

Информатика и её применения

Том 6 Выпуск 4 Год 2012

Тематический выпуск

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

СОДЕРЖАНИЕ

Предисловие И. А. Соколов, В. Ю. Королев	3
Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с автокоррелированными шумами И. Н. Сеницын	4
О точности некоторых математических моделей катастрофически накапливающихся эффектов при прогнозировании риска экстремальных событий И. А. Дучицкий, В. Ю. Королев, И. А. Соколов	9
Об адаптивных стратегиях и условиях их существования М. Г. Коновалов	18
Оценки в нуль-эргодическом случае для некоторых систем обслуживания А. И. Зейфман, А. В. Коротышева, Я. А. Сатин, С. Я. Шоргин	27
Обобщенное распределение Лапласа как предельное для случайных сумм и статистик, построенных по выборкам случайного объема В. Ю. Королев, В. Е. Бенинг, Л. М. Закс, А. И. Зейфман	34
Нижние оценки устойчивости смесей нормальных распределений к возмущениям смешивающих распределений А. Л. Назаров	40
Предварительная обработка при распознавании текстов по изображению низкого качества М. П. Кривенко	49
Модель случайных графов для описания взаимодействий в сети А. А. Грушо, Е. Е. Тимонина	57
Об оптимальном корректном перекодировании целочисленных данных в распознавании Е. В. Дюкова, А. В. Сизов, Р. М. Сотнезов	61
Оценивание гиперпараметров линейных регрессионных моделей при отборе шумовых и коррелирующих признаков А. А. Токмакова, В. В. Стрижов	66
Holographic coding by Walsh–Hadamard transformation of randomized and permuted data S. Dolev, S. Frenkel, and A. Cohen	76

Информатика и её применения

Том 6 Выпуск 4 Год 2012

Тематический выпуск

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

СОДЕРЖАНИЕ

Mathematical foundation, application, and comparison of general data assimilation method based on diffusion approximation with other data assimilation schemes <i>K. P. Belyaev, C. A. S. Tanajura, and N. P. Tuchkova</i>	84
Complete convergence for arrays of negatively dependent random variables <i>S. H. Sung, K. Budsaba, and A. Volodin</i>	95
Abstracts	103
Об авторах	107
About Authors	108
Авторский указатель за 2012 г.	110
2012 Author Index	114

Технический редактор *Л. Кокушкина*

Художественный редактор *М. Седакова*

Сдано в набор 01.11.12. Подписано в печать 19.12.12. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 12,75. Уч.-изд. л. 10,5. Тираж 100 экз.
Заказ № 2071

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

Отпечатано в Академиздатцентре «Наука» РАН с готовых файлов

Москва 121099, Шубинский пер., д. 6.

Предисловие

Вниманию читателей журнала «Информатика и её применения» предлагается очередной тематический выпуск «Вероятностно-статистические методы и задачи информатики и информационных технологий». Предыдущие тематические выпуски журнала по данному направлению вышли в 2008 (т. 2, вып. 2), 2009 (т. 3, вып. 3), 2010 (т. 4, вып. 2) и 2011 гг. (т. 5, вып. 3).

Статьи, собранные в данном журнале, посвящены разработке и совершенствованию вероятностно-статистических и смежных с ними методов, ориентированных на применение к решению конкретных задач информатики и информационных технологий, а также — в ряде случаев — и других прикладных задач. Проблематика, охватываемая публикуемыми работами, в значительной степени развивается в рамках научного сотрудничества между Институтом проблем информатики Российской академии наук (ИПИ РАН) и факультетом вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова в ходе работ над совместными научными проектами (в том числе в рамках функционирования Научно-образовательного центра «Вероятностно-статистические методы анализа рисков»). Многие из авторов статей, включенных в данный номер журнала, являются активными участниками традиционного международного семинара по проблемам устойчивости стохастических моделей, руководимого В. М. Золотаревым и В. Ю. Королевым; регулярные сессии этого семинара проводятся под эгидой МГУ и ИПИ РАН (в 2012 г. указанный семинар проводился в сентябре в г. Светлогорске Калининградской области РФ).

Наряду с представителями ИПИ РАН и МГУ в число авторов данного выпуска журнала входят ученые из Вычислительного центра им. А. А. Дородницына Российской академии наук, Института океанологии им. П. П. Ширшова Российской академии наук, Московского физико-технического института, Вологодского государственного педагогического университета, отдела моделирования и математической статистики Альфа-банка, а также Федерального университета штата Баия (Бразилия), Университета им. Бен-Гуриона в Негеве (Израиль), Университета Реджайны (Канада), Университета Пай Чай (Республика Корея), Университета Таммасат (Таиланд).

Тематика статей данного выпуска включает вопросы математического моделирования и анализа реальных процессов, в том числе моделирование стохастических систем с автокоррелированными шумами; анализ катастрофически накапливающихся эффектов при прогнозировании риска экстремальных событий; построение и исследование моделей некоторых специальных систем и сетей передачи информации; вывод оценок точности аппроксимаций распределений, описывающих реальные процессы; построение адаптивных стратегий оптимального управления на основе наблюдений, доступных в процессе управления; математические методы и алгоритмы распознавания образов.

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

Заместитель главного редактора журнала «Информатика и её применения»,
директор ИПИ РАН, академик

И. А. Соколов

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

В. Ю. Королев

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ С ИНВАРИАНТНОЙ МЕРОЙ В СТОХАСТИЧЕСКИХ СИСТЕМАХ С АВТОКОРРЕЛИРОВАННЫМИ ШУМАМИ*

И. Н. Сеницын¹

Аннотация: Для многомерных нелинейных гауссовских (нормальных) дифференциальных систем с некоррелированными и автокоррелированными помехами на базе метода нормальной аппроксимации разработаны корреляционные алгоритмы аналитического моделирования стохастических режимов с инвариантной мерой. На тестовых примерах с помощью инструментального программного обеспечения в среде MATLAB показана достаточная для многих приложений точность алгоритмов.

Ключевые слова: автокоррелированная помеха; аналитическое моделирование; корреляционный алгоритм; метод нормальной аппроксимации; многомерная нелинейная дифференциальная стохастическая система; распределение с инвариантной мерой

1 Введение

Будем рассматривать в общем случае нестационарный стохастический режим $Z = Z(t)$, являющийся сильным решением следующего нормального (гауссовского) стохастического дифференциального уравнения Ито [1, 2]:

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

Здесь Z — k -мерный вектор состояния $Z \in \Delta$ (Δ — многообразие состояний), $a = a(Z, t)$ и $b = b(Z, t)$ — детерминированные $(k \times 1)$ - и $(k \times m)$ -функции отмеченных аргументов, $V = V(t)$ — m -мерный вектор нормально распределенных белых шумов с нулевыми математическими ожиданиями и $(m \times m)$ -матрицей интенсивностей $v = v(t)$ и представляющий собой среднеквадратичную производную винеровского процесса $W = W(t)$, $V = \dot{W}$. Начальное состояние Z_0 будем считать нормальной (гауссовой) случайной величиной, не зависящей от приращений винеровского процесса $W(t)$ для $t > t_0$.

Если существуют все многомерные плотности вектора состояния Z , то, определив сначала одномерную плотность $f_1 = f_1(z; t)$ и переходную плотность $f = f(z; t|\xi; \tau)$ путем интегрирования уравнения Фоккера–Планка–Колмогорова (ФПК) с соответствующими начальными условиями [1, 2]:

$$\frac{\partial f_1}{\partial t} = -\frac{\partial^T}{\partial z} (a f_1) + \frac{1}{2} \text{tr} \left[\frac{\partial}{\partial z} \frac{\partial^T}{\partial z} (\sigma f_1) \right],$$

$$\sigma = b v b^T, \quad f_1(z; t_0) = f_0(z); \quad (2)$$

$$\frac{\partial f}{\partial t} = -\frac{\partial^T}{\partial z} (a f) + \frac{1}{2} \text{tr} \left[\frac{\partial}{\partial z} \frac{\partial^T}{\partial z} (\sigma f) \right],$$

$$f(z; t|\xi; \tau) = \delta(z - \xi), \quad (3)$$

можно найти все многомерные плотности $f_n = f_n(z_1, \dots, z_n; t_1, \dots, t_n)$ по рекуррентной формуле:

$$f_n = f_n(z_1, \dots, z_n; t_1, \dots, t_n) =$$

$$= f_1(z_1; t_1) f(z_2; t_2|z_1; t) \cdots f(z_n; t_n|z_{n-1}; t_{n-1}),$$

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

Для нахождения стационарных в узком смысле одно- и многомерных распределений стохастических режимов в стохастических системах (СтС), определяемых (2), в (3) следует положить $\partial f_1 / \partial t = 0$.

В [1–12] рассмотрены точные методы аналитического моделирования одно- и многомерных плотностей стохастических стационарных и нестационарных режимов, основанные на построении интегральных инвариантов специально подобранных обыкновенных дифференциальных уравнений. В [13] на основе теории потенциала предложены методы расчета стационарных распределений с инвариантной мерой для гамильтоновых СтС в стохастической среде.

Стохастические системы с автокоррелированными помехами обычно описывают следующими уравнениями Ито [1, 2]:

$$\dot{Z} = a(Z, t) + b_U(z, t)U. \quad (4)$$

* Работа поддержана РФФИ (проект № 10-07-00021) и Программой «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект 1.7).

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

Здесь векторная помеха U удовлетворяет следующему стохастическому линейному дифференциальному уравнению фильтра, формирующего U из гауссового (нормального) белого шума V интенсивности $\nu = \nu(t)$:

$$\sum_{i=0}^l \alpha_i U^{(i)} = \sum_{j=0}^l \beta_j V^{(j)} \quad (h < l), \quad (5)$$

где α_i и β_j — коэффициенты формирующего фильтра (ФФ), в общем случае зависящие от времени.

Поставим задачу разработки на основе метода нормальной аппроксимации (МНА) [1, 2] корреляционных алгоритмов аналитического моделирования одно- и двумерных нормальных (гауссовских) плотностей стохастических режимов $Z = Z(t)$ в СтС (1) и (4) с инвариантной мерой, т. е. удовлетворяющих условию:

$$\frac{\partial f_1(z; t)}{\partial t} + \frac{\partial^T}{\partial z} [a(z, t) f_1(z; t)] = 0, \quad (6')$$

где $f_1 = f_1(z; t)$ — одномерная плотность распределения. Для стационарных режимов условие (6') принимает вид

$$\frac{\partial^T}{\partial z} [a^*(z) f_1^*(z)] = 0, \quad (6'')$$

Рассмотрим случаи некоррелированной помехи (разд. 2) и автокоррелированной помехи, являющейся менее (более) гладкой, чем стохастический режим (разд. 3).

2 Аналитическое моделирование при некоррелированных помехах

Пусть нелинейная СтС (1) допускает применение МНА [1, 2]. Тогда одномерная нормальная плотность $f_1^N(z; t)$, вектор математического ожидания $m_t = MZ(t)$, ковариационная матрица $K_t = MX^{0T}(t)Z^0(t)$ и матрица ковариационных функций $K(t_1, t_2) = MZ^{0T}(t_1)Z^0(t_2)$ ($t_1 < t_2$) определяются следующими уравнениями [1, 2]:

$$f_1^N(z; t, m_t, K_t) = [(2\pi)^k |K_t|]^{-1/2} \times \exp \left\{ -\frac{1}{2} (z^T - m_t^T) K_t^{-1} (z - m_t) \right\}; \quad (7)$$

$$\begin{aligned} \dot{m}_t &= a_1(m_t, K_t, t) = \\ &= \int_{-\infty}^{\infty} a(z, t) f_1^N(z; t, m_t, K_t) dz; \quad (8) \end{aligned}$$

$$\begin{aligned} \dot{K}_t &= a_2(m_t, K_t, t) = \int_{-\infty}^{\infty} [a(z, t)(z^T - m_t^T) + \\ &+ (z - m_t)a^T(z, t) + \sigma(z, t)] f_1^N(z, t, m_t, K_t) dz, \quad (9) \end{aligned}$$

где

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

$$\begin{aligned} \frac{\partial K(t_1, t_2)}{\partial t_2} &= a_3(m_{t_1}, m_{t_2}, K(t_1, t_2)) = \\ &= [(2\pi)^{2k} |\overline{K}_2|]^{-1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{t_1}) a(z_{t_1}, t_2) \times \\ &\times \exp \left\{ - \left([z_1^T \ z_2^T] - \overline{m}_2^T \right) \overline{K}_2^{-1} \times \right. \\ &\left. \times \left([z_1^T \ z_2^T]^T - \overline{m}_2 \right) \right\} dz_1 dz_2, \quad (10) \end{aligned}$$

где

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

Для стационарной системы (1) уравнения для $m_t = m^* = const$, $K_t = K^* = const$ получаются из (8)–(10) при условиях $\dot{m}_t = 0$, $\dot{K}_t = 0$:

$$a_1^*(m^*, K^*) = 0; \quad a_2^*(m^*, K^*) = 0; \quad (11)$$

$$\left. \begin{aligned} \frac{dk(\tau)}{d\tau} &= k_1^a(m^*, K^*) k(\tau); \\ k(\tau) &= k(-\tau)^T; \quad k(0) = K. \end{aligned} \right\} \quad (12)$$

Здесь $k_1^a = k_1^a(m_t, K_t)$ представляет собой коэффициент статистической гауссовской линеаризации нелинейной функции

$$a(Z, t) = a_1(m_t, K_t) + k_1^a(m_t, K_t)(Z - m_t),$$

причем при $m_t = m^*$, $K_t = K^*$ он определяется формулой:

$$k_1^a = a_{21}(m_t, K_t) K_t^{-1},$$

где

$$\begin{aligned} a_{21} &= a_{21}(m_t, K_t) = \\ &= \int_{-\infty}^{\infty} a(z, t) (z^T - m_t^T) f_1^N(z; t, m_t, K_t) dz = \left[\frac{\partial}{\partial m_t} a_1^T \right]^T. \end{aligned}$$

При этом условия (6') и (6'') примут следующий вид:

$$\frac{\partial f_1^N(z; t, m_t, K_t)}{\partial t} + \frac{\partial^T}{\partial z} \{ [a_1(t, m_t, K_t, t) + k_1^a(t, m_t, K_t, t)(z - m_t)] f_1^N(z; t, m_t, K_t) \} = 0, \quad (13')$$

$$\frac{\partial^T}{\partial z} \{ [a_1^*(m^*, K^*) + k_1^a(m^*, K^*)(z - m^*)] f_1^N(z; m^*, K^*) \} = 0. \quad (13'')$$

Таким образом, имеет место утверждение.

Теорема 1. Если матрица $k_1^a = k_1^a(m_t, K_t)$ асимптотически устойчива, то корреляционный алгоритм аналитического моделирования нестационарных режимов МНА в СтС (1) определяется уравнениями (7)–(11), (13'), а для стационарных режимов — уравнениями (7)–(9), (12), (13'').

3 Аналитическое моделирование при автокоррелированных помехах

Сначала рассмотрим СтС (4) и (5) при условиях (6') и (6'') в предположении, что стохастический режим $Z(t)$ является более гладким, чем помеха (имеет производные более высокого порядка, чем помеха). Преобразуем уравнение (4), записанное в виде:

$$\Theta(Z, U, t) = \dot{Z} = a(Z, t) + b_U U, \quad (14)$$

системой, обратной ФФ (5). Как известно [1, 2], система, обратная к ФФ (5), представляет собой параллельное соединение системы, осуществляющей линейную дифференциальную операцию

$$L = \sum_{k=0}^{l-h} \gamma_k D^k, \quad D = \frac{d}{dt},$$

и системы, описываемой дифференциальным уравнением

$$\sum_{i=0}^h \beta_i y^{(i)} = \sum_{j=0}^{h-l} \tilde{\alpha}_j x^{(j)}. \quad (15)$$

Здесь введены следующие обозначения:

$$\gamma_{l-h} = \beta_h^{-1} \alpha_l;$$

$$\gamma_k = \beta_h^{-1} \left[\alpha_{h+k} - \sum_{r=\max(0, k-l+2h)}^{h-1} \sum_{j=0}^{h-r} C_{r+j}^r \beta_{r+j} \gamma_{h+k-r}^{(j)} \right] \quad (k = 0, 1, \dots, l-h-1);$$

$$\tilde{\alpha}_k = \alpha_k - \sum_{r=\max(0, k-l+2h)}^k \sum_{j=0}^{h-r} C_{r+j}^r \beta_{r+j} \gamma_{k-r}^{(j)} \quad (k = 0, 1, \dots, l-h-1).$$

Пропустив сигнал $\Theta(Z, t)$, определяемый уравнением (14), через систему, состоящую из усилителя с коэффициентом b_U^{-1} и системы, обратной ФФ (5), получим на выходе сигнал

$$\Theta_1 = L [b_U^{-1} a(Z, t)] + Z'_1 + V,$$

где Z'_1 — выходной сигнал системы (15) при $\Theta(Z, t) = b_U^{-1} a(Z, t)$:

$$\sum_{i=0}^h \beta_i Z_1^{(i)} = \sum_{j=0}^{h-1} \tilde{\alpha}_j (b_U^{-1} a(Z, t))^{(j)}. \quad (16)$$

Приведа уравнение (16) к системе уравнений первого порядка согласно [1, 2], получим дифференциальное уравнение, определяющее вектор $Z' = [Z_1^{T'} \dots Z_h^{T'}]^T$:

$$\dot{Z}' = c(Z, Z').$$

Наконец, введя расширенный вектор состояния $\bar{Z} = [Z^T \ Z'^T]^T$, получим уравнение вида (1):

$$\dot{\bar{Z}} = \bar{a}(\bar{Z}, t) + \bar{b}V, \quad (17)$$

где

$$\bar{a} = \begin{bmatrix} a \\ c \end{bmatrix}; \quad \bar{b} = \begin{bmatrix} b_U \\ 0 \end{bmatrix}. \quad (18)$$

Таким образом, имеем следующий результат.

Теорема 2. Если матрица $k_1^{\bar{a}}$ асимптотически устойчива, а стохастический режим в (4) является более гладким, чем помеха, то корреляционный алгоритм аналитического моделирования МНА определяется теоремой 1 для уравнения (17) при условии (18).

В случае, когда стохастический режим не менее гладкий, чем помеха U , достаточно применить процедуру расширения вектора состояния путем дифференцирования (14) и исключения помехи U и ее производных, не содержащих белого шума V , из уравнений ФФ (5) с помощью уравнения (14) и уравнений, полученных из него дифференцированием.

С этой целью продифференцируем уравнение (14), умноженное слева на b_U^{-1} , до появления в нем белого шума. Тогда получим:

$$\begin{aligned} \frac{d^i}{dt^i} [b_U^{-1}(t)\Theta(Z, U, t)] = \\ = \frac{d^i}{dt^i} [b_U^{-1}(t)a(Z, t)] + U_{i+1} \quad (i = 0, 1, \dots, s), \quad (19) \end{aligned}$$

если

$$\frac{d^s}{dt^s} [b_U^{-1}(t)a(Z, t)] \text{ при } s < l - h$$

содержит белый шум, и

$$\left. \begin{aligned} \frac{d^i}{dt^i} [b_U^{-1}(t)\Theta(Z, U, t)] &= \\ &= \frac{d^i}{dt^i} [b_U^{-1}(t)a(Z, t)] + U_{i+1} \\ &\quad (i = 0, 1, \dots, l - h - 1); \\ \frac{d^{l-h}}{dt^{l-h}} [b_U^{-1}(t)\Theta(Z, U, t)] &= \\ = \frac{d^{l-h}}{dt^{l-h}} [b_U^{-1}(t)a(Z, t)] + U_{l-h+1} + q_{l-h}V, \end{aligned} \right\} (20)$$

если ни при каком $s < l - h$ производная $d^s [b_U^{-1}(t)a(Z, t)] / dt^s$ не содержит белого шума V .

В первом случае, если согласно [1, 2] привести (5) к системе уравнений первого порядка:

$$\left. \begin{aligned} U_1 &= U; \\ \dot{U}_i &= U_{i+1} \quad (i = 1, \dots, l - h - 1); \\ \dot{U}_i &= U_{i+1} + q_i V \quad (i = l - h, \dots, l - 1); \\ \dot{U}_l &= -\alpha_i^{-1} \sum_{i=1}^{l-1} \alpha_i U_{i+1} + q_l V, \end{aligned} \right\} (21)$$

где q_i и q_l определены в [1, 2] и зависят от α_i и β_j , можно исключить помехи U_1, \dots, U_s с помощью (14) и уравнений, полученных $(s - 1)$ -кратным дифференцированием.

Во втором случае следует из (21) исключить помехи U_1, \dots, U_{l-h-1} и белый шум V с помощью (14) и всех уравнений, полученных из него дифференцированием.

В обоих случаях получим уравнения U_{s+1}, \dots, U_l ($s \leq l - h$). При этом можно расчлнить каждую из этих переменных на две, одна из которых зависит от Z , а другая зависит от $\Theta(Z, t)$ и его производных, и тогда будем иметь

$$U_{s+k} = Z'_k - Y_k, \quad (22)$$

где Z'_1, \dots, Z'_{l-s} определяются дифференциальными уравнениями (17):

$$\dot{Z}' = c', \quad c' = c'(Z, t), \quad Z' = [Z'_1, \dots, Z'_{l-s}]^T, \quad (23)$$

а Y'_1, \dots, Y'_{l-s} — уравнениями

$$\left. \begin{aligned} \dot{Y}' &= c'', \quad c'' = c''(\Theta, \dot{\Theta}, \dots, \Theta^{(l-s)}); \\ Y' &= [Y'_1, \dots, Y'_{l-s}]^T. \end{aligned} \right\} (24)$$

Правые части c' и c'' определяются согласно (21) и (22).

Расширив вектор состояния Z для $\bar{Z} = [Z^T \ Z_1^T \ \dots \ Z_{l-s}^T]^T$, придем к окончательным уравнениям (17), (18) при $c = c'$.

Таким образом, получен следующий результат.

Теорема 3. Если матрица $k_1^{\bar{\alpha}}$ асимптотически устойчива, а стохастический режим в (4) не менее гладкий, чем помеха, то корреляционный алгоритм аналитического моделирования МНА определяется уравнениями теоремы 1 для системы (17) при условиях (19)–(24).

4 Тестовые примеры

1. Стохастическое уравнение Дуффинга [1, 2]

$$\ddot{X} + \omega^2 X - \mu X^3 = -\delta \dot{X} + U_0 + V \quad (25)$$

при $U_0 = 0$ допускает режим стационарных стохастических колебаний, определяемых формулой Гиббса

$$f_1(x, \dot{x}) = C \exp \left[-\frac{2\delta}{\gamma} H(x, \dot{x}) \right],$$

где $H(x, \dot{x}) = \dot{x}^2/2 + \omega^2 x^2/2 - \mu x^4/4$. В основе корреляционного алгоритма аналитического моделирования стационарных и нестационарных режимов лежат следующие уравнения [1, 2]:

$$\begin{aligned} X^3 &\approx m_X(m_X^2 + 3D_X) + 3(m_X^2 + D_X)X^0 = \\ &= k_0 m_X + k_1 X^0 \quad (X^0 = X - m_X); \end{aligned}$$

$$\dot{m}_X = m_{\dot{X}}, \quad \dot{m}_X = U_0 - \omega_3^2 m_X - \delta m_{\dot{X}};$$

$$\dot{D}_X = 2K_{X\dot{X}}, \quad \dot{D}_X = \nu - 2(\omega_{1\alpha}^2 K_{X\dot{X}} + \delta D_{\dot{X}});$$

$$\dot{K}_{X\dot{X}} = D_{\dot{X}} - \omega_{1\alpha}^2 D_X - \delta K_{X\dot{X}},$$

где

$$\left. \begin{aligned} \omega_3^2 &= \omega^2 \left[1 - \frac{\mu(m_X^2 + 3D_X)}{\omega^2} \right]; \\ \omega_{1\alpha}^2 &= \omega^2 \left[1 - \frac{3\mu(m_X^2 + D_X)}{\omega^2} \right] \quad (\omega_3 > \omega_{1\alpha}). \end{aligned} \right\} (26)$$

При $U_0 = 0$ в стационарном случае $m_X^* = 0$, $m_{\dot{X}}^* = 0$, $K_{X\dot{X}}^* = 0$, $D_X^* = \nu/(2\delta)$, а D_X^* определяются путем решения уравнения $\nu/(2\delta) = \omega_{1\alpha}^2 (D_X^*) D_X^*$.

Таким образом, МНА для D_X^* дает решение, совпадающее с точным, а для D_X^* — приближенное, точность которого зависит от коэффициента μ . Процесс установления стационарных колебаний происходит в два этапа: сначала устанавливается D_X^* , а затем D_X^* .

Условие (6'') требует консерватизма статистически линеаризованной системы левой части (25). Для значений μ , отвечающих колебаниям, точность составляет 10%.

2. Для системы

$$\ddot{X} + \omega^2 X - \mu X^3 = -\delta \dot{X} + U_0 + U, \quad \dot{U} = -\gamma U + V \quad (27)$$

уравнения МНА для $Z = [X \dot{X} U]^T$ имеют вид (8) и (9) при

$$a_1 = \begin{bmatrix} m_{\dot{X}} \\ -\omega_s^2 m_X - \delta m_{\dot{X}} + U_0 \\ -m_U \end{bmatrix}; \quad \alpha = \begin{bmatrix} 0 & 1 & 0 \\ -\omega_{1s}^2 & -\delta & 0 \\ 0 & 0 & -\gamma \end{bmatrix};$$

$$\beta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad a_2 = \alpha K_t + K_t \alpha^T + \beta \nu \beta^T,$$

где ν — интенсивность белого шума V ; ω_s , ω_{1s} определяются (26). Отсюда аналитическим моделированием определяются стационарные режимы стохастических колебаний, а также режимы установления. Условие (6'') выполняется в силу консерватизма левой части (27).

Сравнивая оба примера, заключаем, что точность метода за счет «профильрованности» помех значительно повышается и достигает 2%–4%.

3. В инструментальном программном обеспечении CStS-Analysis тестовые примеры описаны в [14, 15].

5 Заключение

Для многомерных нелинейных дифференциальных нормальных (гауссовских) систем с некоррелированными и автокоррелированными гауссовскими помехами на базе метода нормальной аппроксимации разработаны корреляционные алгоритмы аналитического моделирования стохастических режимов с инвариантной мерой. Результаты допускают обобщение на случай негауссовских помех, а также недифференцируемых функций $a = a(Z, t)$ в уравнениях (1) и (4).

С помощью разработанного в среде MATLAB инструментального программного обеспечения на тестовых примерах показана достаточная для многих приложений точность корреляционных алгоритмов.

Литература

1. Пугачев В. С., Синецын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — 2-е изд., доп. — М.: Наука, 1990.
2. Пугачев В. С., Синецын И. Н. Теория стохастических систем. — 2-е изд. — М.: Логос, 2004.
3. Moshchuk N. K., Sinityn I. N. On stationary distributions in nonlinear stochastic differential systems: Preprint. — Coventry, CV4 7AL, UK: Mathematics Institute, University of Warwick, 1989. 15 p.
4. Moshchuk N. K., Sinityn I. N. On stochastic nonholonomic systems: Preprint. — Coventry, CV4 7AL, UK: Mathematics Institute University of Warwick, 1989. 32 p.
5. Мошук Н. К., Синецын И. Н. О стохастических неголономных системах // Прикладная механика и математика, 1990. Т. 54. Вып. 2. С. 213–223.
6. Moshchuk N. K., Sinityn 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., Синецын И. Н. О стационарных и приводимых к стационарному режиму в нормальных стохастических системах // Прикладная механика и математика, 1991. Т. 55. Вып. 6. С. 895–903.
8. Мошук Н. К., Синецын И. Н. Распределения с инвариантной мерой в механических стохастических нормальных системах // Докл. АН СССР, 1992. Т. 322. № 4. С. 662–667.
9. Синецын И. Н. Конечномерные распределения с инвариантной мерой в стохастических механических системах // Докл. РАН, 1993. Т. 328. № 3. С. 308–310.
10. Синецын И. Н. Конечномерные распределения с инвариантной мерой в стохастических нелинейных дифференциальных системах. — М.: Диалог МГУ, 1997. С. 129–140.
11. Синецын 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 (С&Т'2002): Тр. II Междунар. научно-технич. конф. — Воронеж: Саквое, 2002. С. 124–131.
12. Синецын 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.
13. Soize C. The Fokker–Plank equation for stochastic dynamical systems and its explicit steady state solutions. — Singapore: World Scientific, 1994.
14. Sinityn I. N. Lectures on PC-based nonlinear stochastic mechanical systems research: Učební Texty únavu Termomechaniky. — Praha: ČAV, 1992. 63 p.
15. Синецын И. Н. Стохастические информационные технологии для исследования нелинейных круговых систем // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 78–99.

О ТОЧНОСТИ НЕКОТОРЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ КАТАСТРОФИЧЕСКИ НАКАПЛИВАЮЩИХСЯ ЭФФЕКТОВ ПРИ ПРОГНОЗИРОВАНИИ РИСКА ЭКСТРЕМАЛЬНЫХ СОБЫТИЙ*

И. А. Дучицкий¹, В. Ю. Королев², И. А. Соколов³

Аннотация: Построены оценки точности приближения распределений экстремумов специальных случайных сумм масштабными смесями полунормальных законов и обсуждается возможность использования этих результатов при прогнозировании риска экстремальных событий, вызванных катастрофически накапливающимися неблагоприятными эффектами.

Ключевые слова: неоднородные потоки событий; дважды стохастический пуассоновский процесс; отрицательное биномиальное распределение; гамма-распределение; оценка скорости сходимости

1 Введение

По мере возрастания технической и информационной оснащенности человечества вопросы оценивания и прогнозирования надежности технических и информационных систем, а стало быть, и рисков, связанных с их отказами, приобретают все более важное значение. Цивилизация становится все более и более зависимой от возможностей, предоставляемых современными техническими (транспортными, энергетическими, оборонными) и информационными (телекоммуникационными, вычислительными) системами. Одновременно возрастает и зависимость человечества от *надежности и безопасности* технических и информационных систем. В силу этих обстоятельств, естественно, возрастают требования к адекватности математических моделей, используемых для вычисления соответствующих показателей надежности и возможных рисков.

Многие классические методы оценки риска или показателей надежности, разработанные, как правило, в середине XX в., основаны на идеальных предположениях о том, что параметры, характеризующие, скажем, воздействие внешней среды, имеют нормальное распределение, а параметры, характеризующие надежность составных частей изучаемой системы, например время жизни (наработки на отказ), имеют показательное (экспоненциальное) распределение и более общее распределение Вейбулла–Гнеденко. Однако, к со-

жалению, зачастую применение классических методов приводит к недооценке риска катастроф или отказов. Причины иногда имеющей место несостоятельности классических моделей могут быть разными. К примеру, если показатели надежности вычисляются на основе статистических данных, накопленных за определенное время, то, как показано, например, в [1], существенную роль играет однородность потока событий, в результате которых накапливаются статистические данные.

Другими словами, критичным для адекватности классических моделей является асимптотическое постоянство отношения количества зарегистрированных в течение определенного интервала времени экстремальных событий к длине этого интервала времени при неограниченном увеличении последней.

Если асимптотическое постоянство указанного отношения, т. е. его сближение с некоторым числом, имеет место, то классические модели могут давать адекватные результаты. Однако если такого сближения не наблюдается и указанное отношение сильно колеблется, оставаясь случайным (т. е. непредсказуемым), то классические модели неадекватны и могут приводить к весьма существенной недооценке риска. В частности, вместо ожидаемого в соответствии с классической теорией нормального закона в подобных ситуациях (например, если упомянутое выше отношение ведет себя как гамма-распределенная случайная величина) могут возникать, скажем, функции распределения ущерба

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-01-12026-офи-м, 12-07-00109а, 12-07-00115а) и Министерством образования и науки (госконтракт 16.740.11.0133).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; duchik@gmail.com

² Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Российской академии наук; vkorolev@cs.msu.su

³ Институт проблем информатики Российской академии наук; ipiran@ipiran.ru

типа распределения Стьюдента с произвольно малым числом степеней свободы или так называемые дисперсионные гамма-распределения, в частности распределение Лапласа [2–4].

Негативный эффект нежелательных воздействий может проявиться мгновенно, а может накапливаться постепенно. Это приводит к необходимости наряду с самим потоком экстремальных событий рассматривать также и процесс, описывающий «накопленные», негативные воздействия.

Под *катастрофой* (катастрофическим событием) разумно подразумевать превышение модельным процессом некоторого критического уровня, который, например, может быть определен как минимальный уровень, превышение которого накопленным негативным воздействием на рассматриваемую систему (информационную, техническую, экологическую, социальную) ведет к необратимым ее изменениям, в частности к невозможности выполнять системой свои функции в прежнем режиме.

Как показано в [5], если интенсивность потока информативных событий случайна и имеет, скажем, стандартное экспоненциальное распределение, то величина максимума накопленных эффектов также имеет экспоненциальное распределение (но с параметром $\sqrt{2}$), а не ожидаемое в соответствии с классической теорией распределение $G(x) = 2\Phi(\max\{0, x\}) - 1$ максимума винеровского процесса на единичном отрезке, хвост которого убывает при $x \rightarrow \infty$ как $\sqrt{2/\pi}x^{-1}e^{-x^2/2}$. Хвост же упомянутой выше экспоненциальной функции распределения, естественно, убывает существенно медленнее — как $e^{-\sqrt{2}x}$. Это приводит к существенному различию в оценках вероятностей катастрофических воздействий и размеров самих катастрофических воздействий, получаемых на основе этих двух функций распределения. К примеру, квантиль порядка 0,99 функции распределения $G(x)$ равна 2,576. В то же время квантиль того же порядка экспоненциальной функции распределения с параметром $\sqrt{2}$ примерно равна 3,256. Вероятность превышения порога 2,576, «критического» при функции распределения $G(x)$, максимальным суммарным воздействием, имеющим указанную показательную функцию распределения, превышает 0,026, т. е. оказывается более чем в два с половиной раза выше, чем предполагаемая по классической модели.

Этот пример является еще одной наглядной иллюстрацией того, насколько можно недооценить риск катастроф, вызванных накапливающимися эффектами неблагоприятных воздействий, если не принимать во внимание стохастический характер интенсивности потока информативных событий.

Естественно предположить, что негативные эффекты накапливаются как результат неких событий, хаотически рассредоточенных по временной оси, т. е. образующих так называемый хаотический поток событий. Как известно (см., например, [5]), наилучшей моделью однородного хаотического потока событий является пуассоновский процесс, характеризуемый тем обстоятельством, что интервалы времени между событиями потока независимы и имеют одинаковое показательное распределение. Привлекательность пуассоновского процесса в качестве модели однородного дискретного хаоса обусловлена как минимум двумя обстоятельствами.

Во-первых, показательное распределение интервалов времени между событиями пуассоновского потока обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на всей положительной полуоси и имеющих конечное математическое ожидание, а энтропия, как известно, является очень удобной численной характеристикой неопределенности.

Во-вторых, точки (события) пуассоновского потока равномерно распределены на оси времени в том смысле, что для любого конечного интервала времени $[t_1, t_2]$ условное совместное распределение точек пуассоновского потока, попавших в интервал $[t_1, t_2]$, при условии, что в этот интервал попало фиксированное число, скажем n , точек, совпадает с совместным распределением вариационного ряда, построенного по независимой однородной выборке объема n из равномерного на $[t_1, t_2]$ распределения. Равномерное же распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на конечных интервалах, и очень хорошо соответствует общепринятому представлению об абсолютно непредсказуемой ограниченной случайной величине.

Однако в реальных «хаотических» системах хаос практически никогда не бывает однородным в пространстве или времени. Неоднородный и даже стохастический характер интенсивности потока информативных событий может быть обусловлен случайно возникающими (не поддающимися абсолютно надежному прогнозированию) причинами. Как известно, наиболее разумными стохастическими моделями неоднородных хаотических точечных процессов являются *дважды стохастические пуассоновские процессы*, иначе называемые *процессами Кокса* (см., например, [6]).

Для эффективной реализации мер, направленных на повышение надежности и катастрофоустойчивости технических и информационных систем, необходимо уметь вычислять вероятностные харак-

теристики возможных экстремальных воздействий на рассматриваемую систему. Как уже отмечалось выше, при непостоянной (и тем более при стохастической) интенсивности потока экстремальных событий статистические закономерности нежелательных воздействий существенно отличны от того, какими они были бы в однородной ситуации, описываемой классической теорией экстремальных значений. Для вычисления таких мер риска, как квантильные (типа показателей VaR — «Value at Risk»), необходимо иметь более или менее точные аппроксимации для вероятностных распределений величин экстремальных воздействий.

