $y \in \mathbb{R}$. Пусть \mathcal{Q} — некоторое семейство функций распределения. Обозначим

$$\mathcal{F} = \left\{ F(x) = \int_{-\infty}^{\infty} H(x; y) dQ(y), x \in \mathbb{R} : Q \in \mathcal{Q} \right\}. \quad (4)$$

Семейство \mathcal{F} называется *идентифицируемым*, если из равенства

$$\int_{-\infty}^{\infty} H(x; y) dQ_1(y) = \int_{-\infty}^{\infty} H(x; y) dQ_2(y), \quad x \in \mathbb{R},$$

с $Q_1 \in \mathcal{Q}, Q_2 \in \mathcal{Q}$ вытекает, что $Q_1(y) \equiv Q_2(y).$

Рассмотрим некоторые достаточные условия идентифицируемости смесей распределений из однопараметрических семейств. Хорошо известно, что в общем случае сдвиг-масштабные смеси нормальных законов не являются идентифицируемыми. Однако однопараметрические сдвиг-масштабные смеси нормальных законов типа (1) обладают этим свойством, как будет показано при доказательстве теоремы 1.

Семейство функций распределения $\{H(x; y) : y > 0\}$ называется *аддитивно замкнутым*, если для любых $y_1 > 0, y_2 > 0$ справедливо соотношение:

$$H(x; y_1) * H(x; y_2) \equiv H(x; y_1 + y_2). \quad (5)$$

Здесь символ $*$ обозначает свертку. Иногда свойство (5) семейств распределений вероятностей называется *воспроизводимостью* по параметру y .

Следующий результат принадлежит Г. Тейчере [32].

Лемма 2. *Предположим, что множество \mathcal{Q} состоит из всех функций распределения $Q(y)$ с $Q(0) = 0$. Пусть семейство функций распределения $\{H(x; y) : y > 0\}$ аддитивно замкнуто. Тогда семейство смесей (4) является идентифицируемым.*

Основным результатом данного раздела является следующее утверждение.

Теорема 1. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что*

$$P(S_{n,k_n} < x) \implies \Phi(x - \alpha). \quad (6)$$

Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к некоторой функции распределения $F(x)$:

$$P(S_{n,N_n} < x) \implies F(x),$$

необходимо и достаточно, чтобы существовала функция распределения $Q(x)$ такая, что $Q(0) = 0,$

$$F(x) = \int_0^{\infty} \Phi\left(\frac{x - \alpha z}{\sqrt{z}}\right) dQ(z); \quad (7)$$

$$P(N_n < x k_n) \implies Q(x). \quad (8)$$

Доказательство. Зафиксируем вещественное число α и положим

$$H_\alpha(x; y) = \Phi\left(\frac{x - \alpha y}{\sqrt{y}}\right). \quad (9)$$

Убедимся, что семейство так определенных функций распределения аддитивно замкнуто по $y > 0$. Действительно, функции распределения (9) соответствует характеристическая функция

$$\int_{-\infty}^{\infty} e^{itx} d_x H_\alpha(x; y) = \exp\left\{y\left(it\alpha - \frac{t^2}{2}\right)\right\}.$$

Следовательно, свертке функций распределения $H_\alpha(x; y_1)$ и $H_\alpha(x; y_2)$ соответствует характеристическая функция

$$\begin{aligned} \exp \left\{ y_1 \left(it\alpha - \frac{t^2}{2} \right) \right\} \exp \left\{ y_2 \left(it\alpha - \frac{t^2}{2} \right) \right\} &= \\ = \exp \left\{ (y_1 + y_2) \left(it\alpha - \frac{t^2}{2} \right) \right\} &= \\ = \int_{-\infty}^{\infty} e^{itx} d_x H_\alpha(x; y_1 + y_2), \end{aligned}$$

что и означает аддитивную замкнутость по $y > 0$ семейства функций распределения (9). Таким образом, согласно лемме 2 семейство однопараметрических сдвиг-масштабных смесей нормальных законов

$$\mathcal{F}_\alpha = \left\{ \int_0^\infty H_\alpha(x; y) dQ(y), x \in \mathbb{R} : Q \in \mathcal{Q} \right\},$$

где \mathcal{Q} — множество всех функций распределения $Q(y)$ с $Q(0) = 0$, идентифицируемо. В свою очередь, это означает, что для любой функции распределения $F(x)$ множество $\mathcal{M}(F|H_\alpha)$ содержит не более одного элемента. Действительно, обозначим $h_\alpha(t) = \exp\{it\alpha - t^2/2\}$ (характеристическая функция $h_\alpha(t)$ соответствует функции распределения $\Phi(x - \alpha)$). Тогда $h_\alpha^y(t) = \exp\{y(it\alpha - t^2/2)\}$, и, стало быть, характеристические функции $\int_0^\infty h_\alpha^y(t) dQ(y)$, фигурирующие в определении множества $\mathcal{M}(F|H_\alpha)$, соответствуют функциям распределения $\int_0^\infty H_\alpha(x; y) dQ(y)$, составляющим семейство \mathcal{F} . При этом условие (7) означает, что $Q \in \mathcal{M}(F(x)|\Phi(x - \alpha))$. Теперь остается сослаться на лемму 1, в которой роль условий (i) и (ii) играют соответственно (7) и (8). Теорема доказана.

Замечание 1. Условие (6) выполняется в следующей довольно общей ситуации. Предположим, что случайные величины $X_{n,j}$ имеют конечные дисперсии. Также предположим, что величины $X_{n,j}$ могут быть представлены в виде

$$X_{n,j} = X_{n,j}^* + \alpha_n,$$

где $\alpha_n \in \mathbb{R}$, а $X_{n,j}^*$ — случайная величина с $EX_{n,j}^* = 0$, $DX_{n,j}^* = \sigma_n^2 < \infty$, так что $EX_{n,1} = \alpha_n$ и $DX_{n,1} = \sigma_n^2$. Предположим, что $\alpha_n k_n \rightarrow a$ и $k_n \sigma_n^2 \rightarrow 1$ при $n \rightarrow \infty$. Тогда вследствие хорошо известного результата о необходимых и достаточных условиях сходимости к нормальному закону распределений сумм независимых случайных величин с конечными дисперсиями в схеме серий (см., например, [18])

можно заметить, что соотношение (6) имеет место тогда и только тогда, когда выполнено условие Линдеберга: для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} k_n E(X_{n,1}^*)^2 \mathbb{I}(|X_{n,1}^*| \geq \varepsilon) = 0$$

(здесь $\mathbb{I}(A)$ — индикаторная функция множества (события) A), т. е. квадратичные хвосты распределений слагаемых должны убывать достаточно быстро.

Следствие 1. Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что имеет место сходимость (6). Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к обобщенным дисперсионным гамма-распределениям:

$$P(S_{n,N_n} < x) \implies W(x; \alpha, \sigma, \nu, \kappa, \delta), \quad (10)$$

необходимо и достаточно, чтобы

$$P(N_n < x k_n) \implies F(x; \nu, \kappa, \delta), \quad (11)$$

где $F(x; \nu, \kappa, \delta)$ — функция распределения обобщенного гамма-распределения, соответствующая плотности $f(x; \nu, \kappa, \delta)$ (см. (16)).

Доказательство. Из определения обобщенного дисперсионного гамма-распределения вытекает, что множество $\mathcal{M}(W(x; \alpha, \sigma, \nu, \kappa, \delta) | \Phi(x - \alpha))$ состоит из единственного элемента $F(x; \nu, \kappa, \delta)$. Поэтому, применяя теорему 1 с $F(x) = W(x; \alpha, \sigma, \nu, \kappa, \delta)$, замечаем, что в рассматриваемом случае условие (8) принимает вид (11). Теорема доказана.

Замечание 2. В соотношениях (6), (10) и (11) предельные функции распределения непрерывны, поэтому в этих соотношениях сходимость по распределению эквивалентна равномерной сходимости функций распределения.

3 Оценки скорости сходимости распределений случайных сумм к обобщенным дисперсионным гамма-распределениям

В данном разделе будет использоваться специальная и довольно естественная конструкция случайных блужданий, удовлетворяющая комплексу условий, указанному в замечании 1.

Пусть ξ_1, ξ_2, \dots — независимые одинаково распределенные случайные величины с $E\xi_1 = 0$, $D\xi_1 = 1$, $\beta^3 = E|\xi_1|^3 < \infty$, $a \in \mathbb{R}$, n — натуральное число.

Положим

$$X_{n,j} = \frac{\xi_j}{\sqrt{n}} + \frac{\alpha}{n}. \quad (12)$$

В терминах случайных блужданий случайные величины $X_{n,j}$, определенные соотношением (12), могут быть интерпретированы как элементарные приращения процесса, при этом рассматриваемая их конструкция (12) предполагает *одинаковый порядок малости* элементарных трендов и *дисперсий*, что характерно, например, для приращений винеровского процесса со сносом. Обозначим

$$S_n = \sum_{j=1}^n X_{n,j} \left(= \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j + \alpha \right).$$

В силу классической центральной предельной теоремы имеем:

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |P(S_n < x) - \Phi(x - \alpha)| = 0,$$

т. е. так определенные случайные величины $X_{n,j}$ удовлетворяют условию (6) с $k_n = n$.

В книге [31] и статьях [33, 34] предложено моделировать эволюцию неоднородных хаотических стохастических процессов, в частности динамику цен финансовых активов, с помощью обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса). Этот подход получил дополнительное обоснование и развитие в [3, 29, 35, 36]. В работах [36, 37] 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 $M(\Lambda(t))$, где $M(t)$, $t \geq 0$, — однородный пуассоновский процесс с единичной интенсивностью, а $\Lambda(t)$, $t \geq 0$, — независимый от $M(t)$ 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 $M(\Lambda(t))$, $t \geq 0$, 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 соответствии с такой моделью в каждый момент времени t распределение случайной величины $M(\Lambda(t))$ является смешанным пуассоновским. С практической точки зрения для описания статистических закономерностей поведения интенсивности потока информативных событий удобно использовать такую гибкую модель, как обобщенное гамма-распределение. Для большей наглядности рассмотрим случай, когда в рассматриваемой модели время t остается фиксированным, а $\Lambda(t) = nU_{\nu,\kappa,\delta}$, где n — вспомогательный параметр, $U_{\nu,\kappa,\delta}$ — случайная величина, имеющая обобщенное гамма-распределение $F(x; \nu, \kappa, \delta)$ и независимая от стандартного пуассоновского процесса $M(t)$, $t \geq 0$. При этом асимптотика $n \rightarrow \infty$ может интерпретироваться как то, что (случайная) интенсивность потока информативных событий считается очень большой, а при использовании подобных моделей в задачах финансовой математики распределение случайной величины $U_{\nu,\kappa,\delta}$ довольно естественно отождествляется со статистическими закономерностями поведения (случайной) волатильности. Для каждого натурального n положим

$$N_n = M(nU_{\nu,\kappa,\delta}).$$

Очевидно, что так определенная случайная величина N_n имеет смешанное пуассоновское распределение:

$$P(N_n = k) = P(M(nU_{\nu,\kappa,\delta}) = k) = \int_0^\infty e^{-nz} \frac{(nz)^k}{k!} f(z; \nu, \kappa, \delta) dz \quad k = 0, 1, \dots \quad (13)$$

Обозначим $A_n(z) \equiv A_n(z; \nu, \kappa, \delta) = P(N_n < nz)$, $z \geq 0$ ($A_n(z) = 0$ при $z < 0$). Несложно видеть, что

$$A_n(z) \implies F(x; \nu, \kappa, \delta) \quad (n \rightarrow \infty).$$

Действительно, как известно, если $\Pi(x; \ell)$ — функция распределения Пуассона с параметром $\ell > 0$ и $E(x; c)$ — функция распределения с единственным единичным скачком в точке $c \in \mathbb{R}$, то

$$\Pi(\ell x; \ell) \implies E(x; 1) \quad (\ell \rightarrow \infty).$$

Так как для $x \in \mathbb{R}$

$$A_n(x) = \int_0^\infty \Pi(nx; nz) dF(z; \nu, \kappa, \delta),$$

то по теореме Лебега о мажорируемой сходимости при $n \rightarrow \infty$

$$A_n(x) \Rightarrow \int_0^\infty E(xz^{-1}; 1) dF(z; \nu, \kappa, \delta) = \int_0^x dF(z; \nu, \kappa, \delta) = F(x; \nu, \kappa, \delta),$$

т.е. так определенные случайные величины N_n удовлетворяют условию (11) с $k_n = n$. Впредь будем считать, что при каждом $n \geq 1$ случайная величина N_n независима от последовательности $\{\xi_j\}_{j \geq 1}$, что гарантирует независимость случайных величин $N_n, X_{n,1}, X_{n,2}, \dots$.