2 Предельное поведение экстремумов обобщенных дважды стохастических пуассоновских процессов

В качестве базовой модели «накапливающегося воздействия» в данной статье рассматриваются экстремумы обобщенных дважды стохастических пуассоновских процессов, представляющих собой суммы случайного числа независимых одинаково распределенных случайных величин, в которых число слагаемых определяется значением дважды стохастического пуассоновского процесса. При этом под катастрофическим неблагоприятным воздействием будет пониматься превышение случайным блужданием, порожденным базовой последовательностью независимых случайных величин, некоторого заданного уровня. Простейшей задачей оценивания или прогнозирования рисков, связанных с такими событиями, является задача вычисления вероятности превышения возможным значением максимума накапливающихся эффектов критического уровня в течение некоторого фиксированного интервала времени. Целесообразность рассмотрения подобных моделей при прогнозировании рисков катастроф диктуется следующими примерами.

Пример 1. Эволюция финансовых индексов хорошо описывается (неоднородным) случайным блужданием. При этом, как известно, если изменение этого индекса в течение биржевого дня будет слишком большим, то, чтобы избежать слишком больших потерь (т.е. финансовых катастроф), торги автоматически прекращаются. Другими словами, если экстремум процесса, описывающего динамику финансового индекса, достигает критического значения, то торги прекращаются. В такой задаче временным горизонтом, на который строится про-

гноз, равен операционному дню, а элементарные слагаемые в сумме — это приращения процесса за единицы времени (например, минуты). На сегодняшний день весьма актуальным примером подобной задачи является вопрос принятия Банком России (Центробанком) решения о проведении валютной (долларовой) интервенции, чтобы остановить процесс падения курса рубля относительно доллара при опасном приближении курса к верхней границе объявленного допустимого коридора. Здесь представляет интерес возможность приближения курса к границе не в какой-то определенный момент времени, а в течение операционного дня, при этом риск, связанный с таким приближением, оценивается заранее (например, после получения информации об эволюции курса в течение предыдущего операционного дня).

Пример 2. Для снабжения некоторой отрасли некоторого региона в течение фиксированного периода времени (например, квартала или зимнего периода), на склад (хранилище) выделяется определенное количество некоторого ресурса (скажем, топлива). Если *суммарные, накопленные* расходы этого ресурса в течение указанного времени превысят выделенный лимит, то в данной отрасли в данном регионе наступит катастрофический коллапс. При этом естественно предположить, что ресурс отпускается потребителям партиями, вообще говоря, случайного объема согласно запросам, возникающим, вообще говоря, в случайные моменты времени. При этом горизонтом прогнозирования, естественно, считается интервал времени между поставками (например, квартал).

Пример 3. При проектировании дамб и водохранилищ необходимо учитывать то обстоятельство, что количество воды в рассматриваемом резервуаре (водохранилище, бассейне реки, озере и т.п.) изменяется случайным образом: оно увеличивается за счет выпадения осадков (в случайные моменты времени) и уменьшается за счет испарения. Если экстремум уровня превысит критический уровень, то происходят события катастрофического характера. Избыток воды вызывает наводнения, ее недостаток — засуху. Ясно, что в силу естественных циклических причин в качестве горизонта прогнозирования разумно взять год или солнечный цикл (11 лет). При этом уровень воды в каждый момент времени является суммой приращений этого уровня за каждый из предыдущих дней.

Определение 1. Случайный процесс $\Lambda(t)$, $t \geq 0$, с неубывающими непрерывными справа траекториями, удовлетворяющий условиям $\Lambda(0) = 0$, $P(\Lambda(t) < \infty) = 1$ ($0 < t < \infty$), называется *случайной мерой*.

Определение 2. Пусть $N_1(t)$ — стандартный пуассоновский процесс, $\Lambda(t)$ — случайная мера, независимая от $N_1(t)$. Случайный процесс $N(t) = N_1(\Lambda(t))$ называется *дважды стохастическим пуассоновским процессом* (или *процессом Кокса*). В таком случае говорят, что процесс Кокса $N(t)$ управляется процессом $\Lambda(t)$ (или что процесс $\Lambda(t)$ контролирует процесс Кокса $N(t)$).

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 $N(t)$. Поэтому иногда процесс $\Lambda(t)$, управляющий процессом Кокса $N(t)$, будет называться *накопленной интенсивностью* процесса $N(t)$.

Несложно убедиться, что для процесса Кокса $N(t)$, управляемого процессом $\Lambda(t)$, справедливы соотношения

$$EN(t) = E\Lambda(t), \quad DN(t) = E\Lambda(t) + D\Lambda(t).$$

Определение 3. Пусть X_1, X_2, \dots — одинаково распределенные случайные величины. Предположим, что при каждом $t \geq 0$ случайные величины $N(t), X_1, X_2, \dots$ независимы. Процесс

$$S(t) = \sum_{j=1}^{N(t)} X_j, \quad t \geq 0, \quad (1)$$

назовем обобщенным процессом Кокса (при этом для определенности считаем, что $\sum_{j=1}^0 = 0$).

Процессы вида (1) играют чрезвычайно важную роль во многих прикладных задачах. Достаточно сказать, что при $\Lambda(t) \equiv \lambda t$ с $\lambda > 0$ процесс $S(t)$ превращается в классический обобщенный пуассоновский процесс, широко используемый при моделировании многих явлений в физике, теории надежности, финансовой и актуарной деятельности, биологии и т. д. Большое число разнообразных прикладных задач, приводящих к обобщенным пуассоновским процессам, описано в книгах [3, 6–8].

Для вычисления риска катастрофических превышений критического уровня случайным блужданием в течение рассматриваемого интервала времени (горизонта планирования) необходимо знать распределение вероятностей максимума суммы

приращений блуждания. К сожалению, практически никогда распределение элементарных приращений не известно, поэтому точное вычисление этого распределения невозможно. Даже когда есть основания принять определенную модель такого распределения, вычисление исключительно трудоемко. Поэтому на практике точное распределение заменяют его аппроксимацией, в качестве которой рассматривается асимптотическая аппроксимация, основанная на (функциональной) центральной предельной теореме, — так называемое полунормальное распределение — распределение модуля нормально распределенной случайной величины, совпадающее с распределением максимума стандартного винеровского процесса на единичном отрезке.

Итак, рассмотрим обобщенный процесс Кокса, определяемый соотношением (1) с $EX_1 = 0$, $0 < \sigma^2 = DX_1 < \infty$. Обозначим

$$\bar{S}(t) = \max_{0 \leq \tau \leq t} S(\tau); \quad \underline{S}(t) = \min_{0 \leq \tau \leq t} S(\tau).$$

Приведем необходимые и достаточные условия слабой сходимости одномерных распределений случайных процессов $\bar{S}(t)$ и $\underline{S}(t)$, скачки которых обладают указанными выше свойствами.

Стандартную нормальную функцию распределения и ее плотность будем обозначать $\Phi(x)$ и $\phi(x)$ соответственно:

$$\Phi(x) = \int_{-\infty}^x \phi(z) dz, \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

Функцию распределения максимума стандартного винеровского процесса на отрезке $[0, 1]$ обозначим $G(\cdot)$,

$$G(x) = 2\Phi(\max\{0, x\}) - 1, \quad x \in \mathbb{R}.$$

Несложно видеть, что если X — случайная величина со стандартным нормальным распределением, то $G(x) = P(|X| < x)$. Символ \implies будет обозначать сходимость по распределению.

Рассмотрим для начала независимые необязательно одинаково распределенные случайные величины Y_1, Y_2, \dots с $EY_i = 0$ и $0 < \sigma_i^2 = DY_i < \infty$, $i \geq 1$. Для $k \geq 1$ обозначим

$$S_k = Y_1 + \dots + Y_k, \quad B_k^2 = \sigma_1^2 + \dots + \sigma_k^2, \\ \bar{S}_k = \max_{1 \leq i \leq k} S_i, \quad \underline{S}_k = \min_{1 \leq i \leq k} S_i.$$

Предположим, что случайные величины Y_1, Y_2, \dots удовлетворяют условию Линдберга: для любого $\alpha > 0$

$$\lim_{k \rightarrow \infty} \frac{1}{B_k^2} \sum_{i=1}^k \int_{|x| \geq \alpha B_k} x^2 dP(Y_i < x) = 0.$$

Хорошо известно, что при таких предположениях

$$\begin{aligned} P\left(\frac{\bar{S}_k}{B_k} < x\right) &\implies G(x), \\ P\left(\frac{S_k}{B_k} < x\right) &\implies 1 - G(-x), \quad k \rightarrow \infty \end{aligned}$$

(это одно из проявлений так называемого принципа инвариантности). Приведем аналог этого результата для обобщенных дважды стохастических пуассоновских процессов с нулевым средним.

Пусть $d(t)$ — положительная функция, неограниченно возрастающая при $t \rightarrow \infty$.

Теорема 1. *Предположим, что $\Lambda(t) \rightarrow \infty$ по вероятности при $t \rightarrow \infty$. Одномерные распределения нормированного процесса экстремумов обобщенного процесса Кокса слабо сходятся к некоторому распределению, т. е.*

$$\frac{\bar{S}(t)}{\sigma\sqrt{d(t)}} \implies \bar{Z}, \quad \frac{S(t)}{\sigma\sqrt{d(t)}} \implies \underline{Z} \quad (t \rightarrow \infty),$$

тогда и только тогда, когда существует неотрицательная случайная величина U такая, что

$$\frac{\Lambda(t)}{d(t)} \implies U \quad (t \rightarrow \infty). \quad (2)$$

При этом

$$\begin{aligned} P(\bar{Z} < x) &= EG\left(\frac{x}{\sqrt{U}}\right), \\ P(\underline{Z} < x) &= 1 - EG\left(-\frac{x}{\sqrt{U}}\right) \quad x \in \mathbb{R}. \end{aligned}$$

Доказательство теоремы 1 приведено в книге [5].

Следствие 1. *В условиях теоремы 1*

$$P\left(\frac{\bar{S}(t)}{\sigma\sqrt{d(t)}} < x\right) \implies G(x)$$

и

$$P\left(\frac{S(t)}{\sigma\sqrt{d(t)}} < x\right) \implies 1 - G(-x)$$

при $t \rightarrow \infty$ тогда и только тогда, когда

$$\frac{\Lambda(t)}{d(t)} \implies 1 \quad (t \rightarrow \infty).$$

Это утверждение является непосредственным следствием теоремы 1 с учетом идентифицируемости семейства масштабных смесей функций распределения G (и, следовательно, $1 - G(-x)$).

3 Оценки скорости сходимости экстремумов обобщенных дважды стохастических пуассоновских процессов

В нестационарных потоках экстремальных событий, описываемых обобщенными дважды стохастическими процессами, согласно теореме 1 полунормальное распределение $G(x)$ трансформируется в его масштабную смесь, в которой смешивающее распределение отражает статистические закономерности поведения случайной интенсивности потока событий. При замене точного распределения указанной асимптотической аппроксимацией, естественно, возникает погрешность, без точного знания которой невозможно гарантированно оценить риски, которые, как правило, характеризуются его *квантильной мерой* VaR (Value at risk). При этом гарантированные риски вычисляются как квантили уточненного порядка, равного базовой допустимой вероятности нежелательного экстремального события плюс погрешность аппроксимации «истинного» распределения максимума обобщенного дважды стохастического пуассоновского процесса его предельным вариантом.

Перейдем к уточнению простейших и наиболее популярных моментных оценок погрешности, использующих информацию о первых моментах элементарных слагаемых, которые можно просто вычислить (оценить) на основе статистической информации. При этом основным объектом исследования будут модели неоднородных хаотических процессов, в которых интенсивность потока информативных событий имеет гамма-распределение с параметром формы, меньшим единицы. Именно такие потоки, к примеру, характеризуют потоки событий на крупных биржах и в итоге приводят к дисперсионным гамма-распределениям (Variance Gamma distributions) приращений базовых финансовых индексов [4].

Как отмечено в работе [9], потоки отказов авиационной техники также характеризуются именно таким распределением интервалов времени между отказами. Более того, в этой же работе показано, что процесс восстановления с гамма-распределенными интервалами времени между восстановлениями является смешанным пуассоновским тогда и только тогда, когда параметр формы гамма-распределения не превосходит единицы.

Лемма 1. *Пусть $\beta_3 = E|X_1|^3 < \infty$ и N_λ — случайная величина, имеющая пуассоновское распределение*

с параметром $\lambda > 0$ и независимая от последовательности $\{X_j\}_{j \geq 1}$. Тогда существует конечная положительная постоянная $C > 0$ такая, что для всех $\lambda \geq 1$ справедливо неравенство

$$\sup_x \left| \mathbb{P} \left(\frac{1}{\sigma \sqrt{\lambda}} \max_{1 \leq i \leq N_\lambda} \sum_{j=0}^i X_j < x \right) - G(x) \right| \leq \frac{C \beta_3}{\sqrt{\lambda} \sigma^3}.$$

Доказательство. В работе [10] доказано утверждение, частным случаем которого является следующая оценка. Пусть X_1, X_2, \dots — независимые одинаково распределенные случайные величины с $\mathbb{E}X_1 = 0$, $0 < \mathbb{D}X_1 = \sigma^2 < \infty$ и $\mathbb{E}|X_1|^3 = \beta_3 < \infty$. Пусть N — целочисленная неотрицательная случайная величина, независимая от последовательности X_1, X_2, \dots . Для $k \geq 1$ положим

$$\bar{S}_k = \max_{1 \leq n \leq k} X_1 + \dots + X_n.$$

Тогда существуют конечные положительные абсолютные постоянные C' и C'' такие, что

$$\begin{aligned} \sup_x \left| \mathbb{P} \left(\frac{\bar{S}_N}{\sigma \sqrt{\mathbb{E}N}} < x \right) - G(x) \right| &\leq \\ &\leq \frac{C'}{\sqrt{\mathbb{E}N}} \frac{\beta_3}{\sigma^3} + C'' \mathbb{E} \left| \frac{N}{\mathbb{E}N} - 1 \right|. \end{aligned}$$

В случае $N = N_\lambda$ по неравенству Маркова имеем

$$\mathbb{E} \left| \frac{N_\lambda}{\mathbb{E}N_\lambda} - 1 \right| \leq \frac{\sqrt{\mathbb{D}N_\lambda}}{\mathbb{E}N_\lambda} = \frac{1}{\sqrt{\lambda}}.$$

Отсюда с учетом всегда имеющего место неравенства $\beta_3/\sigma^3 \geq 1$ вытекает желаемый результат.

Метод оценивания точности аппроксимации распределений экстремумов обобщенных процессов Кокса масштабными смесями функции распределения $G(x)$, используемый далее, основан на лемме 1 и следующем представлении распределения экстремумов обобщенного дважды стохастического пуассоновского процесса $\bar{S}(t)$, управляемого случайной мерой $\Lambda(t)$, справедливым в силу стохастической независимости всех случайных величин и процессов, вовлеченных в определение обобщенного процесса Кокса:

$$\begin{aligned} \mathbb{P}(\bar{S}(t) < x) &= \\ &= \int_0^\infty \mathbb{P} \left(\max_{1 \leq n \leq N_1(\lambda)} \sum_{j=1}^n X_j < x \right) d\mathbb{P}(\Lambda(t) < \lambda), \\ & \quad x \in \mathbb{R}. \end{aligned}$$

Всюду далее будем предполагать, что $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$, $\mathbb{E}|X_1|^3 \equiv \beta^3 < \infty$. Предположим, что

функция $d(t)$ положительна и неограниченно возрастает при $t \rightarrow \infty$. Обозначим

$$\begin{aligned} \rho_t &= \sup_x \left| \mathbb{P} \left(\frac{\bar{S}(t)}{\sqrt{d(t)}} < x \right) - \int_0^\infty G \left(\frac{x}{\sqrt{\lambda}} \right) d\mathbb{P}(\Lambda < \lambda) \right|; \\ \Delta_t &= \sup_x \left| \mathbb{P} \left(\frac{\Lambda(t)}{d(t)} < x \right) - \mathbb{P}(\Lambda < x) \right|. \end{aligned}$$

Теорема 2. Пусть выполнено условие $\mathbb{E}|X_1|^3 < \infty$. Тогда для любого $t > 0$

$$\rho_t \leq C \beta^3 \mathbb{E}[\Lambda(t)]^{-1/2} + \frac{1}{2} \Delta_t,$$

где C — абсолютная постоянная из леммы 1.

Доказательство теоремы 2 приведено в книге [5].

В качестве примера применения теоремы 2 рассмотрим ситуацию, когда при каждом $t > 0$ случайная величина $N(t)$ имеет отрицательное биномиальное распределение с параметрами $r > 0$ и $p \in (0, 1)$:

$$\begin{aligned} \mathbb{P}(N(t) = n) &= \\ &= C_{r+n-1}^n p^r (1-p)^n, \quad n = 0, 1, 2, \dots \quad (3) \end{aligned}$$

Здесь $r > 0$ и $p \in (0, 1)$ — параметры и для нецелых r величина C_{r+n-1}^n определяется как

$$C_{r+n-1}^n = \frac{\Gamma(r+n)}{n! \Gamma(r)}.$$

В частности, при $r = 1$ соотношение (3) задает геометрическое распределение. Отрицательное биномиальное распределение с натуральным r допускает наглядную интерпретацию в терминах испытаний Бернулли. А именно случайная величина, имеющая отрицательное биномиальное распределение с параметрами $r > 0$ и p , — это число испытаний Бернулли, проведенных до осуществления r -й по счету неудачи, если вероятность успеха в одном испытании равна $1 - p$.

Плотность гамма-распределения с параметром формы $r > 0$ и параметром масштаба $s > 0$, как известно, имеет вид:

$$g_{r,s}(x) = \frac{s^r}{\Gamma(r)} e^{-sx} x^{r-1}, \quad x > 0.$$

Как известно, отрицательное биномиальное распределение с параметрами $r > 0$ и $p \in (0, 1)$ представляет собой смешанное пуассоновское распределение, в котором параметр имеет гамма-распределение с параметром масштаба $s = p/(1-p)$ и параметром формы r .

Функцию гамма-распределения с параметром масштаба s и параметром формы r обозначим $G_{r,s}(x)$,

$$G_{r,s}(x) = \int_0^x g_{r,s}(z) dz.$$

Несложно убедиться, что

$$G_{r,s}(x) \equiv G_{r,1}(sx). \quad (4)$$

Случайную величину с функцией распределения $G_{r,s}(x)$ обозначим $U(r, s)$. Хорошо известно, что

$$EU(r, s) = \frac{r}{s}.$$

Будем считать, что параметр r фиксирован, и в качестве случайной величины $\Lambda(t)$ возьмем величину $U(r, s)$, считая, что $t = s^{-1}$:

$$\Lambda(t) = U(r, t^{-1}).$$

В качестве функции $d(t)$ возьмем

$$d(t) \equiv E\Lambda(t) = EU(r, t^{-1}).$$

Несложно видеть, что в терминах новой параметризации

$$EU(r, t^{-1}) = rt.$$

Тогда с учетом (4)

$$\begin{aligned} P\left(\frac{\Lambda(t)}{d(t)} < x\right) &= P(U(r, t^{-1}) < xrt) = \\ &= P(U(r, 1) < xr) = P(U(r, r) < x) = G_{r,r}(x). \end{aligned}$$

Обратим внимание, что функция распределения, стоящая в правой части последнего соотношения, не зависит от t . Поэтому при указанном выборе функции $d(t)$ условие (2) выполняется тривиальным образом, причем для всех $t > 0$

$$\Delta_t = 0.$$

Вычислим $E[\Lambda(t)]^{-1/2}$, предполагая, что выполнено условие $r > 1/2$. Имеем

$$\begin{aligned} E[\Lambda(t)]^{-1/2} &= E[U(r, t^{-1})^{-1/2}] = \\ &= \int_0^\infty \frac{e^{-x/t} x^{r-1-1/2}}{t^r \Gamma(r)} dx = \frac{\Gamma(r-1/2)}{t^{1/2} \Gamma(r)}. \end{aligned}$$

Таким образом, приходим к следующему утверждению, являющемуся частным случаем теоремы 2.

Следствие 2. Пусть случайная величина $N(t)$ имеет отрицательное биномиальное распределение с параметрами $r > 0$ и $p = (1+t)^{-1}$, где $t > 0$. Предположим, что $E|X_1|^3 \equiv \beta^3 < \infty$. Тогда для каждого $t > 0$:

1°. При $r > 1/2$

$$\begin{aligned} \sup_x \left| P(\bar{S}(t) < x\sqrt{rt}) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| &\leq \\ &\leq C \frac{\Gamma(r-1/2)}{\Gamma(r)} \frac{\beta^3}{\sqrt{t}}, \end{aligned}$$

где C — абсолютная постоянная из леммы 1.

2°. При $r < 1/2$

$$\begin{aligned} \sup_x \left| P(\bar{S}(t) < x\sqrt{rt}) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| &\leq \\ &\leq \frac{C^{2r} \beta^{6r}}{\Gamma(r)} \left(\frac{1}{r} + \frac{1}{(1/2-r)} \right) \frac{1}{t^r}, \end{aligned}$$

где C — абсолютная постоянная из леммы 1.

3°. При $r = 1/2$

$$\begin{aligned} \sup_x \left| P\left(\bar{S}(t) < x\sqrt{\frac{t}{2}}\right) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{1/2,1/2}(y) \right| &\leq \\ &\leq \frac{C\beta^3}{\sqrt{\pi}} \left(2 + \ln\left(1 + \frac{t}{C^2\beta^6}\right) \right) \frac{1}{\sqrt{t}}, \end{aligned}$$

где C — абсолютная постоянная из леммы 1.

Доказательство. Пункт 1°, по сути, уже доказан. Докажем пункт 2°. Пусть, как и ранее, $g_{r,r}(\lambda)$ — плотность гамма-распределения с параметрами r, r ; C — константа из леммы 1. Из приведенных выше рассуждений следует, что

$$\begin{aligned} \rho_t &= \sup_x \left| P(\bar{S}(t) < x\sqrt{rt}) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| \leq \\ &\leq \int_0^\infty \sup_x \left| P\left(\frac{1}{\sqrt{\lambda rt}} \sum_{j=1}^{N_{\lambda rt}} X_j < x\right) - G(x) \right| g_{r,r}(\lambda) d\lambda, \end{aligned}$$

где $N_{\lambda rt}$ — пуассоновская случайная величина с параметром λrt . Заметим, что оценка равномерного расстояния между функциями распределения, даваемая леммой 1, при малых λ заведомо больше

единицы, что тривиально для равномерного расстояния между функциями распределения. Используя этот факт, получаем

$$\begin{aligned} & \int_0^\infty \sup_x \left| \mathbb{P} \left(\frac{1}{\sqrt{\lambda r t}} \sum_{j=1}^{N_{\lambda r t}} X_j < x \right) - \right. \\ & \quad \left. - G(x) \right| g_{r,r}(\lambda) d\lambda \leq \\ & \leq \int_0^\infty \min \left\{ 1, \frac{C\beta^3}{\sqrt{\lambda r t}} \right\} g_{r,r}(\lambda) d\lambda = \\ & = \int_0^{C^2\beta^6/(rt)} g_{r,r}(\lambda) d\lambda + \int_{C^2\beta^6/(rt)}^\infty \frac{C\beta^3}{\sqrt{\lambda r t}} g_{r,r}(\lambda) d\lambda \equiv \\ & \equiv I_1 + I_2. \end{aligned}$$

Оценим первый интеграл:

$$\begin{aligned} I_1 &= \int_0^{C^2\beta^6/(rt)} g_{r,r}(\lambda) d\lambda = \\ &= \frac{1}{\Gamma(r)} \int_0^{C^2\beta^6/(rt)} r^r e^{-r\lambda} \lambda^{r-1} d\lambda = \\ &= \frac{1}{\Gamma(r)} \int_0^{C^2\beta^6/t} e^{-u} u^{r-1} du \leq \\ &\leq \frac{1}{\Gamma(r)} \int_0^{C^2\beta^6/t} u^{r-1} du = \frac{1}{r\Gamma(r)} \frac{C^{2r}\beta^{6r}}{t^r}. \end{aligned}$$

Оценим второй интеграл. При $r < 1/2$ имеем:

$$\begin{aligned} I_2 &= \frac{C\beta^3}{\Gamma(r)\sqrt{rt}} \int_{C^2\beta^6/(rt)}^\infty \frac{r^r e^{-r\lambda} \lambda^{r-1}}{\sqrt{\lambda}} d\lambda = \\ &= \frac{C\beta^3}{\Gamma(r)\sqrt{t}} \int_{C^2\beta^6/t}^\infty e^{-u} u^{r-3/2} du \leq \\ &\leq \frac{C\beta^3}{\Gamma(r)\sqrt{t}} e^{-C^2\beta^6/t} \int_{C^2\beta^6/t}^\infty u^{r-3/2} du \leq \\ &\leq \frac{C\beta^3}{\Gamma(r)\sqrt{t}} \left(r - \frac{1}{2} \right)^{-1} u^{r-1/2} \Big|_{u=C^2\beta^6/t}^{u=\infty} = \\ &= \frac{C\beta^3}{\Gamma(r)\sqrt{t}} \frac{1}{1/2-r} \frac{C^{2r-1}\beta^{6r-3}}{t^{r-1/2}} = \\ &= \frac{C^{2r}\beta^{6r}}{\Gamma(r)(1/2-r)} \frac{1}{t^r}. \end{aligned}$$

Таким образом,

$$\begin{aligned} \sup_x \left| \mathbb{P}(\bar{S}(t) < x\sqrt{rt}) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| &\leq \\ &\leq \frac{C^{2r}\beta^{6r}}{\Gamma(r)} \left(\frac{1}{r} + \frac{1}{(1/2-r)} \right) \frac{1}{t^r}, \end{aligned}$$

что и требовалось доказать.

Докажем пункт 3°. Несложно видеть, что в этом случае

$$\begin{aligned} I_2 &= \frac{C\beta^3}{\Gamma(r)\sqrt{rt}} \int_{C^2\beta^6/(rt)}^\infty \frac{r^r e^{-r\lambda} \lambda^{r-1}}{\sqrt{\lambda}} d\lambda = \\ &= \frac{C\beta^3}{\Gamma(r)\sqrt{t}} \int_{C^2\beta^6/t}^\infty e^{-u} u^{r-3/2} du = \\ &= \frac{C\beta^3}{\Gamma(1/2)\sqrt{t}} \int_{C^2\beta^6/t}^\infty e^{-u} u^{-1/2} du = \\ &= \frac{C\beta^3}{\sqrt{\pi t}} E_1\left(\frac{C^2\beta^6}{t}\right) \leq \frac{C\beta^3}{\sqrt{\pi t}} \ln\left(1 + \frac{t}{C^2\beta^6}\right), \end{aligned}$$

где $E_1(\cdot)$ — интегральная показательная функция. Тогда

$$\begin{aligned} \sup_x \left| \mathbb{P}(\bar{S}(t) < x\sqrt{rt}) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| &\leq \\ &\leq \frac{C\beta^3}{\sqrt{\pi}} \left(2 + \ln\left(1 + \frac{t}{C^2\beta^6}\right) \right) \frac{1}{\sqrt{t}}. \end{aligned}$$

Теорема доказана.

Несложно видеть, что при $r = 1$ предельная функция распределения представляет собой показательное распределение с параметром $\sqrt{2}$.

С помощью теоремы 3 и следствия 2 можно получить следующую оценку точности приближения распределения экстремумов геометрических случайных сумм показательным распределением.

Следствие 3. Пусть случайная величина $N(t)$ имеет геометрическое распределение с параметром $p = (1+t)^{-1}$, где $t > 0$. Предположим, что $E|X_1|^3 \equiv \beta^3 < \infty$. Тогда для каждого $t > 0$

$$\sup_{x \geq 0} |\mathbb{P}(\bar{S}(t) < x\sqrt{t}) - (1 - e^{-\sqrt{2}x})| \leq C\sqrt{\pi} \frac{\beta^3}{\sqrt{t}},$$

где C — абсолютная постоянная из леммы 1.

4 Обсуждение

Как показывает следствие 2, при $t \rightarrow \infty$

$$\sup_x \left| P(\bar{S}(t) < x\sqrt{rt}) - \int_0^\infty G\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| = \begin{cases} O\left(\frac{1}{\sqrt{t}}\right), & \text{если } r > \frac{1}{2}; \\ O\left(\frac{\ln t}{\sqrt{t}}\right), & \text{если } r = \frac{1}{2}; \\ O\left(\frac{1}{t^r}\right), & \text{если } r < \frac{1}{2}. \end{cases}$$

В работе [11] показано, что именно такая зависимость от параметра r присуща скорости сходимости распределений «асимптотически нормальных» в классическом смысле статистик к распределению Стьюдента с $2r$ степенями свободы при замене объема выборки случайной величиной с отрицательным биномиальным распределением (3) с $p = (1 + t)^{-1}$ при $t \rightarrow \infty$. Распределение Стьюдента с $2r$ степенями свободы является масштабной смесью нормальных законов с нулевым средним, в которой смешивающим является гамма-распределение $G_{r,r}$. В работе [12] на примере сумм случайного числа независимых случайных величин с индексом, имеющим указанное отрицательное биномиальное распределение, показано, что такой порядок скорости сходимости является правильным. Таким образом, результаты данной статьи вполне согласуются с упомянутыми работами и распространяют указанную закономерность на «асимптотически полунормальные» статистики, каковыми являются максимальные суммы.

Литература

1. 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. 248 с.
2. 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.
3. Королев В. Ю., Бенинг В. Е., Шоргин С. Я. Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011. 620 с.
4. 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.
5. Королев В. Ю., Соколов И. А. Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008.
6. Bening V., Korolev V. Generalized poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002.
7. Gnedenko B. V., Korolev V. Yu. Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996.
8. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 510 с.
9. Gleser L. J. The gamma distribution as a mixture of exponential distributions: Technical Report # 87-28. — West Lafayette: Purdue University, 1987. 6 p.
10. Korolev V. Yu., Selivanova D. O. Convergence rate estimates in some limit theorems for maximum random sums // J. Math. Sci., 1995. Vol. 76. No. 1. P. 2163–2168.
11. 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: Межвузовский сб. научных тр. — Пермь: ПГУ, 2006. С. 118–134.
12. Нефедова Ю. С. Оценки скорости сходимости в предельной теореме для отрицательных биномиальных случайных сумм // Статистические методы оценивания и проверки гипотез: Межвузовский сб. научных тр. — Пермь: ПГУ, 2011. С. 46–61.

ОБ АДАПТИВНЫХ СТРАТЕГИЯХ И УСЛОВИЯХ ИХ СУЩЕСТВОВАНИЯ*

М. Г. Коновалов¹

Аннотация: Рассматривается задача оптимального управления в отсутствие априорной информации об управляемом объекте. Решением задачи является построение адаптивных стратегий на основе наблюдений, доступных в процессе управления. Изучаются некоторые условия адаптивной управляемости объекта. В качестве математической модели используются управляемые случайные последовательности.

Ключевые слова: управляемые случайные последовательности; адаптивные стратегии; условия существования

1 Введение

Тема статьи относится к области адаптивных методов обработки информации с целью принятия оптимальных решений. Потребность в адаптивном подходе возникает в задачах с большой информационной неопределенностью, что наиболее характерно для телекоммуникационных систем, автоматизированных производственных процессов, робототехники и других сфер, неразрывно связанных с компьютерной обработкой информации. Понятие неопределенности многозначно и связано с отсутствием априорных сведений, недетерминированностью, а также с неполнотой наблюдений. К перечисленным факторам в нарастающей степени добавляется «избыточность» информации, которая порождается чрезмерно прогрессирующими объемами передаваемой и хранимой информации и обусловлена экспоненциальным ростом пропускной способности телекоммуникационных сетей, а также емкостей носителей информации.

Идея адаптации (приспособления, самоорганизации), заимствованная из биологического мира, начала активно эксплуатироваться в науке примерно с середины прошлого века. Кратко, она заключается в том, чтобы, целенаправленно взаимодействуя с окружающей средой, отбирать и использовать поступающую информацию, необходимую для принятия оптимальных решений с точки зрения поставленной цели.

Данная статья посвящена теоретическим аспектам адаптации. В качестве исходного представления использована схема, которая опирается на представление о паре «объект–субъект», взаимодействующей в дискретном времени путем попеременного обмена сигналами. При этом субъект воздействует на объект с помощью управлений, по-

лучая в ответ сигналы, называемые наблюдениями. Действия субъекта преследуют цель, выраженную в наличии определенных свойств у траектории наблюдений.

Основная отличительная особенность заключается в предположении, что действия субъекта происходят при недостаточной информации об объекте. В качестве математической модели объекта взята конструкция управляемой случайной последовательности. В терминах этого аппарата легко очерчиваются четыре аспекта информационной неопределенности:

- (1) недетерминированность понимается как стохастичность;
- (2) недостаток информации об объекте трактуется как неполное знание вероятностного распределения, задающего процесс;
- (3) неполнота наблюдений означает, что состояния процесса наблюдаются лишь частично;
- (4) недостаток знаний выражается в неумении найти или рассчитать ту или иную характеристику, связанную со случайной последовательностью, даже при наличии априорной информации о распределении процесса и полной его наблюдаемости.

Субъект ассоциируется с алгоритмом, согласно которому выбираются управления, регулирующие траекторию случайной последовательности. Такой алгоритм принято называть стратегией управления. Задача заключается в том, чтобы выбрать стратегию, достигающую цели в ситуации, когда информация субъекта об объекте ограничена. По-другому можно сказать, что речь идет о построении стратегии, достигающей цели (в данном случае —

* Работа выполнена при поддержке РФФИ, грант № 11-07-00112.

¹ Институт проблем информатики Российской академии наук, mkonov@ipiran.ru

максимизации предельного среднего дохода) для любого процесса из некоторого заданного класса объектов. Такие стратегии называют адаптивными по отношению к заданному классу объектов [1].

В разд. 2 даются формальные определения объекта, цели и адаптивной стратегии управления.

В разд. 3 анализируются условия существования адаптивной стратегии. В качестве необходимых условий обсуждаются два требования, которые, как представляется, должны выполняться из интуитивных соображений.

Первое из необходимых условий связано с принципиальной особенностью адаптивных стратегий, которые, прежде чем выйти на «оптимальный режим», должны затратить некоторое время на «обучение». (На самом деле в рассматриваемой постановке процесс обучения для адаптивных стратегий длится даже неограниченно долго.) Естественно предположить, что подобные стратегии могут реализоваться, только если в процессе обучения не будут совершены «непоправимые ошибки». Это соображение раскрывается на примерах и получает формальное описание.

Второе необходимое условие является менее очевидным. Оно связано с гипотезой о том, что адаптивная стратегия управления классом случайных последовательностей существует лишь тогда, когда для данного класса возможно построение так называемой адаптивной стратегии перебора. Это выражается в том, что существует и заранее известно некоторое счетное множество вариантов поведения, среди которого для данного класса обязательно найдется оптимальный или близкий к нему вариант. Данное соображение также иллюстрировано примерами и приведена теорема о критерии существования адаптивной стратегии для определенного класса объектов.

Подход, использованный в статье, а также полученные результаты являются продолжением направления, представленного в работе [2].

2 Постановка задачи адаптивного управления

Пусть время t пробегает значения $0, 1, \dots$ и пусть заданы измеримые пространства (X, \mathbf{X}) , (Y, \mathbf{Y}) , (Z, \mathbf{Z}) (соответственно пространства *состояний, управлений и наблюдений*).

Общая траектория процесса упорядочена в виде последовательности $x_0, y_1, z_1, x_1, \dots, x_{t-1}, y_t, z_t, x_t, \dots$. Предыстория процесса до момента t включительно обозначается как

$$x^t = x_0^t = (x_0, \dots, x_{t-1}); \quad y^t = y_1^t = (y_1, \dots, y_{t-1}); \\ z^t = z_1^t = (z_1, \dots, z_{t-1}).$$

Траектории процесса определяются последовательностями условных вероятностных распределений μ, ν и σ .

Последовательность $\mu = (\mu_0, \mu_1, \dots, \mu_t, \dots)$ задает механизм смены состояний. В этой последовательности μ_0 — вероятностное распределение на (X, \mathbf{X}) ; $\mu_t = \mu_t(A|x^{t-1}, y^t)$, $t > 0$ — условная (переходная) вероятность, которая при любых наборах (x^{t-1}, y^t) является вероятностной мерой на (X, \mathbf{X}) и при любом $A \in X$ является измеримой функцией относительно x^{t-1}, y^t .

Последовательность $\nu = (\nu_1, \dots, \nu_t, \dots)$ задает механизм появления наблюдений. В этой последовательности каждый элемент $\nu_t = \nu_t(C|x^{t-1}, y^t)$, $t > 0$, представляет собой условное распределение, которое при любом условии является вероятностной мерой на (Z, \mathbf{Z}) и для любого $C \in Z$ является измеримой функцией относительно переменных, стоящих в условии. Пара $o = (\mu, \nu)$ называется объектом.

Последовательность $\sigma = (\sigma_1, \dots, \sigma_t, \dots)$ называется (допустимой) *стратегией* и определяет выбор управлений. В этой последовательности: $\sigma_1 = \sigma_1(\cdot)$ — вероятностная мера на (Y, \mathbf{Y}) ; $\sigma_{t+1} = \sigma_{t+1}(B|y^t, z^t)$, $t > 0$, — условная вероятность, которая при любых y^t, z^t является вероятностной мерой на (Y, \mathbf{Y}) и при любом $B \in Y$ является измеримой функцией относительно y^t, z^t . Элементы последовательности σ называются (допустимыми) *правилами*.

Введем обозначение для прямых произведений множеств:

$$\Omega_0 = X; \quad \Omega_t = X^{t+1} \times Y^t \times Z^t, \quad t > 0,$$

а также для наименьших σ -алгебр, порожденных соответствующими σ -алгебрами:

$$\mathbf{F}_0 = \mathbf{X}; \quad \mathbf{F}_t = \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \otimes \mathbf{X} \otimes \dots \otimes \mathbf{Y} \otimes \mathbf{Z} \otimes \mathbf{X}$$

(\mathbf{X} повторяется $t + 1$ раз, \mathbf{Y} и \mathbf{Z} — t раз, $t > 0$).

Положим

$$\Omega = \prod_{t \geq 0} \Omega_t; \quad \mathbf{F} = \otimes_{t \geq 0} \mathbf{F}_t.$$

Согласно общей теории [3] последовательности $o = (\mu, \nu)$ и σ порождают на пространстве (Ω, \mathbf{F}) вероятностную меру $\mathbf{P} = \mathbf{P}_{o, \sigma} = \mathbf{P}_{\mu, \nu, \sigma}$, которая согласована с элементами этих последовательностей следующим образом. Случайные последовательности

$$x_t = x_t(\omega); \quad y_{t+1} = y_{t+1}(\omega); \\ z_{t+1} = z_{t+1}(\omega), \quad \omega \in \Omega, \quad t \geq 0,$$

удовлетворяют соотношениям:

$$\begin{aligned} \mathbf{P}(x_0(\omega) \in A_0) &= \int_{A_0} \mu_0(dx_0); \\ \mathbf{P}(x_0(\omega) \in A_0, y_1(\omega) \in B_1, z_1(\omega) \in C_1, \dots \\ &\dots, y_t(\omega) \in B_t, z_t(\omega) \in C_t, x_t(\omega) \in A_t) = \\ &= \int_{A_0} \mu_0(dx_0) \int_{B_1} \sigma_1(dy_1) \int_{C_1} \nu_1(dz_1|x_0, y_1) \dots \\ &\dots \int_{B_t} \sigma_t(dy_t|y^{t-1}, z^{t-1}) \int_{C_t} \nu_t(dz_t|x^{t-1}, y^t) \times \\ &\quad \times \int_{A_t} \mu_t(dx_t|x^{t-1}, y^t) \end{aligned}$$

для любых $A_t \in X, B_{t+1} \in Y, C_{t+1} \in Z, t \geq 0$.

По определению стратегии, ее правила зависят от предыдущих управлений и наблюдений, но не от предыдущих состояний. Это соответствует предположению о том, что состояния объекта не наблюдаемы в ходе процесса управления. В частных случаях объект $o = (\mu, \nu)$ может, конечно, описывать полностью наблюдаемый процесс. Например, если все множества X_t содержат один и тот же единственный элемент. Другой простой пример — когда наблюдения тождественны состояниям. Однако на самом деле, как показывает лемма 1, с формальной точки зрения рассмотрение объекта с «ненаблюдаемой» компонентой всегда можно заменить изучением полностью наблюдаемого процесса.

Лемма 1. Для любого объекта $o = (\mu, \nu)$ условная вероятность $\mathbf{P}(dz_t|y^t, z^{t-1})$ не зависит от стратегии σ при любых $t > 0$.

Доказательство. Согласно отмеченной выше согласованности условных распределений μ, ν, o и порождаемой ими меры \mathbf{P} имеем соотношения:

$$\begin{aligned} I_1 &= \mathbf{P}(y_1(\omega) \in B_1, z_1(\omega) \in C_1) = \\ &= \mathbf{P}(x_0(\omega) \in X_0, y_1(\omega) \in B_1, z_1(\omega) \in C_1) = \\ &= \int_{X_0} \int_{B_1} \int_{C_1} \mu_0(dx_0) \sigma_1(dy_1) \nu_1(dz_1|x_0, y_1) = \\ &= \int_{B_1} \int_{C_1} \sigma_1(dy_1) \int_{X_0} \mu_0(dx_0) \nu_1(dz_1|x_0, y_1), \end{aligned}$$

справедливые при любых $B_1 \in Y$ и $C_1 \in Z$. Кроме того, по определению условной вероятности

$$I_1 = \int_{B_1} \int_{C_1} \sigma_1(dy_1) \mathbf{P}(dz_1|y_1).$$

Сравнивая оба выражения для I_1 , получаем, что

$$\mathbf{P}(dz_1|y_1) = \int_{X_0} \mu_0(dx_0) \nu_1(dz_1|x_0, y_1),$$

т.е. утверждение леммы справедливо для $t = 1$. Пусть оно верно для $n = 1, 2, \dots, t-1$. Для любых $B_1 \in Y, C_1 \in Z, \dots, B_{t-1} \in Y, C_t \in Z$ имеем:

$$\begin{aligned} I_t &= \mathbf{P}(y_1(\omega) \in B_1, z_1(\omega) \in C_1, \dots \\ &\dots, y_t(\omega) \in B_t, z_t(\omega) \in C_t) = \\ &= \mathbf{P}(x_0(\omega) \in X, y_1(\omega) \in B_1, z_1(\omega) \in C_1, \dots \\ &\dots, x_{t-1}(\omega) \in X, y_t(\omega) \in B_t, z_t(\omega) \in C_t) = \\ &= \int_X \int_{B_1} \int_{C_1} \dots \\ &\dots \int_X \int_{B_t} \int_{C_t} \mu_0(dx_0) \sigma_1(dy_1) \nu_1(dz_1|x_0, y_1) \dots \\ &\dots \mu_{t-1}(dx_{t-1}|x^{t-2}, y^{t-1}) \sigma_t(dy_t|y^{t-1}, z^{t-1}) \times \\ &\quad \times \nu_t(dz_t|x^{t-1}, y^t) = \\ &= \int_{B_1} \sigma_1(dy_1) \int_{C_1} \int_{B_2} \sigma_2(dy_2|z_1) \dots \\ &\dots \int_{C_{t-1}} \int_{B_t} \sigma_t(dy_t|y^{t-1}, z^{t-1}) \times \\ &\quad \times \int_X \mu_0(dx_0) \nu_1(dz_1|x_0, y_1) \dots \\ &\dots \int_{X_{t-1}} \mu_{t-1}(dx_{t-1}|x^{t-2}, y^{t-1}) \nu_t(dz_t|x^{t-1}, y^t) = \\ &= \int_{B_1} \sigma_1(dy_1) \int_{C_1} \int_{B_2} \sigma_2(dy_2|z_1) \dots \\ &\dots \int_{C_{t-1}} \int_{B_t} \sigma_t(dy_t|y^{t-1}, z^{t-1}) \int_{C_t} \mathbf{P}(dz_t|y_1) \dots \\ &\quad \dots \mathbf{P}(dz_{t-1}|y^{t-1}, z^{t-2}) \mathbf{P}(dz_t|y^t, z^{t-1}). \end{aligned}$$

Отсюда получаем, что

$$\begin{aligned} \mathbf{P}(dz_t|y_1) \dots \mathbf{P}(dz_{t-1}|y^{t-1}, z^{t-2}) \mathbf{P}(dz_t|y^t, z^{t-1}) &= \\ &= \int_X \mu_0(dx_0) \nu_1(dz_1|x_0, y_1) \dots \\ &\dots \int_X \mu_{t-1}(dx_{t-1}|x^{t-2}, y^{t-1}) \nu_t(dz_t|x^{t-1}, y^t). \end{aligned}$$

Следовательно, по предположению индукции $\mathbf{P}(dz_t|y^t, z^{t-1})$ не зависит от σ .

Таким образом, не уменьшая общности, можно ограничиться (что и будет сделано в оставшейся части текста) рассмотрением полностью наблюда-

емых объектов $o = \mu$, управляемых (допустимыми) стратегиями σ с правилами вида

$$\sigma_1 = \sigma_1(\cdot); \quad \sigma_{t+1} = \sigma_{t+1}(\cdot | y^t, x^t), \quad t > 0.$$

(Множество всех таких стратегий при заданных пространствах состояний и управлений далее обозначается через Σ .) В этом случае вероятностная мера $\mathbf{P} = \mathbf{P}_{\mu, \sigma}$ определена на пространстве (Ω, \mathbf{F}) , в котором $\Omega = \prod_{t \geq 0} X^{t+1} \times Y^t$, $\mathbf{F} = \otimes_{t \geq 0} \mathbf{F}_t$, где $\mathbf{F}_0 = \mathbf{X}$; $\mathbf{F}_t = \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{X} \otimes \mathbf{X} \otimes \dots \otimes \mathbf{Y} \otimes \mathbf{X}$ и согласована с последовательностями μ и σ . Через \mathbf{F}_t обозначена σ -алгебра, порожденная предысторией (x^t, y^t) до момента t включительно.

В то же время необходимо заметить, что предположение о наличии «двухступенчатой» структуры у объектов (состояние—наблюдение) может принести пользу при их изучении. Так происходит, например, в теории частично наблюдаемых управляемых марковских процессов.

Предположим далее, что на наблюдаемой части траектории процесса задан одношаговый доход (в момент t), и будем считать, что этот доход имеет вид $g_t = g(x_t)$, где $g : X \rightarrow (0, 1) \subset \mathbb{R}$ — измеримая числовая функция со значениями из интервала $(0, 1)$.

Обозначим через $v_{t,s} = s^{-1} \sum_{n=1}^s g_{t+n}$ среднее арифметическое доходов на промежутке от $t+1$ до $t+s$ ($t \geq 0, s \geq 1$).

Если объект μ управляется согласно стратегии σ , то число

$$w_t(\mu, \sigma) = \sup \left\{ c : \mathbf{P}_{\mu, \sigma} \left(\lim_{s \rightarrow \infty} v_{t,s} > c \right) = 1 \right\}$$

характеризует получаемый при этом гарантированный предельный средний доход начиная с момента $t = 1$. Поскольку $\lim_{s \rightarrow \infty} v_{t,s}$ не зависит от t , то $w_0(\mu, \sigma) = w_1(\mu, \sigma) = w_2(\mu, \sigma) = \dots$. Величина $w(\mu, \sigma) = w_0(\mu, \sigma)$ играет в дальнейшем роль целевой функции и называется просто *доходом* (при управлении объектом μ с помощью стратегии σ).

Из определения дохода следует, что для любого $t > 0$ выполняется условие

$$\mathbf{P}_{\mu, \sigma} \left(\lim_{s \rightarrow \infty} v_{t,s} \geq w(\mu, \sigma) | \mathbf{F}_{t-1} \right) = 1$$

почти наверное. Столь общее определение дохода, без предположений об эргодичности, оказывается полезным в теоретических рассуждениях, однако на практике все же среднее арифметическое ведет себя более или менее регулярным образом. Поэтому введем следующее определение.

Стратегия σ называется *эргодической* по отношению к классу M , если для любого объекта $\mu \in M$ и

любого $\varepsilon > 0$ выполняется условие $\sum_{s=1}^{\infty} a_s < \infty$, где $a_s = a_s(\mu, \sigma, \varepsilon) = \sup_{t \geq 0} \mathbf{P}_{\mu, \sigma} (|v_{t,s} - w(\mu, \sigma)| > \varepsilon | \mathbf{F}_t)$.

Обозначим еще

$$W = W(\mu) = \sup_{\sigma} w(\mu, \sigma),$$

где точная верхняя грань берется по всем допустимым стратегиям. Стратегия σ называется *ε -оптимальной*, если выполняется неравенство

$$w(\mu, \sigma) \geq W - \varepsilon, \quad \varepsilon \geq 0.$$

Далее объекты будут объединяться в множества объектов (классы объектов). При этом без дополнительных оговорок всюду предполагается, что

- все объекты из класса имеют одинаковые пространства состояний, управлений (и наблюдений);
- в качестве множества допустимых стратегий берется определенное выше множество Σ ;
- функция одношаговых доходов g одна и та же для всех объектов.

Пусть M — класс объектов. Стратегия σ является равномерно ε -оптимальной относительно этого класса, если последнее неравенство выполняется для всех $\mu \in M$. Такую стратегию будем называть также *ε -адаптивной* по отношению к классу M . Класс объектов, для которого существует ε -адаптивная стратегия, называется *ε -адаптивно управляемым*. (Если $\varepsilon = 0$, то приставка « ε -» в этих определениях опускается.)

Основная задача адаптивного управления заключается в построении адаптивных стратегий для различных классов объектов.

К настоящему времени получено много решений для многочисленных вариантов этой задачи. Подобные результаты являются фактически достаточными условиями адаптивной управляемости. Ниже, однако, будет уделено внимание также необходимым условиям существования адаптивных стратегий. Подчеркнем, что рассматриваемая постановка задачи предполагает, по сути, наличие лишь минимальной априорной информации об объекте управления — необходимо знать множество управлений Y .

3 Некоторые условия адаптивной управляемости

Пусть $\mu \in M$ — фиксированный объект, а $\sigma \in \Sigma$ — фиксированная стратегия из некоторой среды. Набор, состоящий из первых t правил стратегии σ , будем обозначать через $\sigma^t = (\sigma_1, \dots, \sigma_t)$.

Таким образом, $\sigma = (\sigma^t, \sigma_{t+1}, \sigma_{t+2}, \dots)$. Положим

$$w_t^*(\mu, \sigma) = w_t^*(\mu, \sigma^t) = \sup_{\sigma_{t+1}, \sigma_{t+2}, \dots} w_t(\mu, \sigma),$$

где верхняя грань берется по всем допустимым правилам начиная с момента $t+1$. В этих обозначениях $w_0^*(\mu, \sigma) = W(\mu)$. Ясно, что $W(\mu) \geq w_1^*(\mu, \sigma) \geq w_2^*(\mu, \sigma) \geq \dots$

Стратегию σ назовем ε -повреждающей для объекта μ , если

$$\inf \{t : w_t^*(\mu, \sigma) < W(\mu) - \varepsilon\} < \infty, \quad \varepsilon > 0.$$

Пример 1 показывает, что существуют объекты, для которых каждая стратегия — ε -повреждающая (с разными значениями ε).

Пример 1. Множество X состояний объекта μ образовано точками с неотрицательными целочисленными координатами на плоскости, $X = \{(i, j), i \geq 0, j \geq 0\}$. Множество управлений $Y = \{1; 2\}$. Начальное состояние $x_0 = (0, 0)$. Детерминированные переходы между состояниями заданы следующим образом ($t > 0, i \geq 0$):

$$\begin{aligned} \mu_t(x_t = (i+1, 0) | x_{t-1} = (i, 0), y_t = 1) &= 1; \\ \mu_t(x_t = (i, j+1) | x_{t-1} = (i, j), y_t = 1) &= 1, \quad j > 0; \\ \mu_t(x_t = (i, j+1) | x_{t-1} = (i, j), y_t = 2) &= 1, \quad j \geq 0. \end{aligned}$$

Одношаговые доходы определены как $g(i, 0) = 0, g(i, j) = 1 - 2^{-i}$ для $i \geq 0, j > 0$.

Стратегия, состоящая из бесконечного повторения управления 1, приносит доход 0. Стратегия, в которой управление 2 первый раз применяется (детерминировано) в момент t , приносит доход $1 - 2^{t-1}$, что меньше максимально возможного на 2^{t-1} . Рандомизация правил и их зависимость от предыстории не вносит принципиальных изменений — каждая стратегия остается ε -повреждающей относительно предельно наибольшего, но недостижимого значения 1.

В примере 2 оптимальная стратегия для любого объекта из класса является повреждающей для остальных объектов.

Пример 2. Пусть $X = \{0, 1, 2, \dots\} \cup \{a, b\}$; $Y = \{0; 1\}$; $g(a) = 1; g(b) = g(i) = 0, i \geq 0$. Задан счетное множество объектов $M = \{\mu^{(k)}, k = 0, 1, \dots\}$. Пусть для всех k :

$$\begin{aligned} \mu^{(k)}(x_0 = 0) &= 1; \\ \mu^{(k)}(x_{t+1} = i+1 | x_t = i, y_t = 0) &= 1, \quad i \geq 0; \\ \mu^{(k)}(x_{t+1} = a | x_t = k, y_t = 1) &= 1; \\ \mu^{(k)}(x_{t+1} = b | x_t = i, y_t = 1) &= 1, \quad i \neq k; \\ \mu^{(k)}(x_{t+1} = a | x_t = a, y_t = j) &= \\ = \mu^{(k)}(x_{t+1} = b | x_t = b, y_t = j) &= 1, \quad j = 0 \vee 1. \end{aligned}$$

Таким образом, состояния a и b — поглощающие, причем в состоянии a , приносящее максимальный доход, объект $\mu^{(k)}$ может попасть, только если применить управление 1, находясь в состоянии k . Первые (существенные) правила оптимальной стратегии для объекта $\mu^{(k)}$ требуют применения управления 0 до достижения состояния k , а затем применения в этом состоянии управления 1. Однако такая стратегия является повреждающей для всех остальных объектов. Следовательно, для класса M не существует равномерно оптимальной стратегии.

Пусть M — класс объектов. Обозначим через $\Sigma_\varepsilon(\mu)$ множество ε -повреждающих стратегий для объекта $\mu, \mu \in M$. Положим $\Sigma_\varepsilon(M) = \bigcap_{\mu \in M} (\Sigma \setminus \Sigma_\varepsilon(\mu))$.

Лемма 2. Для того чтобы существовала ε -адаптивная стратегия, необходимо, чтобы $\Sigma_\varepsilon(M) \neq \emptyset$.

Доказательство. Если $\Sigma_\varepsilon \neq \emptyset$, то любая допустимая стратегия хотя бы для одного из объектов является ε -повреждающей и, следовательно, не является ε -оптимальной, а потому не может быть равномерно ε -оптимальной по отношению к классу M .

В примере 3, несмотря на наличие повреждающих стратегий, адаптивная стратегия существует.

Пример 3. Пусть $X = Y = \{1, \dots, K\}$ и пусть задана детерминированная функция $f : X \rightarrow X$, которая представляет собой циклическую подстановку на множестве X , т.е. $f(i) \neq f(j)$, если $i \neq j$; $i, j = 1, \dots, K$. Рассмотрим следующий неоднородный во времени детерминированный объект. Положим

$$\begin{aligned} \mu_0(x_0 = 1) &= 1; \\ \mu_t(x_t = f(k) | x^{t-1}, y^t) &= I_{\{y_t = k\}}, \quad 0 < k, t \leq K; \\ \mu_t(x_t = f(k) | x^{t-1}, y^t) &= I_{\{y_{K+1} = k\}}, \\ & \quad 0 < k \leq K, \quad t > K \end{aligned}$$

(I_A — индикатор события A).

Одношаговые доходы определим как $g(i) = i, i \in X$.

Так определенный объект обозначим через μ^f . Ясно, что для этого объекта траектория управляемого процесса, начиная с момента $K+1$, и, следовательно, доход зависят исключительно от управления, примененного в момент $K+1$. Доход будет максимален (и равен K) тогда и только тогда, когда $y_{K+1} = k' = k^*(f) = \arg \max_{1 \leq k \leq K} f(k)$.

Пусть $M = \{\mu^f\}$ — совокупность всех объектов данного вида (которая содержит $K!$ элементов).

Очевидно, для класса M существует равномерно оптимальная стратегия, доставляющая доход, равный K . Например, достаточно вначале в моменты $t = 1, \dots, K$ по одному разу применить каждое из управлений, а затем в момент $K + 1$ применить управление k^* , которое будет выявлено путем наблюдения за полученными одношаговыми доходами. Таким образом, на первых тактах необходимо совершить «обучение» — выявить управление, приносящее наибольший одношаговый доход. В то же время существуют и повреждающие стратегии. Например, стратегия, в которой первые K правил заключаются в применении управления 1. Правило σ_{K+1} такой стратегии может быть построено только в виде зависимости от управления 1 и от значения $f(1)$, поэтому при любом его определении найдется объект μ^f , для которого в момент $K + 1$ будет с положительной вероятностью предписано применение неоптимального управления, и, следовательно, доход будет меньше K .

В примере 3 «обучение» оказалось возможным только благодаря знанию структуры процессов. Если бы заранее не было известно, что необходимо на первых тактах по разу «испробовать» все управления, то легко можно было пропустить период, когда возможно обучение, и совершить тем самым «непоправимую ошибку». Следовательно, для того чтобы конструктивно построить равномерно оптимальную стратегию, необходима дополнительная информация. Это противоречит избранному принципу постановки задачи — минимальности априорной информации об объекте.

Введем более жесткое определение адаптивной стратегии, которое, в частности, устраняет указанное несоответствие.

Пусть M — некоторый класс объектов. Эргодическая стратегия σ (ее определение дано в конце разд. 2) называется *устойчивой* по отношению к классу M , если для любого объекта $\mu \in M$ стратегия $\tilde{\sigma}$, полученная из стратегии σ путем произвольной (допустимой) замены конечного числа правил, (1) имеет одинаковый со стратегией доход $w(\mu, \sigma) = w(\mu, \tilde{\sigma})$ и (2) является эргодической по отношению к классу M .

Адаптивная стратегия для класса M называется *строго адаптивной*, если она устойчивая по отношению к этому классу.

Пример 4. Легко показать, что строго адаптивными являются многочисленные адаптивные стратегии для класса управляемых конечных связанных марковских цепей [1, 2].

Рассмотрим еще один мотив, выдвигаемый в качестве необходимого условия адаптивной управляемости.

Пример 5. Пусть класс объектов состоит из функций вещественного аргумента u вида $\mu^y = \mu^y(u) = I_{\{u=y\}}$, $y \in [0, 1]$. (В терминах управляемых случайных последовательностей: $X = \{0; 1\}$, $Y = [0, 1]$; $\mu_t(x_t | x^{t-1}, y^t) = x_t I_{\{y_t=y\}} + (1-x_t) I_{\{y_t=y\}}$; $g(x) = x$, $x \in X$.) Интуитивно представляется очевидным, что невозможно найти максимум такой функции за счетное число шагов, если не знать значение, в котором она обращается в единицу. В то же время формально для каждого объекта μ^y существует оптимальная стратегия. Например, можно постоянно повторять управление y . Однако не существует стратегии, равномерно оптимальной по отношению к классу $M = \{\mu^y\}$. В такой стратегии для каждого $y \in [0, 1]$ необходимо должно было бы выполняться следующее условие: $\sigma_t(y_t = y | \cdot) > 0$ хотя бы для одного значения t . Но это невозможно, поскольку для фиксированного значения t данное неравенство может быть выполнено лишь для счетного множества значений y , а t также пробегает счетное множество значений. Счетное объединение счетных множеств само счетно, поэтому необходимое неравенство не может быть выполнено для всех точек на отрезке $[0, 1]$.

Аналогичные рассуждения показывают, что в данном примере не существует счетного множества стратегий, обладающего тем свойством, что для любого объекта найдется ε -оптимальная стратегия из этого множества.

Конечное или счетное множество стратегий $\Sigma = \{\sigma(1), \sigma(2), \dots\}$ назовем *базовым* по отношению к классу объектов $M \in \mathcal{M}$, если:

- (1) для любого объекта из M и любого $\varepsilon > 0$ существует оптимальная стратегия из множества Σ ;
- (2) любая стратегия $\sigma(i)$ является устойчивой по отношению к классу M .

Теорема. *Строго адаптивная стратегия для класса объектов M существует тогда и только тогда, когда для этого класса существует базовое множество стратегий Σ .*

Доказательство теоремы.

Необходимость условий в данном случае является тривиальной, поскольку строго адаптивная стратегия, если она существует, образует базовое множество стратегий Σ , состоящее из одного элемента.

Докажем достаточность. Определим с помощью стратегий из Σ новую стратегию a следующим образом. Обозначим

$$\theta_{t,n} = \text{Int} \left((1 - v_{t,n})^{-n} \right),$$

где $\text{Int}(a)$ означает целую часть числа a , и зададим последовательность марковских моментов $\tau = \{\tau_n\}$ с помощью рекуррентных соотношений

$$\tau_0 = 0, \quad \tau_n = \tau_{n-1} + n + \theta_n,$$

где $\theta_n = \theta_{\tau_{n-1}, n}$. Соответствующие σ -алгебры обозначим $\mathbf{F}_{(n)} = \mathbf{F}_{\tau_{n-1}}$.

Будем считать, что на пространстве (Ω, \mathbf{F}) задана последовательность случайных величин $\beta = \{\beta_n\}$, независимых относительно $\mathbf{F}_{(n)}$. Каждая случайная величина имеет одно и то же невырожденное распределение $\{b_i\}$ на множестве номеров стратегий из Σ .

Определим правила стратегии $a = a(\Sigma, \beta)$ формулой

$$a_t = \sum_{n=1}^{\infty} \sigma_t(\beta_n) I_{\{\tau_{n-1} < t \leq \tau_n\}},$$

где $\sigma_t(\beta_n)$ — правило стратегии $\sigma(i) \in \Sigma$ в момент t , если $\beta_n = i$.

Наглядно работа стратегии a выглядит следующим образом. Процесс управления разбивается на этапы. Этап с номером n начинается в момент $\tau_{n-1} + 1$ и оканчивается в момент τ_n ; $\tau_0 = 0$. В момент, предшествующий началу очередного этапа, определяется номер стратегии в множестве Σ , из которой будут взяты правила для применения на данном этапе. Этот номер равен значению случайной величины β_n . Продолжительность n -го этапа равна $n + \theta_n$ и зависит, следовательно, от номера этапа и от оценки качества применяемой стратегии, полученной в течение первых n тактов этапа. Стратегия a называется стратегией перебора [2]. Таким образом, последовательность β определяет на каждом этапе выбор стратегии из множества Σ , правила из которой применяются на этом этапе.

Пусть задан объект $\mu \in M$ и пусть $W = W(\mu)$ — точная верхняя грань доходов для этого объекта, взятая по всем допустимым стратегиям, и пусть

$$W_i = w(\mu, \sigma(i)); \quad v_n^{(1)} = v_{\tau_{n-1}, n};$$

$$v_n^{(2)} = v_{\tau_{n-1}, n+\theta_n}; \quad \Delta_n = \tau_n - \tau_{n-1} = n + \theta_n.$$

Для произвольного $\varepsilon > 0$ определим множества

$$A_n^{(k)}(\varepsilon) = \left\{ v_n^{(k)} \geq W - \varepsilon \right\},$$

обозначая их дополнения $\overline{A_n^{(k)}(\varepsilon)}$, $k = 1, 2$.

Обозначим

$$s_n^{(1)} = \sum_{l=1}^n I_{A_l^{(1)}(\varepsilon) \cap A_l^{(2)}(2\varepsilon)} \Delta_l;$$

$$s_n^{(2)} = \sum_{l=1}^n I_{\overline{A_l^{(1)}(\varepsilon)} \cap A_l^{(2)}(2\varepsilon)} \Delta_l;$$

$$s_n^{(3)} = \sum_{l=1}^n I_{\overline{A_l^{(1)}(\varepsilon)}} \Delta_l,$$

так что $\tau_n = \sum_{l=1}^n \Delta_l = s_n^{(1)} + s_n^{(2)} + s_n^{(3)}$.

С помощью введенных обозначений запишем оценку для усредненного дохода к моменту τ_n :

$$\begin{aligned} w_n &= \frac{1}{\tau_n} \sum_{t=1}^{\tau_n} g_t = \frac{\sum_{l=1}^n v_l^{(2)} \Delta_l}{\sum_{l=1}^n \Delta_l} \geq \\ &\geq (W - 2\varepsilon) \frac{s_n^{(1)}}{s_n^{(1)} + s_n^{(2)} + s_n^{(3)}}. \end{aligned} \quad (1)$$

Для оценки суммы $s_n^{(1)}$ запишем неравенство

$$s_n^{(1)} \geq \Delta_{v_n},$$

в котором обозначено

$$v_n = \max \left\{ l : l \leq n, A_l^{(1)}(\varepsilon) \cap A_l^{(2)}(2\varepsilon) \right\}.$$

Оценим вероятность события $B_n = \{v_n \leq n - \ln n\}$, для которого выполняется включение

$$B_n \subset \bigcap_{n - \ln n < l \leq n} \overline{A_l^{(1)}(\varepsilon)} \cap \overline{A_l^{(2)}(2\varepsilon)}.$$

Согласно определениям эргодической стратегии, базового множества стратегий и семейства случайных величин β имеем:

$$\begin{aligned} \mathbf{P}_a \left(\overline{A_l^{(1)}(\varepsilon)} \cup \overline{A_l^{(2)}(2\varepsilon)} \mid \mathbf{F}_{(l)} \right) &\leq \\ &\leq \sum_{\substack{i \in \mathcal{I}; \\ W_i \leq W - \varepsilon/2}} \mathbf{P}_a(\beta_l = i \mid \mathbf{F}_{(l)}) + \\ &+ \sum_{\substack{i \in \mathcal{I}; \\ W_i \leq W - \varepsilon/2}} \mathbf{P}_a \left(\overline{A_l^{(1)}(\varepsilon)}, \beta_l = i \mid \mathbf{F}_{(l)} \right) \leq \\ &\leq \sum_{\substack{i \in \mathcal{I}; \\ W_i \leq W - \varepsilon/2}} \mathbf{P}_a(\beta_l = i) + \\ &+ \sum_{\substack{i \in \mathcal{I}; \\ W_i > W - \varepsilon/2}} \mathbf{P}_a \left(v_l^{(1)} \leq W_i - \frac{\varepsilon}{2}, \beta_l = i \mid \mathbf{F}_{(l)} \right) \leq \\ &\leq \sum_{\substack{i \in \mathcal{I}; \\ W_i \leq W - \varepsilon/2}} b_i + a_l \left(\frac{\varepsilon}{2} \right) \leq q < 1 \end{aligned}$$

при всех достаточно больших l . Отсюда следует, что для всех достаточно больших значений n выполняется неравенство

$$\mathbf{P}_a(B_n) \leq q^{n - \ln n}.$$

Следовательно, согласно лемме Бореля–Кантелли

$$\mathbf{P}_a \left(\overline{\lim}_{n \rightarrow \infty} B_n \right) = 0. \quad (2)$$

Это означает, что

$$s_n^{(1)} \geq \Delta_{v_n} \geq (1 - W - \varepsilon)^{-n + \ln n}.$$

Оценим сумму $s_n^{(2)}$. Обозначив $C_n = A_n^{(1)}(\varepsilon) \cap \overline{A_n^{(2)}}(2\varepsilon)$ и $W_{(n)} = \sum_{i \in I} W_i I_{\{\beta_n = i\}}$, получим:

$$\begin{aligned} \mathbf{P}_a(C_n | \mathbf{F}_{(n)}) &= \mathbf{P}_a \left(C_n, W_{(n)} < W - \frac{3\varepsilon}{2} | \mathbf{F}_{(n)} \right) + \\ &+ \mathbf{P}_a \left(C_n, W_{(n)} \geq W - \frac{3\varepsilon}{2} | \mathbf{F}_{(n)} \right) \leq \\ &\leq \mathbf{P}_a \left(v_n^{(1)} > W - \varepsilon, W_{(n)} < W - \frac{3\varepsilon}{2} | \mathbf{F}_{(n)} \right) + \\ &+ \mathbf{P}_a \left(v_n^{(2)} \leq W - 2\varepsilon, W_{(n)} \geq W - \frac{3\varepsilon}{2} | \mathbf{F}_{(n)} \right) \leq \\ &\leq \sum_{i \in \mathcal{I}; W_i \leq W - \varepsilon/2} \mathbf{P}_a \left(v_{\tau_n, n} > W_i + \frac{\varepsilon}{2}, \beta_l = i | \mathbf{F}_{(n)} \right) + \\ &+ \sum_{\substack{i \in \mathcal{I}; \\ W_i > W - 3\varepsilon/2}} \mathbf{P}_a \left(v_{\tau_n, n + \theta_n} \leq W_i - \frac{\varepsilon}{2}, \beta_l = i | \mathbf{F}_{(n)} \right) \leq \\ &\leq a_n \left(\frac{\varepsilon}{2} \right). \end{aligned}$$

Из определения базового множества стратегий следует, что

$$\sum_{n=1}^{\infty} \mathbf{P}_a(C_n) < \infty,$$

поэтому согласно лемме Бореля–Кантелли получаем:

$$\mathbf{P}_a \left(\overline{\lim}_{n \rightarrow \infty} C_n \right) = 0. \quad (3)$$

Отсюда следует, что

$$\sup_n s_n^{(2)} \leq c < \infty.$$

Для суммы $s_n^{(3)}$ имеем следующую оценку:

$$s_n^{(3)} \geq \sum_{l=1}^n (n + (1 - W + \varepsilon)^{-l}) < n^2 + n(1 - W + \varepsilon)^{-n}.$$

Подставляя оценки, полученные для сумм $s_n^{(k)}$, в неравенство (1), получаем:

$$\begin{aligned} w_n &\geq (W - \varepsilon) \left(1 + \frac{s_n^{(2)} + s_n^{(3)}}{s_n^{(1)}} \right)^{-1} \geq \\ &\geq (W - \varepsilon) \left(1 + \frac{c + n^2 + n(1 - W + \varepsilon)^{-n}}{(1 - W - \varepsilon/2)^{-n + \ln n}} \right)^{-1} \geq \\ &\geq W - 3\varepsilon \end{aligned}$$

для всех достаточно больших значений n . Отсюда

$$\underline{\lim}_{n \rightarrow \infty} w_n \geq W. \quad (4)$$

Рассмотрим далее множество

$$\Omega' = \left\{ \lim_{n \rightarrow \infty} w_n = W \right\} \cap \overline{B} \cap \overline{C},$$

где \overline{B} и \overline{C} означают соответственно дополнения к множествам $B = \overline{\lim}_{n \rightarrow \infty} B_n$ и $C = \overline{\lim}_{n \rightarrow \infty} C_n$.

Согласно формулам (2)–(4)

$$\mathbf{P}_a(\Omega') = 1.$$

Определим следующие события:

$$D_{n,t}^{(1)} = \{\tau_{n-1} < t \leq \tau_{n-1} + n\} \cap \Omega';$$

$$D_{n,t}^{(2)} = \{\tau_{n-1} + n < t \leq \tau_n\} \cap \Omega';$$

$$D_{n,t}^{(3)} = \{\tau_{n-1} < t \leq \tau_n\} \cap \Omega'.$$

На множестве $D_{n,t}^{(1)}$ усредненный доход $v_t = v_{0,t} = t^{-1} \sum_{s=1}^t g_s$ оценивается с помощью формулы (1) как

$$v_t \geq \frac{\tau_{n-1} w_n}{\tau_{n-1} + n + \theta_n} \geq W - \varepsilon_n^{(1)},$$

где $\varepsilon_n^{(1)} \rightarrow 0$ при $n \rightarrow \infty$.

Пусть событие $D_{n,t}^{(2)}$ имеет место. Тогда $\theta_n \geq (1 - W + \varepsilon)^{-n}$. Кроме того, из определения событий $B_n, B, D_{n,t}^{(2)}$ следует, что для всех достаточно больших значений n выполняется неравенство $v_n > n - \ln n$. Следовательно, на множестве $D_n^{(2)}$ справедлива оценка

$$v_t \geq \frac{\tau_{n-1} w_n}{\tau_{n-1} + n + \theta_n} \geq W - \varepsilon_n^{(2)},$$

где $\varepsilon_n^{(2)} \rightarrow 0$ при $n \rightarrow \infty$.

Из определения событий $C_n, C, D_{n,t}^{(3)}$ вытекает, что

$$D_{n,t}^{(3)} \subset \left\{ \min_{n < m \leq n + \theta_n} v_{n,m} \geq W - 2\varepsilon \right\},$$

поэтому на множестве $D_n^{(3)}$ справедливы неравенства:

$$\begin{aligned} v_t &\geq \frac{\tau_{n-1} w_n}{t} + \left(1 - \frac{\tau_{n-1}}{t} \right) (1 - \tau_n)^{-1} \sum_{s=\tau_{n-1}+1}^t g_s \geq \\ &\geq \frac{\tau_{n-1} w_n}{t} + \left(1 - \frac{\tau_{n-1}}{t} \right) (W - 2\varepsilon) \geq W - 2\varepsilon - \varepsilon_n^{(3)}, \end{aligned}$$

где $\varepsilon_n^{(3)} \rightarrow 0$ при $n \rightarrow \infty$.