Таким образом, в силу непрерывности функции обобщенного дисперсионного гамма-распределения $W(x; \alpha, \sigma, \nu, \kappa, \delta)$ из следствия 1 вытекает, что

$$D_n \equiv \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\sum_{j=1}^{N_n} X_{n,j} < x \right) - W(x; \alpha, \sigma, \nu, \kappa, \delta) \right| \rightarrow 0 \quad (n \rightarrow \infty)$$

(см. замечание 2).

Скорость стремления D_n к нулю описывается следующим утверждением.

Теорема 2. Для любого $n \geq 1$ справедлива оценка

$$D_n \leq 0,4532 \frac{\beta^3}{\sqrt{n}} EU_{\nu, \kappa, \delta}^{-1/2} + 0,1210 \frac{\alpha^2}{n}.$$

Доказательство. Как уже было показано, распределение случайной величины N_n является смешанным пуассоновским (см. (13)). Следовательно, по теореме Фубини

$$\mathbb{P} \left(\sum_{j=1}^{N_n} X_{n,j} < x \right) = \mathbb{P} \left(\sum_{j=1}^{M(nU_{\nu, \kappa, \delta})} X_{n,j} < x \right) = \int_0^\infty \mathbb{P} \left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) f(z; \nu, \kappa, \delta) dz.$$

При этом

$$EX_{n,j} = \frac{\alpha}{n}; \quad DX_{n,j} = \frac{1}{n}; \quad E|X_{n,j} - EX_{n,j}|^3 = \frac{\beta^3}{n^{3/2}}.$$

Таким образом, при каждом $z \in (0, \infty)$

$$E \sum_{j=1}^{M(nz)} X_{n,j} = \alpha z; \\ D \sum_{j=1}^{M(nz)} X_{n,j} = nz \left(\frac{\alpha^2}{n^2} + \frac{1}{n} \right) = z \left(1 + \frac{\alpha^2}{n} \right).$$

Из (1) вытекает, что

$$D_n = \sup_x \left| \int_0^\infty f(z; \nu, \kappa, \delta) \left[\mathbb{P} \left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) - \Phi \left(\frac{x - \alpha z}{\sqrt{z(1 + \alpha^2/n)}} \right) + \Phi \left(\frac{x - \alpha z}{\sqrt{z(1 + \alpha^2/n)}} \right) - \Phi \left(\frac{x - \alpha z}{\sqrt{z}} \right) \right] dz \right| \leq I_1 + I_2, \quad (14)$$

где

$$I_1 = \int_0^\infty f(z; \nu, \kappa, \delta) \sup_x \left| \mathbb{P} \left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) - \Phi \left(\frac{x - \alpha z}{\sqrt{z(1 + \alpha^2/n)}} \right) \right| dz; \\ I_2 = \int_0^\infty f(z; \nu, \kappa, \delta) \sup_x \left| \Phi \left(\frac{x - \alpha z}{\sqrt{z(1 + \alpha^2/n)}} \right) - \Phi \left(\frac{x - \alpha z}{\sqrt{z}} \right) \right| dz.$$

В дальнейшем понадобится следующее утверждение.

Лемма 3. Пусть случайные величины X_1, X_2, \dots одинаково распределены. Пусть N_ℓ — пуассоновская случайная величина с параметром $\ell > 0$. Предположим, что случайные величины N_ℓ, X_1, X_2, \dots независимы в совокупности. Обозначим

$$S_\ell = X_1 + \dots + X_{N_\ell}.$$

Тогда

$$\sup_x \left| \mathbb{P}(S_\ell < x) - \Phi \left(\frac{x - ES_\ell}{\sqrt{DS_\ell}} \right) \right| \leq \frac{0,4532}{\sqrt{\ell}} \frac{E|X_1 - EX_1|^3}{(DX_1)^{3/2}}.$$

Доказательство леммы 3 приведено в работе [38] (см. также [3, с. 144]).

Продолжим доказательство теоремы 2. Рассмотрим I_1 . Применяя лемму 3, получаем

$$I_1 \leq 0,4532 \frac{\beta^3}{\sqrt{n}} \int_0^\infty \frac{f(z; \nu, \kappa, \delta)}{\sqrt{z}} dz = 0,4532 \frac{\beta^3}{\sqrt{n}} EU_{\nu, \kappa, \delta}^{-1/2}. \quad (15)$$

Рассмотрим I_2 . Сформулируем еще одно вспомогательное утверждение.

Лемма 4. Пусть $b \in \mathbb{R}$, $0 < c < \infty$, $0 < d < \infty$. Тогда

$$\sup_y |\Phi(y) - \Phi(cy)| \leq \frac{1}{\sqrt{2\pi e}} \left| \max \left\{ c, \frac{1}{c} \right\} - 1 \right|; \quad (16)$$

$$\sqrt{1+d} - 1 \leq \frac{d}{2}. \quad (17)$$

Элементарное доказательство неравенств (16) и (17) можно получить, например, с помощью формулы Лагранжа.

Продолжим доказательство теоремы 2. В лемме 4 положим

$$y = \frac{x - \alpha z}{\sqrt{z(1 + \alpha^2/n)}}, \quad c = \sqrt{1 + \frac{\alpha^2}{n}}.$$

Тогда $c \geq 1$ и в силу утверждения (16) леммы 4 имеем:

$$I_2 \leq \frac{1}{\sqrt{2\pi e}} \left(\sqrt{1 + \frac{\alpha^2}{n}} - 1 \right).$$

При этом в силу утверждения (17) леммы 4

$$\sqrt{1 + \frac{\alpha^2}{n}} - 1 \leq \frac{\alpha^2}{2n}.$$

Окончательно получаем:

$$I_2 \leq \frac{\alpha^2}{2\sqrt{2\pi e}n}. \quad (18)$$

Подставляя (15) и (18) в (14), получаем утверждение теоремы. Теорема доказана.

Для симметричных обобщенных дисперсионных гамма-распределений (т. е. $\alpha = 0$) оценку, представленную в теореме 2, можно уточнить не только за счет того, что в таком случае обнуляется второе слагаемое в правой части, но и за счет уменьшения коэффициента при первом слагаемом. Аналогом леммы 3 здесь может служить следующее утверждение.

Лемма 5. В условиях леммы 3 справедливо неравенство:

$$\sup_x \left| P(S_\ell < x) - \Phi \left(\frac{x - ES_\ell}{\sqrt{DS_\ell}} \right) \right| \leq \frac{0,3041}{\sqrt{\ell}} \frac{E|X_1|^3}{(EX_1^2)^{3/2}}.$$

Доказательство этого утверждения приведено в [39] (см. также [3, теорема 2.4.3]).

Если в случае $\alpha = 0$ в доказательстве теоремы 2 вместо леммы 3 воспользоваться леммой 5, то в результате получится следующее утверждение.

Следствие 2. Пусть в дополнение к условиям теоремы 2 $\alpha = 0$. Тогда для любого $n \geq 1$ справедлива оценка

$$D_n \leq 0,3041 \frac{\beta^3}{\sqrt{n}} EU_{\nu, \kappa, \delta}^{-1/2}.$$

Литература

1. *Seshadri V.* Halphen's laws // Encyclopedia of Statistical Sciences, Update Vol. 1. / Eds. S. Kotz, C. B. Read, D. L. Banks. — New York: Wiley, 1997. P. 302–306.
2. *Sichel H. S.* Statistical evaluation of diamondiferous deposits // J. South Afr. Inst. Min. Metall., 1973. Vol. 76. P. 235–243.
3. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — 2-е изд., перераб. и дополн. — М.: Физматлит, 2011. 620 с.
4. *Barndorff-Nielsen O. E.* Exponentially decreasing distributions for the logarithm of particle size // Proc. R. Soc. L. Ser. A, 1977. Vol. A(353). P. 401–419.
5. *Barndorff-Nielsen O. E.* Hyperbolic distributions and distributions of hyperbolae // Scand. J. Statist., 1978. Vol. 5. P. 151–157.
6. *Barndorff-Nielsen O. E.* Models for non-Gaussian variation, with applications to turbulence // Proc. R. Soc. L. Ser. A, 1979. Vol. A(368). P. 501–520.
7. *Madan D. B., Seneta E.* The variance gamma (V.G.) model for share market return // J. Business, 1990. Vol. 63. P. 511–524.
8. *Eberlein E., Keller U.* Hyperbolic distributions in finance // Bernoulli, 1995. Vol. 1. No. 3. P. 281–299.
9. *Prause K.* Modeling financial data using generalized hyperbolic distributions. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1997. Preprint No. 48.
10. *Carr P. P., Madan D. B., Chang E. C.* The Variance Gamma process and option pricing // Eur. Finance Rev., 1998. Vol. 2. P. 79–105.
11. *Eberlein E., Keller U., Prause K.* New insights into smile, mispricing and value at risk: The hyperbolic model // J. Business, 1998. Vol. 71. P. 371–405.
12. *Barndorff-Nielsen O. E.* Processes of normal inverse Gaussian type // Finance and Stochastics, 1998. Vol. 2. P. 41–18.
13. *Eberlein E., Prause K.* The generalized hyperbolic model: Financial derivatives and risk measures. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1998. Preprint No. 56.
14. *Шуряев А. Н.* Основы стохастической финансовой математики. Т. 1. Факты. Модели. — М.: Фазис, 1998.
15. *Eberlein E.* Application of generalized hyperbolic Lévy motions to finance. — Freiburg: Universität Freiburg,

- Institut für Mathematische Stochastic, 1999. Preprint No. 64.
16. *Barndorff-Nielsen O. E., Blsild P., Schmiegel J.* A parsimonious and universal description of turbulent velocity increments // *Eur. Phys. J.*, 2004. Vol. B. 41. P. 345–363.
 17. *LeCam L.* Maximum likelihood; an introduction // *Intern. Stat. Rev.*, 1990. Vol. 58. P. 153–171.
 18. *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. — М.—Л.: ГИТТЛ, 1949.
 19. *Barndorff-Nielsen O. E., Kent J., Sørensen M.* Normal variance-mean mixtures and z -distributions // *Intern. Stat. Rev.*, 1982. Vol. 50. No. 2. P. 145–159.
 20. *Rényi A.* On the central limit theorem for the sum of a random number of independent random variables // *Acta Math. Acad. Sci. Hung.*, 1960. Vol. 11. P. 97–102.
 21. *Гнеденко Б. В., Фахим Х.* Об одной теореме переноса // *Докл. АН СССР*, 1969. Т. 187. № 1. С. 15–17.
 22. *Королев В. Ю., Соколов И. А.* Скошенные распределения Стьюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации // *Информатика и её применения*, 2012. Т. 6. Вып. 1. С. 2–10.
 23. *Королев В. Ю.* О взаимосвязи обобщенного распределения Стьюдента и дисперсионного гамма-распределения при статистическом анализе выборок случайного объема // *Докл. РАН*, 2012 (в печати).
 24. *Королев В. Ю.* Обобщенные гиперболические распределения как предельные для случайных сумм // *Теория вероятностей и ее применения*, 2013. Т. 58. Вып. 1.
 25. *Королев В. Ю., Шоргин С. Я.* Математические методы анализа стохастической структуры информационных потоков. — М.: ИПИ РАН, 2011. 130 с.
 26. *Stacy E. W.* A generalization of the gamma distribution // *Annals Math. Statistics*, 1962. Vol. 33. P. 1187–1192.
 27. *Гумбель Э.* Статистика экстремальных значений. — М.: Мир, 1965.
 28. *Kalashnikov V. V.* Geometric sums: Bounds for rare events with applications. — Dordrecht: Kluwer Academic Pubs., 1997.
 29. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008.
 30. *Королев В. Ю.* О распределении размеров частиц при дроблении // *Информатика и её применения*, 2009. Т. 3. Вып. 3. С. 60–68.
 31. *Gnedenko B. V., Korolev V. Yu.* Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996.
 32. *Teicher H.* Identifiability of mixtures // *Ann. Math. Stat.*, 1961. Vol. 32. P. 244–248.
 33. *Королев В. Ю.* Построение моделей распределений биржевых цен при помощи методов асимптотической теории случайного суммирования // *Обзорные промышленной и прикладной математики. Сер. Финансовая и страховая математика*, 1997. Т. 4. Вып. 1. С. 86–102.
 34. *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.
 35. *Bening V., Korolev V.* Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002.
 36. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011.
 37. *Stochastic models of structural plasma turbulence / Eds. V. Korolev, N. Skvortsova.* — Utrecht: VSP, 2006.
 38. *Королев В. Ю., Шевцова И. Г., Шоргин С. Я.* О неравенствах типа Берри–Эссеена для пуассоновских случайных сумм // *Информатика и её применения*, 2011. Т. 5. Вып. 3. С. 64–66.
 39. *Korolev V., Shevtsova I.* An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // *Scandinavian Actuarial J.*, 2012. No. 2. P. 81–105. DOI:10.1080/03461238.2010.485370.

ОБ ОЦЕНКАХ ФУНКЦИЙ КОНЦЕНТРАЦИИ РЕГУЛЯРНЫХ СТАТИСТИК, ПОСТРОЕННЫХ ПО ВЫБОРКАМ СЛУЧАЙНОГО ОБЪЕМА*

В. Е. Бенинг¹, Н. К. Галиева², В. Ю. Королев³

Аннотация: Приведены оценки функций концентрации (ф.к.) регулярных статистик, построенных по выборкам случайного объема.

Ключевые слова: функция концентрации; случайная сумма; асимптотически нормальная статистика; распределение Стьюдента; распределение Лапласа

1 Введение

Асимптотическим свойствам распределений сумм случайного числа случайных величин (с.в.) посвящено много работ (см., например, [1–6]). Такого рода суммы находят широкое применение в страховании, экономике, биологии и т. п. [3, 5, 6]. Однако 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. При этом в статистике часто возникают ситуации, когда размер выборки не является заранее определенным и может рассматриваться как случайный. Например, при испытании на надежность число отказавших приборов за определенное время является случайной величиной. Вообще, в подавляющем большинстве ситуаций, связанных с анализом и обработкой экспериментальных данных, можно считать, что число случайных факторов, влияющих на наблюдаемые величины, само является случайным и изменяется от наблюдения к наблюдению. Поэтому вместо различных вариантов центральной предельной теоремы, обосновывающих нормальность предельного распределения в классической статистике, в таких ситуациях следует опираться на их аналоги для выборок случайного объема. Это делает естественным изучение асимптотического поведения распределений статистик достаточно общего вида, основанных на выборках случайного объема. Примерами

могут служить работы [7, 8], в которых рассматривались асимптотические свойства распределений выборочных квантилей, построенных по выборкам случайного объема.

В данной работе получены оценки для функций концентрации статистик, построенных по выборкам случайного объема. Эти оценки непосредственно зависят от скорости сходимости функций распределения таких статистик к предельному закону.

В работе приняты следующие обозначения: \mathbb{R} — множество вещественных чисел; \mathbb{N} — множество натуральных чисел; $\Phi(x)$ и $\varphi(x)$ — соответственно функция распределения и плотность стандартного нормального закона.

В разд. 2 рассмотрены случаи предельных распределений Стьюдента и Лапласа, получены аппроксимации для функций концентрации в этих случаях.

Как известно, ф.к. с.в. Z называется функция (см., например, [9, с. 53])

$$Q_Z(\lambda) = \sup_{x \in \mathbb{R}} P(x \leq Z \leq x + \lambda), \quad \lambda \geq 0. \quad (1)$$

Очевидно, ф.к. $Q_Z(\lambda)$ — неубывающая функция λ , удовлетворяющая неравенству

$$0 \leq Q_Z(\lambda) \leq 1$$

для любого $\lambda \geq 0$. Из ее определения следует оценка

$$\sup_{x \in \mathbb{R}} P(Z = x) = Q_Z(0) \leq Q_Z(\lambda), \quad \lambda \geq 0,$$

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м и 12-07-00115а), Министерством образования и науки (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@cs.msu.su

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.su

позволяющая оценить максимальную вероятность отдельного значения с.в. Z . Из определения (1) ф.к. следует, что для любого числа $a \in \mathbb{R}$ и любого $b > 0$ справедливы тождества:

$$Q_{Z+a}(\lambda) \equiv Q_Z(\lambda); \quad Q_{bZ}(\lambda) \equiv Q_Z\left(\frac{\lambda}{b}\right).$$

Применениям функций концентрации к проблемам слабой сходимости посвящены гл. 3 и 4 книги [10].

Лемма 1.1 Пусть ξ и η — две с.в., тогда

$$\sup_{\lambda \geq 0} |Q_\xi(\lambda) - Q_\eta(\lambda)| \leq 4 \sup_{x \in \mathbb{R}} |P(\xi < x) - P(\eta < x)|.$$

Доказательство. Обозначим

$$\delta = \sup_{x \in \mathbb{R}} |P(\xi < x) - P(\eta < x)|.$$

Тогда для любого $\lambda \geq 0$ имеем:

$$\begin{aligned} Q_\xi(\lambda) &= \\ &= \sup_{x \in \mathbb{R}} [P(\xi = x + \lambda) + P(\xi < x + \lambda) - P(\xi < x)] = \\ &= \sup_{x \in \mathbb{R}} [(P(\eta = x + \lambda) + P(\eta < x + \lambda) - P(\eta < x)) + \\ &+ (P(\xi = x + \lambda) - P(\eta = x + \lambda)) + (P(\xi < x + \lambda) - \\ &- P(\eta < x + \lambda)) + (P(\eta < x) - P(\xi < x))] \leq \\ &\leq \sup_{x \in \mathbb{R}} |P(\eta = x + \lambda) + P(\eta < x + \lambda) - P(\eta < x)| + \\ &+ \sup_{x \in \mathbb{R}} |P(\xi = x + \lambda) - P(\eta = x + \lambda)| + \\ &+ \sup_{x \in \mathbb{R}} |P(\xi < x + \lambda) - P(\eta < x + \lambda)| + \\ &+ \sup_{x \in \mathbb{R}} |P(\eta < x) - P(\xi < x)| \leq \\ &\leq Q_\eta(\lambda) + 2\delta + \sup_{y \in \mathbb{R}} |P(\xi = y) - P(\eta = y)|. \quad (2) \end{aligned}$$

Далее, для любого $y \in \mathbb{R}$ имеем:

$$\begin{aligned} |P(\xi = y) - P(\eta = y)| &= |(P(\xi \leq y) - \\ &- P(\xi < y)) - (P(\eta \leq y) - P(\eta < y))| \leq \\ &\leq |P(\xi \leq y) - P(\eta \leq y)| + |P(\xi < y) - P(\eta < y)| = \\ &= \left| \lim_{\epsilon \downarrow 0} P(\xi < y + \epsilon) - \lim_{\epsilon \downarrow 0} P(\eta < y + \epsilon) \right| + \\ &+ |P(\xi < y) - P(\eta < y)| = \\ &= \left| \lim_{\epsilon \downarrow 0} [P(\xi < y + \epsilon) - P(\eta < y + \epsilon)] \right| + \\ &+ |P(\xi < y) - P(\eta < y)| \leq \\ &\leq \lim_{\epsilon \downarrow 0} |P(\xi < y + \epsilon) - P(\eta < y + \epsilon)| + \delta \leq \\ &\leq \limsup_{\epsilon \downarrow 0} \sup_y |P(\xi < y + \epsilon) - P(\eta < y + \epsilon)| + \delta \leq 2\delta. \end{aligned}$$

Поэтому

$$\sup_{y \in \mathbb{R}} |P(\xi = y) - P(\eta = y)| \leq 2\delta. \quad (3)$$

Более того, пример двух случайных величин ξ и η таких, что $P(\xi = 0) = P(\xi = 1) = 1/2$ и $P(\eta = 1/2) = 1$ показывает, что оценка (3) неуплучшаема.

Используя соотношения (3) и (2), получим:

$$Q_\xi(\lambda) \leq Q_\eta(\lambda) + 4\delta.$$

Точно так же убеждаемся, что справедливо неравенство

$$Q_\eta(\lambda) \leq Q_\xi(\lambda) + 4\delta.$$

Лемма доказана.

Замечание 1.1. Если функцию концентрации определить как

$$\tilde{Q}_\xi(\lambda) = \sup_{x \in \mathbb{R}} P(x \leq \xi < x + \lambda), \quad \lambda > 0,$$

то, как несложно убедиться, аналог неравенства, устанавливаемого леммой 1.1, будет иметь вид:

$$\sup_{\lambda > 0} |\tilde{Q}_\xi(\lambda) - \tilde{Q}_\eta(\lambda)| \leq 2 \sup_{x \in \mathbb{R}} |P(\xi < x) - P(\eta < x)|,$$

т. е. коэффициент в правой части будет в два раза меньше.

Напомним определение унимодальности распределения вероятности по Хинчину (см., например, [11, с. 186] или [10, с. 12], [12, с. 118]). Согласно этому определению, с.в. ξ имеет унимодальное (одновершинное) распределение, если существует точка x_0 такая, что функция распределения (ф.р.) $F_\xi(x)$ случайной величины ξ выпукла при $x < x_0$, а функция $1 - F_\xi(x)$ выпукла при $x > x_0$. При этом точка x_0 называется модой случайной величины ξ . Точка x_0 может быть точкой разрыва ф.р. $F_\xi(x)$, но вне точки x_0 одновершинность предполагает существование у ф.р. $F_\xi(x)$ плотности, которая монотонна в интервалах $x < x_0$ и $x > x_0$. Несложно убедиться, что любая унимодальная функция распределения непрерывна всюду, быть может, за исключением моды.