Таким образом, на множестве

$$D_{n,t} = \bigcup_{k=1}^3 D_{n,t}^{(k)} = \{\tau_{n-1} < t \leq \tau_n\} \cap \Omega'$$

имеет место оценка $v_n \geq W - \varepsilon - \varepsilon_n$, где $\varepsilon_n \rightarrow 0$ при $n \rightarrow \infty$. Достаточность утверждения теоремы следует из соотношений $\Omega = \bigcup_{n=1}^{\infty} \{\tau_{n-1} < t \leq \tau_n\}$ и $\lim_{t \rightarrow \infty} I_{D_{n,t}} = 0$.

4 Заключение

Адаптивные стратегии, позволяющие достигать цели в условиях информационной неопределенности, основываясь на «обучении» в процессе взаимодействия с объектом, находят все более широкое практическое применение.

В этой работе было уделено внимание теоретическим аспектам адаптивного подхода. Сформулированы определения адаптивных стратегий и

приведена формальная постановка задачи адаптивного управления. Сформулированы и доказаны некоторые утверждения о необходимых условиях и достаточных условиях адаптивной управляемости.

Продолжение исследований в данном направлении позволит найти ответы на принципиальные вопросы, в каких ситуациях можно рассчитывать на «приспособление к неизвестной среде» и сколь универсальными могут быть «обучающиеся» алгоритмы.

Литература

1. *Sragovich V. G.* Mathematical theory of adaptive control. — Singapore: World Scientific, 2006.
2. *Коновалов М. Г.* Методы адаптивной обработки информации и их приложения. — М.: ИПИ РАН, 2007.
3. *Неве Ж.* Математические основы теории вероятностей. — М.: Мир, 1969.

ОЦЕНКИ В НУЛЬ-ЭРГОДИЧЕСКОМ СЛУЧАЕ ДЛЯ НЕКОТОРЫХ СИСТЕМ ОБСЛУЖИВАНИЯ*

А. И. Зейфман¹, А. В. Коротышева², Я. А. Сатин³, С. Я. Шоргин⁴

Аннотация: Рассматриваются модели обслуживания с групповым поступлением и обслуживанием требований. Получены оценки скорости сходимости в нуль-эргодическом случае. Рассмотрен пример конкретного класса таких систем обслуживания.

Ключевые слова: нестационарные системы обслуживания с групповым поступлением и обслуживанием требований; нуль-эргодичность; оценки

1 Введение

Напомним, что марковская цепь $X = X(t)$, $t \geq 0$, с непрерывным временем и дискретным пространством состояний называется нуль-эргодичной, если при любых начальных условиях и каждом i выполняется условие $\Pr\{X(t) = i\} \rightarrow 0$ при $t \rightarrow \infty$.

Понятие нуль-эргодичности (null ergodicity) и соответствующие оценки скорости сходимости изучались для моделей массового обслуживания, описываемых процессами рождения и гибели начиная с 1990-х гг. [1–5].

В настоящей работе оценки, относящиеся к нуль-эргодической ситуации, будут исследованы для более общего класса марковских систем обслуживания с групповым поступлением и обслуживанием требований. Изучение этого класса моделей начато в работах [6, 7]. Отметим еще, что в таких системах, как будет описано далее, возможны и катастрофические сбои, такого рода системы массового обслуживания (СМО с катастрофами, queues with disasters, queues with catastrophes) в разных ситуациях и при разных предположениях изучались ранее (см., например, [8–17]).

Пусть $X = X(t)$, $t \geq 0$, — число требований в системе обслуживания ($0 \leq X(t) < \infty$).

Обозначим через

$$p_{ij}(s, t) = \Pr\{X(t) = j | X(s) = i\}, \\ i, j \geq 0, \quad 0 \leq s \leq t,$$

переходные вероятности процесса $X = X(t)$, а через $p_i(t) = \Pr\{X(t) = i\}$ — его вероятности состояний.

Будем всюду далее предполагать, что

$$\Pr(X(t+h) = j | X(t) = i) = \\ = \begin{cases} q_{ij}(t)h + \alpha_{ij}(t, h) & \text{при } j \neq i; \\ 1 - \sum_{k \neq i} q_{ik}(t)h + \alpha_i(t, h) & \text{при } j = i, \end{cases}$$

причем все $\alpha_i(t, h)$ есть $o(h)$ равномерно по i , т. е. $\sup_i |\alpha_i(t, h)| = o(h)$.

При этом предполагаем, что при $k > 0$ $q_{i, i+k}(t) = \lambda_k(t)$, $q_{i+k, i}(t) = \mu_k(t)$.

Другими словами, будем исследовать системы обслуживания, в которых интенсивности поступления и обслуживания k требований в момент t в системе обслуживания ($\lambda_k(t)$ и $\mu_k(t)$ соответственно) не зависят от числа требований, находящихся в системе в момент t , причем $\lambda_{k+1}(t) \leq \lambda_k(t)$ и $\mu_{k+1}(t) \leq \mu_k(t)$ при всех k и почти при всех $t \geq 0$.

Отметим, что в рассматриваемой системе, как легко видеть, интенсивности обслуживания $\mu_k(t)$ являются и интенсивностями катастроф, так что если при некоторых k, t выполняется неравенство $\mu_k(t) > 0$, то при этих k, t возможен катастрофический сбой системы обслуживания с потерей всех требований.

Далее в соответствии со стандартным подходом (см. подробное описание в работах [3–5, 18]) предположим дополнительно, что все интенсивности являются линейными комбинациями конечно-го числа локально интегрируемых на $[0, \infty)$ функций. Кроме того, будем предполагать, что почти при всех $t \geq 0$

$$L_\lambda(t) + L_\mu(t) = L(t) < \infty,$$

* Исследование поддержано РФФИ, гранты 11-01-12026-офи-м, 12-07-00115-а, 12-07-00109-а.

¹ Вологодский государственный педагогический университет; Институт проблем информатики Российской академии наук и ИСЭРТ РАН, a_zeifman@mail.ru

² Вологодский государственный педагогический университет, a_korotysheva@mail.ru

³ Вологодский государственный педагогический университет, yacovi@mail.ru

⁴ Институт проблем информатики Российской академии наук, SShorgin@ipiran.ru

где

$$L_\lambda(t) = \sum_{i=1}^{\infty} \lambda_i(t); \quad L_\mu(t) = \sum_{i=1}^{\infty} \mu_i(t).$$

Тогда для описания вероятностной динамики процесса получаем прямую систему Колмогорова в виде дифференциального уравнения в пространстве последовательностей l_1 :

$$\frac{d\mathbf{p}}{dt} = A(t)\mathbf{p}(t), \quad (1)$$

где

$$A(t) = \begin{pmatrix} a_{00}(t) & \mu_1(t) & \mu_2(t) & \mu_3(t) & \cdots & \mu_r(t) & \cdots \\ \lambda_1(t) & a_{11}(t) & \mu_1(t) & \mu_2(t) & \cdots & \mu_{r-1}(t) & \cdots \\ \lambda_2(t) & \lambda_1(t) & a_{22}(t) & \mu_1(t) & \cdots & \mu_{r-2}(t) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda_r(t) & \lambda_{r-1}(t) & \lambda_{r-2}(t) & \cdots & \lambda_1(t) & a_{rr}(t) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

причем $a_{ii}(t) = -\sum_{k=1}^i \mu_k(t) - \sum_{k=1}^{\infty} \lambda_k(t)$.

Далее будем обозначать через $\|\bullet\|$ l_1 -норму, т.е. $\|\mathbf{x}\| = \sum |x_i|$, а $\|B\| = \sup_j \sum_i |b_{ij}|$, если $B = (b_{ij})_{i,j=0}^{\infty}$, и пусть Ω — множество всех векторов из l_1 с неотрицательными координатами и единичной нормой.

Тогда имеем $\|A(t)\| = 2 \sum_{k=1}^{\infty} (\lambda_k(t) + \mu_k(t)) \leq 2L(t)$ почти при всех $t \geq 0$, следовательно, операторная функция $A(t)$ ограничена почти при всех $t \geq 0$ и локально интегрируема на $[0; \infty)$.

Как известно (см., например, [19]), задача Коши для уравнения (1) тогда имеет единственное решение при любых начальных условиях, и из $\mathbf{p}(s) \in \Omega$ вытекает $\mathbf{p}(t) \in \Omega$ при $t \geq s \geq 0$.

Через $E(t, k) = E\{X(t) | X(0) = k\}$ будем далее обозначать математическое ожидание процесса (среднее число требований) в момент t при условии, что в нулевой момент времени он находится в состоянии k .

2 Общий подход и оценки

Положим

$$\nu(t) = \inf_{i \geq 0} \left(|a_{ii}(t)| - \sum_{k=1}^i \frac{d_{i-k}}{d_i} \mu_k(t) - \sum_{k=1}^{\infty} \frac{d_{i+k}}{d_i} \lambda_k(t) \right). \quad (2)$$

Теорема 1. Пусть для процесса $X(t)$ можно найти последовательность $\{d_i\}$ положительных чисел такую, что $d_{-1} = d_0 = 1$, $\sup_{i \geq 1} d_i = d < \infty$, и выполнено условие

$$\int_0^{\infty} \nu(t) dt = +\infty.$$

Тогда $X(t)$ нуль-эргодичен и при любых $0 \leq s \leq t$ и любом n выполняется неравенство

$$\sum_{i=0}^{\infty} d_i p_i(t) \leq d e^{-\int_s^t \nu(\tau) d\tau}.$$

Доказательство. Введем диагональную матрицу следующего вида:

$$D = \text{diag}(d_0, d_1, \dots),$$

тогда, очевидно,

$$D^{-1} = \text{diag}(d_0^{-1}, d_1^{-1}, \dots).$$

Рассмотрим теперь пространство последовательностей \mathcal{B} таких, что

$$\|\mathbf{x}\|_{\mathcal{B}} = \|D\mathbf{x}\|_1 = \sum_{i=0}^{\infty} d_i |x_i| < \infty,$$

и прямую систему Колмогорова (1) как уравнение в \mathcal{B} . Имеем

$$\|A(t)\|_{\mathcal{B}} = \|DA(t)D^{-1}\|_1$$

и с учетом (2) и структуры матрицы $A(t)$ получаем:

$$\begin{aligned} \|A(t)\|_{\mathcal{B}} &= \sup_{i \geq 0} \left(|a_{ii}(t)| + \sum_{k=1}^i \frac{d_{i-k}}{d_i} \mu_k(t) + \sum_{k=1}^{\infty} \frac{d_{i+k}}{d_i} \lambda_k(t) \right) = \\ &= \sup_{i \geq 0} \left(a_{ii}(t) + \sum_{k=1}^i \frac{d_{i-k}}{d_i} \mu_k(t) + \sum_{k=1}^{\infty} \frac{d_{i+k}}{d_i} \lambda_k(t) + \right. \\ &\left. + 2|a_{ii}(t)| \right) \leq -\nu(t) + 2 \sup_{i \geq 0} |a_{ii}(t)| = -\nu(t) + \|A(t)\|_1. \end{aligned}$$

Следовательно, операторная функция $A(t)$ ограничена почти при всех $t \geq 0$ и в пространстве последовательностей \mathcal{B} .

Для получения нужных оценок потребуется понятие логарифмической нормы операторной функции и связанные с ней оценки оператора Коши линейного дифференциального уравнения в банаховом пространстве. Соответствующий подход к исследованию в случае процессов рождения и гибели подробно описан в [3, 5], а для конечномерного стационарного случая см. также обсуждение и интересные исторические замечания в [20].

Рассмотрим логарифмическую норму $\gamma(A(t))$ в \mathcal{B} :

$$\begin{aligned} \gamma(A(t))_1 &= \gamma(DA(t)D^{-1})_{\mathcal{B}} = \\ &= \sup_{i \geq 0} \left(a_{ii}(t) + \sum_{k=1}^i \frac{d_{i-k}}{d_i} \mu_k(t) + \sum_{k=1}^{\infty} \frac{d_{i+k}}{d_i} \lambda_k(t) \right) = \\ &= -\nu(t). \end{aligned}$$

Как известно (см., например, [5]), отсюда вытекает оценка:

$$\|U(t, s)\|_{\mathcal{B}} \leq e^{-\int_s^t \nu(\tau) d\tau},$$

где $U(t, s)$ — оператор Коши уравнения (1). А тогда при любых $0 \leq s \leq t$, если $\mathbf{p}(s) \in \Omega$, получаем неравенство:

$$\begin{aligned} \|\mathbf{p}(t)\|_{\mathcal{B}} &= \sum_{i=0}^{\infty} d_i p_i(t) \leq \\ &\leq \|\mathbf{p}(s)\|_{\mathcal{B}} e^{-\int_s^t \nu(\tau) d\tau} \leq d e^{-\int_s^t \nu(\tau) d\tau}. \end{aligned}$$

Следствие 1. Пусть выполнены условия теоремы 1. Тогда при любых $0 \leq s \leq t$ и любых n, k справедливы следующие оценки для вероятности числа требований в системе обслуживания:

$$\begin{aligned} \Pr(X(t) \leq n) &\leq \frac{d}{\partial_n} e^{-\int_s^t \nu(\tau) d\tau}; \\ \Pr(X(t) \leq n / X(s) = k) &\leq \frac{d_k}{\partial_n} e^{-\int_s^t \nu(\tau) d\tau}, \end{aligned} \quad (3)$$

где $\partial_n = \min_{i \leq n} d_i$.

Для доказательства достаточно отметить, что

$$\partial_n \Pr(X(t) \leq n) \leq \sum_{i=0}^n d_i p_i(t) \leq \|\mathbf{p}(t)\|_{\mathcal{B}}$$

при любых начальных условиях и

$$\|\mathbf{p}(s)\|_{\mathcal{B}} = \sum_{i=0}^{\infty} d_i p_i(s) = d_k,$$

если $X(s) = k$, соответственно.

Следствие 2. Пусть выполнены условия теоремы 1. Тогда при любом $t \geq 0$ и любых k, n справедлива следующая оценка для среднего количества требований в системе обслуживания:

$$E(t; k) \geq (n+1) \left(1 - \frac{d_k}{\partial_n} e^{-\int_0^t \nu(\tau) d\tau} \right).$$

Доказательство. Из (3) получаем оценку:

$$\sum_{i \geq n+1} p_i(t) = 1 - \sum_{i=0}^n p_i(t) \geq 1 - \frac{d_k}{\partial_n} e^{-\int_0^t \nu(\tau) d\tau},$$

и теперь остается лишь заметить, что

$$E(t; k) \geq (n+1) \sum_{i \geq n+1} p_i(t).$$

Интересно отметить, что при выполнении простого дополнительного условия математическое ожидание числа требований можно ограничить и сверху.

Следствие 3. Пусть почти при всех $t \geq 0$ выполнено условие

$$\sum_{k=1}^{\infty} k \lambda_k(t) = L(t) < \infty.$$

Тогда при любом $t \geq 0$ и любом k справедлива следующая оценка для среднего количества требований в системе обслуживания:

$$E(t; k) \leq E(0; k) + \int_0^t L(\tau) d\tau.$$

Доказательство. С учетом сходимости рассматриваемых рядов из системы (1) получаем неравенство:

$$\begin{aligned} \frac{dE(t; k)}{dt} &= \\ &= p_0 \sum_{k=1}^{\infty} k \lambda_k(t) + p_1 \left(-\mu_1(t) + \sum_{k=1}^{\infty} k \lambda_k(t) \right) + \\ &+ p_2 \left(-2\mu_2(t) - \mu_1(t) + \sum_{k=1}^{\infty} k \lambda_k(t) \right) + \dots \leq \\ &\leq \sum_{k=0}^{\infty} p_k L(t) = L(t), \end{aligned}$$

откуда и вытекает требуемая оценка.

Рассмотрим отдельно простейший стационарный случай.

Следствие 4. Пусть интенсивности поступления и обслуживания требований не зависят от времени. Пусть для процесса $X(t)$ можно найти последовательность $\{d_i\}$ положительных чисел такую, что $d_{-1} = d_0 = 1$, $\sup_{i \geq 1} d_i = d < \infty$, и выполнено условие

$$\nu = \inf_{i \geq 0} \left(|a_{ii}| - \sum_{k=1}^i \frac{d_{i-k}}{d_i} \mu_k - \sum_{k=1}^{\infty} \frac{d_{i+k}}{d_i} \lambda_k \right) > 0.$$

Тогда $X(t)$ нуль-эргодичен и при любых $t \geq 0$, n и k выполняются оценки:

$$\sum_{i=0}^{\infty} d_i p_i(t) \leq d e^{-\nu t},$$

$$\Pr(X(t) \leq n) \leq \frac{d}{\partial_n} e^{-\nu t};$$

$$\Pr(X(t) \leq n/X(s) = k) \leq \frac{d_k}{\partial_n} e^{-\nu t}$$

$$E(t; k) \geq (n + 1) \left(1 - \frac{d_k}{\partial_n} e^{-\nu t}\right).$$

3 Система с групповым поступлением и обслуживанием в специальном случае

Рассмотрим здесь более подробно свойства эргодичности для системы массового обслуживания следующего типа.

Будем предполагать, что требования в систему поступают группами, интенсивность поступления новых требований (не более m одновременно) равна $\lambda(t)$, а обслуживаются также группами (не более n требований одновременно) с интенсивностями $\mu(t)$.

Тогда число требований в СМО описывается рассмотренной в предыдущем параграфе марковской цепью $X(t)$ с интенсивностями

$$\lambda_k(t) = \begin{cases} \lambda(t) & \text{при } k \leq m; \\ 0 & \text{при } k > m; \end{cases}$$

$$\mu_k(t) = \begin{cases} \mu(t) & \text{при } k \leq n; \\ 0 & \text{при } k > n. \end{cases}$$

Теорема 2. Пусть для процесса $X(t)$ можно найти число $\delta < 1$ такое, что

$$\int_0^{\infty} \nu^*(t) dt = +\infty, \tag{4}$$

где

$$\nu^*(t) = (1 - \delta) \left(\{1 + (1 + \delta) + \dots + (1 + \delta + \dots + \delta^{m-1})\} \lambda(t) - \left\{ \frac{1}{\delta} + \frac{1 + \delta}{\delta^2} + \dots + \frac{1 + \delta + \dots + \delta^{n-1}}{\delta^n} \right\} \mu(t) \right).$$

Тогда $X(t)$ нуль-эргодичен и при любых $0 \leq s \leq t$ и любых k, r выполняются неравенства

$$\sum_{i=0}^{\infty} d_i p_i(t) \leq e^{-\int_s^t \nu^*(\tau) d\tau},$$

$$\Pr(X(t) \leq k) \leq \delta^{-k} e^{-\int_s^t \nu^*(\tau) d\tau};$$

$$\Pr(X(t) \leq k/X(s) = r) \leq \delta^{r-k} e^{-\int_s^t \nu^*(\tau) d\tau}.$$

Доказательство. Достаточно положить $d_k = \delta^k$, тогда выполнены условия теоремы 1 и следствий, причем $d = 1$, $\partial_k = \delta^k$.

Замечание 1. Если интенсивности поступления и обслуживания требований не зависят от времени (стационарный случай), то условие (4), гарантирующее нуль-эргодичность, равносильно выполнению неравенства

$$m(m + 1)\lambda - n(n + 1)\mu > 0, \tag{5}$$

а при 1-периодических интенсивностях — тому, что

$$m(m + 1) \int_0^1 \lambda(t) dt - n(n + 1) \int_0^1 \mu(t) dt > 0. \tag{6}$$

Действительно, при выполнении неравенств (5) и (6) условие (4) заведомо будет выполнено при достаточно малых $1 - \delta > 0$.

Отметим теперь, что полученные условия нуль-эргодичности рассмотренной системы оказываются достаточно точными, как показывает следующее утверждение.

Теорема 3. Пусть для процесса $X(t)$ можно найти число $d > 1$ такое, что

$$\int_0^{\infty} \beta^*(t) dt = +\infty, \tag{7}$$

где

$$\beta^*(t) = (d - 1) \left(\left\{ \frac{1}{d} + \frac{1 + d}{d^2} + \dots + \frac{1 + d + \dots + d^{m-1}}{d^m} \right\} \mu(t) - \{1 + (1 + d) + \dots + (1 + d + \dots + d^{m-1})\} \lambda(t) \right).$$

Тогда $X(t)$ слабо эргодичен и

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_1 \leq 4e^{-\int_s^t \beta^*(u) du} \sum_{i \geq 1} g_i |p_i^*(s) - p_i^{**}(s)| \tag{8}$$

при любых начальных условиях $\mathbf{p}^*(s), \mathbf{p}^{**}(s)$ и любых $s, t, 0 \leq s \leq t$, где $g_i = \sum_{n=1}^i d^{n-1}$.

Доказательство. Воспользуемся методикой, описанной в [6, 7].

Полагая $p_0 = 1 - \sum_{i \geq 1} p_i$, из уравнения (1) получим

$$\frac{d\mathbf{z}}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t), \quad (9)$$

где вектор $\mathbf{f}(t) = (\lambda, \dots, \lambda, 0, \dots)^T$ имеет m первых ненулевых координат, элементы первых m строк матрицы $B(t) = (b_{ij}(t))_{i,j=1}^{\infty}$ получаются вычитанием $\lambda(t)$ из соответствующих элементов $A(t)$, а элементы остальных строк совпадают с соответствующими элементами матрицы исходной системы.

Положим $d_k = d^{k-1}$ и рассмотрим треугольную матрицу вида

$$H = \begin{pmatrix} d_1 & d_1 & d_1 & \dots \\ 0 & d_2 & d_2 & \dots \\ 0 & 0 & d_3 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix},$$

а также пространство последовательностей

$$l_{1H} = \{ \mathbf{z} = (p_1, p_2, \dots)^T / \|\mathbf{z}\|_{1H} \equiv \|H\mathbf{z}\|_1 < \infty \}.$$

Рассмотрим (9) как дифференциальное уравнение в пространстве l_H , при этом несложно убедиться, что $B(t)$ и $\mathbf{f}(t)$ локально интегрируемы при $t \geq 0$.

Оценим логарифмическую норму

$$\gamma(B(t))_{1H} = \gamma(HB(t)H^{-1})_1.$$

Имеем

$$HBH^{-1} = \begin{pmatrix} a_{11} & (\mu_1 - \mu_2) \frac{d_1}{d_2} & (\mu_2 - \mu_3) \frac{d_1}{d_3} & \dots & (\mu_{r-1} - \mu_r) \frac{d_1}{d_r} & \dots \\ \lambda_1 \frac{d_2}{d_1} & a_{22} & (\mu_1 - \mu_3) \frac{d_2}{d_3} & \dots & (\mu_{r-2} - \mu_r) \frac{d_2}{d_r} & \dots \\ \lambda_2 \frac{d_3}{d_1} & \lambda_1 \frac{d_3}{d_2} & a_{33} & \dots & (\mu_{r-3} - \mu_r) \frac{d_3}{d_r} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \lambda_{r-1} \frac{d_r}{d_1} & \lambda_{r-2} \frac{d_r}{d_2} & \lambda_{r-3} \frac{d_r}{d_3} & \dots & a_{rr} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

где $\lambda_k(t) = \lambda(t)$, $1 \leq k \leq m$, $\mu_k(t) = \mu(t)$, $1 \leq k \leq n$, а остальные $\lambda_k(t)$, $\mu_k(t)$ — тождественные нули.

Тогда получаем:

$$\begin{aligned} \gamma(B(t))_{1H} &= \sup \left(\mathbf{b}_{ii}(t) + \sum_{j \neq i} \mathbf{b}_{ji}(t) \right) = \\ &= - (d-1) \left(\left\{ \frac{1}{d} + \frac{1+d}{d^2} + \dots \right. \right. \\ &\dots + \frac{1+d+\dots+d^{n-1}}{d^n} \left. \right\} \mu(t) - \{ 1 + (1+d) + \dots \\ &\dots + (1+d+\dots+d^{m-1}) \} \lambda(t) \Big) = -\beta^*(t). \end{aligned}$$

Как известно (см., например [7]), отсюда вытекает оценка:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1H} \leq e^{-\int_s^t \beta^*(u) du} \|\mathbf{p}^*(s) - \mathbf{p}^{**}(s)\|_{1H},$$

справедливая при любых $s, t, 0 \leq s \leq t$, и любых начальных условиях $\mathbf{p}^*(s), \mathbf{p}^{**}(s)$. Отсюда с учетом неравенства $\|\mathbf{z}\|_1 \leq 2\|\mathbf{z}\|_{1H}$, полученного ранее при сравнении норм, и получаем требуемую оценку (8).

Замечание 1. Если интенсивности поступления и обслуживания требований не зависят от времени (стационарный случай), то условие (7), гарантирующее в этом случае сильную эргодичность, равносильно выполнению неравенства:

$$m(m+1)\lambda - n(n+1)\mu < 0.$$

При 1-периодических интенсивностях условие (7) гарантирует слабую эргодичность, существование предельного 1-периодического режима и 1-периодического предельного среднего и эквивалентно выполнению неравенства

$$m(m+1) \int_0^1 \lambda(t) dt - n(n+1) \int_0^1 \mu(t) dt < 0.$$

Пример 1.

Рассмотрим конкретный пример системы этого класса при конкретных значениях интенсивностей и допустимых размерах групп требований с оценками в нуль-эргодичном случае и слабо эргодичном случае.

Пусть интенсивности поступления и обслуживания требований есть $\lambda(t) = 1 + \sin 2\pi t$ и $\mu(t) = 1 + \cos 2\pi t$ соответственно.

Рассмотрим два случая.

Случай 1. Пусть $m = 3, n = 2$. Полагая, как описано ранее, $d_k = \delta^k$, получаем:

$$\nu^*(t) = (3 - \delta - \delta^2 - \delta^3) \lambda(t) + (2 - \delta^{-1} - \delta^{-2}) \mu(t).$$

Отметим, что наибольшее значение выражения $f(\delta) = 5 - \delta - \delta^2 - \delta^3 - \delta^{-1} - \delta^{-2}$ на нужном проме-

жутке достигается при $\delta^* \approx 0,84$ и равно $f^* \approx 0,254$. При этом получаем:

$$\nu^*(t) \approx 0,254 + (3 - \delta^* - (\delta^*)^2 - (\delta^*)^3) \sin 2\pi t + (2 - (\delta^*)^{-1} - (\delta^*)^{-2}) \cos 2\pi t,$$

причем

$$\int_0^t \nu^*(\tau) d\tau \geq 0,25t + (3 - \delta^* - (\delta^*)^2 - (\delta^*)^3) \times \left(\int_0^{[t]} \sin 2\pi u du + \int_{[t]}^t \sin 2\pi u du \right) + (2 - (\delta^*)^{-1} - (\delta^*)^{-2}) \left(\int_0^{[t]} \cos 2\pi u du + \int_{[t]}^t \cos 2\pi u du \right) \geq 0,25t - (5 - \delta^* - (\delta^*)^2 - (\delta^*)^3 - (\delta^*)^{-1} - (\delta^*)^{-2}) \geq 0,25t - 0,3, \quad (10)$$

откуда заведомо

$$e^{-\int_0^t \nu^*(\tau) d\tau} \leq 2e^{-0,25t}.$$

Следовательно, процесс $X(t)$ (число требований в системе) нуль-эргодичен и при любых $t \geq 0$ и любых n, r выполняются неравенства

$$\sum_{i=0}^{\infty} \delta^{*k} p_k(t) \leq 2e^{-0,25t},$$

а также

$$\Pr(X(t) \leq n) \leq 2\delta^{*n} e^{-0,25t};$$

$$\Pr(X(t) \leq n / X(0) = r) \leq 2\delta^{*r-n} e^{-0,25t}.$$

Случай 2. Пусть $m = 2, n = 3$.

Аналогично предыдущему случаю рассмотрим теперь выражение

$$\beta^*(t) = (3 - d - d^2) \lambda(t) + (2 - d^{-1} - d^{-2} - d^{-3}) \mu(t).$$

Отметим, что наибольшее значение выражения $g(d) = 5 - d - d^2 - d^{-1} - d^{-2} - d^{-3}$ на нужном промежутке достигается при $d^* = (\delta^*)^{-1} \approx 1,2$ и равно $g^* = f^* \approx 0,254$. С учетом оценки (10) получаем теперь

$$e^{-\int_0^t \beta^*(\tau) d\tau} \leq 2e^{-0,25t}.$$

Тогда $X(t)$ слабо эргодичен и

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_1 \leq 8e^{-0,25t} \sum_{i \geq 1} g_i |p_i^*(0) - p_i^{**}(0)|$$

при любых начальных условиях $\mathbf{p}^*(0), \mathbf{p}^{**}(0)$ и любом $t \geq 0$, где $g_i = \sum_{n=1}^i (d^*)^{n-1}$.

Кроме того, $X(t)$ имеет предельный 1-периодический режим и 1-периодическое предельное среднее, которые можно построить, пользуясь методикой, описанной в [7, 18].

Литература

1. *Zeifman A. I.* Some estimates of the rate of convergence for birth and death processes // *J. Appl. Prob.*, 1991. Vol. 28. P. 268–277.
2. *Zeifman A. I.* On the estimation of probabilities for birth and death processes // *J. Appl. Prob.*, 1995. Vol. 32. P. 623–634.
3. *Zeifman A. I.* Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes // *Stoch. Proc. Appl.*, 1995. Vol. 59. P. 157–173.
4. *Granovsky B. L., Zeifman A. I.* Nonstationary queues: Estimation of the rate of convergence // *Queueing Syst.*, 2004. Vol. 46. P. 363–388.
5. *Зейфман А. И., Бенинг В. Е., Соколов И. А.* Марковские цепи и модели с непрерывным временем. — М.: Элекс-КМ, 2008.
6. *Сатин Я. А., Зейфман А. И., Коротышева А. В., Шоргин С. Я.* Об одном классе марковских систем обслуживания // *Информатика и её применения*, 2011. Т. 5. Вып. 4. С. 18–24.
7. *Сатин Я. А., Зейфман А. И., Коротышева А. В.* О скорости сходимости и усечения для одного класса марковских систем обслуживания // *Теория вероятностей и ее применения*, 2012 (в печати).
8. *Dudin A., Nishimura S.* A BMAP/SM/1 queueing system with Markovian arrival input of disasters // *J. Appl. Prob.*, 1999. Vol. 36. P. 868–881.
9. *Dudin A., Karolik A.* BMAP/SM/1 queue with Markovian input of disasters and non-instantaneous recovery // *Perform. Eval.*, 2001. Vol. 45. P. 19–32.
10. *Di Crescenzo A., Giorno V., Nobile A. G., Ricciardi L. M.* On the M/M/1 queue with catastrophes and its continuous approximation // *Queueing Syst.*, 2003. Vol. 43. P. 329–347.
11. *Dudin A., Semenova O.* Stable algorithm for stationary distribution calculation for a BMAP—SM—1 queueing system with markovian input of disasters // *J. Appl. Prob.*, 2004. Vol. 42. P. 547–556.
12. *Zeifman A., Satin Ya., Chegodaev A., Bening V., Shorgin V.* Some bounds for $M(t)/M(t)/S$ queue with catastrophes // *4th Conference (International) on Performance Evaluation Methodologies and Tools Proceedings*. Athens, Greece, 2008.

13. Зейфман А. И., Сатин Я. А., Чегодаев А. В. О нестационарных системах обслуживания с катастрофами // Информатика и её применения, 2009. Т. 3. Вып. 1. С. 47–54.
14. Зейфман А. И., Сатин Я. А., Коротышева А. В., Терешина Н. А. О предельных характеристиках системы обслуживания $M(t)/M(t)/S$ с катастрофами // Информатика и её применения, 2009. Т. 3. Вып. 3. С. 16–22.
15. Zeifman A., Satin Ya., Shorgin S., Bening V. On $M_n(t)/M_n(t)/S$ queues with catastrophes // 4th Conference (International) on Performance Evaluation Methodologies and Tools Proceedings. Pisa, Italy, 2009.
16. Зейфман А. И., Коротышева А. В., Панфилова Т. Л., Шоргин С. Я. Оценки устойчивости для некоторых систем обслуживания с катастрофами // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 27–33.
17. Zeifman A., Korotysheva A. Perturbation bounds for $M_t/M_t/N$ queue with catastrophes // Stochastic Models, 2012. Vol. 28. P. 49–62.
18. Zeifman A., Leorato S., Orsingher E., Satin Ya., Shilova G. Some universal limits for nonhomogeneous birth and death processes // Queueing Syst., 2006. Vol. 52. P. 139–151.
19. Далецкий Ю. Л., Крейн М. Г. Устойчивость решений дифференциальных уравнений в банаховом пространстве. — М.: Наука, 1970.
20. Ван Доорн Э. А., Зейфман А. И., Панфилова Т. Л. Оценки и асимптотика скорости сходимости для процессов рождения и гибели // Теория вероятностей и ее применения, 2009. Т. 54. С. 18–38.

ОБОБЩЕННОЕ РАСПРЕДЕЛЕНИЕ ЛАПЛАСА КАК ПРЕДЕЛЬНОЕ ДЛЯ СЛУЧАЙНЫХ СУММ И СТАТИСТИК, ПОСТРОЕННЫХ ПО ВЫБОРКАМ СЛУЧАЙНОГО ОБЪЕМА*

В. Ю. Королев¹, В. Е. Бенинг², Л. М. Закс³, А. И. Зейфман⁴

Аннотация: Доказываются предельные теоремы, устанавливающие критерии сходимости распределений случайных сумм и статистик, построенных по выборкам случайного объема, к обобщенному распределению Лапласа.

Ключевые слова: обобщенное распределение Лапласа; симметричное устойчивое распределение; одно-стороннее устойчивое распределение; масштабная смесь нормальных законов; случайная сумма; выборка случайного объема; смешанное пуассоновское распределение

1 Обобщенное распределение Лапласа

Пусть $0 < \alpha \leq 2$. *Обобщенным распределением Лапласа* назовем абсолютно непрерывное распределение вероятностей, задаваемое плотностью

$$f_{\alpha}(x) = \frac{\alpha}{2\Gamma(1/\alpha)} e^{-|x|^{\alpha}}, \quad -\infty < x < \infty. \quad (1)$$

Для упрощения дальнейших выкладок и обозначений здесь и далее в представлении (1) будет использоваться лишь один параметр α , который является «характеристическим», определяя форму обобщенного распределения Лапласа. При $\alpha = 1$ соотношение (1) определяет классическое распределение Лапласа (с нулевым средним и дисперсией 2). При $\alpha = 2$ соотношение (1) определяет нормальное (гауссово) распределение (с нулевым средним и дисперсией 1/2).

Класс распределений (1) был введен и изучен М. Ф. Субботиным в 1923 г. [1]. В разных источниках распределения этого класса называются по-разному. Например, на с. 74–76 книги [2] эти распределения называются *обобщенными распределениями ошибок*, в книге [3] они названы *экспоненциально-степенными* (exponential power distributions). В [4] используется термин *обобщенные показательные распределения*, тогда как в статьях [5, 6] эти

распределения названы соответственно *обобщенными гауссовыми* и *обобщенными нормальными*. Распределения типа (1) широко применяются в разнообразных областях от астрономических измерений и обработки изображений до байесовского анализа.

В работе [7] было замечено, что при $0 < \alpha \leq 2$ распределения (1) представимы в виде масштабных смесей нормальных законов (этот результат также цитируется в [8]). Для удобства приведем свое доказательство этого результата.

Функцию распределения и плотность строго устойчивого распределения с характеристическим показателем α и параметром θ , задаваемого характеристической функцией

$$f_{\alpha, \theta}(t) = \exp \left\{ -|t|^{\alpha} \exp \left\{ -\frac{i\pi\theta\alpha}{2} \operatorname{sign} t \right\} \right\}, \quad t \in \mathbb{R}, \quad (2)$$

где $0 < \alpha \leq 2$, $|\theta| \leq \theta_{\alpha} = \min\{1, 2/\alpha - 1\}$ будем обозначать соответственно $G_{\alpha, \theta}(x)$ и $g_{\alpha, \theta}(x)$ (см., например, [9]). Функцию распределения и плотность стандартного нормального закона будем обозначать соответственно $\Phi(x)$ и $\phi(x)$,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \quad \Phi(x) = \int_{-\infty}^x \phi(z) dz.$$

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-12026-офи-м, 11-01-00515а, 11-07-00112а и 12-07-00115а), а также Министерством образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики РАН, vkorolev@cs.msu.ru

² Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики РАН, bening@cs.msu.ru

³ Альфа-банк, отдел моделирования и математической статистики, lily.zaks@gmail.com

⁴ Вологодский государственный педагогический университет; Институт проблем информатики Российской академии наук; ВНКЦ ЦЭМИ РАН, a_zeifman@mail.ru

Символ $\stackrel{d}{=}$ будет обозначать совпадение распределений.