Лемма 1.2. Пусть ξ — с.в. с симметричным унимодальным распределением. Тогда для $\lambda > 0$

$$Q_\xi(\lambda) = P\left(|\xi| < \frac{\lambda}{2}\right).$$

Доказательство. В условиях леммы достаточно показать, что

$$\arg \max_{x \in \mathbb{R}} P(x \leq \xi \leq x + \lambda) = -\frac{\lambda}{2}.$$

Пусть $F_\xi(x)$ — ф.р. с.в. ξ . В силу симметричности распределения с.в. ξ ее мода равна нулю. Пусть $0 < a < \lambda/2$. Тогда точки $\pm\lambda/2$, $-\lambda/2 + a$ и $\lambda/2 + a$ являются точками непрерывности функции распределения $F_\xi(x)$. Поэтому имеем:

$$\begin{aligned} & \mathbb{P}\left(-\frac{\lambda}{2} \leq \xi \leq \frac{\lambda}{2}\right) - \mathbb{P}\left(-\frac{\lambda}{2} + a \leq \xi \leq \frac{\lambda}{2} + a\right) = \\ & = F_\xi\left(\frac{\lambda}{2}\right) - F_\xi\left(-\frac{\lambda}{2}\right) - F_\xi\left(\frac{\lambda}{2} + a\right) + \\ & + F_\xi\left(-\frac{\lambda}{2} + a\right) = F_\xi\left(-\frac{\lambda}{2} + a\right) - F_\xi\left(-\frac{\lambda}{2}\right) + \\ & + F_\xi\left(\frac{\lambda}{2}\right) - F_\xi\left(\frac{\lambda}{2} + a\right) = 1 - F_\xi\left(\frac{\lambda}{2} - a\right) - 1 + \\ & + F_\xi\left(\frac{\lambda}{2}\right) + F_\xi\left(\frac{\lambda}{2}\right) - F_\xi\left(\frac{\lambda}{2} + a\right) = \left[F_\xi\left(\frac{\lambda}{2}\right) - \right. \\ & \left. - F_\xi\left(\frac{\lambda}{2} - a\right)\right] - \left[F_\xi\left(\frac{\lambda}{2} + a\right) - F_\xi\left(\frac{\lambda}{2}\right)\right] = \\ & = 2F_\xi\left(\frac{\lambda}{2}\right) - \left[F_\xi\left(\frac{\lambda}{2} - a\right) + F_\xi\left(\frac{\lambda}{2} + a\right)\right]. \quad (4) \end{aligned}$$

Здесь третье равенство имеет место в силу симметричности распределения с.в. ξ . Поскольку распределение с.в. ξ симметрично и унимодально, ф.р. $F_\xi(z)$ вогнута при $z > 0$, т. е. для любых $0 < x < y < \infty$ и $\alpha \in (0, 1)$

$$F_\xi(\alpha x + (1 - \alpha)y) \geq \alpha F_\xi(x) + (1 - \alpha)F_\xi(y). \quad (5)$$

Положим в неравенстве (5) $x = \lambda/2 - a$, $y = \lambda/2 + a$, $\alpha = 1/2$. Получим:

$$2F_\xi\left(\frac{\lambda}{2}\right) \geq F_\xi\left(\frac{\lambda}{2} - a\right) + F_\xi\left(\frac{\lambda}{2} + a\right).$$

Продолжая (4), можно заметить, что для рассматриваемых значений a справедливо неравенство:

$$\begin{aligned} & \mathbb{P}\left(-\frac{\lambda}{2} \leq \xi \leq \frac{\lambda}{2}\right) - \mathbb{P}\left(-\frac{\lambda}{2} + a \leq \xi \leq \frac{\lambda}{2} + a\right) \geq \\ & \geq 0. \quad (6) \end{aligned}$$

В силу симметричности распределения случай $-\lambda/2 < a < 0$ рассматривается аналогично. В случае $a > \lambda/2$ неравенство (6) имеет место в силу вогнутости $F_\xi(x)$ при $x > 0$, а в случае $a < -\lambda/2$ неравенство (6) выполнено в силу выпуклости $F_\xi(x)$ при $x < 0$. Лемма доказана.

Замечание 1.2. Заметим, что если с.в. ξ имеет непрерывное симметричное унимодальное распределение, то $Q_\xi(0) = 0$.

Пусть X_1, X_2, \dots — независимые одинаково распределенные с.в. с $EX_1 = \mu$, $0 < DX_1 = \sigma^2$ и $E|X_1 - \mu|^3 = \beta^3 < \infty$. Для натурального n обозначим

$$S_n = X_1 + \dots + X_n.$$

Теорема 1.1. Для любого $n \in \mathbb{N}$ имеет место неравенство

$$\sup_{\lambda \geq 0} \left| Q_{S_n}(\lambda) - 2\Phi\left(\frac{\lambda}{2\sigma\sqrt{n}}\right) + 1 \right| \leq 1,8992 \frac{\beta^3}{\sigma^3\sqrt{n}}.$$

Доказательство. В лемме 1.1 положим $\mathbb{P}(\xi < x) = \Phi(x)$,

$$\eta = \eta_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

так что

$$\begin{aligned} Q_{S_n}(\lambda) &= \sup_{x \in \mathbb{R}} \mathbb{P}(x \leq S_n \leq x + \lambda) = \\ &= \sup_{x \in \mathbb{R}} \mathbb{P}\left(\frac{x - n\mu}{\sigma\sqrt{n}} \leq \eta_n \leq \frac{x - n\mu + \lambda}{\sigma\sqrt{n}}\right) = \\ &= \sup_{y \in \mathbb{R}} \mathbb{P}\left(y \leq \eta_n \leq y + \frac{\lambda}{\sigma\sqrt{n}}\right) = Q_{\eta_n}\left(\frac{\lambda}{\sigma\sqrt{n}}\right). \end{aligned}$$

При этом в силу неравенства Берри–Эссеена (см. [13])

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\eta_n < x) - \Phi(x)| \leq 0,4748 \frac{\beta^3}{\sigma^3\sqrt{n}}.$$

Теперь утверждение теоремы следует из лемм 1.1 и 1.2 в силу симметричности и унимодальности нормального распределения. Теорема доказана.

Следствие 1.1. Для любых $\lambda \geq 0$ и $n \in \mathbb{N}$ имеют место неравенства

$$\begin{aligned} 2\Phi\left(\frac{\lambda}{2\sigma\sqrt{n}}\right) - 1 - 1,8992 \frac{\beta^3}{\sigma^3\sqrt{n}} &\leq Q_{S_n}(\lambda) \leq \\ &\leq 2\Phi\left(\frac{\lambda}{2\sigma\sqrt{n}}\right) - 1 + 1,8992 \frac{\beta^3}{\sigma^3\sqrt{n}}. \end{aligned}$$

Будем говорить, что статистика T_n (т. е. измеримая функция от наблюдений X_1, \dots, X_n) асимптотически нормальна, если существуют $\delta > 0$, $\nu \in \mathbb{R}$ и $\mu \in \mathbb{R}$ такие, что для любого $x \in \mathbb{R}$ справедливо соотношение:

$$\mathbb{P}(\delta n^\nu (T_n - \mu) < x) \longrightarrow \Phi(x), \quad n \rightarrow \infty. \quad (7)$$

Предположим, что известна оценка скорости сходимости в (7) вида

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\delta n^\nu (T_n - \mu) < x) - \Phi(x)| \leq \frac{C}{n^\gamma}, \quad (8)$$

где $C > 0$, $\gamma > 0$.

Теорема 1.2. *Предположим, что статистика T_n удовлетворяет соотношению (8). Тогда для любого $n \in \mathbb{N}$ имеет место неравенство:*

$$\sup_{\lambda \geq 0} \left| Q_{T_n}(\lambda) - 2\Phi\left(\frac{\lambda\delta n^\nu}{2}\right) + 1 \right| \leq \frac{4C}{n^\gamma}.$$

Доказательство. Положим $P(\xi < x) = \Phi(x)$, $\eta = \eta_n = \delta n^\nu(T_n - \theta)$. Тогда

$$\begin{aligned} Q_{T_n}(\lambda) &= \sup_{x \in \mathbb{R}} P(x \leq T_n \leq x + \lambda) = \\ &= \sup_{x \in \mathbb{R}} P(\delta n^\nu(x - \mu) \leq \eta_n \leq \delta n^\nu(x + \lambda - \mu)) = \\ &= \sup_{y \in \mathbb{R}} P(y \leq \eta_n \leq y + \lambda\delta n^\nu) = Q_{\eta_n}(\lambda\delta n^\nu). \end{aligned}$$

Теперь утверждение теоремы вытекает из лемм 1.1 и 1.2. Теорема доказана.

Следствие 1.2. *Для любых $\lambda \geq 0$ и $n \in \mathbb{N}$ имеют место неравенства:*

$$\begin{aligned} 2\Phi\left(\frac{\lambda\delta n^\nu}{2}\right) - 1 - \frac{4C}{n^\gamma} &\leq Q_{T_n}(\lambda) \leq \\ &\leq 2\Phi\left(\frac{\lambda\delta n^\nu}{2}\right) - 1 + \frac{4C}{n^\gamma}. \end{aligned}$$

2 Статистики, построенные по выборкам случайного объема

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 $EN_n = n$ и, значит, 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 одинаково распределенными и имеющими ф.р. $F(x)$. Обозначим через $T_n = T_n(X_1, \dots, X_n)$ 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

$$T_{N_n}(\omega) \equiv T_{N_n(\omega)}(X_1(\omega), \dots, X_{N_n(\omega)}(\omega)), \quad \omega \in \Omega.$$

Таким образом, T_{N_n} — это статистика, построенная на основе статистики T_n по выборке случайного объема N_n .

Справедливо следующее утверждение.

Теорема 2.1. *Предположим, что для некоторых $\mu \in \mathbb{R}$, $C > 0$, $\sigma > 0$, $\nu \in \mathbb{R}$ и симметричной унимодальной ф.р. $G(x)$ статистика T_{N_n} удовлетворяет соотношению:*

$$\sup_{x \in \mathbb{R}} |P(\sigma n^\nu(T_{N_n} - \mu) < x) - G(x)| \leq \frac{C}{n^\gamma}.$$

Тогда для любого $n \in \mathbb{N}$ имеет место неравенство:

$$\sup_{\lambda \geq 0} \left| Q_{T_{N_n}}(\lambda) - 2G\left(\frac{\lambda\sigma n^\nu}{2}\right) + 1 \right| \leq \frac{4C}{n^\gamma}.$$

Доказательство. Непосредственно следует из доказательства теоремы 1.2.

Следствие 2.1. *Для любых $\lambda \geq 0$ и $n \in \mathbb{N}$ имеют место неравенства*

$$\begin{aligned} 2G\left(\frac{\lambda\sigma n^\nu}{2}\right) - 1 - \frac{4C}{n^\gamma} &\leq Q_{T_{N_n}}(\lambda) \leq \\ &\leq 2G\left(\frac{\lambda\sigma n^\nu}{2}\right) - 1 + \frac{4C}{n^\gamma}. \end{aligned}$$

2.1 Распределение Стьюдента

В работе [14] показано, что если случайный объем выборки N_n имеет отрицательное биномиальное распределение с параметрами $p = 1/n$ и $r > 0$, т.е. (при $r = 1$ имеем геометрическое распределение)

$$P(N_n = k) = \frac{(k+r-2) \cdots r}{(k-1)!} \frac{1}{n^r} \left(1 - \frac{1}{n}\right)^{k-1}, \quad k \in \mathbb{N},$$

то для асимптотически нормальной статистики T_n справедливо предельное соотношение [14, следствие 2.1]:

$$P(\sigma\sqrt{n}(T_{N_n} - \mu) < x) \longrightarrow G_{2r}(x\sqrt{r}), \quad n \rightarrow \infty,$$

где $G_{2r}(x)$ — функция распределения Стьюдента с параметром $\gamma = 2r$, т.е. имеющего плотность вида

$$p_\gamma(x) = \frac{\Gamma(\gamma+1/2)}{\sqrt{\pi}\Gamma(\gamma/2)} \left(1 + \frac{x^2}{\gamma}\right)^{-(\gamma+1)/2}, \quad x \in \mathbb{R},$$

где $\Gamma(\cdot)$ — эйлерова гамма-функция, а $\gamma > 0$ — параметр формы (если параметр γ натурален, то он называется числом степеней свободы). В рассматриваемой ситуации он может быть произвольно

мал, что соответствует типичному тяжелохвостому распределению. Отметим, что распределение Стьюдента является симметричным унимодальным непрерывным распределением. Если $\gamma = 2$, 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 $\gamma = 1$ ($r = 1/2$) имеем распределение Коши: $G_1(x) = 1/2 + (1/\pi) \operatorname{arctg} x$.

Если для ф.р. статистики T_n справедлива оценка скорости сходимости вида

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\sigma\sqrt{n}(T_n - \mu) < x) - \Phi(x)| \leq \frac{C_0}{\sqrt{n}},$$

$$C_0 > 0, \quad n \in \mathbb{N}, \quad (9)$$

где C_0 — величина, не зависящая от n , то, как показано в работе [15] (также см. [16, теорема 6.11]), при $r \in (0, 1/2)$ справедлива оценка скорости сходимости статистики T_{N_n} вида

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\sigma\sqrt{n}(T_{N_n} - \mu) < x) - G_{2r}(x\sqrt{r})| \leq \frac{C_1}{n^r},$$

$$n \in \mathbb{N}.$$

Для случая $r = 1/2$ неравенство имеет вид:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\sigma\sqrt{n}(T_{N_n} - \mu) < x) - \frac{1}{\pi} \operatorname{arctg}(x\sqrt{r}) - \frac{1}{2} \right| \leq C_2 \frac{\log n}{\sqrt{n}}, \quad n > 1.$$

Если же $r > 1/2$, то

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\sigma\sqrt{n}(T_{N_n} - \mu) < x) - G_{2r}(x\sqrt{r})| \leq \frac{C_3}{\sqrt{n}},$$

$$n \in \mathbb{N}.$$

В частности, если $r = 1$, т. е. если N_n имеет геометрическое распределение с параметром $1/n$, то

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\sigma\sqrt{n}(T_{N_n} - \mu) < x) - \frac{x}{2\sqrt{2+x^2}} - \frac{1}{2} \right| \leq \frac{C_3}{\sqrt{n}}, \quad n \in \mathbb{N}.$$

Здесь C_i , $i = 1, 2, 3$, — величины, не зависящие от n , но, возможно, зависящие от r и других параметров задачи. В работе [17] на примере статистики $T_n = (1/n)(X_1 + \dots + X_n)$ показано, что порядки убывания правых частей в неравенствах, приведенных выше, неумлучшаемы.

Учитывая эти неравенства и теорему 2.1, непосредственно получаем следующее утверждение.

Теорема 2.2. *Предположим, что для некоторых $\mu \in \mathbb{R}$, $C_0 > 0$ и $\sigma > 0$ ф.р. статистики T_n удовлетворяет соотношению (9). Тогда при $r \in (0, 1/2)$ и любом $n \in \mathbb{N}$ имеет место неравенство:*

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - 2G_{2r}\left(\frac{\lambda\sigma\sqrt{rn}}{2}\right) + 1 \right| \leq \frac{4C_1}{n^r}.$$

Если $r = 1/2$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - \frac{2}{\pi} \operatorname{arctg}\left(\frac{\lambda\sigma\sqrt{n}}{2\sqrt{2}}\right) \right| \leq 4C_2 \frac{\log n}{\sqrt{n}},$$

$$n > 1.$$

Если $r > 1/2$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - 2G_{2r}\left(\frac{\lambda\sigma\sqrt{rn}}{2}\right) + 1 \right| \leq \frac{4C_3}{\sqrt{n}}.$$

В частности, если $r = 1$, т. е. если N_n имеет геометрическое распределение с параметром $1/n$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - \frac{\lambda\sigma\sqrt{n}}{\sqrt{8 + \lambda^2\sigma^2n}} \right| \leq \frac{4C_3}{\sqrt{n}}.$$

2.2 Примеры

Пусть X_1, X_2, \dots — независимые одинаково распределенные случайные величины с $EX_1 = \mu$ и $0 < DX_1 = \sigma^2$. Для натурального n обозначим

$$T_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Тогда, используя неравенство (5) и теорему 2.2, при $r \in (0, 1/2)$ имеем:

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - 2G_{2r}\left(\frac{\lambda\sigma\sqrt{rn}}{2}\right) + 1 \right| \leq \frac{4C_1}{n^r}, \quad n \in \mathbb{N}.$$

Если $r = 1/2$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - \frac{2}{\pi} \operatorname{arctg}\left(\frac{\lambda\sigma\sqrt{n}}{2\sqrt{2}}\right) \right| \leq 4C_2 \frac{\log n}{\sqrt{n}},$$

$$n > 1.$$

Если $r > 1/2$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - 2G_{2r}\left(\frac{\lambda\sigma\sqrt{rn}}{2}\right) + 1 \right| \leq \frac{4C_3}{\sqrt{n}}, \quad n \in \mathbb{N}.$$

В частности, если $r = 1$, т. е. если N_n имеет геометрическое распределение с параметром $1/n$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - \frac{\lambda \sigma \sqrt{n}}{\sqrt{8 + \lambda^2 \sigma^2 n}} \right| \leq \frac{4C_3}{\sqrt{n}}, \quad n > 1.$$

Рассмотрим теперь U -статистики. Пусть X_1, X_2, \dots — независимые одинаково распределенные наблюдения и $h(x_1, x_2)$ — симметричная измеримая функция двух переменных такая, что

$$\mathbb{E}h(X_1, X_2) = 0, \quad \mathbb{E}|h(X_1, X_2)|^p < \infty, \quad p > \frac{5}{3}.$$

Определим U -статистику:

$$U_n = \sum_{1 \leq i < j \leq n} h(X_i, X_j).$$

Предположим также, что

$$\tau^2 \equiv \mathbb{E}g^2(X_1) > 0, \quad \mathbb{E}|g(X_1)|^3 < \infty,$$

где

$$g(x) = \mathbb{E}(h(X_1, X_2) | X_1 = x).$$

Тогда из теоремы 2.2 и теоремы 2.1 работы [18] при $r \in (0, 1/2)$ получаем неравенство:

$$\begin{aligned} \sup_{\lambda \geq 0} \left| \tilde{Q}_{U_{N_n}}(\lambda) - 2G_{2r} \left(\frac{\lambda \sqrt{r}}{2(n-1)\sqrt{n}\tau} \right) + 1 \right| &\leq \\ &\leq \frac{4C_1}{n^r}, \quad n \in \mathbb{N}. \end{aligned}$$

Если $r = 1/2$, то

$$\begin{aligned} \sup_{\lambda \geq 0} \left| \tilde{Q}_{U_{N_n}}(\lambda) - \frac{2}{\pi} \operatorname{arctg} \left(\frac{\lambda}{2\sqrt{2n}(n-1)\tau} \right) \right| &\leq \\ &\leq 4C_2 \frac{\log n}{\sqrt{n}}, \quad n > 1. \end{aligned}$$

Если $r > 1/2$, то

$$\begin{aligned} \sup_{\lambda \geq 0} \left| \tilde{Q}_{U_{N_n}}(\lambda) - 2G_{2r} \left(\frac{\lambda \sqrt{r}}{2(n-1)\sqrt{n}\tau} \right) + 1 \right| &\leq \\ &\leq \frac{4C_3}{\sqrt{n}}, \quad n \in \mathbb{N}. \end{aligned}$$

В частности, если $r = 1$, т. е. если N_n имеет геометрическое распределение с параметром $1/n$, то

$$\begin{aligned} \sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - \frac{\lambda}{\sqrt{\lambda^2 + 8(n-1)^2 n \tau^2}} \right| &\leq \frac{4C_3}{\sqrt{n}}, \\ &n > 1. \end{aligned}$$

Рассмотрим теперь L -статистики. Пусть X_1, X_2, \dots — независимые одинаково распределенные наблюдения с ф.р. $F(x)$ и пусть $J(s)$ — измеримая функция, заданная на интервале $(0, 1)$ и удовлетворяющая условию Липшица. Рассмотрим числа вида

$$c_{in} = J \left(\frac{i}{n+1} \right), \quad i = 1, \dots, n,$$

и определим L -статистику

$$L_n = \frac{1}{n} \sum_{i=1}^n c_{in} X_{(i:n)},$$

где $X_{(1:n)} \leq \dots \leq X_{(n:n)}$ — вариационный ряд, построенный по выборке X_1, \dots, X_n . Предположим также, что

$$\mathbb{E}|X_1|^3 < \infty;$$

$$\begin{aligned} \sigma_J^2 \equiv \int_{-\infty}^{\infty} J(F(x))J(F(y))(\min(F(x), F(y)) - \\ - F(x)F(y)) dx dy > 0 \end{aligned}$$

Тогда из теоремы 2.2 и теоремы 3.1.2 работы [19] при $r \in (0, 1/2)$ получаем неравенство:

$$\begin{aligned} \sup_{\lambda \geq 0} \left| \tilde{Q}_{L_{N_n}}(\lambda) - 2G_{2r} \left(\frac{\lambda \sqrt{rn}}{2\sigma_J} \right) + 1 \right| &\leq \frac{4C_1}{n^{r/2}}, \\ &n \in \mathbb{N}. \end{aligned}$$

Если $r = 1/2$, то

$$\begin{aligned} \sup_{\lambda \geq 0} \left| \tilde{Q}_{L_{N_n}}(\lambda) - \frac{2}{\pi} \operatorname{arctg} \left(\frac{\lambda \sqrt{n}}{2\sqrt{2}\sigma_J} \right) \right| &\leq 4C_2 \frac{\log n}{\sqrt{n}}, \\ &n > 1. \end{aligned}$$

Если $r > 1/2$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{L_{N_n}}(\lambda) - 2G_{2r} \left(\frac{\lambda \sqrt{rn}}{2\sigma_J} \right) + 1 \right| \leq \frac{4C_3}{\sqrt{n}}, \quad n \in \mathbb{N}.$$

В частности, если $r = 1$, т. е. если N_n имеет геометрическое распределение с параметром $1/n$, то

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n}}(\lambda) - \frac{\lambda \sqrt{n}}{\sqrt{\lambda^2 n + 8\sigma_J^2}} \right| \leq \frac{4C_3}{\sqrt{n}}, \quad n \geq 1.$$

2.3 Распределение Лапласа

Рассмотрим распределение Лапласа с ф.р. $\Lambda_\gamma(x)$ и плотностью

$$\lambda_s(x) = \frac{1}{s\sqrt{2}} \exp\left\{-\frac{\sqrt{2}|x|}{s}\right\}, \quad s > 0, \quad x \in \mathbb{R}.$$

В работе [20] была построена последовательность с.в. $N_n(m)$, зависящая от параметра $m \in \mathbb{N}$ следующего вида. Пусть Y_1, Y_2, \dots — независимые одинаково распределенные с.в., имеющие непрерывную ф.р. Определим с.в.

$$N(m) = \min\left\{i \geq 1: \max_{1 \leq qj \leq qm} Y_j < \max_{m+1 \leq qk \leq qm+i} Y_k\right\}.$$

Хорошо известно, что так определенные с.в. имеют распределение вида

$$P(N(m) \geq k) = \frac{m}{m+k-1}, \quad k \geq 1. \quad (10)$$

Пусть теперь $N^{(1)}(m), N^{(2)}(m), \dots$ — независимые одинаково распределенные с.в., имеющие распределение (10). Определим с.в.

$$N_n(m) = \max_{1 \leq qj \leq qn} N^{(j)}(m).$$

Как показано в работе [20],

$$\lim_{n \rightarrow \infty} P\left(\frac{N_n(m)}{n} < x\right) = e^{-m/x}, \quad x > 0,$$

и для асимптотически нормальной статистики T_n справедливо соотношение:

$$P(\sigma\sqrt{n}(T_{N_n(m)} - \mu) < x) \rightarrow \Lambda_{1/m}(x), \quad n \rightarrow \infty,$$

где $\Lambda_{1/m}(x)$ — функция распределения Лапласа с параметром $s = 1/m$. Отметим, что распределение Лапласа является симметричным унимодальным непрерывным распределением. Если для ф.р. статистики T_n справедлива оценка скорости сходимости типа (9), то в работе [21] получена оценка скорости сходимости ф.р. статистики $T_{N_n(m)}$

$$\sup_{x \in \mathbb{R}} \left| P(\sigma\sqrt{n}(T_{N_n(m)} - \mu) < x) - \Lambda_{1/m}(x) \right| \leq \frac{C_4}{\sqrt{n}}, \quad n \in \mathbb{N}.$$

Учитывая это неравенство и теорему 2.1, непосредственно получаем следующее утверждение.