Лемма 1. *Обобщенное распределение Лапласа является масштабной смесью нормальных законов.*

Доказательство. Из (2) вытекает, что характеристическая функция симметричного ($\theta = 0$) строго устойчивого распределения имеет вид

$$f_{\alpha,0}(t) = e^{-|t|^\alpha}, \quad t \in \mathbb{R}. \quad (3)$$

С другой стороны, хорошо известно, что симметричное строго устойчивое распределение с параметром α является масштабной смесью нормальных законов, в которой смешивающим распределением является односторонний устойчивый закон ($\theta = 1$) с параметром $\alpha/2$:

$$G_{\alpha,0}(x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{z}}\right) dG_{\alpha/2,1}(z), \quad x \in \mathbb{R} \quad (4)$$

(см., например, теорему 3.3.1 в [9]). Запишем соотношение (4) в терминах характеристических функций с учетом (3):

$$e^{-|t|^\alpha} = \int_0^\infty \exp\left\{-\frac{t^2 z}{2}\right\} g_{\alpha/2,1}(z) dz. \quad (5)$$

Обозначим

$$h_{\alpha/2}(z) = \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}} \frac{g_{\alpha/2,1}(z)}{\sqrt{z}}, \quad z \geq 0.$$

Тогда, переобозначив аргумент $t \mapsto x$ и выполнив несколько формальных тождественных преобразований равенства (5), будем иметь:

$$\begin{aligned} \ell_\alpha(x) &= \frac{\alpha}{2\Gamma(1/\alpha)} e^{-|x|^\alpha} = \\ &= \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}} \int_0^\infty \frac{\sqrt{z}}{\sqrt{2\pi}} \exp\left\{-\frac{x^2 z}{2}\right\} \frac{g_{\alpha/2,1}(z)}{\sqrt{z}} dz = \\ &= \int_0^\infty \sqrt{z} \phi(x\sqrt{z}) h_{\alpha/2}(z) dz. \quad (6) \end{aligned}$$

Можно убедиться, что $h_{\alpha/2}(z)$ — плотность распределения неотрицательной случайной величины. Действительно, при каждом $z > 0$

$$\int_{-\infty}^\infty \sqrt{z} \phi(x\sqrt{z}) dx = 1.$$

Поэтому из (6) вытекает, что

$$\begin{aligned} 1 &= \int_{-\infty}^\infty \ell_\alpha(x) dx = \\ &= \int_{-\infty}^\infty \int_0^\infty \sqrt{z} \phi(x\sqrt{z}) h_{\alpha/2}(z) dz dx = \\ &= \int_0^\infty h_{\alpha/2}(z) \left(\int_{-\infty}^\infty \sqrt{z} \phi(x\sqrt{z}) dx \right) dz = \\ &= \int_0^\infty h_{\alpha/2}(z) dz. \end{aligned}$$

Если Z_α — случайная величина, имеющая обобщенное распределение Лапласа с параметром α , то соотношение (6) означает, что

$$Z_\alpha \stackrel{d}{=} X \sqrt{U_{\alpha/2}},$$

где X и $U_{\alpha/2}$ — независимые случайные величины, причем X имеет стандартное нормальное распределение,

$$U_{\alpha/2} \stackrel{d}{=} \frac{1}{V_{\alpha/2}},$$

а $V_{\alpha/2}$ — неотрицательная абсолютно непрерывная случайная величина с плотностью $h_{\alpha/2}(z)$. Лемма доказана.

Поскольку функция $h_{\alpha/2}(x)$, введенная выше, является плотностью, то из ее определения вытекает следующее интересное утверждение, позволяющее вычислить $E U_{\alpha/2,1}^{-1/2}$, несмотря на то что плотность $g_{\alpha/2,1}(z)$, вообще говоря, нельзя выписать в явном виде в терминах элементарных функций, однако не имеющее прямого отношения к теме данной статьи.

Следствие 1. *Пусть $Y_{\alpha,1}$ — неотрицательная случайная величина, имеющая одностороннее устойчивое распределение с характеристическим показателем $\alpha \in (0, 1)$. Тогда*

$$E Y_{\alpha,1}^{-1/2} = \frac{\Gamma(1/(2\alpha))}{\alpha\sqrt{2\pi}}.$$

Замечание 1. Соотношению (6) можно придать несколько иной вид. Как несложно убедиться, плотность $w_{\alpha/2}$ случайной величины $U_{\alpha/2}$ имеет вид:

$$\begin{aligned} w_{\alpha/2}(z) &= z^{-2} h_{\alpha/2}(z^{-1}) = \\ &= \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}} \frac{g_{\alpha/2,1}(z^{-1})}{z^{3/2}}, \quad z > 0. \end{aligned}$$

Тогда (6) можно записать в эквивалентном виде

$$\begin{aligned} \ell_\alpha(x) &= \frac{\alpha}{\Gamma(1/\alpha)} \sqrt{\frac{\pi}{2}} \int_0^\infty \frac{1}{\sqrt{2\pi z}} \times \\ &\times \exp\left\{-\frac{x^2}{2z}\right\} \frac{g_{\alpha/2,1}(z^{-1})}{z^{3/2}} dz = \\ &= \int_0^\infty \frac{1}{\sqrt{z}} \phi\left(\frac{x}{\sqrt{z}}\right) w_{\alpha/2}(z) dz. \end{aligned} \quad (7)$$

Пример 1. Рассмотрим случай $\alpha = 1$. Тогда, как известно, $G_{1/2,1}(x)$ — это распределение Леви (распределение момента времени достижения стандартным винеровским процессом единичного уровня, являющееся также частным случаем обратного нормального распределения). Ему соответствует плотность

$$g_{1/2,1}(z) = \frac{1}{z^{3/2}\sqrt{2\pi}} \exp\left\{-\frac{1}{2z}\right\}, \quad z > 0.$$

В этом случае

$$\begin{aligned} w_{1/2}(z) &= \sqrt{\frac{\pi}{2}} \frac{g_{1/2,1}(z^{-1})}{z^{3/2}} = \\ &= \frac{\sqrt{\pi} z^{3/2} e^{-z/2}}{\sqrt{2}\sqrt{2\pi} z^{3/2}} = \frac{1}{2} e^{-z/2}, \end{aligned}$$

т.е. в рассматриваемом случае $w_{1/2}(z) = (1/2)e^{-z/2}$ — плотность экспоненциального распределения с параметром $1/2$. При этом в соответствии (7)

$$\ell_1(x) = \frac{1}{2} e^{-|x|} = \int_0^\infty \frac{1}{\sqrt{z}} \phi\left(\frac{x}{\sqrt{z}}\right) \frac{e^{-z/2}}{2} dz,$$

что согласуется с хорошо известным свойством классического распределения Лапласа (см., например, лемму 12.7.1 в [10]).

В прикладной теории вероятностей хорошо известен принцип, восходящий, по-видимому, к работе [11], согласно которому та или иная модель может считаться в достаточной мере обоснованной только тогда, когда она является *асимптотической аппроксимацией*, т.е. когда существует довольно простая предельная теорема, в которой рассматриваемая модель выступает в качестве предельного распределения.

В книге [12] прослежена глубокая связь этого принципа с универсальным принципом неубывания энтропии в замкнутых системах. Как было показано, при $0 < \alpha \leq 2$ обобщенное распределение

Лапласа имеет вид масштабных смесей нормальных законов. Как известно, нормальное распределение обладает максимальной (дифференциальной) энтропией среди всех распределений, носителем которых является вся числовая прямая, и имеющих конечный второй момент. Если бы моделируемая сложная система была информационно изолирована от окружающей среды, то в соответствии с принципом неубывания энтропии, который в теории вероятностей проявляется в виде предельных теорем [12], наблюдаемые статистические распределения ее характеристик были бы неотличимы от нормального. Но поскольку любая математическая модель по своему определению не может учесть все факторы, влияющие на состояние или эволюцию моделируемой системы, то параметры этого нормального закона изменяются в зависимости от состояния среды, внешней по отношению к моделируемой системе. Другими словами, эти параметры являются случайными и изменяются под влиянием информационных потоков между системой и внешней средой. Таким образом, во многих ситуациях разумные модели статистических закономерностей изменения параметров сложных систем должны иметь вид смесей нормальных законов, частным случаем которых является обобщенное распределение Лапласа (1).

По-видимому, до сих пор обобщенное распределение Лапласа использовалось во многих задачах прежде всего в силу относительной простоты его аналитического представления. «Асимптотического» обоснования адекватности подобной модели пока дано не было. В данной работе будет показано, что обобщенное распределение Лапласа может выступать в качестве предельного в довольно простых предельных теоремах для регулярных статистик, в частности в схеме случайного суммирования случайных величин, и, следовательно, наряду с нормальным законом может считаться асимптотической аппроксимацией для распределений многих процессов, например сходных с неоднородными случайными блужданиями.

Функции распределения, соответствующие плотностям $\ell_\alpha(x)$, $h_{\alpha/2}(z)$ и $w_{\alpha/2}(z)$, будем обозначать соответствующими заглавными латинскими буквами: $L_\alpha(x)$, $H_{\alpha/2}(z)$ и $W_{\alpha/2}(z)$. Тогда, как несложно видеть, соотношения (6) и (7) эквивалентны соотношениям:

$$L_\alpha(x) = \int_0^\infty \Phi(x\sqrt{z}) dH_{\alpha/2}(z); \quad (8)$$

$$L_\alpha(x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{z}}\right) dW_{\alpha/2}(z). \quad (9)$$

2 Критерий сходимости распределений случайных сумм к обобщенному распределению Лапласа

Всюду далее символ \implies обозначает сходимость по распределению.

Рассмотрим последовательность независимых одинаково распределенных случайных величин X_1, X_2, \dots , заданных на некотором вероятностном пространстве $(\Omega, \mathfrak{A}, P)$. Будем предполагать, что $EX_1 = 0$, $0 < \sigma^2 = DX_1 < \infty$. Для натурального $n \geq 1$ положим

$$S_n = X_1 + \dots + X_n.$$

Пусть N_1, N_2, \dots — последовательность целочисленных неотрицательных случайных величин, заданных на том же самом вероятностном пространстве так, что при каждом $n \geq 1$ случайная величина N_n независима от последовательности X_1, X_2, \dots . Всяду далее для определенности будем считать, что $\sum_{j=1}^0 = 0$.

Принято считать, что случайная последовательность N_1, N_2, \dots неограниченно возрастает ($N_n \rightarrow \infty$) по вероятности, если для любого $m \in (0, \infty)$ $P(N_n \leq m) \rightarrow 0$ при $n \rightarrow \infty$.

Лемма 1. *Предположим, что случайные величины X_1, X_2, \dots и N_1, N_2, \dots удовлетворяют указанным выше условиям, причем $N_n \rightarrow \infty$ по вероятности при $n \rightarrow \infty$. Для того чтобы существовала такая функция распределения $F(x)$, что*

$$P\left(\frac{S_{N_n}}{\sigma\sqrt{n}} < x\right) \implies F(x) \quad (n \rightarrow \infty),$$

необходимо и достаточно, чтобы существовала функция распределения $Q(x)$, удовлетворяющая условиям:

$$Q(0) = 0; \quad F(x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{y}}\right) dQ(y);$$

$$x \in \mathbb{R}; \quad P(N_n < nx) \implies Q(x) \quad (n \rightarrow \infty).$$

Доказательство. Данная лемма доказана в работе [13].

Теорема 2. *Предположим, что случайные величины X_1, X_2, \dots и N_1, N_2, \dots удовлетворяют указанным выше условиям, причем $N_n \rightarrow \infty$ по вероятности при $n \rightarrow \infty$. Для того чтобы*

$$P\left(\frac{S_{N_n}}{\sigma\sqrt{n}} < x\right) \implies L_\alpha(x) \quad (n \rightarrow \infty),$$

необходимо и достаточно, чтобы

$$P(N_n < nx) \implies W_{\alpha/2}(x) \quad (n \rightarrow \infty).$$

Доказательство. Данное утверждение является непосредственным следствием леммы 1 с $Q(x) = W_{\alpha/2}(x)$ и представления (9).

3 Критерий сходимости распределений регулярных статистик, построенных по выборкам случайного объема, к обобщенному распределению Лапласа

Рассмотрим традиционную для математической статистики постановку задачи. Пусть $T_n = T_n(X_1, \dots, X_n)$ — некоторая статистика, то есть измеримая функция от случайных величин X_1, \dots, X_n . Для каждого $n \geq 1$ определим случайную величину T_{N_n} , положив

$$T_{N_n}(\omega) = T_{N_n(\omega)}(X_1(\omega), \dots, X_{N_n(\omega)}(\omega))$$

для каждого элементарного исхода $\omega \in \Omega$.

Будем говорить, что статистика T_n асимптотически нормальна, если существуют $\delta > 0$ и $\theta \in \mathbb{R}$ такие, что

$$P(\delta\sqrt{n}(T_n - \theta) < x) \implies \Phi(x) \quad (n \rightarrow \infty). \quad (10)$$

Лемма 2. *Предположим, что $N_n \rightarrow \infty$ по вероятности. Пусть статистика T_n асимптотически нормальна в смысле (10). Для того чтобы существовала такая функция распределения $F(x)$, что*

$$P(\delta\sqrt{n}(T_{N_n} - \theta) < x) \implies F(x) \quad (n \rightarrow \infty),$$

необходимо и достаточно, чтобы существовала функция распределения $Q(x)$, удовлетворяющая условиям:

$$Q(0) = 0; \quad F(x) = \int_0^\infty \Phi(x\sqrt{y}) dQ(y);$$

$$x \in \mathbb{R}; \quad P(N_n < nx) \implies Q(x) \quad (n \rightarrow \infty).$$

Доказательство. Данная лемма, по сути, является частным случаем теоремы 3 из [14], доказательство которой, в свою очередь, основано на общих теоремах о сходимости суперпозиций независимых случайных последовательностей [13, 15] (см. также теорему 3.3.2 в [12]).

Теорема 3. *Предположим, что $N_n \rightarrow \infty$ по вероятности. Пусть статистика T_n асимптотически нормальна в смысле (10). Для того чтобы*

$$P(\delta\sqrt{n}(T_{N_n} - \theta) < x) \implies L_\alpha(x) \quad (n \rightarrow \infty),$$

необходимо и достаточно, чтобы

$$P(N_n < nx) \implies H_{\alpha/2}(x) \quad (n \rightarrow \infty).$$

Доказательство. Данное утверждение является непосредственным следствием леммы 2 с $Q(x) = H_{\alpha/2}(x)$ и представления (8).

4 Обсуждение

В теоремах 2 и 3 главным условием является сходимость распределений нормированных индексов N_n к распределениям $W_{\alpha/2}$ и $H_{\alpha/2}$ соответственно. Приведем довольно общий пример, показывающий, когда эти условия можно считать выполненными.

В книге [12] 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 книгах [10, 16–18]. В книгах [18, 19] 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. Для большей наглядности рассмотрим случай, когда в рассматриваемой модели время t остается фиксированным, а $\Lambda(t) = nU_{\alpha/2}$, где n — вспомогательный параметр, $U_{\alpha/2}$ — случайная величина с функцией распределения $W_{\alpha/2}(x)$, 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. Для каждого натурального n положим

$$N_n = M(nU_{\alpha/2}).$$

Очевидно, что так определенная случайная величина N_n имеет смешанное пуассоновское распределение

$$\begin{aligned} P(N_n = k) &= P(M(nU_{\alpha/2}) = k) = \\ &= \int_0^\infty e^{-nz} \frac{(nz)^k}{k!} w_{\alpha/2}(z) dz, \quad k = 0, 1, \dots \end{aligned}$$

Так определенная случайная величина N_n может быть интерпретирована как число событий, зарегистрированных к моменту времени n в пуассоновском процессе со случайной интенсивностью, имеющей плотность $w_{\alpha/2}(z)$. Предположим, что случайная величина $U_{\alpha/2}$ и пуассоновский процесс $M(t)$ независимы от последовательности X_1, X_2, \dots . Тогда, очевидно, при каждом n случайная величина N_n также будет независима от этой последовательности.

Обозначим $A_n(z) = P(N_n < nz), z \geq 0$ ($A_n(z) = 0$ при $z < 0$). Несложно видеть, что $A_n(x) \implies W_{\alpha/2}(x) (n \rightarrow \infty)$. 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 $x \in \mathbb{R}$

$$A_n(x) = \int_0^\infty \Pi(nx; nz) dW_{\alpha/2}(z),$$

то по теореме Лебега о мажорируемой сходимости при $n \rightarrow \infty$

$$\begin{aligned} A_n(x) &\implies \int_0^\infty E(x/z; 1) dW_{\alpha/2}(z) = \\ &= \int_0^x dW_{\alpha/2}(z) = W_{\alpha/2}(x), \end{aligned}$$

т. е. так определенные случайные величины N_n удовлетворяют условию, фигурирующему в лемме 1, с $Q(x) = W_{\alpha/2}(x)$.

Аналогично пусть $V_{\alpha/2}$ — случайная величина с функцией распределения $H_{\alpha/2}(x)$, независимая от стандартного пуассоновского процесса $M(t)$, $t \geq 0$. Для натурального n положим $N_n = M(nV_{\alpha/2})$. Несложно видеть, что распределение случайной величины N_n является смешанным пуассоновским и имеет вид:

$$P(N_n = k) = \frac{1}{k!} \int_0^{\infty} e^{-nz} (nz)^k h_{\alpha/2}(z) dz, \\ k = 0, 1, 2, \dots$$

Так определенная случайная величина N_n может быть интерпретирована как число событий, зарегистрированных к моменту времени n в пуассоновском процессе со случайной интенсивностью, имеющей плотность $h_{\alpha/2}(z)$. Предположим, что случайная величина $V_{\alpha/2}$ и пуассоновский процесс $M(t)$ независимы от последовательности X_1, X_2, \dots . Тогда, очевидно, при каждом n случайная величина N_n также будет независима от этой последовательности.

Как и выше, легко убедиться, что $P(N_n < nz) \Rightarrow H_{\alpha/2}(z)$ ($n \rightarrow \infty$), т. е. так определенные случайные величины N_n удовлетворяют условию, фигурирующему в лемме 2, с $Q(x) = H_{\alpha/2}(x)$.

Литература

1. *Subbotin M. T.* On the law of frequency of error // Математический сборник, 1923. Т. 31. Вып. 2. С. 296–301.
2. *Evans M., Hastings N., Peacock B.* Statistical distributions. — 3rd ed. — New York: Wiley, 2000.
3. *Box G., Tiao G.* Bayesian inference in statistical analysis. — Reading: Addison–Wesley, 1973.
4. RiskMetrics Technical Document. — New York: Risk-Metric Group, J. P. Morgan, 1996.
5. *Varanasi M. K., Aazhang B.* Parametric generalized Gaussian density estimation // J. Acoustic Soc. Amer., 1989. Vol. 86. No. 4. P. 1404–1415.
6. *Nadaraja S.* A generalized normal distribution // J. Appl. Stat., 2005. Vol. 32. No. 7. P. 685–694.
7. *West M.* On scale mixtures of normal distributions // Biometrika, 1987. Vol. 74. No. 3. P. 646–648.
8. *Choy S. T. B., Smith A. F. F.* Hierarchical models with scale mixtures of normal distributions // TEST, 1997. Vol. 6. P. 205–221.
9. *Золотарев В. М.* Одномерные устойчивые распределения. — М.: Наука, 1983.
10. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011. 620 с.
11. *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. — М.—Л.: ГИТТЛ, 1949.
12. *Gnedenko B. V., Korolev V. Yu.* Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996.
13. *Королев В. Ю.* Сходимость случайных последовательностей с независимыми случайными индексами. I // Теория вероятностей и ее применения, 1994. Т. 39. Вып. 2. С. 313–333.
14. *Королев В. Ю.* Сходимость случайных последовательностей с независимыми случайными индексами. II // Теория вероятностей и ее применения, 1995. Т. 40. Вып. 4. С. 907–910.
15. *Korolev V. Yu.* A general theorem on the limit behavior of superpositions of independent random processes with applications to Cox processes // J. Math. Sci., 1996. Vol. 81. No. 5. P. 2951–2956.
16. *Bening V., Korolev V.* Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002. 434 p.
17. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС-ПРЕСС, 2008.
18. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011. 510 с.
19. *Stochastic models of structural plasma turbulence / Eds. V. Korolev, N. Skvortsova.* — Utrecht: VSP, 2006. 400 p.

НИЖНИЕ ОЦЕНКИ УСТОЙЧИВОСТИ СМЕСЕЙ НОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ К ВОЗМУЩЕНИЯМ СМЕШИВАЮЩИХ РАСПРЕДЕЛЕНИЙ*

А. Л. Назаров¹

Аннотация: Исследуется устойчивость смесей нормальных распределений к возмущениям смешивающих распределений. Рассматриваются неравенства, описывающие близость смешивающих распределений через близость соответствующих смесей. Доказана теорема существования оценок устойчивости для подклассов масштабных и сдвиговых смесей нормальных законов. Оценка для сдвиговых смесей выписана в явном виде. При этом приведен пример, показывающий, что полученный результат не может быть принципиально улучшен без дополнительных предположений.

Ключевые слова: смеси нормальных распределений; устойчивость стохастических моделей; преобразование Фурье; теорема Планшереля; теорема Прохорова; метрика Леви; нижние оценки устойчивости смесей

1 Введение

Изучение проблем устойчивости стохастических моделей представляет большой интерес для специалистов, занимающихся теорией вероятностей и математической статистикой. Данные задачи важны как для исследователей, работающих над прикладными задачами, так и для тех, кто занимается теоретическими вопросами. Зачастую исследование устойчивости необходимо для обоснования применимости того или иного метода к анализу реальных данных.

В качестве примера можно привести задачу обоснования применимости сеточных методов разделения смесей вероятностных распределений [1–4], использующихся для оценки смешивающего распределения по реализациям смеси. При работе этих алгоритмов оценка смешивающего распределения ищется в классе распределений, сосредоточенных в конечном множестве фиксированных точек, покрывающих область, содержащую носитель истинного распределения. Для доказательства допустимости такого подхода используются оценки устойчивости смесей вероятностных распределений к возмущениям смешивающих распределений.

В статье рассматриваются вопросы устойчивости смесей нормальных распределений. Смеси нормальных распределений используются для изучения и моделирования многих явлений реаль-

ного мира. Доказаны предельные теоремы, описывающие сходимость одномерных распределений обобщенных процессов Кокса к сдвиг-масштабным смесям нормальных законов [4]. Примеры случайных процессов, имеющих в качестве одномерных распределений смеси нормальных распределений, можно найти в [4] и работах из библиографии к ней.

Оценкам устойчивости различных классов смесей нормальных распределений к возмущениям смешивающих распределений был посвящен ряд работ. Результат для частного случая — простой модели загрязнения (контаминации), предложенной Тьюки [5], был получен в [3]. В [6] можно найти решение этой задачи для оценки расстояния между нормальным распределением и масштабной смесью при выполнении некоторых условий, наложенных на смешивающее распределение. Работа [7] посвящена верхним оценкам устойчивости нормальных смесей. Выписаны верхние оценки близости сдвиг-масштабных смесей через близость смешивающих распределений в метрике Леви–Прохорова. Кроме этого рассмотрены аналогичные оценки в метрике Леви для масштабных и сдвиговых смесей.

Целью данного исследования стала попытка установить, как близость двух смесей вероятностных распределений влияет на близость соответствующих им смешивающих распределений в некоторой вероятностной метрике. Для этого рассматривались нижние оценки близости смесей

* Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м, 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, nazarov.vmik@gmail.com

через расстояние между смешивающими распределениями. Полученные результаты используются в [8] для исследования асимптотических свойств оценок, полученных с помощью сеточных методов разделения смесей.

Не ограничивая общности, если это специально не оговорено, будем считать, что все случайные величины и элементы, рассматриваемые в данной статье, определены на одном вероятностном пространстве (Ω, \mathcal{A}, P) .

2 Используемые метрики

Для описания близости распределений в данной статье используется несколько метрик. Условимся здесь и далее обозначать метрику Леви через L . Напомним определение метрики Леви.

Пусть X и Y — две случайные величины с функциями распределения F и G соответственно. Расстояние (метрика) Леви $L(X, Y)$ между случайными величинами X и Y определяется следующим образом:

$$L(X, Y) = \inf \{ \varepsilon > 0 \mid F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon, \forall x \in \mathbb{R} \}.$$

Расстояние Леви допускает наглядную геометрическую интерпретацию: оно равно длине стороны наибольшего квадрата, который можно вписать между графиками функций распределения F и G так, чтобы его стороны были параллельны координатным осям. Так как в определении метрики Леви задействованы не сами случайные величины X и Y , а только их функции распределения F и G , то выражения $L(X, Y)$ и $L(F, G)$ не будут различаться:

$$L(X, Y) \equiv L(F, G).$$

Метрику в L^p будем обозначать через $\| \cdot \|_p$:

$$\|g\|_p = \left(\int_{-\infty}^{\infty} |g(s)|^p ds \right)^{1/p}.$$

Для оценки близости распределений смесей будет использоваться расстояние между соответствующими функциями распределения в метрике L^2 . Данное расстояние метризует слабую сходимость для некоторого подкласса рассматриваемых смесей.

Чтобы расстояние между функциями распределения в L^2 было конечно, достаточно потребовать от соответствующих случайных величин существования конечного первого момента. Действительно, пусть X_1, X_2 — случайные величины, имеющие конечный первый момент, F_1, F_2 — соответствующие функции распределения.

$$\begin{aligned} \|F_1 - F_2\|_2^2 &= \int_{-\infty}^{\infty} |F_1(s) - F_2(s)|^2 ds \leq \\ &\leq \int_{-\infty}^{\infty} |F_1(s) - F_2(s)| ds \leq \int_{-\infty}^0 (F_1(s) + F_2(s)) ds + \\ &+ \int_0^{\infty} (1 - F_1(s) + 1 - F_2(s)) ds = E|X_1| + E|X_2| < \infty. \end{aligned}$$

Очевидно, что из сходимости функций распределений в L^2 напрямую следует сходимость в метрике Леви, которая, в свою очередь, метризует слабую сходимость. Однако легко видеть, что обратное утверждение неверно даже для класса распределений, имеющих конечный первый момент. Наложим на рассматриваемые функции распределения дополнительные условия. Справедливо следующее утверждение.

Лемма 1. *Рассмотрим класс K случайных величин, имеющих конечный первый момент. Пусть $\{F_n\}_{n=1}^{\infty}$ — последовательность функций распределения случайных величин из K , слабо сходящихся к F . Если семейство K является равномерно интегрируемым, то последовательность $\{F_n\}_{n=1}^{\infty}$ сходится к F в метрике L^2 .*

Доказательство. Воспользовавшись методом одного (единого) вероятностного пространства [9] можно показать, что для функций распределения $\{F_n\}_{n=1}^{\infty}$ и F существует вероятностное пространство и последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$, заданных на нем, которая сходится почти наверное к X , заданной на нем же. Далее, воспользовавшись тем, что $\{X_n\}_{n=1}^{\infty}$ равномерно интегрируема и сходится почти всюду к X , получаем (см., например, [10]), что $\{X_n\}_{n=1}^{\infty}$ сходится к X в среднем.

Заметим, что доказательство существования последовательности $\{X_n\}_{n=1}^{\infty}$ в [10] конструктивно. Представленная последовательность обладает тем свойством, что для любых двух ее элементов Y и Z , имеющих функции распределения G и H , случайные величины $\max\{Y, Z\}$ и $\min\{Y, Z\}$ имеют функции распределения $\min\{G, H\}$ и $\max\{G, H\}$.

Зафиксируем некоторый номер $n \in \mathbb{N}$. Справедлива следующая цепочка неравенств:

$$\begin{aligned} 0 \leq |F_n - F|_2^2 &= \int_{-\infty}^{\infty} |F_n(s) - F(s)|^2 ds \leq \\ &\leq \int_{-\infty}^{\infty} |F_n(s) - F(s)| ds = \end{aligned}$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} (\max\{F_n(s), F(s)\} - \min\{F_n(s), F(s)\}) ds = \\ &= - \int_{-\infty}^0 (\min\{F_n(s), F(s)\} - \max\{F_n(s), F(s)\}) ds + \\ &\quad + \int_0^{\infty} ((1 - \min\{F_n(s), F(s)\}) - \\ &\quad - (1 - \max\{F_n(s), F(s)\})) ds = \\ &= E(\max\{X_n, X\} - \min\{X_n, X\}) = E|X_n - X|. \end{aligned}$$

Учитывая, что последовательность $\{X_n\}_{n=1}^{\infty}$ сходится в среднем, получаем утверждение леммы. \square

Таким образом, $\|\cdot\|_2$ метризует слабую сходимую для равномерно интегрируемых последовательностей случайных величин.

Наиболее интересным для исследования рассматриваемой задачи является класс сдвиг-масштабных смесей нормальных законов. В общем случае искомым оценкам для данного класса не существует. Это связано с тем, что без дополнительных ограничений класс сдвиг-масштабных смесей нормальных законов не является идентифицируемым [4]. Можно привести пример двух смесей, совпадающих по распределению, но имеющих различные смешивающие распределения. Поэтому в работе рассматриваются нижние оценки для идентифицируемых классов — сдвиговых и масштабных смесей нормальных законов.

Выпишем интересующие оценки для класса сдвиговых смесей нормальных законов, т.е. для свертки некоторых смешивающих распределений со стандартным нормальным законом.

3 Оценки устойчивости для сдвиговых смесей нормальных распределений

Рассмотрим класс сдвиговых смесей нормальных законов, имеющих конечный первый момент. Оценим, насколько близость распределений случайных величин

$$Y_1 = X + U_1 \text{ и } Y_2 = X + U_2, \quad X \sim \mathcal{N}(0, 1)$$

(случайные величины X и U_i стохастически независимы для $i = 1, 2$), очевидно, являющихся сдвиговыми смесями нормальных законов, влияет на близость распределений U_1 и U_2 .

Докажем сначала вспомогательное утверждение.

Лемма 2. Пусть F_1 и F_2 — произвольные функции распределения, G — функция симметричного распределения. Существуют функции распределения \tilde{F}_1 и \tilde{F}_2 , такие что $f(x) = \tilde{F}_1(x) - \tilde{F}_2(x)$, $x \in \mathbb{R}$, принимает одинаковые значения в парах точек непрерывности, симметричных относительно нуля, и

$$\begin{aligned} L(F_1, F_2) &\leq L(\tilde{F}_1, \tilde{F}_2), \quad \|F_1 * G - F_2 * G\|_p \geq \\ &\geq \|\tilde{F}_1 * G - \tilde{F}_2 * G\|_p, \quad p \geq 1. \end{aligned}$$

Доказательство. Ниже будем использовать следующее обозначение. Если некоторая случайная величина Z имеет функцию распределения H , то распределение случайной величины $-Z$ будем обозначать через $H'(x) \equiv (1 - H)(-x)$, $x \in \mathbb{R}$.

Пусть $L(F_1, F_2) = \delta$. Так как рассматриваемые расстояния инвариантны относительно сдвига, без ограничения общности можно считать, что

$$\sup \left\{ \varepsilon > 0 \mid F_1 \left(-\frac{\varepsilon}{2} \right) - \varepsilon \geq F_2 \left(\frac{\varepsilon}{2} \right) \right\} = \delta.$$

Положим $\tilde{F}_1(x) = (1/2)(F_1(x) + (1 - F_2)(-x))$, $\tilde{F}_2(x) = (1/2)(F_2(x) + (1 - F_1)(-x))$, $x \in \mathbb{R}$.

Очевидно,

$$\begin{aligned} \sup \left\{ \varepsilon > 0 \mid (1 - F_2) \left(\frac{\varepsilon}{2} \right) - \varepsilon \geq \right. \\ \left. \geq (1 - F_1) \left(-\frac{\varepsilon}{2} \right) \right\} = \delta. \end{aligned}$$

Следовательно, $L(\tilde{F}_1, \tilde{F}_2) \geq \delta$. Далее, воспользовавшись неравенством треугольника и линейностью оператора свертки, получаем

$$\begin{aligned} &\|\tilde{F}_1 * G - \tilde{F}_2 * G\|_p = \\ &= \frac{1}{2} \|F_1 * G + (1 - F_2) * G - F_2 * G - (1 - F_1) * G\|_p \leq \\ &\leq \frac{1}{2} \|F_1 * G - F_2 * G\|_p + \\ &+ \frac{1}{2} \|(1 - F_2) * G - (1 - F_1) * G\|_p. \quad (1) \end{aligned}$$

Учитывая, что G — функция симметричного распределения,

$$\begin{aligned} (1 - F_i) * G(x) &= \int_{-\infty}^{\infty} (1 - F_i)(-x + y) dG(y) = \\ &= 1 - \int_{-\infty}^{\infty} F_i(-x - y) dG(y) = \\ &= (1 - F_i * G)(-x), \quad \forall x \in \mathbb{R}, \end{aligned}$$

для $i = 1, 2$. Следовательно:

$$\|(1 - F_2) * G - (1 - F_1) * G\|_p = \|F_1 * G - F_2 * G\|_p.$$

Подставляя равенство в (1), получаем утверждение леммы. \square

Воспользовавшись результатом леммы 2, получим интересные оценки для класса сдвиговых смесей.

Теорема 1. Пусть X, U_1, U_2 — случайные величины; Φ, F_1, F_2 — соответствующие функции распределения, пары случайных величин X, U_1 и X, U_2 независимы, $X \sim \mathcal{N}(0, 1)$, у случайных величин U_1 и U_2 существуют конечные первые моменты. Если $L(F_1, F_2) \geq \delta$, то

$$\|F_1 * \Phi - F_2 * \Phi\|_2 \geq \sqrt{\frac{\delta^3}{2}} \exp\left\{-\frac{2}{\pi^2 \delta^6}\right\}.$$

Доказательство. Пусть $L(U_1, U_2) \geq \delta$, тогда $\|F_1 - F_2\|_2^2 \geq \delta^3$. Учитывая результат леммы 2, без ограничения общности будем считать, что $f(x) = F_1(x) - F_2(x)$, $x \in \mathbb{R}$, принимает одинаковые значения в парах точек непрерывности, симметричных относительно нуля. Так как у случайных величин U_1 и U_2 существуют конечные первые моменты, легко показать, что $f(x) \in L_1(-\infty, \infty)$. Следовательно, существует преобразование Фурье:

$$g(t) = \int_{-\infty}^{\infty} f(x) e^{-itx} dx, \quad t \in \mathbb{R}.$$

Воспользовавшись симметричностью почти всюду и ограниченностью sf , легко показать, что $|g(t)| \leq (2/|t|)$, $t \in \mathbb{R}$.

Из теоремы Планшереля следует, что $\|g\|_2^2 \geq 2\pi\delta^3$. Возьмем $A = 2/(\pi\delta^3)$. Получим:

$$\int_{-A}^A |g(t)|^2 dt \geq \|g\|_2^2 - 2 \int_A^{\infty} \frac{1}{t^2} dt = 2\pi\delta^3 - \pi\delta^3 = \pi\delta^3.$$

Далее,

$$\begin{aligned} \|F_1 * \Phi - F_2 * \Phi\|_2^2 &= \|f * \Phi\|_2^2 = \frac{1}{2\pi} \|g \cdot \varphi\|_2^2 \geq \\ &\geq \frac{1}{2\pi} \int_{-A}^A |g(t)|^2 e^{-t^2} dt \geq \\ &\geq \frac{e^{-A^2}}{2\pi} \int_{-A}^A |g(t)|^2 dt \geq \frac{\delta^3 e^{-A^2}}{2}. \end{aligned}$$

Здесь φ — характеристическая функция стандартного нормального закона. Подставляя значение A в последнее соотношение, получаем утверждение теоремы. \square

Заметим, что требование существования конечных первых моментов является достаточным, но не необходимым.

Оценка, полученная в теореме 1, на первый взгляд может показаться достаточно грубой. Ниже построен пример двух смесей, показывающий, что без дополнительных условий полученная оценка не может быть принципиально улучшена. Справедливо следующее утверждение.