Теорема 2.3. *Предположим, что для некоторых $\mu \in \mathbb{R}$, $C_0 > 0$ и $\sigma > 0$ ф.р. статистики T_n удовлетворяет соотношению (9). Тогда при любом $n \in \mathbb{N}$ имеет место неравенство:*

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n(m)}}(\lambda) - 2\Lambda_{1/m}\left(\frac{\lambda\sigma\sqrt{n}}{2}\right) + 1 \right| \leq \frac{4C_4}{\sqrt{n}}.$$

2.4 Примеры

Пусть X_1, X_2, \dots — независимые одинаково распределенные случайные величины с $EX_1 = \mu$ и $0 < DX_1 = \sigma^2$. Для натурального n обозначим

$$T_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Тогда, используя неравенство (5) и теорему 2.3, имеем оценку:

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{T_{N_n(m)}}(\lambda) - 2\Lambda_{1/m}\left(\frac{\lambda\sigma\sqrt{n}}{2}\right) + 1 \right| \leq \frac{4C_4}{\sqrt{n}}.$$

Рассмотрим теперь U -статистики. Пусть X_1, X_2, \dots — независимые одинаково распределенные наблюдения и $h(x_1, x_2)$ — симметричная измеримая функция двух переменных такая, что

$$Eh(X_1, X_2) = 0, \quad E|h(X_1, X_2)|^p < \infty, \quad p > \frac{5}{3}.$$

Определим U -статистику

$$U_n = \sum_{1 \leq i < j \leq n} h(X_i, X_j).$$

Предположим также, что

$$\tau^2 \equiv E g^2(X_1) > 0, \quad E|g(X_1)|^3 < \infty,$$

где

$$g(x) = E(h(X_1, X_2)|X_1 = x).$$

Тогда из теоремы 2.3 и теоремы 2.1 работы [18] получаем неравенство

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{U_{N_n(m)}}(\lambda) - 2\Lambda_{1/m}\left(\frac{\lambda}{2\tau(n-1)\sqrt{n}}\right) + 1 \right| \leq \frac{4C_4}{\sqrt{n}}.$$

Рассмотрим теперь L -статистики. Пусть X_1, X_2, \dots — независимые одинаково распределенные наблюдения с ф.р. $F(x)$ и пусть $J(s)$ — измеримая функция, заданная на интервале $(0, 1)$ и удовлетворяющая условию Липшица. Рассмотрим числа вида

$$c_{in} = J\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n,$$

и определим L -статистику

$$L_n = \frac{1}{n} \sum_{i=1}^n c_{in} X_{(i:n)},$$

где $X_{(1:n)} \leq \dots \leq X_{(n:n)}$ — вариационный ряд, построенный по выборке X_1, \dots, X_n . Предположим также, что

$$E|X_1|^3 < \infty;$$

$$\sigma_J^2 \equiv \int_{-\infty}^{\infty} J(F(x))J(F(y))(\min(F(x), F(y)) - F(x)F(y)) dx dy > 0.$$

Тогда из теоремы 2.3 и теоремы 3.1.2 работы [19] получаем неравенство:

$$\sup_{\lambda \geq 0} \left| \tilde{Q}_{L_{N_n(m)}}(\lambda) - 2\Lambda_{1/m} \left(\frac{\lambda\sqrt{n}}{2\sigma_J} \right) + 1 \right| \leq \frac{4C_4}{\sqrt{n}}.$$

Литература

1. Гнеденко Б. В., Фихим Х. Об одной теореме переноса // Докл. АН СССР, 1969. Т. 187. С. 15–17.
2. Von Ghossy R., Rappl G. Some approximation methods for the distribution of random sums // Insurance: Mathematics and Economics, 1983. Vol. 2. P. 251–270.
3. Гнеденко Б. В. Курс теории вероятностей. — М.: Наука, 1988.
4. Круглов В. М., Королев В. Ю. Предельные теоремы для случайных сумм. — М.: Изд-во Московского ун-та, 1990.
5. Gnedenko B. V., Korolev V. Yu. Random summation. Limit theorems and applications. — Boca Raton: CRC Press, 1996.
6. Королев В. Ю., Бенинг В. Е., Шоргин С. Я. Математические основы теории риска. — 2-е изд., перераб. и дополн. — М.: Физматлит, 2011.
7. Гнеденко Б. В. Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // Тр. Тбилисского матем. ин-та, 1989. Т. 92. С. 146–150.
8. Королев В. Ю. Асимптотические свойства выборочных квантилей, построенных по выборкам случайного объема // Теория вероятностей и ее применения, 1999. Т. 44. Вып. 2. С. 440–445.
9. Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972.
10. Хенгартнер В., Теодореску Р. Функции концентрации. — М.: Наука, 1980.
11. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. — М.: Мир, 1984.
12. Лукач Е. Характеристические функции. — М.: Наука, 1979.
13. Shevtsova I. On the absolute constants in the Berry–Esseen inequalities for identically distributed summands. arXiv: 1111.6554v1, 2011.
14. 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.
15. Гавриленко С. В., Зубов В. Н., Королев В. Ю. Оценка скорости сходимости распределений регулярных статистик, построенных по выборкам случайного объема с отрицательным биномиальным распределением, к распределению Стьюдента // Статистические методы оценивания и проверки гипотез. — Пермь: Изд-во Пермского гос. ун-та, 2006. С. 118–134.
16. 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.
17. Нефедова Ю. С. Оценки скорости сходимости в предельной теореме для отрицательных биномиальных случайных сумм // Статистические методы оценивания и проверки гипотез. — Пермь: Изд-во Пермского гос. ун-та, 2011. С. 46–61.
18. Helmers R., Van Zwet W. R. The Berry–Esseen bound for U -statistics // Statistical Decision Theory and Related Topics 3, 1982. Vol. 1. P. 497–512.
19. Helmers R. Edgeworth expansions for linear combinations of order statistics. — Amsterdam: Mathematical Center Tracts 105, 1984. 137 p.
20. 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.
21. 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.

ОБ АБСОЛЮТНЫХ КОНСТАНТАХ В НЕРАВЕНСТВЕ БЕРРИ–ЭССЕЕНА И ЕГО СТРУКТУРНЫХ И НЕРАВНОМЕРНЫХ УТОЧНЕНИЯХ*

И. Г. Шевцова¹

Аннотация: Для равномерного расстояния Δ_n между функцией распределения (ф.р.) стандартного нормального закона и ф.р. нормированной суммы n независимых случайных величин (с.в.) X_1, \dots, X_n с $EX_j = 0$, $EX_j^2 = \sigma_j^2$, $j = 1, \dots, n$, при всех $n \geq 1$ приведены оценки

$$\Delta_n \leq \min\{0,5583\ell_n, 0,3723(\ell_n + 0,5\tau_n), 0,3057(\ell_n + \tau_n)\},$$

$$\Delta_n \leq \min\{0,4690\ell_n, 0,3322(\ell_n + 0,429\tau_n), 0,3031(\ell_n + 0,646\tau_n)\}, \text{ если } X_1 \stackrel{d}{\dots} \stackrel{d}{=} X_n,$$

где $\ell_n = \sum_{j=1}^n E|X_j|^3$, $\tau_n = \sum_{j=1}^n \sigma_j^3$, $\sum_{j=1}^n \sigma_j^2 = 1$. Получены уточненные результаты для случая симметричного распределения слагаемых. Также показано, что в неравенстве Нагаева–Бикялиса (неравномерном аналоге неравенства Берри–Эссена) абсолютная константа не превосходит 21,82 в общем случае и 17,36 в случае одинаково распределенных слагаемых.

Ключевые слова: центральная предельная теорема; оценка скорости сходимости; нормальная аппроксимация; неравенство Берри–Эссена; неравенство Нагаева–Бикялиса; абсолютная константа

Обозначим \mathcal{F}_3 множество всех ф.р. F с.в. X с $EX = 0$ и $E|X|^3 < \infty$. Через $\mathcal{F}_{3,s}$ обозначим множество всех симметричных ф.р. из \mathcal{F}_3 . Пусть X_1, X_2, \dots, X_n — независимые с.в. с ф.р. $F_1, \dots, F_n \in \mathcal{F}_3$ соответственно. Положим $\sigma_j^2 = EX_j^2$, $\beta_{3,j} = E|X_j|^3$, $j = 1, \dots, n$, $\sum_{j=1}^n \sigma_j^2 = 1$, $\ell_n = \sum_{j=1}^n \beta_{3,j}$, $\tau_n = \sum_{j=1}^n \sigma_j^3$, $\bar{F}_n(x) = P(X_1 + \dots + X_n < xB_n)$. Пусть $\phi(x)$ и $\Phi(x)$ — соответственно плотность и ф.р. стандартного нормального закона,

$$\Delta_n(x) = |\bar{F}_n(x) - \Phi(x)|, \quad x \in \mathbb{R};$$

$$\Delta_n = \sup_x |\bar{F}_n(x) - \Phi(x)|, \quad n = 1, 2, \dots$$

В работе [1] с использованием техники преобразования смещения формы и новой оценки точности аппроксимации характеристической функции первыми членами ее разложения в ряд Тейлора для всех $n \geq 1$ были получены оценки:

$$\Delta_n \leq \min\{0,5584\ell_n, 0,36266(\ell_n + 0,54\tau_n), 0,3129(\ell_n + 0,922\tau_n)\}, \quad F_1, \dots, F_n \in \mathcal{F}_3;$$

$$\Delta_n \leq \min\{0,4693\ell_n, 0,3322(\ell_n + 0,429\tau_n), 0,3031(\ell_n + 0,646\tau_n)\}, \quad F_1 = \dots = F_n \in \mathcal{F}_3;$$

$$\Delta_n \leq \min\{0,3425(\ell_n + 0,63\tau_n), 0,2895(\ell_n + \tau_n)\}, \quad F_1, \dots, F_n \in \mathcal{F}_{3,s};$$

$$\Delta_n \leq \min\{0,29489(\ell_n + 0,587\tau_n), 0,2730(\ell_n + 0,732\tau_n)\}, \quad F_1 = \dots = F_n \in \mathcal{F}_{3,s}.$$

С помощью алгоритма, использованного в [1], и оценки

$$|f'(t)| \leq \sigma \sin\left(\sigma|t| \wedge \frac{\pi}{2}\right), \quad t \in \mathbb{R},$$

для производной характеристической функции $f(t) = Ee^{itX}$ с $EX = 0$, $\sigma^2 \equiv EX^2 < \infty$, вытекающей из результатов работ [2, 3], указанные результаты можно уточнить и получить следующие оценки, справедливые при всех $n \geq 1$:

$$\Delta_n \leq \min\{0,5583\ell_n, 0,3723(\ell_n + 0,5\tau_n), 0,3057(\ell_n + \tau_n)\}, \quad F_1, \dots, F_n \in \mathcal{F}_3;$$

*Работа поддержана грантом МК–2256.2012.1, а также Российским фондом фундаментальных исследований (проекты 12-01-31125-а, 11-01-00515а, 11-07-00112а, 12-07-00115а).

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

$$\Delta_n \leq \min\{0,4690\ell_n, 0,3322(\ell_n + 0,429\tau_n), \\ 0,3031(\ell_n + 0,646\tau_n)\}, \quad F_1 = \dots = F_n \in \mathcal{F}_3;$$

$$\Delta_n \leq \min\{0,34245(\ell_n + 0,63\tau_n), \\ 0,2873(\ell_n + \tau_n)\}, \quad F_1, \dots, F_n \in \mathcal{F}_{3,s};$$

$$\Delta_n \leq \min\{0,29353(\ell_n + 0,593\tau_n), \\ 0,2730(\ell_n + 0,729\tau_n)\}, \quad F_1 = \dots = F_n \in \mathcal{F}_{3,s}.$$

Кроме того, исправляя неточность, допущенную в [4], и используя приведенные выше неравенства, можно показать, что справедливы следующие неравномерные оценки:

$$\sup_{x \in \mathbb{R}} (1 + |x|^3) \Delta_n(x) \leq \min\{21,82\ell_n, 18,19(\ell_n + \tau_n)\}, \\ F_1, \dots, F_n \in \mathcal{F}_3;$$

$$\sup_{x \in \mathbb{R}} (1 + |x|^3) \Delta_n(x) \leq \min\{17,36\ell_n, \\ 15,70(\ell_n + 0,646\tau_n)\}, \quad F_1 = \dots = F_n \in \mathcal{F}_3.$$

Более того, исправленный метод работы [4] позволяет построить монотонно убывающую функцию $C(t)$ с $\lim_{t \rightarrow \infty} C(t) = 1 + e = 3,71 \dots$ и такую, что

$$\sup_{|x| \geq t} |x|^3 \Delta_n(x) \leq C(t)\ell_n, \quad t \geq 0, \quad n \geq 1, \\ F_1, \dots, F_n \in \mathcal{F}_3.$$

В частности, для $C(t)$ справедливы верхние оценки $C(0) \leq 21,26$, $C(4) \leq 17,19$, $C(5) \leq 12,35$, $C(10) \leq 7,36$ для любых $F_1, \dots, F_n \in \mathcal{F}_3$; $C(0) \leq 16,90$, $C(4) \leq 14,58$, $C(5) \leq 11,56$, $C(10) \leq 5,85$, если $F_1 = \dots = F_n \in \mathcal{F}_3$.

С помощью методов, использованных в данной работе, можно получить аналогичные равномерные и неравномерные оценки для случая, когда слагаемые обладают моментами порядка лишь $2 + \delta$ с некоторым $0 < \delta \leq 1$.

Литература

1. *Shevtsova I.* On the absolute constants in the Berry–Esseen-type inequalities // 30th Seminar (International) on Stability Problems for Stochastic Models (Svetlogorsk, 2012): Book of Abstracts. — М.: ИПИ РАН, 2012. С. 71–72.
2. *Roßberg H.-J.* Positiv definite Verteilungsdichten // Appendix to: *Gnedenko B. W.* Einführung in die Wahrscheinlichkeitstheorie. — 9th ed. — Berlin: Akademie-Verlag, 1991.
3. *Matysiak W., Szablowski P. J.* Some inequalities for characteristic functions // J. Math. Sci., 2001. Vol. 105. No. 6. P. 2594–2598.
4. *Нефедова Ю. С., Шевцова И. Г.* О неравномерных оценках скорости сходимости в центральной предельной теореме // Теория вероятн. и ее примен., 2012, Т. 57. Вып. 1. С. 62–97.

ANALYTICAL MODELING OF INVARIANT MEASURE DISTRIBUTIONS IN STOCHASTIC SYSTEMS WITH DISCONTINUOUS CHARACTERISTICS

I. N. Sinitsyn

IPI RAN, sinitsin@dol.ru

Based on normal approximation and statistical linearization, exact and approximate algorithms for distributions with invariant measure analytical modeling in nongaussian stochastic systems (StS) with discontinuous characteristics are developed. Peculiarities in Poisson StS are considered. Test examples confirm the practical accuracy of algorithms.

Keywords: analytical modeling; autocorrelated noise; distribution with invariant measure; method of normal approximation; method of statistical linearization; Poisson stochastic system; Pugachev integrodifferential equations, stochastic regime; stochastic system in Ito sense

PROBABILITY AND STATISTICAL MODELING OF INFORMATION FLOWS IN COMPLEX FINANCIAL SYSTEMS BASED ON HIGH-FREQUENCY DATA

V. Yu. Korolev¹, A. V. Chertok², A. Yu. Korchagin³, and A. K. Gorshenin⁴

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; Euphoria Group LLC, a.v.chertok@gmail.com

³Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, sasha.korchagin@gmail.com

⁴IPI RAN, a.k.gorshenin@gmail.com

A microstructure model is proposed for information flows in complex financial systems and stochastic nature of intensities of flows of claims which determines market price formation. The outer information flow with random intensity is considered separately within the proposed and statistically verified multiplicative model. This model makes it possible to analyze characteristics related to the intensities of the flows of claims and instant relation between the forces of buyers and sellers without modeling of exogenous information background which is practically impossible to predict. Also, the generalized price process model is proposed which makes account of all the available information on flows of claims and admits further analytical interpretation.

Keywords: financial markets; information flows; market price formation; intensity of the flow of claims; limit order book; mixture of probability distributions; generalized price

STATIONARY CHARACTERISTICS OF THE QUEUEING SYSTEM WITH LIFO SERVICE, PROBABILISTIC PRIORITY, AND HYSTERIC POLICY

T. A. Milovanova¹ and A. V. Pechinkin²

¹People's Friendship University of Russia, tmilovanova77@mail.ru

²IPI RAN, apechinkin@ipiran.ru

Consideration is given to single server queueing system with LIFO service discipline, probabilistic priority, and single-threshold hysteric policy. Expressions for main stationary characteristics, including stationary waiting and sojourn times distributions in terms of Laplace–Stieltjes transform, are obtained.

Keywords: queueing system; last-in-first-out; probabilistic priority; hysteric policy

ON CONVERGENCE IN THE SPACE L_p OF THE WORKLOAD MAXIMUM FOR A CLASS OF GAUSSIAN QUEUEING SYSTEMS

O. V. Lukashenko¹ and E. V. Morozov²

¹Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences; Petrozavodsk State University, lukashenko-oleg@mail.ru

²Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences; Petrozavodsk State University, emorozov@karelia.ru

A class of queueing systems fed by an input containing linear deterministic component and a random component described by a centered Gaussian process is considered. The variance of the input is a regularly varying at infinity function with exponent $0 < V < 2$. The conditions are found under which the maximum of stationary workload (remaining work) over time interval $[0, t]$ converges in the space L_p as $t \rightarrow \infty$ (and under an appropriate scaling) to an explicitly given constant a . Asymptotics of the workload maximum in nonstationary regime is also given. The asymptotics of the hitting time of an increasing value b by the workload process is obtained.

Keywords: Gaussian queue; workload maximum; fractional Brownian motion; asymptotical analysis; regular varying

ALGORITHMS FOR INDUCTIVE GENERATION OF SUPERPOSITIONS FOR APPROXIMATION OF EXPERIMENTAL DATA

G. I. Rudoy¹ and V. V. Strijov²

¹Moscow Institute of Physics and Technology, rudoy@forecsys.ru

²Dorodnicyn Computing Centre of RAS, strijov@ccas.ru

The paper presents an algorithm which inductively generates admissible nonlinear models. An algorithm to generate all admissible superpositions of given complexity in finite number of iterations is proposed. The proof of its correctness is stated. The proposed approach is illustrated by a computational experiment on synthetic data.

Keywords: symbolic regression; nonlinear models; inductive generation; models complexity

STATISTICAL TECHNIQUES OF BANS DETERMINATION OF PROBABILITY MEASURES IN DISCRETE SPACES

A. A. Grusho¹, N. A. Grusho², and E. E. Timonina³

IPI RAN, mkriivenko@ipiran.ru

The method of statistical bans determination of probability measures in the discrete spaces is suggested. Consistency of the constructed estimates is shown. The application of the received estimates for testing the statistical hypotheses in the discrete spaces is constructed. It is shown that estimates of bans can generate consistent sequences of criteria in some sense.

Keywords: consistent sequences of criteria; bans of probability measures in the discrete spaces; consistency of estimates

OPERATIONS ON THE TREE REPRESENTATIONS OF PIECEWISE QUASI-AFFINE FUNCTIONS

S. A. Guda

Faculty of Mathematics, Mechanics and Computer Science, Southern Federal University, gudasergey@gmail.com

The piecewise quasi-affine (PQAF) functions and PQAF-sets are considered. Tree representation of PQAF-function and PQAF-set and its complexity are introduced. The algorithms of union, intersection, empty check,

sum, image/preimage calculation, inversion, composition, and comparison are given. The algorithms are provided with complexity estimates of the resulting objects. A theorem about the type and complexity of lexicographical extrema in the PQAF-set, depending on the parameters, has been proved.

Keywords: piecewise quasi-affine function; quasi-affine selection tree; lexicographical extremum; Z-polytope

METHODOLOGICAL BASIS FOR THE CREATION OF INFORMATION SYSTEMS FOR THE CALCULATION OF INDICATORS OF THEMATIC LINKAGES BETWEEN SCIENCE AND TECHNOLOGY

V. A. Minin¹, I. M. Zatsman², M. G. Kruzhkov³, and T. P. Norekyan⁴

¹RFBR, minin@rfbr.ru

²IPI RAN, iz_ipi@a170.ipi.ac.ru

³IPI RAN, magnit75@yandex.ru

⁴IPI RAN, izzittami@gmail.com

International experience of indicators' calculation of thematic linkages between science and technology is analyzed. The purpose of the analysis is to develop creation principles of information systems for calculating indicators of linkages with an allowance for structure of inherited scientific and patent information resources, which have been historically formed in Russia. This type of information systems is new to the Russian scientific and technical sphere. Their creation is necessary for evaluation of R&D programs and decision-making at all stages of program activities. In the paper, the methodology of indicators' calculation of thematic linkages in the Russian scientific and technical sphere is outlined to develop creation principles of these information systems.