Утверждение 1. Пусть $X \sim \mathcal{N}(0, 1)$, $\delta \in (0, (4 - \pi)/(8 + \pi))$. Существуют случайные величины U_1, U_2 такие, что пары случайных величин X, U_1 и X, U_2 независимы, расстояние Леви $L(U_1, U_2) \geq \delta$, а

$$L(X + U_1, X + U_2) \leq \left(\frac{21}{(4 - \pi)^2}\right)^{1/3} \delta \exp\left(-\frac{\pi^2}{3\delta^2}\right).$$

Доказательство. Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, где Ω — объединение счетного числа непересекающихся множеств $B_0, B_1, B_{-1}, A_n, n \in \mathbb{Z}$:

$$\begin{aligned} \mathcal{F} &= \sigma(\{B_0, B_1, B_{-1}, A_n, n \in \mathbb{Z}\}); \\ \mathbb{P}(B_{-1}) &= \mathbb{P}(B_1) = \frac{\pi\delta}{4 - \pi}; \quad \mathbb{P}(B_0) = 1 - \delta \frac{8 + \pi}{4 - \pi}; \\ \mathbb{P}(A_0) &= \delta; \\ \mathbb{P}(A_n) &= \frac{4\delta}{4 - \pi} \frac{1}{4n^2 - 1}, \quad n = 1, -1, 2, -2, \dots \end{aligned}$$

Пусть на указанном вероятностном пространстве заданы случайные величины U_1 и U_2 , имеющие функции распределения F_1 и F_2 соответственно:

$$U_1(\omega) = \begin{cases} \left(-n + \frac{(-1)^{n+1}}{2}\right) \delta, & \omega \in A_{-n}, n \in \mathbb{N}; \\ -\delta, & \omega \in B_{-1}; \\ -\frac{\delta}{2}, & \omega \in B_0; \\ \frac{\delta}{2}, & \omega \in A_0; \\ \frac{\delta}{2}, & \omega \in B_1; \\ \left(n + \frac{(-1)^{n+1}}{2}\right) \delta, & \omega \in A_n, n \in \mathbb{N}; \end{cases}$$

$$U_2(\omega) = \begin{cases} \left(-n + \frac{(-1)^n}{2}\right) \delta, & \omega \in A_{-n}, n \in \mathbb{N}; \\ -\frac{\delta}{2}, & \omega \in B_{-1}; \\ -\frac{\delta}{2}, & \omega \in B_0; \\ -\frac{\delta}{2}, & \omega \in A_0; \\ \delta, & \omega \in B_1; \\ \left(n + \frac{(-1)^n}{2}\right) \delta, & \omega \in A_n, n \in \mathbb{N}. \end{cases}$$

Разность функций распределения случайных величин U_1 и U_2 имеет вид:

$$F_1(x) - F_2(x) = \begin{cases} \frac{4\delta}{4-\pi} \frac{(-1)^n}{4n^2-1}, & x \in \left[\delta \left(n - \frac{1}{2} \right), \delta \left(n + \frac{1}{2} \right) \right), \\ & n \leq -2; \\ -\frac{4\delta}{4-\pi} \frac{1}{3}, & x \in \left[-\frac{3\delta}{2}, -\delta \right); \\ \frac{\pi\delta}{4-\pi} - \frac{4\delta}{4-\pi} \frac{1}{3}, & x \in \left[-\delta, -\frac{\delta}{2} \right); \\ -\delta, & x \in \left[-\frac{\delta}{2}, \frac{\delta}{2} \right); \\ \frac{\pi\delta}{4-\pi} - \frac{4\delta}{4-\pi} \frac{1}{3}, & x \in \left[\frac{\delta}{2}, \delta \right); \\ -\frac{4\delta}{4-\pi} \frac{1}{3}, & x \in \left[\delta, \frac{3\delta}{2} \right); \\ \frac{4\delta}{4-\pi} \frac{(-1)^n}{4n^2-1}, & x \in \left[\delta \left(n - \frac{1}{2} \right), \delta \left(n + \frac{1}{2} \right) \right), \\ & n \geq 2. \end{cases}$$

Заметим, что расстояние Леви между функциями распределения равно δ .

Выпишем преобразование Фурье разности функций распределения. Воспользовавшись тем, что для любого $a > 0$

$$\int_{-a}^a e^{-itx} dx = \frac{2 \sin at}{t}; \quad \left| \frac{\sin at}{t} \right| \leq a;$$

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \left| \sum_{k=n}^{\infty} \frac{(-1)^k}{4k^2-1} (I_{[\delta(-k-1/2), \delta(-k+1/2)]}(x) + I_{[\delta(k-1/2), \delta(k+1/2)]}(x)) \right| dx \leq$$

$$\leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \frac{2\delta}{4k^2-1} = 0;$$

$$\limsup_t \left| \frac{2}{t} \sum_{k=n}^{\infty} \frac{(-1)^k}{4k^2-1} \left(\sin \left(k + \frac{1}{2} \right) \delta t - \sin \left(k - \frac{1}{2} \right) \delta t \right) \right| =$$

$$= \limsup_t \left| \frac{4}{t} \sum_{k=n}^{\infty} \frac{(-1)^k}{4k^2-1} \left(\sin \frac{\delta}{2} t \cos k\delta t \right) \right| \leq$$

$$\leq 2\delta \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \frac{1}{4k^2-1} = 0,$$

и тем, что из сходимости последовательности функций в $L_1(-\infty, \infty)$ следует равномерная схо-

димость последовательности образов Фурье этих функций [11], получаем:

$$f(t) = \int_{-\infty}^{\infty} (F_1(x) - F_2(x)) e^{-itx} dx =$$

$$= \frac{2}{t} \left(-\delta \sin \frac{\delta}{2} t + \frac{\pi\delta}{4-\pi} \left(\sin \delta t - \sin \frac{\delta}{2} t \right) + \frac{4\delta}{4-\pi} \times \right.$$

$$\left. \times \sum_{n=1}^{\infty} \frac{(-1)^n}{4n^2-1} \left(\sin \left(n + \frac{1}{2} \right) \delta t - \sin \left(n - \frac{1}{2} \right) \delta t \right) \right).$$

Далее,

$$f(t) = \frac{2\delta \sin(\delta/2)t}{t} \left(-1 + \frac{\pi}{4-\pi} \left(2 \cos \frac{\delta}{2} t - 1 \right) - \right.$$

$$\left. - \frac{8}{4-\pi} \sum_{n=1}^{\infty} \frac{\cos \pi n}{1-4n^2} \cos n\delta t \right) =$$

$$= \frac{2\delta \sin(\delta/2)t}{t} \left(\frac{\pi}{4-\pi} 2 \cos \frac{\delta}{2} t - \frac{4}{4-\pi} - \right.$$

$$\left. - \frac{8}{4-\pi} \sum_{n=1}^{\infty} \frac{\cos \pi n}{1-4n^2} \cos n\delta t \right) =$$

$$= \frac{4\pi}{4-\pi} \frac{\delta \sin(\delta/2)t}{t} \left(\cos \frac{\delta}{2} t - \frac{2}{\pi} - \right.$$

$$\left. - \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\cos \pi n}{1-4n^2} \cos n\delta t \right).$$

Разложим функцию $g(t) = \cos(\delta/2)t$ в ряд Фурье на отрезке $[0, \pi/\delta]$ по системе функций

$$\left\{ \sqrt{\frac{\delta}{\pi}}, \sqrt{\frac{2\delta}{\pi}} \cos n\delta t, n = 1, 2, 3, \dots \right\}.$$

Учитывая, что

$$\int_0^{\pi/\delta} \cos \left(\frac{\delta}{2} x \right) \cos(n\delta x) dx = \frac{2}{\delta} \frac{\cos \pi n}{1-4n^2}, \quad n \geq 0,$$

получаем

$$\cos \frac{\delta}{2} t = \frac{2}{\pi} + \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\cos \pi n}{1-4n^2} \cos n\delta t, \quad t \in [0, \frac{\pi}{\delta}].$$

Так как умножение на ограниченную функцию сохраняет сходимость в $L_2[0, \pi/\delta]$,

$$\int_0^{\pi/\delta} (f(t))^2 dt = \int_0^{\pi/\delta} \left(\frac{4\pi}{4-\pi} \frac{\delta \sin(\delta/2)t}{t} \right)^2 \times$$

$$\times \left(\cos \frac{\delta}{2} t - \frac{2}{\pi} - \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\cos \pi n}{1-4n^2} \cos n\delta t \right)^2 dt = 0.$$

Перейдем теперь к оценке расстояния между функциями распределения случайных величин $X + U_1$, $X + U_2$. Заметим, что

$$\begin{aligned} \|F_1(x) - F_2(x)\|_2^2 &= 2 \sum_{n=2}^{\infty} \frac{16\delta^3}{(4-\pi)^2} \frac{1}{(4n^2-1)^2} + \\ &+ \frac{16\delta^3}{(4-\pi)^2} \frac{1}{9} + \left(\frac{\pi\delta}{4-\pi} - \frac{4\delta}{4-\pi} \frac{1}{3} \right)^2 \delta + \delta^3 \leq \\ &\leq \sum_{n=1}^{\infty} \frac{16\delta^3}{(4-\pi)^2} \frac{2}{(4n^2-1)} + \left(\frac{2}{4-\pi} \right)^2 \delta^3 + \delta^3 \leq \\ &\leq \frac{21\delta^3}{(4-\pi)^2}. \end{aligned}$$

Воспользовавшись теоремой Планшереля, оценим расстояние между функциями распределения смесей

$$\begin{aligned} \|\Phi * F_1(x) - \Phi * F_2(x)\|_2^2 &= \frac{1}{2\pi} \left\| \int_{-\infty}^{\infty} f(t) e^{-t^2/2} dt \right\|_2^2 = \\ &= 2 \frac{1}{2\pi} \int_0^{\pi/\delta} (f(t))^2 e^{-t^2} dt + 2 \frac{1}{2\pi} \int_{\pi/\delta}^{\infty} (f(t))^2 e^{-t^2} dt \leq \\ &\leq 2 \frac{e^{-\pi^2/\delta^2}}{2\pi} \int_{\pi/\delta}^{\infty} (f(t))^2 dt \leq \frac{e^{-\pi^2/\delta^2}}{2\pi} \|f(t)\|_2^2 \leq \\ &\leq e^{-\pi^2/\delta^2} \frac{21\delta^3}{(4-\pi)^2}. \end{aligned}$$

Далее получаем

$$\begin{aligned} L(\Phi * F_1, \Phi * F_2) &\leq \|\Phi * F_1(x) - \Phi * F_2(x)\|_2^{2/3} \leq \\ &\leq \left(\frac{21}{(4-\pi)^2} \right)^{1/3} \delta e^{-\pi^2/(3\delta^2)}. \end{aligned}$$

Заметим, что

$$\left(\frac{21}{(4-\pi)^2} \right)^{1/3} \approx 3,0545. \quad \square$$

Таким образом, вид оценок, полученных выше, не может быть принципиально улучшен без добавления ограничений на смешивающие распределения.

4 Оценки устойчивости для масштабных смесей нормальных распределений

Перейдем к исследованию устойчивости масштабных смесей нормальных законов. Верхние

оценки устойчивости, описывающие близость смесей через близость соответствующих им смешивающих распределений, получены в работе [7] с использованием метрики Леви. К сожалению, аналогичные результаты для нижних оценок устойчивости не могут быть получены без дополнительных ограничений на смешивающее распределение.

Лемма 3. Пусть X, V_1, V_2 — случайные величины, удовлетворяющие условиям $V_1 > 0, V_2 > 0$, пары случайных величин X, V_1 и X, V_2 независимы, случайная величина X имеет стандартное нормальное распределение. Справедливы следующие утверждения.

1. Для любых $\varepsilon \in (0, 1), \delta > 0$ существуют V_1, V_2 , такие, что $L(V_1, V_2) \geq \varepsilon, L(XV_1, XV_2) < \delta$.
2. Для любых $\varepsilon \in (0, 1), \delta > 0$ существуют V_1, V_2 , такие, что $L(\ln V_1, \ln V_2) \geq \varepsilon, L(XV_1, XV_2) < \delta$.

Доказательство.

1. Докажем первое утверждение. Пусть V_1, V_2 — случайные величины, вырожденные в точках $a, a + \varepsilon$ соответственно. Очевидно, $L(V_1, V_2) = \varepsilon$. Покажем, что для любого $\delta > 0$ существует $a = a(\delta)$, такое, что $L(XV_1, XV_2) < \delta$. Обозначим x_{\max} — рациональное число, такое, что $2(1 - \Phi(x_{\max})) < \delta$. Справедлива следующая цепочка равенств:

$$\begin{aligned} L(XV_1, XV_2) &= \\ &= \sup_{x \geq 0} \left\{ |\gamma| : \Phi\left(\frac{x}{a}\right) = \Phi\left(\frac{x+\gamma}{a+\varepsilon}\right) + \gamma \right\} = \\ &= \sup_{x \in [0, (a+\varepsilon)x_{\max}]} \left\{ |\gamma| : \Phi\left(\frac{x}{a}\right) = \Phi\left(\frac{x+\gamma}{a+\varepsilon}\right) + \gamma \right\} \vee \\ &\vee \sup_{x \in ((a+\varepsilon)x_{\max}, +\infty)} \left\{ |\gamma| : \Phi\left(\frac{x}{a}\right) = \Phi\left(\frac{x+\gamma}{a+\varepsilon}\right) + \gamma \right\}. \end{aligned}$$

Далее,

$$\begin{aligned} \sup_{x \in [0, (a+\varepsilon)x_{\max}]} \left\{ |\gamma| : \Phi\left(\frac{x}{a}\right) = \Phi\left(\frac{x+\gamma}{a+\varepsilon}\right) + \gamma \right\} &\leq \\ &\leq \sup_{x \in [0, (a+\varepsilon)x_{\max}]} \left(\Phi\left(\frac{x}{a}\right) - \Phi\left(\frac{x}{a+\varepsilon}\right) \right) \leq \\ &\leq \sup_{x \in [0, (a+\varepsilon)x_{\max}]} \left(\phi(0) \left(\frac{x}{a} - \frac{x}{a+\varepsilon} \right) \right) \leq \\ &\leq \frac{1}{\sqrt{2\pi}} \frac{x_{\max}\varepsilon}{a}. \end{aligned}$$

Рассмотрим теперь второе выражение:

$$\begin{aligned} \sup_{x \in ((a+\varepsilon)x_{\max}, +\infty)} \left\{ |\gamma| : \Phi\left(\frac{x}{a}\right) = \Phi\left(\frac{x+\gamma}{a+\varepsilon}\right) + \gamma \right\} &\leq \\ &\leq \sup_{x \in ((a+\varepsilon)x_{\max}, +\infty)} \left(\Phi\left(\frac{x}{a}\right) - \Phi\left(\frac{x}{a+\varepsilon}\right) \right) \leq \end{aligned}$$

$$\begin{aligned} &\leq \sup_{x \in ((a+\varepsilon)x_{\max}, +\infty)} \left(1 - \Phi\left(\frac{x}{a}\right) + \right. \\ &\quad \left. + 1 - \Phi\left(\frac{x}{a+\varepsilon}\right) \right) = \\ &= 1 - \Phi\left(x_{\max} + \frac{x_{\max}\varepsilon}{a}\right) + 1 - \Phi(x_{\max}) < \\ &\quad < 2(1 - \Phi(x_{\max})). \end{aligned}$$

Получается, что при $a > x_{\max}\varepsilon/(\sqrt{2\pi}\delta)$

$$L(XV_1, XV_2) < \delta \vee \delta = \delta.$$

2. Докажем второе утверждение. Аналогично пусть V_1, V_2 — случайные величины, вырожденные в точках $a, a \exp(\varepsilon)$ соответственно. Очевидно, $L(\ln V_1, \ln V_2) = \varepsilon$. Покажем, что для любого $\delta > 0$ существует $a = a(\delta)$, такое, что $L(XV_1, XV_2) < \delta$. Пусть Z — случайная величина, вырожденная в нуле. Воспользуемся неравенством треугольника:

$$L(XV_1, XV_2) \leq L(XV_1, Z) + L(Z, XV_2).$$

Далее,

$$\begin{aligned} (L(XV_1, Z))^2 &\leq \int_0^\infty \left(1 - \Phi\left(\frac{x}{a}\right) \right) dx = \frac{1}{\sqrt{2\pi}} a, \\ (L(Z, XV_2))^2 &\leq \frac{1}{\sqrt{2\pi}} a \exp(\varepsilon). \end{aligned}$$

Таким образом, при $a < \sqrt{2\pi}(\delta^2/4) \exp(-\varepsilon)$

$$L(XV_1, XV_2) < \frac{\delta}{2} + \frac{\delta}{2} \exp\left(-\frac{\varepsilon}{2}\right) < \delta. \quad \square$$

Из леммы 3 следует, что расстояние Леви между масштабными смесями нормальных законов невозможно оценить как через расстояние Леви между случайными величинами с соответствующими смешивающими распределениями, так и через расстояние Леви между их логарифмами. Таким образом, выписать оценки, аналогичные оценкам для сдвиговых смесей, без дополнительных ограничений не представляется возможным.

5 О существовании нижних оценок устойчивости смесей вероятностных распределений

В утверждениях, рассмотренных выше, на смешивающие распределения не накладывались никакие дополнительные ограничения, кроме технических (необходимых для того, чтобы расстояния в

используемых метриках были конечны). Однако если потребовать выполнения некоторых дополнительных условий, можно доказать существование искомым оценок для сдвиговых и масштабных смесей нормальных законов.

Приведенное ниже утверждение сформулировано для произвольного класса смесей, порожденных некоторым ядром. Это связано с тем, что при доказательстве никак не используются ни свойства какого-то конкретного ядра, ни конкретный вид смешивающих распределений. Смеси нормальных законов являются частным случаем класса, описываемого в приведенной теореме.

Рассмотрим семейство смесей $\mathfrak{A} = \{P_t, t \in T\}$ и семейство соответствующих смешивающих распределений $\mathfrak{B} = \{Q_t, t \in T\}$, заданных на измеримых пространствах $(S_{\mathfrak{A}}, \mathcal{F}_{\mathfrak{A}})$, $(S_{\mathfrak{B}}, \mathcal{F}_{\mathfrak{B}})$ соответственно. Предполагается, что T — некоторое параметрическое множество и для любого $t \in T$ смеси P_t соответствует смешивающее распределение Q_t . Через $\bar{\mathfrak{B}}$ обозначим семейство, содержащее все распределения, заданные на $(S_{\mathfrak{B}}, \mathcal{F}_{\mathfrak{B}})$ и являющиеся слабыми пределами последовательностей из \mathfrak{B} . Ему соответствуют семейство смесей $\bar{\mathfrak{A}}$ и параметрическое множество \bar{T} . Пусть α и β метризуют слабую сходимость в $\bar{\mathfrak{A}}$ и $\bar{\mathfrak{B}}$ соответственно.

Теорема 2. Пусть $\bar{\mathfrak{A}}$ идентифицируемо, \mathfrak{B} плотно. Если для $\bar{\mathfrak{A}}$ существуют верхние оценки устойчивости вида

$$\begin{aligned} \forall \varepsilon > 0 \quad \forall t_1, t_2 \in \bar{T} \quad \beta(Q_{t_1}, Q_{t_2}) \leq \varepsilon \Rightarrow \\ \Rightarrow \alpha(P_{t_1}, P_{t_2}) \leq g_u(\varepsilon), \end{aligned}$$

где g_u — некоторая вещественная функция, определенная на вещественной положительной полуоси,

$$\lim_{x \downarrow 0} g_u(x) = 0,$$

то для этого семейства должны существовать нижние оценки устойчивости вида

$$\begin{aligned} \forall \varepsilon > 0 \quad \forall t_1, t_2 \in T \quad \beta(Q_{t_1}, Q_{t_2}) \geq \varepsilon \Rightarrow \\ \Rightarrow \alpha(P_{t_1}, P_{t_2}) \geq g_l(\varepsilon) > 0, \end{aligned}$$

где g_l — некоторая вещественная функция, определенная на вещественной положительной полуоси.

Замечание 1. В условиях теоремы вместо требования существования верхних оценок можно потребовать от семейства $\bar{\mathfrak{A}}$, чтобы оно было плотным. Возможно, это условие является избыточным для некоторых классов смесей. Однако для смесей нормальных распределений указанные ограничения являются достаточно слабыми.

Доказательство. По условию \mathfrak{A} идентифицируемо. Поэтому из того, что расстояние β между двумя смесями из \mathfrak{B} положительно, следует, что для соответствующих распределений из \mathfrak{A} расстояние α также положительно.

Пусть доказываемое утверждение неверно. Тогда существует $\varepsilon > 0$, такое, что для множества индексов $T_\varepsilon^2 = \{t_1, t_2 \in T | \beta(Q_{t_1}, Q_{t_2}) \geq \varepsilon\}$

$$\inf_{t_1, t_2 \in T_\varepsilon^2} \alpha(P_{t_1}, P_{t_2}) = 0.$$

Рассмотрим последовательность пар распределений из \mathfrak{A} :

$$\{(P_{i_n}, P_{j_n})\}_{n=1}^\infty, \quad \alpha(P_{i_n}, P_{j_n}) < \frac{1}{n}, \\ i_n, j_n \in T_\varepsilon^2, \quad n = 1, 2, 3, \dots$$

Заметим, что последовательности $\{Q_{i_n}\}_{n=1}^\infty$ и $\{Q_{j_n}\}_{n=1}^\infty$ являются плотными и по теореме Прохорова относительно компактными (см., например, [12]). Следовательно, из них можно выделить сходящиеся подпоследовательности. Поэтому без ограничения общности можем считать, что сами последовательности $\{Q_{i_n}\}_{n=1}^\infty$ и $\{Q_{j_n}\}_{n=1}^\infty$ сходятся и их пределы принадлежат \mathfrak{B} . Учитывая, что для \mathfrak{A} существуют верхние оценки устойчивости, соответствующие последовательности смесей будут также сходить к соответствующим пределам. Так как α метризует слабую сходимости в \mathfrak{A} , а β метризует слабую сходимости в \mathfrak{B} , получаем:

$$\lim_{n \rightarrow \infty} P_{i_n} = \lim_{n \rightarrow \infty} P_{j_n}; \\ \lim_{n \rightarrow \infty} Q_{i_n} \neq \lim_{n \rightarrow \infty} Q_{j_n}.$$

Однако в этом случае получается, что \mathfrak{A} не является идентифицируемым. Возникает противоречие, а следовательно, доказываемое утверждение верно. \square

Следствие 1. Очевидно, что указанная функция g_l может быть взята неубывающей, и, следовательно, в условиях теоремы 2

$$\forall \varepsilon > 0 \quad \forall t_1, t_2 \in T \quad \alpha(P_{t_1}, P_{t_2}) \leq \\ \leq \varepsilon \Rightarrow \beta(Q_{t_1}, Q_{t_2}) \leq g_l^{-1}(\varepsilon),$$

где

$$g_l^{-1}(y) = \inf_{x>0} \{g_l(x) > y\} \quad \forall y > 0.$$

Следствие 2. Из условий теоремы 2 напрямую следует, что в случае, когда семейство смешивающих распределений является плотным, нижние оценки устойчивости существуют для сдвиговых и масштабных смесей нормальных распределений.

Заметим, что если опираться на результаты работы [7] для масштабной смеси вида

$$Y = VX, \quad X \sim \mathcal{N}(0, 1), \quad V > 0$$

(случайные величины X и V стохастически независимы), компактность будет требоваться от семейства смешивающих распределений, соответствующих случайным величинам вида $\ln V$, а не V .

6 Заключение

На практике в большинстве случаев носители рассматриваемых смешивающих распределений ограничены в совокупности. Более того, очень часто они сосредоточены в конечном числе точек.

В связи с этим оценки для каждого конкретного случая могут оказаться сильно лучше тех, что рассмотрены выше. Тем не менее полученные оценки имеют существенное значение.

Доказательство состоятельности оценок, полученных с помощью сеточных методов разделения смесей вероятностных распределений, опирается на теоремы существования, доказанные в данной работе. Помимо этого, рассмотренные примеры позволяют лучше понять, насколько сильно могут изменяться оценки смешивающего распределения при работе с реализациями одной и той же смеси.

Литература

1. Королев В. Ю., Непомнящий Е. В., Рыбальченко А. Г., Виноградова А. В. Сеточные методы разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых индексов // Информатика и её применения, 2008. Т. 2. Вып. 2. С. 3–18.
2. Назаров А. Л. Разделение смесей вероятностных распределений сеточным методом максимального правдоподобия при помощи алгоритма условного градиента // Сб. статей молодых ученых факультета ВМиК МГУ, 2009. Вып. 6. С. 128–135.
3. Korolev V., Nazarov A. Separating mixtures of probability distributions with the grid method of moments and the grid maximal likelihood method // Autom. Remote Control, 2010. Vol. 71. No. 3. P. 455–472.
4. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011.

5. *Tukey J. W.* A survey of sampling from contaminated distributions // *Contributions to Probability and Statistics Essays in Honor of Harold Hotelling*, 1960. P. 448–485.
6. *Hall P.* On measures of the distance of a mixture from its parent distribution // *Stochastic Proc. Their Applications*, 1979. Vol. 8. No. 3. P. 357–365.
7. *Назаров А. Л.* Об устойчивости смесей вероятностных законов к возмущениям смешивающих распределений // *Статистические методы оценивания и проверки гипотез: Межвузовский сб. науч. тр.* — Пермь: ПГУ, 2010. Вып. 22. С. 154–172.
8. *Назаров А. Л.* Асимптотические свойства оценок, полученных с помощью сеточных методов разделения смесей вероятностных распределений // *Статистические методы оценивания и проверки гипотез: Межвузовский сб. науч. тр.* — Пермь: ПГУ, 2012. Вып. 24. С. 22–35.
9. *Скоруход А. В.* Предельные теоремы для случайных процессов // *Теория вероятностей и ее применения*, 1956. Т. 1. Вып. 3. С. 289–319.
10. *Ширяев А. Н.* *Вероятность.* — М.: МЦНМО, 2007.
11. *Колмогоров А. Н., Фомин С. В.* *Элементы теории функций и функционального анализа.* — М.: Наука, 1976.
12. *Биллингсли П.* *Сходимость вероятностных мер.* — М.: Наука, 1977.

ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ПРИ РАСПОЗНАВАНИИ ТЕКСТОВ ПО ИЗОБРАЖЕНИЮ НИЗКОГО КАЧЕСТВА

М. П. Кривенко¹

Аннотация: Рассматриваются методы предварительной обработки изображений текста, включающей решение задач коррекции наклона и выделения строк; при этом распознаваемое изображение обладает низким качеством и получено с высоким разрешением. При условии, что яркость пикселей строк знаков, хоть и незначительно, но отличается от яркости пикселей фона, предлагаются и анализируются процедуры коррекции наклона и выделения строк текста.

Ключевые слова: распознавание текста; предварительная обработка изображения; коррекция наклона; выделение строк текста

1 Введение

Задача восстановления текста по изображению документа является крайне сложной, в связи с чем обычно применяется декомпозиция процесса решения задачи путем выделения этапов превращения изображения, содержащего текстовые фрагменты, в наборы текстов. На рис. 1 приведена типовая схема обработки данных (см., например, [1, подразд. 2.1]).

Далее предполагается, что исходное изображение представляет собой набор пикселей, каждый из которых передает градации яркости серого (оттенки серого). Каждая подсистема распознавания текста, получая данные от предыдущего этапа, пре-

образует их в выходные, которые обрабатываются на следующем этапе. Такая последовательная схема обработки изображений обеспечивает возможность получать наилучшие решения для каждого этапа. С целью создания предпосылок для того, чтобы после объединения всех частных, лучших по отдельности подсистем можно было получать эффективные решения, вводится ядро системы — база знаний.

В данной работе рассматривается этап предобработки, при этом из обычного состава этого этапа выделены задачи коррекции наклона (в англоязычной литературе для этого вида аномальности изображения текста принят термин *skew*) и выделение изображений строк текста.



Рис. 1 Традиционная схема распознавания текста по его изображению

¹Институт проблем информатики Российской академии наук, mkcrivenko@ipiran.ru

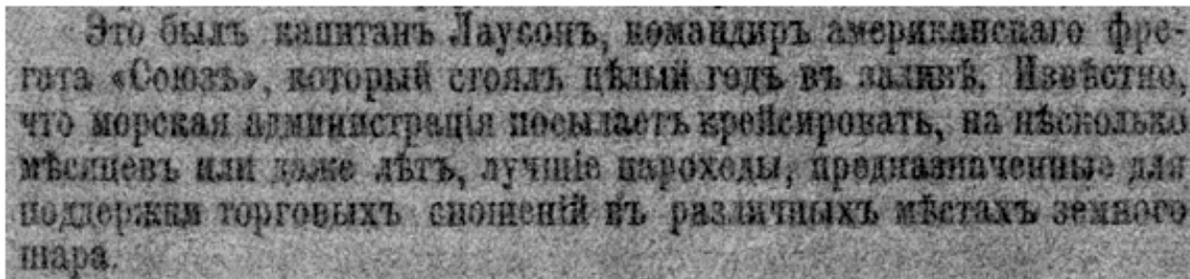


Рис. 2 Фрагмент распознаваемого текста

Коррекция наклона строк. Нормальным изображением считается то, в котором линии документа (вдоль которых идут отдельные знаки) являются горизонтальными или вертикальными в зависимости от языка. Уклонение может быть преднамеренным, чтобы подчеркнуть важные детали в документе, либо неумышленным в силу тех или иных условий получения изображения. В любом случае оно должно быть устранено, иначе точность последующих процессов — сегментации и классификации — резко упадет. В данной работе рассматривается случай, когда для всех блоков текста характерна единая ориентация и без потери общности можно считать, что речь идет о горизонтальных строках. Коррекция уклонения распадается на решение двух задач: определение угла наклона и собственно выравнивание изображения.

Выделение изображений строк (сегментация изображения). Относительно структуры текстовых фрагментов предполагается следующее:

- знаки компонуются в строки — удлиненное (вытянутое) в горизонтальном направлении изображение, пиксели которого темнее пикселей, лежащих вне строк;
- строка имеет практически постоянную высоту, которая определяется расстоянием между двумя ближайшими переходами от более темных к менее темным пикселям;
- начало и конец строки в горизонтальном направлении являются переменными величинами.

Из сделанных предположений следует, что критерием для выделения строк может служить яркость пикселей, а наиболее простой техникой для этого оказывается пороговая. Глобальные алгоритмы пороговой обработки используют единственный порог для всего изображения. Локальные (перестраиваемые, адаптивные) методы вычисляют определяемый порог для каждого пиксела с использованием яркости пикселей по соседству.

Особенности рассматриваемых задач. Далее речь пойдет о предобработке изображений в условиях их низкого качества. В качестве примера рассмотрим реальное изображение (на рис. 2 приведен фрагмент одной из страниц текста), низкое качество которого иллюстрирует рис. 3: нет ярко выраженных экстремумов распределения яркости, соответствующих пикселям знаков и пикселям фона. Количественно это можно выразить следующим образом: аппроксимируем плотность распределения яркости серого пикселей с помощью смеси двух нормальных распределений, а именно: $f^*(u) = 0,28\varphi(u, 180, 202) + 0,72\varphi(u, 197, 56)$. Здесь первый элемент смеси — плотность нормального распределения $\varphi(u, 180, 202)$ со средним 180 и дисперсией 202 — соответствует пикселям знаков, а второй — пикселям фона. Если опереться на это представление и найти порог бинаризации изображения так, чтобы минимизировать ошибку классификации пикселей на черные и белые, то получим значение порога $t = 184$ и оценку ошибки классификации в 14%, т. е. достаточно большой уровень ошибки.

Пытаясь скомпенсировать недостатки объекта анализа, подчас идут на повышение разрешения изображения, при этом высота строчных знаков мо-

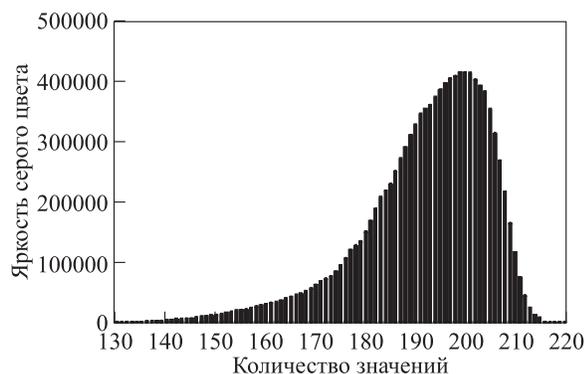


Рис. 3 Распределение значений яркости серого цвета для реального изображения

жет достигать 50 пикселей. Указанные особенности обрабатываемых изображений сразу же ограничивают спектр возможных методов анализа, а также предъявляют повышенные требования к чувствительности используемых подходов.

При построении процедур распознавания предполагается, что яркость пикселей для изображений знаков текста все же отличается от фона (далее для конкретности речь пойдет о темных пикселях знаков на более светлом фоне), но различие является несущественным.

В соответствии с типизацией задач распознавания текстов, а также приводимыми примерами в [2, подразд. 1.1] описанные свойства обрабатываемых текстов позволяют характеризовать задачи их обработки как крайне сложные.

2 Коррекция наклона изображений строк текста

Обычно различают два типа наклона в документах: глобальный и локальный. Первая категория является самой популярной при исследованиях, для нее было предложено множество методов, однако далеко не все проблемы решены, особенно для документов с рисунками, графиками и фигурами или для шрифтов различных размеров (см., например, [3]). С другой стороны, некоторые исследователи разработали методы для локальных уклонений, которые включались в технологию точного восстановления текстов, но они являются весьма затратными с вычислительной точки зрения [4].

Выделение глобальных уклонений обычно делают согласно основному применяемому подходу на четыре основные категории методов: проекционный профиль (projection profile), преобразование Хафа (Hough), группирование ближайших соседей (nearest neighbor clustering), кросс-корреляция между сечениями (interline cross correlation).

Подход на основе проекционных профилей. Традиционный и простой подход в выявлении угла, искажающего изображение документа. В диапазоне ожидаемых углов вычисляются отдельные проекционные профили, из которых выбирается тот, который наиболее ярко отражает различие между особенностями линий с пикселями знаков и линий с пикселями фона. Для снижения высоких вычислительных затрат были предложены различные модификации, суть которых состоит либо в уменьшении количества данных, вовлеченных в вычисление профиля, либо в оптимизации стратегии поиска наилучшего решения.

Подход на основе преобразований Хафа. Он нацелен на обнаружение линий и кривых на цифровых изображениях. Используется в предположении, что линии, характеризующие направленность текста, содержат достаточно большое количество пикселей. Чтобы повысить вычислительную эффективность, применяются разновидности базовой идеи использования преобразования Хафа, которые сокращают число точек в пространстве образов.

Подход на основе ближайших соседей. Методы этого класса используют общие предположения, что знаки в строке текста выровнены и близки друг к другу. Они реализуют восходящий, от частного к общему, процесс формирования представления строки из отдельных знаков или их деталей, а затем использования этого представления для оценки наклона.

Подход на основе сечений изображения. Это название подхода соответствует англоязычному словосочетанию «interline cross correlation clustering» которое достаточно трудно перевести коротко. Сам подход основан на предположении, что изображение текста без наклона строк представляет однородную горизонтальную структуру. Тогда оценка наклона строится путем измерения изменений характеристик пикселей вдоль вертикальных линий изображения.

Краткий обзор происхождения и сути перечисленных подходов, их сравнительный анализ достаточно полно представлен в [5] (54 источника с 1972 по 1997 г.) и в [6] (58 источников с 1986 по 2011 гг.).

Несмотря на примеры эффективного использования методов коррекции наклона строк, можно выделить ряд общих проблем при их применении:

- необходимость фактически использовать двоичное (бинаризованное) или достаточно четкое в оттенках серого изображение;
- возможность получать грубые оценки угла наклона;
- подчас высокая вычислительная сложность;
- явное или неявное использование дополнительных параметров, существенно влияющих на качество получаемых решений;
- ориентация на специфические области применения (в частности, ограничения на размер шрифта, требование к определенному разрешению изображения, зависимость от языка, невысокая эффективность при работе с рукописным текстом);
- оценка эффективности предлагаемых методов проводится на специфических наборах данных, не являющихся общедоступными, как следствие, затруднительно провести сравнительный

анализ эффективности, понять место предлагаемых решений в ряду уже существующих.

В данной работе рассматривается использование проекционного профиля по следующим основным причинам:

- он связан с аккумулярованием информации вдоль горизонтальных линий изображения, и это создает предпосылки для того, чтобы «справиться» с изображением низкого качества;
- метод оценивания угла наклона на основе профиля хорошо зарекомендовал себя с точки зрения чувствительности;
- с помощью проекционного профиля можно строить и процедуры выделения строк.

Введем основные обозначения. Пусть $I(y, x)$ — яркость серого для пиксела, располагающегося на пересечении y -й горизонтальной линии и x -й вертикальной линии изображения (пиксел в x -м столбце на y -й линии изображения). Проекционный профиль по определению есть $PP(y) = \sum_x I(y, x)$, для его нормированной формы примем следующее обозначение:

$$p(y) = \frac{PP(y)}{w},$$

где w — постоянная ширина строки изображения в пикселах. Для последовательности $p(y)$ можно вычислить выборочные характеристики: $\bar{p} = \sum_y p(y)$ и $s^2 = \sum_y (p(y) - \bar{p})^2$.

Для пояснения того, как работают введенные выборочные характеристики при оценивании угла наклона строк текста, на рис. 4 приведен фрагмент графика поведения $p(y) - \bar{p}$ для значений $y = 250, 251, \dots, 360$ для реального изображения в трех случаях: без наклона строк (кривая A), с углом

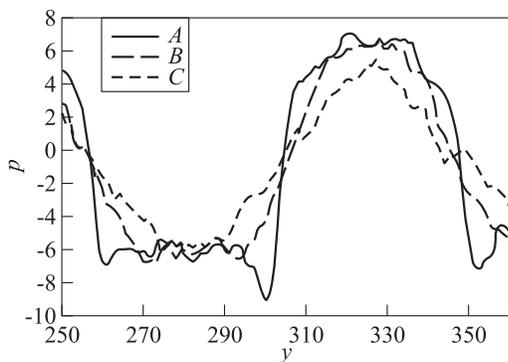


Рис. 4 Зависимость значений нормированного и центрированного проекционного профиля от номера строки для изображения без наклона (кривая A), с наклоном в $0,5^\circ$ (кривая B) и с наклоном в $1,0^\circ$ (кривая C)

Таблица 1 Характеристики проекционных профилей

Угол наклона строк текста	\bar{p}	s^2
$0,0^\circ$	192,1	32,2
$0,5^\circ$	192,1	24,8
$1,0^\circ$	192,1	15,9

наклона строк в $0,5^\circ$ (кривая B), с углом наклона строк в $1,0^\circ$ (кривая C).

Для проекционных профилей указанного изображения в упомянутых случаях табл. 1 содержит значения выборочных характеристик. При увеличении отклонения абсолютного значения угла наклона строк от нулевого значения происходит следующее:

- среднее значение \bar{p} для последовательности $p(y)$ практически не изменяется (незначительное влияние оказывает дискретизация при повороте изображения), поэтому все графики $p(y)$ колеблются около одной и той же прямой;
- локальные экстремальные значения $p(y)$ на интервалах, отвечающих положениям строк ($y = 255, 256, \dots, 305$ на рис. 4) и положениям межстрочных интервалов ($y = 306, 307, \dots, 345$ на рис. 4), убывают;
- поведение $p(y)$ на интервалах, отвечающих положениям строк и положениям межстрочных интервалов, по форме меняется от «прямоугольного» к «треугольному», сливаясь в пределе с осью абсцисс.

Тогда становится понятным, что s^2 может стать мерой горизонтальности строк текста. Также из приведенной иллюстрации следует, что s^2 предпочтительней интерпретировать не как рассеяние, а как интегральную мощность некоторого процесса $p(y)$, при этом нет необходимости в предварительной бинаризации изображения (для сравнения см., например, [1, п. 2.4.2.1]).

Таким образом, получаем, что искомый угол α^* поворота для коррекции наклона будет решением задачи $\alpha^* = \arg \max_{\alpha} s^2(\alpha)$, где мера $s^2(\alpha)$ посчитана для предварительно повернутого на угол α исходного изображения.

Для выяснения возможности практического использования описанной процедуры коррекции наклона на изображениях с рассматриваемыми свойствами был проведен анализ ее качества. Для исследования бралось изображение текста с горизонтальными строками текста, оно искусственно искажалось путем поворота на определенный угол α_0 , а затем сравнивался результат оценивания угла наклона и известное значение. Подобные действия проводились для 10 различных изображений. В качестве итоговых характеристик

Таблица 2 Характеристики оценок угла наклона строк

$-\alpha_0$, пиксел	Псевдореальные данные (10 изображений)		Реальные данные (10 изображений)	
	$\text{Ave} \{\alpha^* - (-\alpha_0)\}$	$\sqrt{\text{Ave} \{(a^* - (-\alpha_0))^2\}}$	$\text{Ave} \{\alpha^* - (-\alpha_0)\}$	$\sqrt{\text{Ave} \{(a^* - (-\alpha_0))^2\}}$
0	0,4	0,5	-0,3	0,9
1	-0,7	0,6	-1,4	0,8
2	0,0	0,0	0,0	0,0
4	0,0	0,0	0,0	0,0
8	0,0	0,0	0,0	0,0

качества оценок наклона рассматривались следующие: $\text{Ave} \{\alpha^* - (-\alpha_0)\}$ и $\sqrt{\text{Ave} \{(a^* - (-\alpha_0))^2\}}$. Принимая во внимание, что реальные изображения подвергались коррекции наклона строк текста, описанный анализ сначала проводился для псевдореальных данных — искусственно созданных изображений со строками текста без наклона, характеристики которых в смысле отличия яркостей пикселей знаков и пикселей фона были похожи на реальные. Угол поворота измерялся числом пикселей, на которое сдвигался по вертикали самый правый пиксел некоторой линии изображения относительно самого левого.

Представленные в табл. 2 результаты позволяют сделать следующие выводы:

- описанный метод обработки проекционного профиля приводит к достаточно хорошим результатам (заметим, что наклон строки в 1 пиксел соответствует всего лишь углу $0,02^\circ$);
- ненулевые значения характеристик оценки наклона для реальных данных при $\alpha_0 = 0$ являются следствием дискретизации при повороте изображения. В частности, для отдельного изображения при его повороте на -1 , а затем снова на 1 пиксел количество пикселей, не совпавших по яркости именно из-за дискретизации, оказалось равным 6947 из общего числа 10^7 (0,07%). Таким образом, дополнительным средством повышения качества процедур коррекции наклона могут стать приемы уточнения процедур поворота, хотя соответствующее повышение не является столь уж существенным.

Наряду с проанализированной процедурой коррекции наклона строк была сделана попытка построить и использовать иные методы:

- разделение элементов смеси при аппроксимации распределения $p(y)$ с помощью двух-элементной смеси нормальных распределений проекционных профилей. При этом в качестве меры отсутствия наклона рассматривалась оценка ошибки классификации для двух классов (двух элементов смеси). Основанием

для принятия решения становилось следующее правило: в случае наилучшего разделения двух классов изображение должно иметь ярко выраженную горизонтальную структуру текста;

- сравнительный анализ фрагментов изображения с помощью проверки однородности подвыборок вдоль горизонтальных линий изображения с последующим анализом числа успехов (принятие предположения об однородности) для последовательности испытаний для отдельных горизонтальных линий изображения.

Судя по проведенным автором экспериментам, явное усложнение алгоритмов обработки данных не дает очевидных преимуществ по сравнению с базовым методом. Более того, оба дополнительных метода (разделение смеси и сравнительный анализ фрагментов) проявляют меньшую устойчивость в области малых углов искажения наклона ($0, \pm 1, \pm 2$ пиксела). По-видимому, основная причина здесь в том, что обрабатываемые изображения обладают существенной зависимостью значений яркости отдельных пикселей, что отражается и на проекционных профилях. Это заставляет с осторожностью относиться к обращению с обрабатываемыми данными как с выборкой.

3 Выделение изображений строк текста

Главным образом, существует три основных категории подходов к решению рассматриваемой задачи: проекционный профиль; преобразование Хафа; заполнение, или замазывание (smearing).

Подход на основе проекционных профилей. Соответствующие методы включают нахождение интегральных характеристик отдельных горизонтальных линий — характеристик, специфичных для строк знаков и межстрочных интервалов. После рассмотрения их значений по вертикали (проекции на вертикальную линию) становится возможным выявлять позиции и размеры по высоте отдельных строк текста.

Подход на основе преобразований Хафа. Обнаружение линий в цифровых изображениях с помощью пространства образов дополняется правилами проверки гипотез в пространстве изображений. При этом отбираются такие образования (группы) точек в пространстве Хафа, которые являются наиболее представительными с точки зрения количества элементов изображения, выстроенных вдоль определенной линии.

Подход на основе заполнения. Для бинаризованных документов метод заполнения ярче всего воплощается в процедуре RLSA (Run-Length Smoothing Algorithm), когда вдоль горизонтального направления последовательность, состоящая из белых пикселей между черными, заполняется черными пикселями, если расстояние между граничными черными пикселями не превосходит определенного порога. Далее строится ограничивающий связанные компоненты прямоугольник, он и определяет текстовую строку. Существуют разновидности этого метода, приспособленного к изображениям с уровнями яркости.

Краткий обзор происхождения и сути описанных подходов, их сравнительный анализ изложен в [7] (60 источников с 1962 по 2004 гг.) и в [8] (42 источника с 1982 по 2008 гг.).

Проблемы при применении описанных подходов практически те же, что указаны в связи с коррекцией наклона строк. В первую очередь к ним следует отнести следующие: необходимость этапа бинаризации изображения, привнесение дополнительных задания параметров, отсутствие сравнительного анализа методов оценки эффективности предлагаемых методов на общедоступных данных.

В данной работе предложена схема выделения строк текста с помощью следующих идей:

- основу составляет использование проекционных профилей. Для повышения качества принимаемых решений изображение делится

(фрагментируется) на несколько вертикальных полос, для каждой из которых строится свой проекционный профиль. С помощью этого приема удастся учитывать возможность появления строк знаков с различной структурой (например, наличие неполных строк). Ширина полос при подсчете проекционных индексов принимается одинаковой (если ширина изображения не кратна числу полос, то либо отдельные пиксели отбрасываются, либо обрабатываются несколько различающиеся значения ширины фрагментов);

- выделение строк осуществляется путем сравнения расстояний между проекционными профилями горизонтальных полос изображения и эталона, задаваемого с помощью шаблона;
- шаблон описывает априорное представление о строках текста. Он может быть простым, состоящим из фиксированного числа полных строк с определенными яркостями пикселей в них, или комбинированным, состоящим из набора фрагментированных строк с определенными яркостями пикселей во фрагментах.

Варианты шаблонов при делении изображения на три вертикальные полосы, рассмотренные в данной работе, схематично представлены в табл. 3. Светлые клетки соответствуют пикселям фона, темные — пикселям знаков текста. Так, комбинированный шаблон из двух строк состоит из строки пикселей фона и строки, содержащей пиксели знаков, причем в последнем случае это может быть один из трех вариантов: темный фрагмент и два светлых фрагмента, два темных и один светлый, все три темных фрагмента.

Шаблон задается с помощью следующих параметров:

- ширина строки, которая определяется размерами изображения и задается до распознавания;

Таблица 3 Варианты шаблонов при делении изображения на 3 вертикальные полосы

Строк в шаблоне \ Тип шаблона	Простой	Комбинированный			
1					
2					
3					

- количество фрагментов, оно определяется сложностью структуры строк текста и задается исследователем до распознавания. Увеличение этого параметра повышает качество предварительной обработки и сопровождается ростом временной сложности алгоритмов обработки данных;
- характеристики нормированного проекционного профиля $p_C(y)$ строки знаков и $p_B(y)$ строки фона, которые подлежат оцениванию по изображениям;
- высота строки знаков h_C и высота строки фона h_B , они подлежат оцениванию по изображениям.

В данной работе характеристиками нормированного проекционного профиля выступали средние значения яркости серого p_C и p_B . Соответствующие оценки находятся с помощью представления $p^*(y)$ в виде смеси двух нормальных плотностей так, что средние элементов смеси упорядочены по возрастанию $\mu_1^* < \mu_2^*$. Тогда $p_C^* = \mu_1^*$ и $p_B^* = \mu_2^*$. Если наряду с изображениями текста до распознавания доступны изображения строк только фона, то можно применить более тонкую процедуру: представить распределение $p(y)$ в виде смеси более чем двух нормальных распределений (тем самым улучшая аппроксимацию реального распределения). Затем расщепить эту смесь на две составляющие, одна из которых соответствует нормированному проекционному профилю знаков, а другая — фона [9]. И, наконец, в качестве характеристик шаблона взять моменты выделенных составляющих смеси.

Оценивание высот строки знаков и строки фона шаблона реализуется с помощью следующих шагов:

- получение проекционного профиля $p(y)$ по всей ширине изображения текста;
- бинаризация последовательности $p(y)$ путем аппроксимации распределения $p(y)$ с помощью смеси двух нормальных распределений, затем нахождение порога для разделения элементов полученной смеси и, наконец, формирование последовательности из 0 и 1;
- выделение серий двух типов (0-серий и 1-серий) и подсчет частот встречаемости длин серий определенного типа;
- оценивание высот строк знаков и строк фона путем выбора длин серий, частоты которых являются наибольшими.

Таким образом, основными шагами обработки исходных изображений становятся следующие:

- определение характеристик шаблона, для чего проводится оценивание высот чередующихся строк из темных (строк знаков) и светлых

(строк фона) пикселей. Предполагается, что строки имеют одинаковую высоту, но не обязательно совпадающую для строк знаков и фона;

- определение положений строк на отдельных изображениях.

Определение положений строк знаков включает следующие действия:

- для каждого возможного положения строки знаков нахождение значения меры близости между фрагментом изображения, соответствующим этой строке, и выбранным эталоном, характеризующим строчную структуру текста;
- отбор из полученных на предыдущем шаге значений мер близости наименьшего значения так, чтобы сформировалось полное изображение страницы.

В качестве меры близости в случае простого шаблона рассматривается квадрат различия значений нормированного проекционного профиля для фрагмента изображения, соответствующего по структуре и размеру выбранному шаблону, и средней яркости серого для элементов шаблона. Для комбинированного шаблона мера близости вычисляется как минимум мер близости для каждого частного вида шаблона.

Наибольшие сложности возникают при реализации этапа отбора наименьших значений мер близости, что связано с наличием множества ложных локальных минимумов. Правда, их число уменьшается с ростом сложности шаблона. Но в любом случае приходится отказываться от последовательной по линиям изображения стратегии обнаружения строк знаков и обращаться к процедуре, шаг за шагом выделяющей наименьшее значение меры близости из оставшихся и исключением из дальнейшего рассмотрения линий изображения, соответствующих найденной строке знаков и линиям фона, обрамляющим эту строку.

В табл. 4 сведены результаты экспериментов для 10 реальных изображений страниц, содержащих в общей сложности 430 строк текста. Для оценивания качества процедуры выделения позиций строк текста использовались следующие характеристики: A — число не найденных строк, B — число несуществующих строк (найденных процедурой позиций строк, не содержащих текст), δ — оценка смещения позиции строк текста, σ — оценка стандартного отклонения смещения позиции строк текста.

Полученные результаты позволяют сделать следующие выводы:

- несмотря на низкое качество изображения, удается получить высокие показатели точности выделения строк текста;

Таблица 4 Характеристики процедуры выделения строк

Строк в шаблоне	Простой шаблон				Комбинированный шаблон			
	<i>A</i>	<i>B</i>	δ	σ	<i>A</i>	<i>B</i>	δ	σ
1	35	2	0,3	3,5	9	2	0,3	3,4
2	32	1	-0,5	2,3	7	1	-0,5	2,3
3	26	0	0,2	2,1	3	0	0,1	2,1

- усложнение шаблона (как за счет увеличения количества входящих в него строк, так и путем фрагментации) приводит к явному улучшению качества выделения позиций строк;
- фрагментация шаблона является мощным средством описания сложной структуры строки текста (появление отступов, наличие неполных строк, выравнивание заголовков и т. п.).

4 Заключение

Предложенные методы позволяют эффективно проводить предварительную обработку распознаваемого текста, несмотря на низкое качество изображения. В ситуации, когда значения яркости серого для пикселей знаков и фона все же различаются, удается использовать идеи проекционных профилей. При этом приходится отказываться от бинаризации исходного изображения, усложнять и вводить новые приемы обработки данных (использование модели смеси распределений при анализе проекционных профилей, введение поэтапного оценивания неизвестных характеристик строчной структуры текста, фрагментацию изображения). Важным моментом предложенных решений является возможность их обобщения на случай, когда средние значения яркости серого для строк текста и фона практически не отличаются и приходится обращаться к анализу локальных взаимосвязей яркости пикселей.

Литература

1. *Cheriet M., Kharma N., Liu C-L., Suen C. Y.* Character recognition systems: A guide for students and practitioners. — Hoboken, New Jersey: Wiley-Interscience, 2007. 326 p.
2. *Mori S., Nishida H., Yamada H.* Optical character recognition. — Hoboken, New Jersey: Wiley, 1999. 560 p.
3. *Sharif A. E., Movahhedinia N.* On skew estimation of Persian/Arabic printed documents // *J. Appl. Sci.*, 2008. Vol. 8. Is. 12. P. 2265–2271.
4. *Saragiotis P., Papamarkos N.* Local skew correction in documents // *Int. J. Pattern Recognition Artificial Intelligence*, 2008. Vol. 22. No. 4. P. 691–710.
5. *Hull J. J.* Document image skew detection: Survey and annotated bibliography // *Document Analysis Systems II*. — Singapore: World Scientific, 1998. P. 40–64.
6. *Rehman A., Saba T.* Document skew estimation and correction: Analysis of techniques, common problems and possible solution // *Appl. Artificial Intelligence*, 2011. Vol. 25. No. 9. P. 769–787.
7. *Likforman-Sulem L., Zahour A., Taconet B.* Text line segmentation of historical documents: A survey // *Int. J. Document Analysis Recognition*, 2006. Vol. 9. No. 2–4. P. 123–128.
8. *Louloudis G., Gatos B., Pratikakis I., Halatsis C.* Text line and word segmentation of handwritten documents // *Pattern Recognition*, 2009. Vol. 42. Is. 12. P. 3169–3183.
9. *Кривенко М. П.* Расщепление смеси вероятностных распределений на две составляющие // *Информатика и её применения*, 2008. Т. 2. Вып. 4. С. 48–56.

МОДЕЛЬ СЛУЧАЙНЫХ ГРАФОВ ДЛЯ ОПИСАНИЯ ВЗАИМОДЕЙСТВИЙ В СЕТИ*

А. А. Грушо¹, Е. Е. Тимонина²

Аннотация: Рассматривается новый класс случайных графов, призванный моделировать функционирование сети во времени. Предполагается, что наблюдения за сетью ведутся с помощью «оконного» метода. С целью выявления аномалий исследуется нормальное поведение степеней, которые можно наблюдать в «окнах» рассматриваемой модели. Исследована асимптотика максимальной степени вершин в графе, который порожден «окном» данного размера.

Ключевые слова: случайные графы; моделирование глобальных сетей; информационная безопасность; аномальное поведение

1 Введение

Возможны различные способы распространения информации в сети. Наиболее известным способом является обращение к информационным ресурсам, выложенным на тематических сайтах. Поиск таких сайтов по нужной тематике — задача интернет-поисковиков. Другое направление — распространение информации по электронной почте или через налаженные и специальным образом формируемые связи. К таким способам относятся спам и управление бот-сетями.

Формирование собрания единомышленников или предупреждение о каком-либо событии могут передаваться любым из перечисленных выше методов. В связи с этим представляет интерес моделирование процессов последовательной передачи информации. Будем рассматривать второй способ передачи информации.

В предположении о дискретности времени будем описывать состояние связей хостов между собой неориентированным случайным графом. Ребрам графа отвечают логические связи хостов. Вместе с тем последовательности связей вершин не обязательно связаны с распространением информации в рамках некоторой корпорации. В этом случае последовательности связей формируются случайно и независимо друг от друга. Выявление корпоративных связей в сети или центров управления связано с анализом случайных связей в рамках большого случайного графа сети.

Будем рассматривать изменения в графе сети с течением времени с помощью «скользящего окна». Граф, который получается при таком рассмотре-

нии, формируется фиксацией логических связей, захватываемых «окном» и отображенных на одном графе. Такой граф получается наложением всех графов в моменты времени, принадлежащие «окну», и объединением параллельных ребер.

В работе определена математическая модель таких случайных графов и исследованы характеристики, связанные с указанными выше прикладными задачами. Полученные результаты носят асимптотический характер при условии, что число хостов стремится к бесконечности.

Работа имеет следующую структуру. В разд. 2 приведены некоторые близкие модели случайных графов. В разд. 3 определена основная динамическая модель. В разд. 4 исследована асимптотика максимальной степени в случайном графе, связанном с «окнами». В разд. 5 подведены итоги и намечены дальнейшие задачи.

2 Модели случайных графов

В научной литературе рассматривались модели случайных графов, связанных с Интернетом. В работе [1] в качестве одного из примеров приводится классическая модель случайного графа $G_{N,p}$ с независимыми ребрами, появляющимися с одной и той же вероятностью p . Этой модели посвящено много работ и книг [2–13]. И хотя основополагающая статья Эрдеша и Реньи [2] связана с несколько другим классом случайных графов, большинство аналогичных результатов было также доказано для графов $G_{N,p}$ в работах [3–6, 8] и др. Фазовые перехо-

* Работа выполнена при поддержке РФФИ (проекты № 10-01-00480, № 11-07-00112).

¹ Институт проблем информатики Российской академии наук; Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, grusho@yandex.ru

² Институт проблем информатики Российской академии наук, eltimon@yandex.ru

ды в структуре таких графов впервые исследованы в работах [3–5]. Первые модели с неравновероятными ребрами исследовались в работах [8]. В перечисленных моделях появление ребра не допускало его дальнейшего исчезновения.

Изменение случайного графа во времени в связи с задачей роста сети Интернет рассматривалось в работах [7, 9, 10].

Специальный класс случайных графов, посвященный исследованию связей в Интернете, объединяет модели графов интернет-типа [11–13]. Они определяются степенями вершин, являющихся независимыми случайными величинами. При этом свободные концы ребер замыкаются друг на друга случайно и равновероятно.

3 Динамическая модель сетевого взаимодействия

Определим детально модель сетевого взаимодействия, кратко изложенную во введении.

Рассмотрим дискретное время $t = 0, 1, 2, \dots$. Множество хостов сети обозначим $A\{a_1, \dots, a_N\}$. Логическая связь хостов a_i и a_j в момент времени t означает либо наличие в этот момент времени сеанса связи по протоколу TCP между a_i и a_j , либо передачу одиночного пакета от одного хоста к другому в этот момент времени по любому протоколу без установления соединения. Для простоты считаем, что время прохождения пакета по сети равно 1. При этом выбрасываются из рассмотрения все промежуточные поддерживающие сетевые службы (провайдеры, маршрутизаторы, адресные службы и т. д.). Из этих допущений получаем модель графа сети. В каждый момент времени t определен неориентированный граф G_t , вершины которого совпадают с множеством A , а ребра соответствуют существующим в момент времени t логическим связям. Из определения логической связи следует, что в соседние моменты времени t и $t + 1$ существование данного ребра в графе G_t и графе G_{t+1} являются зависимыми событиями. Однако процессы появления и исчезновения разных ребер можно считать независимыми.

В простейшем случае полагаем, что процесс, описывающий возникновение и исчезновение одного ребра, является стационарной однородной цепью Маркова с двумя состояниями: 1 — есть ребро, 0 — нет ребра.

Наблюдения за графами $\{G_t\}$ происходят с помощью «оконной» системы. Пусть задано натуральное число r и для любого момента времени t рассматриваются графы G_t, \dots, G_{t+r} , появляющиеся в

моменты времени $[t, t + r]$. Определим операцию объединения этих графов

$$G_{t,r} = \bigcup_{i=0}^r G_{t+i},$$

где из нескольких параллельных ребер оставляет-ся одно ребро. Граф $G_{t,r}$ несет информацию об активности любой вершины в заданный промежуток времени. Эти графы представляют интерес в задачах информационной безопасности. Например, если вершина a_i является центром управления бот-сетью, то использование бот-сети для организации DDoS атаки должно порождать в некоторый промежуток времени $[t, t + r]$ резкое повышение степени вершины a_i в графе $G_{t,r}$. При малых r и очень больших r при неизвестном t этот всплеск активности может оказаться незаметным в масштабах всей сети. Поэтому исследование модели случайных графов может позволить оценить возможности по выявлению неслучайных всплесков активности отдельных вершин и даже дать оценку для центра управления бот-сетью.

Пусть поведение каждого ребра описывается стационарной однородной цепью Маркова с матрицей переходных вероятностей

$$P = \begin{pmatrix} p & 1 - p \\ q & 1 - q \end{pmatrix},$$

$$1 > p = p(N) > 0, \quad 1 > q = q(N) > 0,$$

и стационарным распределением $(p_0 \ 1 - p_0)$.

Тогда вероятность непоявления данного ребра за промежуток времени $[t, t + r]$ равна

$$1 - p_r = (1 - p_0)(1 - q)^r. \quad (1)$$

Эта вероятность не зависит от t , поэтому будем обозначать ее p_r . Из (1) получаем вероятность появления данного ребра в промежуток времени $[t, t + r]$ в графе $G_{t,r}$:

$$p_r = 1 - (1 - p_0)(1 - q)^r.$$

4 Асимптотические оценки максимальной степени в графах $G_{t,r}$

Для $v \in A$ обозначим через $d(v)$ степень вершины v . Определим индикаторную функцию события B :

$$I(B) = \begin{cases} 1, & \text{если событие } B \text{ произошло;} \\ 0 & \text{в противном случае.} \end{cases}$$

Ожидаемое число соединений у фиксированной вершины в ограниченный промежуток времени мало по сравнению с общим числом вершин. Асимптотически это отвечает условию $p_0 N \rightarrow \mu > 0$, $N \rightarrow \infty$. Серии единиц связаны с режимом установления соединения. Поэтому величина $(1-p)$ может не стремиться к 1. Пусть $q \rightarrow 0$, $N \rightarrow \infty$, так что $qN \rightarrow \lambda$. Из условия стационарности следует соотношение:

$$q = \frac{p_0}{1-p_0} (1-p).$$

Таким образом, получаем

$$p_r = \frac{\mu}{N} + \frac{\lambda r}{N} + O\left(\frac{\lambda r}{N^2}\right).$$

Положим $\alpha_r = \mu + \lambda r$ и будем считать, что

$$p_r = \frac{\alpha_r}{N}.$$

Обозначим

$$X = \sum_{v \in A} I(d(v) > d).$$

Тогда

$$\left\{ \max_{v \in A} d(v) \right\} > d = \{X \geq 1\}.$$

Используя неравенство Маркова, получаем:

$$P \left\{ \max_{v \in A} d(v) > d \right\} \leq N \sum_{k>d} \binom{N-1}{k} p_r^k (1-p_r)^{N-1-k}. \quad (2)$$

Пусть $B(N-1, k, p_r)$ — функция распределения биномиального закона, $\bar{B}(N-1, k, p_r) = 1 - B(N-1, k, p_r)$. Заметим, что в формуле (2) справа стоит $N \bar{B}(N-1, d, p_r)$.

Асимптотические оценки проводим в условиях

$$N \rightarrow \infty, \quad d = \frac{C \ln N}{\ln \ln N}, \quad C > 0, \quad p_r = \frac{\alpha_r}{N}.$$

Для оценки функции $\bar{B}(N-1, d, p_r)$ воспользуемся представлением для неполной бета-функции [14]:

$$\bar{B}(N-1, k, p_r) = N \binom{N-2}{k} \int_0^{p_r} z^k (1-z)^{N-k-2} dz.$$

Используя формулу Тейлора, получим при некотором $0 < \theta < 1$ следующее представление для математического ожидания X :

$$EX = N(N-1) \binom{N-2}{d} (p_r \theta)^d p_r (1-p_r \theta)^{N-2-d}.$$

Отсюда получаем следующую асимптотическую формулу:

$$EX = N^{(1-C)(1+o(1))} \alpha_r (1 - e^{-\alpha_r \theta}). \quad (3)$$

Теорема. При $C > 1$, $N \rightarrow \infty$, $d = C \ln N / (\ln \ln N)$, $p_r = \alpha_r / N$

$$P \left\{ \max_{v \in A} d(v) > d \right\} \rightarrow 0.$$

Доказательство следует из (3).

Таким образом, установлена граница для максимальной степени вершины в графе $G_{t,r}$. На основании этого результата можно построить оценку центра управления бот-сетью. Если существует вершина, степень которой превосходит заданную границу, то с вероятностью, близкой к 1, высокая степень этой вершины получена вне условий стационарности и других допущений, которые были сделаны для нормального поведения сети.

Предположим теперь, что $C < 1$ и математическое ожидание $EX \rightarrow \infty$. Построим оценку числа вершин, имеющих степень больше d , при условии, что $EX \rightarrow \infty$. С этой целью оценим и сравним дисперсию DX случайной величины X и $(EX)^2$. Очевидно, что

$$(EX)^2 = N^2 (1 - B(N-1, d, p_r))^2 = N^2 \bar{B}^2(N-1, d, p_r).$$

Случайную величину X можно представить в виде:

$$X = \sum_{i=1}^N I_i, \quad (4)$$

где I_i — индикатор события $d(i) > d$. Тогда из (4) следует:

$$EX^2 = E \left(\sum_{i=1}^N I_i \right) + E \left(2 \sum_{i<j} I_i I_j \right) = N \bar{B}(N-1, d, p_r) + 2 \sum_{i<j} P(I_i I_j = 1).$$

По формуле полной вероятности

$$P(I_1 I_2 = 1) = p_r \bar{B}^2(N-2, d-1, p_r) + (1-p_r) \bar{B}^2(N-2, d, p_r).$$

Рассмотрим разность EX^2 и $(EX)^2$. Несложные вычисления приводят к выражению:

$$\begin{aligned} EX^2 - (EX)^2 &= \\ &= N\overline{B}(N-1, d, p_r)(1 - \overline{B}(N-1, d, p_r)) + \\ &\quad + N(N-1)(1-p_r)p_rb^2(N-2, d, p_r), \end{aligned}$$

где

$$b(N-2, d, p_r) = \binom{N-2}{d} p_r^d (1-p_r)^{N-2-d}.$$

Предполагалось, что $EX \rightarrow \infty$. Тогда

$$\frac{DX}{(EX)^2} = \frac{1}{N\overline{B}(N-1, d, p_r)} (B(N-1, d, p_r) + \alpha_r(1-p_r) \frac{(N-1)b^2(N-2, d, p_r)}{N\overline{B}(N-1, d, p_r)}).$$

Из предыдущих оценок имеем, что

$$\begin{aligned} Nb(N-2, d, p_r) &= O(N^{1-C}); \\ N\overline{B}(N-1, d, p_r) &= N^{(1-C)(1+o(1))} \alpha_r (1 - e^{-\alpha_r \theta}). \end{aligned}$$

Отсюда следует, что

$$\frac{DX}{(EX)^2} = O(N^{C-1}), \quad C < 1.$$

Воспользуемся следствиями 4.32 и 4.33 из работы [15]. Получаем, что с вероятностью, стремящейся к 1, $0 < X$ и отношение $X/(EX) \rightarrow 1$. Это означает, что при $C < 1$ с вероятностью, близкой к 1, существует вершина степени больше d и число таких вершин совпадает с математическим ожиданием X .

5 Заключение

В ходе решения поставленных в данной работе задач появилось много новых направлений, которые заслуживают отдельного внимания. В данной работе исследовано поведение больших степеней в графе, соответствующем фиксированному «окну». Естественно, желательно обобщить эти результаты на случай скользящего «окна».

Поведение графов интернет-типа часто нельзя считать стационарным. Возникает задача анализа «оконных» графов в условиях нестационарного поведения сети. Перечень проблем, возникших при данном исследовании, не исчерпывается данными двумя задачами.

Литература

1. *Kolaczyk E. D.* Statistical analysis of network data: Methods and models. — Springer Science + Business Media, LLC, 2009. 386 p.
2. *Erdős P., Rényi A.* On the evolution of random graphs // Publ. Math. Inst. Hungarian Acad. Sci., Ser. A, 1960. Vol. 5. P. 17–61.
3. *Степанов В. Е.* О вероятности связности случайного графа $g_m(t)$ // Теория вероятностей и ее применения, 1970. Т. 15. № 1. С. 55–67.
4. *Степанов В. Е.* Фазовый переход в случайных графах // Теория вероятностей и ее применения, 1970. Т. 15. № 2. С. 187–203.
5. *Степанов В. Е.* Структура случайных графов $g_n(x|h)$ // Теория вероятностей и ее применения, 1972. Т. 17. № 3. С. 227–242.
6. *Bollobas B.* Random graphs. — London: Academic Press, 1985.
7. *Kleinberg J., Kumar S., Raghavan P., Rajagopalan S., Tomkins A.* The web as a graph: measurements, models, and methods // Conference (International) on Combinatorics and Computing Proceedings — Berlin: Springer, 1999. Lecture Notes in Computer Science. Vol. 1627. P. 1–18.
8. *Колчин В. Ф.* Случайные графы. — М.: Физматлит, 2000. 256 с.
9. *Kumar R., Raghavan P., Rajagopalan S., Sivakumar D., Tomkins A., Upfal E.* Stochastic models for the web graph // 42nd Annual IEEE Symposium on the Foundations of Computer Science Proceedings, 2000. Vol. 41. P. 57–65.
10. *Chung F., Lu L., Dewey T., Galas D.* Duplication models for biological networks // J. Comput. Biology, 2003. Vol. 10. No. 5. P. 677–687.
11. *Павлов Ю. Л., Степанов М. М.* Об асимптотических свойствах случайных графов «интернет-типа» // Обзорение прикладной и промышленной математики, 2005. Т. 12. № 3. С. 677.
12. *Степанов М. М.* О предельных распределениях степеней узлов в случайных графах интернет-типа // Методы математического моделирования и информационные технологии: Тр. Института прикладных математических исследований Карельского научного центра РАН. — Петрозаводск: КарНЦ РАН, 2005. Вып. 6. С. 235–242.
13. *Павлов Ю. Л.* Предельное распределение объема гигантской компоненты в случайном графе интернет-типа // Дискретная математика, 2007. Т. 19. № 3. С. 22–34.
14. *Феллер В.* Введение в теорию вероятностей и ее приложения. — 2-е изд. — М.: Мир, 1967. Т. 1.
15. *Alon N., Spencer J.* The probabilistic method. — 2nd ed. — New York: John Wiley & Sons, 2000.

ОБ ОПТИМАЛЬНОМ КОРРЕКТНОМ ПЕРЕКОДИРОВАНИИ ЦЕЛОЧИСЛЕННЫХ ДАННЫХ В РАСПОЗНАВАНИИ*

Е. В. Дюкова¹, А. В. Сизов², Р. М. Сотнезов³

Аннотация: Исследуются вопросы применения логических процедур распознавания по прецедентам в случае вещественнозначной информации и целочисленной информации высокой значности. Рассмотрена задача корректного понижения значности данных. Разработаны генетические алгоритмы поиска оптимальной корректной перекодировки исходной информации. Проведено тестирование алгоритмов на реальных данных.

Ключевые слова: распознавание образов; корректная перекодировка; покрытие булевой матрицы

1 Введение

Один из подходов к решению задачи распознавания по прецедентам сводится к комбинаторному (логическому) анализу исходных признаковых описаний объектов. При данном подходе для каждого признака определяется бинарная функция близости между его значениями, позволяющая различать объекты и их подописания. Особенно эффективен комбинаторный подход в случае целочисленной информации низкой значности, например бинарной. Поэтому актуальной является задача корректного понижения значности исходных данных.

Пусть $\{x_1, \dots, x_n\}$ — система признаков и $\epsilon_j, \epsilon_j \geq 0$, — точность измерения признака x_j , $j \in \{1, 2, \dots, n\}$. Пусть далее $S_{i_1} = (a_{i_1 1}, \dots, a_{i_1 n})$ и $S_{i_2} = (a_{i_2 1}, \dots, a_{i_2 n})$ — обучающие объекты (здесь $a_{i_t j}$ — значение признака x_j для объекта S_{i_t} , $t \in \{1, 2\}$, $j = 1, 2, \dots, n$). Положим

$$\delta_j(S_{i_1}, S_{i_2}) = \begin{cases} 1, & \text{если } |a_{i_1 j} - a_{i_2 j}| \leq \epsilon_j; \\ 0 & \text{иначе.} \end{cases}$$

Функция $\delta_j(S_{i_1}, S_{i_2})$ называется функцией близости объектов S_{i_1} и S_{i_2} по признаку x_j . Предполагается, что обучающие объекты из разных классов имеют разные описания, т. е. для любых S_{i_1} и S_{i_2} , принадлежащих разным классам, существует хотя бы один признак x_j такой, что $\delta_j(S_{i_1}, S_{i_2}) = 1$. В случае бинарной информации полагают $\epsilon_j = 0$, $j \in \{1, 2, \dots, n\}$.