Keywords: linkages between science and technology; classification of the scientific domains; international patent classification; classifying scientific documents

PARALLEL TEXTS ALIGNMENT STRATEGIES: THE SEMANTIC ASPECTS

E. B. Kozerenko

¹IPI RAN, kozerenko@mail.ru

The paper deals with the problems of design and development of the linguistically motivated mechanisms for parallel texts alignment and grammatical (functional semantic) matches extraction for design of statistical language portraits to be further incorporated into the hybrid models of machine translation. By hybrid models are meant those employing both statistical and rule-based mechanisms within one framework for natural language processing. The presented approach consists in the use of the starting ("seed") grammar to be further enriched by the matches discovered in parallel texts. The seed grammar is used featuring the cognitive and functional characteristics of phrase structures. The grammar formalism employed is the Cognitive Transfer Grammar based on the transfemes (bilingual phrase structures). The interpreters' transformations existing in parallel texts are of primary importance to the research efforts.

Keywords: alignment; parallel texts; syntax; semantics; phrase structures; hybrid models; machine translation

SEMANTIC VECTOR SPACES FOR DIFFERENT KNOWLEDGE DOMAINS

Yu. I. Morozova

IPI RAN, yulia-ipi@yandex.ru

The paper focuses on the actual problems of studying semantics of linguistic units using the methods of corpus linguistics. It gives a review of distributional semantics which is a new area of linguistic research. The paper proposes

an enhancement to the existing distributional models by switching from lexemes to word collocations. The paper describes the methodology used to build semantic vector spaces for different knowledge domains.

Keywords: distributional semantics; vector spaces; meaningful word combinations; collocations

INFORMATION METHOD FOR ASSESSMENT SEMANTIC ADEQUACY OF TEXTS

L. A. Kuznetsov¹ and V. F. Kuznetsova²

¹Lipetsk State Technical University, kuznetsov.leonid48@gmail.com

²Lipetsk State Technical University, kuznetsov@stu.lipetsk.ru

A problem of automated knowledge examination is considered based on intelligent comparison of student's answers with reference knowledge chunks. In order to increase relevance of automatically drawn conclusions, the information theory ground is used to develop the original methodology of textual content closeness evaluation. As a part of this work, this approach is applied to evaluate how close students' essays match reference textual blocks, relating to a subject.

Keywords: semantic similarity texts; probabilistic model of text; information theory; entropy; transmitted information of texts; knowledge estimate automatization

VARIANCE-GENERALIZED-GAMMA-DISTRIBUTIONS AS LIMIT LAWS FOR RANDOM SUMS

L. M. Zaks¹ and V. Yu. Korolev²

¹Department of Modeling and Mathematical Statistics, Alpha-Bank, lily.zaks@gmail.com

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

A general theorem is proved establishing necessary and sufficient conditions for the convergence of the distributions of sums of a random number of independent identically distributed random variables to variance-mean mixtures of normal laws. As a corollary, necessary and sufficient conditions for the convergence of the distributions of sums of a random number of independent identically distributed random variables to variance-generalized-gamma-distributions are obtained. For a special case of continuous-time random walks generated by compound doubly stochastic Poisson processes, convergence rate estimates are presented.

Keywords: random sum; generalized hyperbolic distribution; generalized inverse Gaussian distribution; generalized gamma-distribution; variance-generalized-gamma-distribution; mixture of probability distributions; identifiable mixtures; additively closed family; convergence rate estimate

ON BOUNDS FOR THE CONCENTRATION FUNCTIONS OF REGULAR STATISTICS CONSTRUCTED FROM SAMPLES WITH RANDOM SIZES

V. E. Bening¹, N. K. Galieva², and V. Yu. Korolev³

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, bening@cs.msu.su

²Kazakhstan Branch, M. V. Lomonosov Moscow State University, nurgul_u@mail.ru

³Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

Two-sided bounds for the concentration functions of regular statistics constructed from samples with random sizes are constructed.

Keywords: concentration function; random sum; asymptotically normal statistic; Student distribution; Laplace distribution

ON THE ABSOLUTE CONSTANTS IN THE BERRY–ESSEEN INEQUALITY AND ITS STRUCTURAL AND NONUNIFORM IMPROVEMENTS

I. G. Shevtsova

Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, ishevtsova@cs.msu.su

New improved upper bounds are presented for the absolute constants in the Berry–Esseen inequality and its structural and nonuniform improvements. In particular, it is shown that the absolute constant in the classical Berry–Esseen inequality does not exceed 0.5514 in general case and 0.4690 for the case of identically distributed summands. The corresponding bounds in the Nagaev–Bikelis inequality are 21.82 and 17.32.

Keywords: central limit theorem; convergence rate estimate; normal approximation; Berry–Esseen inequality; Nagaev–Bikelis inequality; absolute constant

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

Галиева Нургуль Кадыржановна (р. 1985) — преподаватель Казахстанского филиала Московского государственного университета им. М. В. Ломоносова

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, член-корреспондент Академии криптографии РФ; ведущий научный сотрудник ИПИ РАН; профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник ИПИ РАН

Гуда Сергей Александрович (р. 1982) — кандидат физико-математических наук, старший преподаватель факультета математики, механики и компьютерных наук Южного федерального университета

Закс Лилия Михайловна (р. 1989) — главный специалист отдела моделирования и математической статистики Альфа-банка

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом ИПИ 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 государ-

ственного университета им. М. В. Ломоносова; ведущий научный сотрудник ИПИ РАН

Корчагин Александр Юрьевич (р. 1989) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Кружков Михаил Григорьевич (р. 1975) — программист ИПИ РАН

Кузнецов Леонид Александрович (р. 1942) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий кафедрой Липецкого государственного технического университета

Кузнецова Вера Федоровна (р. 1948) — кандидат технических наук, доцент Липецкого государственного технического университета

Лукашенко Олег Викторович (р. 1986) — кандидат физико-математических наук, младший научный сотрудник Института прикладных математических исследований КарНЦ РАН; преподаватель Петрозаводского государственного университета

Милованова Татьяна Александровна (р. 1977) — старший преподаватель Российского университета дружбы народов

Минин Владимир Александрович (р. 1941) — доктор физико-математических наук, начальник управления РФФИ

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

Норекия Тамара Павловна (р. 1966) — младший научный сотрудник ИПИ РАН

Печинкин Александр Владимирович (р. 1946) — доктор физико-математических наук, профессор, глав-

ный научный сотрудник ИПИ РАН, профессор Российского университета дружбы народов

Рудой Георгий Игоревич (р. 1991) — студент Московского физико-технического института

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом ИПИ РАН

Стрижов Вадим Викторович (р. 1967) — кандидат физико-математических наук, доцент, научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор; ведущий научный сотрудник ИПИ РАН

Черток Андрей Викторович (р. 1987) — выпускник аспирантуры факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Шевцова Ирина Геннадьевна (р. 1983) — кандидат физико-математических наук; ассистент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

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

Chertok Andrey V. (b. 1987) — graduate PhD student, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University

Galieva Nurgul K. (b. 1985) — lecturer, Kazakhstan Branch of M.V. Lomonosov Moscow State University

Gorshenin Andrey K. — (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, Corresponding member of the Russian Academy of Cryptography; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University

Grusho Nikolai A. (p. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Guda Sergey A. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Faculty of Mathematics, Mechanics and Computer Science, Southern Federal University

Korchagin Alexander Yu. (b. 1989) — PhD student, Faculty of Computational Mathematics and Cybernetics, 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

Kruzhkov Mikhail G. (b. 1975) — programmer, Institute of Informatics Problems, Russian Academy of Sciences

Kuznetsov Leonid A. (b. 1942) — Doctor of Science in technology, professor, honored scientist of RF, Head of Department, Lipetsk State Technical University

Kuznetsova Vera F. (b. 1948) — Candidate of Science (PhD) in technology, associate professor, Lipetsk State Technical University

Lukashenko Oleg V. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, junior scientist, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences; lecturer, Petrozavodsk State University

Milovanova Tatyana A. (b. 1977) — senior lecturer, Peoples' Friendship University of Russia

Minin Vladimir A. (b. 1941) — Doctor of Science in physics and mathematics, Head of Department, Russian Foundation for Basic Research

Morozov Evsei V. (b. 1947) — Doctor of Science in physics and mathematics, professor, leading scientist, Institute of Applied Mathematical Research of 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

Norekyan Tamara P. (b. 1966) — junior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Pechinkin Alexander V. (b. 1946) — Doctor of Science in physics and mathematics; principal scientist, Institute of Informatics Problems of the Russian Academy of Sciences; professor, Peoples' Friendship University of Russia

Rudoy Georgiy I. (b. 1991) — student, Moscow Institute of Physics and Technology

Shevtsova Irina G. (b. 1983) — Candidate of Science (PhD) in physics and mathematics; assistant professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University

ty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences

Sinitsyn 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

Strijov Vadim V. (b. 1967) — Candidate of science (PhD) in physics and mathematics, associate professor, scien-

tist, Dorodnicyn Computing Center, Russian Academy of Sciences

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Zaks Lily M. (b. 1989) — principal officer, Department of Modeling and Mathematical Statistics, Alpha-Bank

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

Правила подготовки рукописей статей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений. Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи, в виде исключения, могут печататься на английском языке. Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях. Публикация не должна нарушать закон об авторских правах. Направляя свою рукопись в редакцию, авторы автоматически передают учредителям и редколлегии неисключительные права на издание данной статьи на русском языке и на ее распространение в России и за рубежом. При этом за авторами сохраняются все права как собственников данной рукописи. В связи с этим авторами должно быть представлено 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)».

Указанное соглашение может быть представлено как в бумажном виде, так и в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности опубликования представленной статьи в открытой печати.

2. Статья подписывается всеми авторами. На отдельном листе представляются данные автора (или всех авторов): фамилия, полные имя и отчество, телефон, факс, e-mail, почтовый адрес. Если работа выполнена несколькими авторами, указывается фамилия одного из них, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет самостоятельную экспертизу присланных статей. Возвращение рукописи на доработку не означает, что статья уже принята к печати. Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.
4. Решение редакционной коллегии о принятии статьи к печати или ее отклонении сообщается авторам. Редколлегия не обязуется направлять рецензию авторам отклоненной статьи; дискуссия с авторами по поводу отклоненных статей не ведется.
5. Корректур статей высылаются авторам для просмотра. Редакция просит авторов присылать свои замечания в кратчайшие сроки.
6. 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 25 страниц указанного формата. Ознакомиться с шаблонами, содержащими примеры оформления, можно по адресу в Интернете: <http://www.ipiran.ru/journal/template.doc>.
7. К рукописи, предоставляемой в 2-х экземплярах, обязательно прилагается электронная версия статьи (как правило, в форматах MS WORD (.doc) или ЛАТЭХ (.tex), а также — дополнительно — в формате .pdf) на дискете, лазерном диске или по электронной почте. Сокращения слов, кроме стандартных, не применяются. Все страницы рукописи должны быть пронумерованы.

8. Статья должна содержать следующую информацию на русском и английском языках: название, Ф.И.О. авторов, места работы авторов и их электронные адреса, подробные сведения об авторах, оформленные в соответствии с форматом, определяемым файлами http://www.ipiran.ru/journal/issues/2011_05_01/authors.asp и http://www.ipiran.ru/journal/issues/2011_01_eng/authors.asp, аннотация (не более 100 слов), ключевые слова. Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в порядке их первого упоминания. В списке литературы не должно быть позиций, на которые нет ссылки в тексте статьи. Все фамилии авторов, заглавия статей, названия книг, конференций и т. п. даются на языке оригинала, если этот язык использует кириллический или латинский алфавит.
9. Присланные в редакцию материалы авторам не возвращаются.
10. При отправке файлов по электронной почте просим придерживаться следующих правил:
- указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в случае больших объемов информации возможно использование общеизвестных архиваторов (ZIP, RAR);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
11. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию с авторов 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