Один из способов понижения значности данных состоит в преобразовании исходной выборки путем

разбиения множества значений каждого признака на интервалы порогами. Значения признаков, попавших в один интервал, считаются близкими и кодируются одним числом. Однако при произвольном выборе порогов обучающие объекты, принадлежащие разным классам, могут стать неразличимыми. При данном способе преобразования информации важным является понятие корректной перекодировки данных, т. е. такого преобразования обучающей информации, при котором объекты из разных классов остаются различимыми.

Ю. И. Журавлевым предложена методика корректного перекодирования исходных данных. Показано, что задача построения корректной перекодировки может быть сведена к построению специального вида покрытия булевой матрицы, которая строится по обучающей выборке.

В [1, 2] предложен подход, позволяющий выбирать наилучшую в смысле качества распознавания корректную перекодировку. Недостаток подхода — его большая вычислительная сложность.

Целью данной работы является развитие методов корректного перекодирования данных и снижение вычислительной сложности этих методов. В работе предложены более эффективные способы оценки качества перекодировок. Для сокращения перебора при поиске оптимальной корректной перекодировки использован генетический подход. Приведены результаты тестирования генетических алгоритмов поиска оптимальной корректной перекодировки на реальных прикладных задачах из репозитория системы «Распознавание», описанной в [3].

* Работа выполнена при финансовой поддержке РФФИ (проект № 10-01-00770) и гранта Президента РФ по поддержке ведущих научных школ НШ-7950.2010.1.

¹Вычислительный центр Российской академии наук им. А. А. Дородницына, edjukova@mail.ru

²Московский государственный университет им. М. В. Ломоносова, box.sizov@gmail.com

³Вычислительный центр Российской академии наук им. А. А. Дородницына, rom.sot@gmail.com

2 Основные обозначения

Рассмотрим задачу распознавания по прецедентам с двумя непересекающимися классами K_1 и K_2 [4]. Пусть $T = (a_{ij})_{m \times n}$ — обучающая таблица, $a_{ij} \in \mathbb{R}$, \mathbb{R} — множество действительных чисел. Столбцам таблицы T соответствуют признаки x_1, x_2, \dots, x_n , а каждая строка является набором значений признаков, описывающим один из обучающих объектов. Предполагается, что в таблице T нет столбцов, состоящих из одинаковых чисел.

Пусть S_{i_1} и S_{i_2} — обучающие объекты, принадлежащие разным классам, $j \in \{1, 2, \dots, n\}$.

Определение 1. Число $(a_{i_2j} + a_{i_1j})/2$ назовем порогом для признака x_j , если в T не существует элемента a_{ij} такого, что $a_{ij} \in (a_{i_1j}, a_{i_2j})$.

Через $D^{(j)}$ обозначим множество всех порогов для признака x_j , $j \in \{1, 2, \dots, n\}$. Суммой двух элементов a_{i_1j} и a_{i_2j} таблицы T по порогу $d \in D^{(j)}$, $j \in \{1, 2, \dots, n\}$, назовем число $(a_{i_1j} \oplus a_{i_2j} | d)$, равное 1, если a_{i_1j} и a_{i_2j} лежат по разные стороны от порога d , и равное 0 в противном случае. Пусть $D^{(j)} = \{d_1^{(j)}, \dots, d_{u_j}^{(j)}\}$.

Через \prod будем обозначать последовательность всех порогов $d_1^{(1)}, \dots, d_{u_1}^{(1)}, d_1^{(2)}, \dots, d_{u_2}^{(2)}, \dots, d_1^{(n)}, \dots, d_{u_n}^{(n)}$, где $u_j = |D^{(j)}|$ при $j = 1, 2, \dots, n$. Суммой двух строк таблицы T с номерами i_1 и i_2 по последовательности порогов \prod назовем строку $(a_{i_11} \oplus a_{i_21} | d_1^{(1)}, \dots, a_{i_11} \oplus a_{i_21} | d_{u_1}^{(1)}, a_{i_12} \oplus a_{i_22} | d_1^{(2)}, \dots, a_{i_12} \oplus a_{i_22} | d_{u_2}^{(2)}, \dots, a_{i_1n} \oplus a_{i_2n} | d_1^{(n)}, \dots, a_{i_1n} \oplus a_{i_2n} | d_{u_n}^{(n)})$.

Пусть m_1 и m_2 — число обучающих объектов из классов K_1 и K_2 соответственно. Построим булеву матрицу L . Матрица L имеет размеры $h \times N$, где $h = m_1 m_2$, $N = |D^{(1)}| + \dots + |D^{(n)}|$. Каждая ее строка образуется в результате попарного сложения строк таблицы T , описывающих объекты из разных классов, по последовательности порогов \prod . Порядок выбора пар может быть задан произвольным образом. Множеству порогов $D^{(j)}$, $j \in \{1, 2, \dots, n\}$, по построению соответствует группа из u_j столбцов матрицы L , обозначаемая через G_j .

Определение 2. Набор столбцов H матрицы L назовем кодирующим покрытием, если выполнены следующие два условия: 1) H является покрытием L , т. е. для любой строки матрицы L в наборе H можно указать хотя бы один столбец, имеющий 1 на пересечении с этой строкой; 2) $H \cap G_j \neq \emptyset$ при $j = 1, 2, \dots, n$.

Определение 3. Кодирующее покрытие назовем неприводимым, если никакое его собственное подмножество кодирующим покрытием не является.

Определение 4. Число $\max_{j \in \{1, 2, \dots, n\}} |H \cap G_j| + 1$ назовем значностью кодирующего покрытия H .

Кодирующее покрытие H задает очевидным образом преобразование таблицы T в таблицу T^H на основе замены элементов T числами из $\{0, 1, \dots, k-1\}$, где k — значность H . Действительно, пусть a_{pj} — произвольный элемент таблицы T и пусть $\{d_1, \dots, d_v\}$ — пороги, соответствующие столбцам из $H \cap G_j$, причем $d_1 < \dots < d_v$ и $v < k$. Возможны три случая:

- 1) $a_{pj} \leq d_1$;
- 2) $d_t < a_{pj} \leq d_{t+1}$, $t \in \{1, 2, \dots, v-1\}$;
- 3) $d_v < a_{pj}$.

В случае 1 элемент a_{pj} кодируется числом 0, в случае 2 — числом t , в случае 3 — числом v . Легко видеть, что в таблице T^H описания объектов из разных классов различны. В дальнейшем T^H будем называть корректной перекодировкой таблицы T .

Таким образом, каждому кодирующему покрытию матрицы L соответствует корректная перекодировка. Мощность множества кодирующих покрытий матрицы L экспоненциально растет с ростом размеров задачи. Поэтому сложной в вычислительном плане является задача выбора наилучшей по качеству распознавания корректной перекодировки. Данная задача рассмотрена в [1]. В указанной работе построен алгоритм КОД1 поиска оптимальной корректной перекодировки.

3 Алгоритм поиска оптимальной корректной перекодировки КОД1

Введем понятие типичного элемента b_{ij} в таблице $T^H = (b_{ij})_{m \times n}$. Пусть $b_{ij} = a$ и q_t , $t \in \{1, 2\}$, — число строк в T^H , имеющих a в пересечении со столбцом с номером j и описывающих объекты из класса K_t . Элемент b_{ij} назовем типичным в T^H для класса K_1 , если

$$\frac{q_1}{|K_1|} - \frac{q_2}{|K_2|} > \mu_j,$$

где $\mu_j, \mu_j \geq 0$, — заданный порог типичности значений признака x_j . Аналогично вводится понятие типичного элемента для класса K_2 .

Положим

$$I_{ij}(a) = \begin{cases} 1, & \text{если } b_{ij} = a; \\ 0 & \text{иначе.} \end{cases}$$

Пусть $D \subseteq D^{(j)}$, Q_j — множество всех типичных элементов j -го столбца T^H .

Для каждого признака x_j зададим целое число k_j , $0 < k_j \leq u_j$, и для каждого p , $p \in \{1, \dots, k_j\}$, поставим задачу максимизации функционала

$$F(D) = \frac{1}{m} \sum_{i=1}^m \sum_{a \in Q_j} I_{ij}(a), \quad |D| = p, D \subseteq D^{(j)}.$$

Таким образом, для каждого признака x_j и для каждого p , $p \in \{1, \dots, k_j\}$, выбираем подмножество порогов D_{pj}^* такое, что $F(D_{pj}^*) = \max F(D)$, $|D| = p$, $D \subseteq D^{(j)}$. Множество перекодировок признака x_j упорядочиваем по убыванию значений $F(D_{pj}^*)$. Будем считать, что перекодировка $H_1 = \{D_{p_1 1}^*, D_{p_2 2}^*, \dots, D_{p_n n}^*\}$ следует за перекодировкой $H_2 = \{D_{q_1 1}^*, D_{q_2 2}^*, \dots, D_{q_n n}^*\}$, если

$$\sum_{j=1}^n F(D_{q_j j}^*) \geq \sum_{j=1}^n F(D_{p_j j}^*).$$

В заданном порядке последовательно просматриваем всевозможные перекодировки таблицы T . Первая по порядку корректная перекодировка считается оптимальной.

Сложность алгоритма КОД1 быстро растет с ростом размеров задачи. В следующем разделе построены алгоритмы, в которых используются другие критерии оптимальности кодирующего покрытия. Для сокращения вычислительной сложности рассмотрен генетический подход.

4 Однокритериальные генетические алгоритмы поиска оптимальной корректной перекодировки КОД2, КОД3, КОД4

Пусть c_j , $j \in \{1, 2, \dots, N\}$, — число единиц в j -м столбце матрицы $L_{h \times N}$, $R_1(H)$ — множество номеров столбцов L , входящих в кодирующее покрытие H , $R_2(H)$ — множество номеров столбцов матрицы L , не входящих в кодирующее покрытие H . Тогда

$$f_1(H) = \sum_{j \in R_2(H)} c_j;$$

$$f_2(H) = \frac{1}{|H|} \sum_{j \in R_1(H)} \frac{1}{c_j}.$$

Построены алгоритмы КОД2, КОД3 и КОД4. В генетическом алгоритме КОД4 особям соответствуют кодирующие покрытия, а в алгоритмах КОД2 и КОД3 особям соответствуют неприводимые кодирующие покрытия. Роль функции приспособленности играет один из двух функционалов:

$f_1(H)$ для КОД2 и $f_2(H)$ для КОД3 и КОД4. Функция приспособленности особи является критерием качества кодирующего покрытия в перечисленных алгоритмах.

Для реализации генетических алгоритмов взята схема генетического алгоритма из [5], которая адаптирована к условиям задачи. Основной особенностью задачи поиска кодирующего покрытия является условие включения в покрытие хотя бы одного столбца из набора столбцов G_j , $j = 1, 2, \dots, n$, матрицы L . Для выполнения этого условия внесены соответствующие изменения в процедуру восстановления допустимости решения.

В работе использован оригинальный оператор мутации с переменным числом мутирующих генов, что позволяет минимизировать влияние данного оператора на особь на ранних этапах работы и усилить его влияние с увеличением числа итераций.

5 Двухкритериальный генетический алгоритм поиска оптимальной корректной перекодировки КОД5

Разработан двухкритериальный генетический алгоритм, основанный на схеме однокритериального генетического подхода. Отличия заключаются в процедуре вычисления функции приспособленности и процедуре добавления особи в популяцию.

На каждом этапе для каждой особи вычисляются значения функционалов $f_1(H)$ и $f_3(H)$, где $f_3(H)$ — значность перекодировки. На основе полученных векторов значений $v(f_1(H), f_3(H))$ производится вычисление рангов. Пусть V — множество векторов v , вычисленных для каждой особи популяции. Рангом вектора $rg(v)$, $v \in V$, называется величина, равная числу векторов, строго доминирующих данный вектор. В случае если таких векторов нет, ранг вектора v принимается равным 1. В качестве функции приспособленности используется

$$f(v) = \max_{w \in V} (rg(w) - rg(v)) + 1.$$

В процедуре обновления популяции заменяется особь, имеющая самый низкий ранг. Алгоритм рассматривает только неприводимые покрытия.

6 Результаты экспериментов

Построенные в настоящей работе генетические алгоритмы были протестированы на реальных данных и сравнивались с алгоритмами КОД1 и градиентным алгоритмом. В качестве распознающего

Таблица 1 Характеристики задач

N	m_1, m_2, n	Z
1	90, 42, 9	112
2	60, 71, 13	97
3	30, 30, 33	51
4	50, 50, 19	78
5	70, 80, 24	133

Таблица 2 Значность оптимальных перекодировок

N	КОД1	КОД2	КОД3	КОД4	КОД5
1	39	31	27	30	24
2	26	24	12	22	11
3	24	20	11	18	10
4	29	23	17	23	15
5	—	21	15	20	13

Таблица 3 Качество распознавания

N	A_0	A_0^*	A_1	A_2	A_3	A_4	A_5
1	54	66	67	67	76	72	73
2	63	83	88	92	93	92	94
3	58	70	72	72	82	80	81
4	53	68	71	73	79	72	75
5	57	66	—	67	76	74	78

Таблица 4 Время счета

N	m_1, m_2, n	A_0^*	A_1	A_3
1	90, 42, 9	6	59	10
2	60, 71, 13	8	244	17
3	30, 30, 33	13	415	29
4	50, 50, 19	9	194	17
5	70, 80, 24	24	—	67

алгоритма использовалась процедура голосования по представительным наборам с ограничением по длине набора. Сравнение проводилось на реальных задачах из репозитория программной системы «Распознавание», описанной в [3].

Результаты счета представлены в табл. 1–4. В этих таблицах введены следующие обозначения:

N — номер задачи;

m_1, m_2 — число объектов в классах;

n — число признаков;

A_0 — алгоритм голосования по представительным наборам, построенным по исходной обучающей выборке;

A_0^* — распознающий алгоритм, примененный к данным, перекодированным градиентным алгоритмом поиска неприводимых кодирующих покрытий;

A_1 — распознающий алгоритм, примененный к перекодированным алгоритмом КОД1 данным;

A_2 — распознающий алгоритм, примененный к перекодированным алгоритмом КОД2 данным;

A_3 — распознающий алгоритм, примененный к перекодированным алгоритмом КОД3 данным;

A_4 — распознающий алгоритм, примененный к перекодированным алгоритмом КОД4 данным;

A_5 — распознающий алгоритм, примененный к перекодированным алгоритмом КОД5 данным;

Z — значность исходных данных.

В табл. 1 приведены характеристики задач: число обучающих объектов в классах и число признаков. В случае вещественнозначной информации значность определяется по числу различных значений признаков в обучающей выборке.

Таблица 2 содержит значность полученных алгоритмами перекодировок. Прочерк означает, что алгоритм был исключен из эксперимента из-за слишком большого времени выполнения (более одного часа). Нетрудно заметить, что наименьшая значность достигнута алгоритмом КОД5. Данный результат обусловлен использованием функционала $f_3(H)$ в этом алгоритме.

В табл. 3 приведено качество распознавания в процентах, полученное алгоритмами на скользящем контроле.

Из приведенных в табл. 3 результатов следует, что все предложенные способы перекодирования данных улучшают качество распознавания. Из тех же результатов следует, что решения алгоритмов A_2 и A_4 не хуже по качеству распознавания, чем решения алгоритмов A_0 , A_0^* и A_1 , а алгоритмы A_3 и A_5 превосходят другие алгоритмы по качеству распознавания.

Нетрудно заметить, что условие неприводимости кодирующего покрытия оказывает существенное влияние как на значность перекодировки, так и на качество распознавания. Согласно данным, приведенным в табл. 3, лучшими функционалами оценки качества перекодировки являются $f_2(H)$ и пара функционалов $f_1(H)$ и $f_3(H)$, так как результаты алгоритмов A_3 и A_5 оказались лучшими во всех рассмотренных задачах.

Алгоритм A_3 превзошел алгоритм A_5 в задачах 1, 3, 4, поэтому можно сделать вывод, что меньшая значность перекодировки (функционал $f_3(H)$) не всегда означает, что перекодировка лучше по качеству распознавания.

В табл. 4 приведено время счета алгоритмов в секундах.

Прочерк означает, что алгоритм был исключен из эксперимента из-за слишком большого времени выполнения (более одного часа). В таблице не представлены алгоритмы A_2 , A_4 и A_5 по той причине, что вычислительная сложность этих алгоритмов аналогична вычислительной сложности алгоритма A_3 , а следовательно, время счета практически не отличается.

Из результатов следует, что распознающие алгоритмы, полученные на основе алгоритмов КОД2, КОД3, КОД4 и КОД5, по сравнению с распознающим алгоритмом, основанным на алгоритме КОД1, выигрывают не только по качеству решения, но и по скорости выполнения. Кроме того, качество решения полученных алгоритмов существенно лучше, чем качество решения алгоритма голосования по представительным наборам, примененного к неперекодированным данным.

Литература

1. Дюкова Е. В., Журавлев Ю. И., Песков Н. В., Сахаров А. А. Обработка вещественнозначной информации логическими процедурами распознавания // Искусственный интеллект, 2004. № 2. С. 80–85.
2. Djukova E., Inyakin A., Peskov N., Sakharov A. Combinatorial (logical) data analysis in pattern recognition problems // Pattern Recognition and Image Analysis, 2005. Vol. 15. No. 1. P. 46–48.
3. Журавлев Ю. И., Рязанов В. В., Сенько О. В. «Распознавание». Математические методы. Программная система. Практические применения. — М.: Фазис, 2006. 176 с.
4. Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики, 1978. Вып. 33. С. 5–68.
5. Sotegov R. M. Genetic algorithms for problems of logical data analysis in discrete optimization and image recognition // Pattern Recognition and Image Analysis, 2009. Vol. 19. No. 3. P. 469–477.

ОЦЕНИВАНИЕ ГИПЕРПАРАМЕТРОВ ЛИНЕЙНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ ПРИ ОТБОРЕ ШУМОВЫХ И КОРРЕЛИРУЮЩИХ ПРИЗНАКОВ*

А. А. Токмакова¹, В. В. Стрижов²

Аннотация: Решается задача отбора признаков при восстановлении линейной регрессии. Принята гипотеза о нормальном распределении вектора зависимой переменной и параметров модели. Для оценки ковариационной матрицы параметров используется аппроксимация Лапласа: логарифм функции ошибки приближается функцией плотности нормального распределения. Исследуется проблема присутствия в выборке шумовых и коррелирующих признаков, так как при их наличии матрица ковариаций параметров модели становится вырожденной. Предлагается алгоритм, производящий отбор информативных признаков. В вычислительном эксперименте приводятся результаты исследования на временном ряде.

Ключевые слова: байесовский вывод; ковариационная матрица; гиперпараметры модели; отбор признаков; регрессия

1 Введение

Часто при анализе временных рядов требуется рассмотрение большого числа признаков. В связи с этим возникают проблемы, связанные с наличием в выборке большого количества мультикоррелирующих признаков или с высокой зашумленностью выборки. В работе выдвинута гипотеза о нормальном распределении вектора зависимой переменной и вектора параметров модели [1, 2]. Необходимо оценить ковариационные матрицы этих распределений и установить связь между пространством данных и пространством параметров, что позволит произвести отбор шумовых и коррелирующих признаков.

Развитие методов отбора признаков имеет богатую историю. Начиная с 1960 г. активно развивались шаговые методы (Stepwise Regression) [3]. Главная идея этих методов состоит в отборе признаков, вносящих наибольший вклад в зависимую переменную. Вводится критерий, на основании которого алгоритм добавляет или удаляет признаки. Широкое применение получили частные случаи шаговой регрессии — алгоритмы LARS (Least Angle Regression) [4] и LASSO (Least Absolute Shrinkage and Selection Operator) [5].

Алгоритм LARS заключается в последовательном добавлении признаков. На каждом шаге веса признаков меняются таким образом, чтобы обеспечить наибольшую корреляцию восстановленного вектора зависимых переменных с вектором ре-

грессионных остатков. Алгоритм позволяет сократить количество свободных переменных и избежать проблемы неустойчивой оценки весов.

Метод LASSO вводит ограничения на норму вектора коэффициентов модели, что приводит к обращению в ноль некоторых коэффициентов модели. Метод приводит к повышению устойчивости модели и позволяет отбирать признаки, оказывающие наибольшее влияние на вектор ответов.

Одной из причин возникновения задачи отбора признаков является их мультиколлинеарность. Первые шаги по решению этой проблемы были сделаны в 1963 г. А. И. Тихоновым, который ввел понятие регуляризации — дополнительного ограничения на задачу [6]. В работе [7] введено понятие регуляризации и описан общий метод решения задач. Но поскольку работы Тихонова были опубликованы на Западе только лишь в 1977 г., в 1970 г. Hoerl и Kennard предложили метод гребневой регрессии [8]. В минимизируемую функцию вводилось дополнительное слагаемое, что повышало устойчивость решения [9], однако не позволяло производить отбор признаков.

Позднее стали появляться методы, использующие качественно иной подход к решению проблемы мультиколлинеарности. Например, Belsley предложил метод для удаления признаков [10], использующий сингулярное разложение матрицы плана. Алгоритм находит коэффициент, характеризующий степень зависимости признаков друг от друга. Позднее появился метод фактора инфляции

* Работа выполнена при поддержке РФФИ, грант № 10-07-00422.

¹Московский физико-технический институт, aleksandra-tok@yandex.ru

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

дисперсии (Variance Inflation Factor) [11], оценивающий увеличение дисперсии заданного коэффициента регрессии, что свидетельствует о высокой корреляции данных.

В данной работе для отбора признаков линейной регрессионной модели предлагается выполнить анализ пространства параметров. Вектор параметров рассматривается как многомерная случайная величина. Оценивается наиболее вероятное значение параметров. При оценке используется связный байесовский вывод [12, 13].

Основываясь на гипотезе о нормальном распределении параметров модели [2], оценивается ковариационная матрица распределения параметров [1, 14]. На ее главной диагонали стоят дисперсии случайных величин, что позволяет установить степень значимости данного конкретного параметра в модели. При таком подходе к отбору признаков не возникает необходимости разбиения выборки на обучение и контроль. Для оценки ковариационной матрицы в работе используется аппроксимация Лапласа [15]. Логарифм функции ошибки приближается функцией плотности нормального распределения, и появляется возможность вычисления нормировочной константы.

2 Постановка задачи

Дана регрессионная выборка: $D = \{\mathbf{x}_i, y_i\}_{i=1}^m = (X, \mathbf{y})$, где $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \dots, m$, — векторы независимой переменной, а $y_i \in \mathbb{R}$, $i = 1, \dots, m$, — значения зависимой переменной. Решается задача восстановления регрессии

$$\mathbf{y} = \mathbf{f}(\mathbf{w}, X), \quad (1)$$

где $\mathbf{f}(\mathbf{w}, X)$ — некоторая параметрическая вектор-функция. Пусть многомерная случайная величина \mathbf{y} имеет нормальное распределение:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \sigma^2 I_m),$$

где \mathbf{f} — вектор-функция, σ^2 — дисперсия распределения, I_m — единичная матрица размерности m .

Требуется приблизить функцию $\mathbf{f}(\mathbf{w}, X)$ параметрической функцией $\hat{\mathbf{f}}(X, \mathbf{w})$ из заданного класса \mathcal{F} (линейные функции), причем $|\mathcal{F}|$ конечно. Отображение $\mathbf{f} : \mathbb{R}^m \times \mathbb{W}^n \rightarrow \mathbb{R}^m$ будем называть моделью. Здесь \mathbb{R}^m — пространство данных, а $\mathbb{W}^n \subseteq \mathbb{R}^n$ — пространство параметров. В задаче линейной регрессии задача приближения функции $\mathbf{f}(\mathbf{w}, X)$ эквивалентна задаче отбора признаков. В данном случае модель определяется параметрами, которые соответствуют множеству индексов активных признаков $\mathcal{A} \subseteq \mathcal{J} = \{1, 2, \dots, n\}$. Таким обра-

зом, при выборе модели требуется найти такое множество индексов \mathcal{A}^* , которое бы обеспечивало минимум функции

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \subseteq \mathcal{J}} S(\mathbf{f}_{\mathcal{A}} | \mathbf{w}^*, D),$$

где $S(\mathbf{f} | \mathbf{w}, D)$ — функция ошибки, $\mathbf{f}_{\mathcal{A}}$ — параметрическая вектор-функция, вычисляемая только на множестве активных признаков, заданном индексами из множества \mathcal{A} . При этом параметры \mathbf{w}^* модели должны обеспечивать минимум функции

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{W}} S(\mathbf{w} | \mathbf{f}_{\mathcal{A}}, D).$$

3 Вид функции ошибки

Пользуясь предположением о том, что вектор зависимой переменной — многомерная случайная величина с нормальным распределением, запишем конкретный вид функции ошибки $S(\mathbf{w})$ для поставленной задачи.

Пусть многомерная случайная величина \mathbf{y} имеет нормальное распределение. Обозначим $\beta^{-1} = \sigma^2$. Тогда распределение зависимой переменной \mathbf{y} можно представить в виде:

$$p(\mathbf{y}) = (2\pi\beta^{-1})^{-m/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{f})^T \beta I (\mathbf{y} - \mathbf{f})\right). \quad (2)$$

Рассмотрим функцию правдоподобия данных, которая имеет вид:

$$p(\mathbf{y} | X, \mathbf{w}, \beta, \mathbf{f}) \stackrel{\text{def}}{=} p(D | \mathbf{w}, \beta, \mathbf{f}) = \frac{\exp(-\beta E_D)}{Z_D(\beta)}. \quad (3)$$

Здесь E_D — функция ошибки. Из выражений (2) и (3) определим ее как

$$E_D = \frac{1}{2}(\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}).$$

Коэффициент Z_D нормирует плотность нормального распределения и равен

$$Z_D(\beta) = (2\pi\beta^{-1})^{m/2}. \quad (4)$$

Рассмотрим равенство (1). Слева стоит многомерная случайная величина \mathbf{y} , имеющая нормальное распределение. Матрица X не является случайной величиной, поэтому предположим, что $\mathbf{w} \in \mathbb{W}^n$ также является многомерной случайной величиной с нормальным распределением. Параметрами этого распределения будут математическое ожидание \mathbf{w}_0 и матрица ковариаций A^{-1} :

$$p(\mathbf{w} | A, \mathbf{f}) = \frac{\exp(-E_{\mathbf{w}})}{Z_{\mathbf{w}}(A)}. \quad (5)$$

Определим функцию-штраф за большое значение параметров модели для принятого распределения как $E_{\mathbf{w}} = (1/2) (\mathbf{w} - \mathbf{w}_0)^T A (\mathbf{w} - \mathbf{w}_0)$. Нормирующая константа $Z_{\mathbf{w}}$ в этом случае равна

$$Z_{\mathbf{w}}(A) = (2\pi)^{n/2} |A^{-1}|^{1/2}. \quad (6)$$

Апостериорное распределение параметров модели для заданных A и β имеет вид:

$$p(\mathbf{w}|D, A, \beta, \mathbf{f}) = \frac{p(D|\mathbf{w}, \beta, \mathbf{f})p(\mathbf{w}|A, \mathbf{f})}{p(D|A, \beta, \mathbf{f})}, \quad (7)$$

$$\frac{p(D|\mathbf{w}, \beta, \mathbf{f})p(\mathbf{w}|A, \mathbf{f})}{p(D|A, \beta, \mathbf{f})} = \frac{\exp(-\beta E_D) \exp(-E_{\mathbf{w}})}{Z_D(\beta) Z_{\mathbf{w}}(A)} = \frac{\exp(-(\beta E_D + E_{\mathbf{w}}))}{Z_D(\beta) Z_{\mathbf{w}}(A)}.$$

В выражении (7) приняты следующие обозначения: $p(\mathbf{w}|D, A, \beta, \mathbf{f})$ — апостериорное распределение параметров; $p(D|\mathbf{w}, \beta, \mathbf{f})$ — функция правдоподобия данных; $p(\mathbf{w}|A, \mathbf{f})$ — априорное распределение параметров; $p(D|A, \beta, \mathbf{f})$ — функция правдоподобия модели. Записывая функцию ошибки как

$$S = E_{\mathbf{w}} + \beta E_D = \frac{1}{2} (\mathbf{w} - \mathbf{w}_0)^T A (\mathbf{w} - \mathbf{w}_0) + \frac{1}{2} (\mathbf{y} - \mathbf{f})^T \beta I (\mathbf{y} - \mathbf{f}), \quad (8)$$

получим следующее выражение для апостериорного распределения параметров:

$$p(\mathbf{w}|D, A, \beta, \mathbf{f}) = \frac{\exp(-S(\mathbf{w}))}{Z_S(A, \beta)},$$

где $Z_S = Z_S(A, \beta)$ — нормирующий коэффициент. Оценка нормировочного коэффициента производится с помощью аппроксимации Лапласа.

4 Аппроксимация Лапласа

Аппроксимация Лапласа позволяет оценить нормировочный коэффициент для ненормированной плотности вероятности. Пусть задано ненормированное распределение $p^*(\mathbf{w})$. Требуется найти нормировочную константу

$$Z = \int p^*(\mathbf{w}) d\mathbf{w},$$

при которой распределение $p(\mathbf{w}) = Z^{-1} p^*(\mathbf{w})$. Предположим, что $p^*(\mathbf{w})$ имеет максимум в точке \mathbf{w}_0 , т. е.

$$\left. \frac{dp(\mathbf{w})}{d\mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_0} = 0.$$

Прологарифмируем и разложим $p^*(\mathbf{w})$ по Тейлору в окрестности \mathbf{w}_0 :

$$\ln p^*(\mathbf{w}) = \ln p^*(\mathbf{w}_0) + 0 - \frac{1}{2} (\mathbf{w} - \mathbf{w}_0)^T A (\mathbf{w} - \mathbf{w}_0) + \dots, \quad (9)$$

где матрица Гессе $A = [\alpha_{ij}]$ определена как

$$\alpha_{ij} = - \left. \frac{\partial^2 \ln p^*(\mathbf{w})}{\partial w_i \partial w_j} \right|_{\mathbf{w}=\mathbf{w}_0},$$

т. е. $A = -\nabla \nabla \ln p^*(\mathbf{w})|_{\mathbf{w}=\mathbf{w}_0}$, где ∇ — градиент функции.

Отбрасывая все члены выше квадратичного в разложении и беря экспоненту обеих частей выражения (9), получим:

$$p^*(\mathbf{w}) \approx p^*(\mathbf{w}_0) \exp\left(-\frac{1}{2} (\mathbf{w} - \mathbf{w}_0)^T A (\mathbf{w} - \mathbf{w}_0)\right).$$

Тогда нормальное распределение $\hat{p}(\mathbf{w})$, приближающее нормированное распределение $p(\mathbf{w})$, имеет вид:

$$\hat{p}(\mathbf{w}) = \mathcal{N}(\mathbf{w}_0, A^{-1}) = \frac{1}{(2\pi)^{n/2} |A^{-1}|^{1/2}} \times \exp\left(-\frac{1}{2} (\mathbf{w} - \mathbf{w}_0)^T A (\mathbf{w} - \mathbf{w}_0)\right).$$

Следовательно, нормировочная константа имеет вид:

$$Z = p^*(\mathbf{w}_0) \frac{(2\pi)^{n/2}}{|A|^{1/2}}.$$

5 Оценка ковариационных матриц

Анализируя функцию ошибки $S(\mathbf{w})$, построим алгоритм, позволяющий выявлять шумовые и коррелирующие признаки.

Пусть нам известен локальный минимум $S(\mathbf{w})$ и он находится в точке \mathbf{w}_0 . Рассмотрим матрицу Гессе функции ошибок $H = -\nabla \nabla S(\mathbf{w})|_{\mathbf{w}=\mathbf{w}_0}$, где ∇ — градиент функции. При появлении в выборке шумовых или коррелирующих признаков происходит резкое возрастание некоторых элементов матрицы H . Необходимо установить связь между компонентами матрицы Гессе и ковариационной матрицей параметров, для того чтобы произвести отбор активных параметров \mathcal{A} и повысить устойчивость решения.

Рассмотрим ряд Тейлора второго порядка логарифма числителя (7):

$$-S(\mathbf{w}) \approx -S(\mathbf{w}_0) - \frac{1}{2} \Delta \mathbf{w}^T H \Delta \mathbf{w}, \quad (10)$$

где $\Delta \mathbf{w} = \mathbf{w} - \mathbf{w}_0$. В выражении (10) нет слагаемого первого порядка, так как предполагается, что точка \mathbf{w}_0 доставляет локальный минимум функции $S(\mathbf{w})$. Следовательно,

$$\left. \frac{\partial S(\mathbf{w})}{\partial w} \right|_{\mathbf{w}=\mathbf{w}_0} = 0.$$

Применяя экспоненту к обеим частям выражения (10), получим необходимое приближение:

$$\begin{aligned} \exp(-S(\mathbf{w})) &\approx \\ &\approx \exp(-S(\mathbf{w}_0)) \exp\left(-\frac{1}{2} \Delta \mathbf{w}^T H \Delta \mathbf{w}\right). \end{aligned} \quad (11)$$

При полученном приближении выражение (11) будет выглядеть следующим образом:

$$\begin{aligned} p(\mathbf{w}|D, A, \beta) &\approx \\ &\approx \frac{\exp(-S(\mathbf{w}_0)) \exp(-(1/2) \Delta \mathbf{w}^T H \Delta \mathbf{w})}{Z_S(A, \beta)}, \end{aligned} \quad (12)$$

где $Z_S(A, \beta)$ выступает в роли нормировочного коэффициента плотности вероятностного распределения. Оценка для коэффициента Z_S получена с помощью аппроксимации Лапласа (пояснения см. в гл. 4):

$$Z_S = \frac{\exp(-S(\mathbf{w}_0)) (2\pi)^{n/2}}{|H|^{1/2}}. \quad (13)$$

Подставив (13) в (12), получим оценку правдоподобия модели, на основании которой будем производить отбор оптимальных гиперпараметров модели

$$p(\mathbf{w}|D, A, \beta) = \frac{|H|^{1/2} \exp(-(1/2) \Delta \mathbf{w}^T H \Delta \mathbf{w})}{(2\pi)^{n/2}}.$$

Выражение (12) определяет выбор наиболее правдоподобной модели. Для нахождения гиперпараметров воспользуемся принципом максимума правдоподобия $p(D|A, \beta)$ относительно A и β . Запишем $p(D|A, \beta)$ в следующем виде:

$$p(D|A, \beta) = \int p(D|\mathbf{w}, A, \beta) p(\mathbf{w}|A) d\mathbf{w}.$$

Используя выражения (5) и (3), перепишем функцию правдоподобия в виде:

$$\begin{aligned} p(D|A, \beta) &= \\ &= \frac{1}{Z_{\mathbf{w}}(A)} \frac{1}{Z_D(\beta)} \int \exp(-S(\mathbf{w})) d\mathbf{w}. \end{aligned} \quad (14)$$

Из соображений нормировки интеграл выражения (7) равен единице, т. е.

$$\int p(\mathbf{w}|D, \beta) d\mathbf{w} = \int \frac{\exp(-S(\mathbf{w}))}{Z_S(A, \beta)} d\mathbf{w} = 1.$$

Следовательно, интеграл в правой части (14) в точности равен Z_S . Поэтому

$$\begin{aligned} p(D|A, \beta) &= \\ &= \frac{1}{Z_{\mathbf{w}}(A)} \frac{1}{Z_D(\beta)} \exp(-S(\mathbf{w}_0)) (2\pi)^{n/2} |H|^{-1/2}. \end{aligned} \quad (15)$$

Подставив значение $Z_{\mathbf{w}}$ из (6) и Z_D из (4) в (15), получим:

$$\begin{aligned} p(D|A, \beta) &= (2\pi)^{-n/2} |A^{-1}|^{-1/2} (2\pi)^{-m/2} (\beta^{-1})^{m/2} \times \\ &\times \exp(-S(\mathbf{w}_0)) (2\pi)^{n/2} |H|^{-1/2}. \end{aligned}$$

Получим оценку логарифма правдоподобия:

$$\begin{aligned} \ln p(D|A, \beta, \mathbf{f}) &= -\frac{1}{2} \ln |A^{-1}| - \frac{m}{2} \ln 2\pi + \\ &+ \frac{m}{2} \ln \beta^{-1} - S(\mathbf{w}_0) - \frac{1}{2} \ln |H|. \end{aligned} \quad (16)$$

Поочередно приравнивая частные производные по A и β выражения (16) к нулю, найдем максимум (16) по гиперпараметрам.

Пусть матрица A диагональна. Введем обозначение $\alpha = [\alpha_1, \dots, \alpha_n]^T$ для вектора, состоящего из элементов диагонали матрицы A . Представим гессиан в виде:

$$\begin{aligned} H &= -\nabla \nabla S(\mathbf{w}) = -\nabla \nabla (\beta E_D + E_{\mathbf{w}}) = \\ &= -\beta \nabla \nabla E_D - \nabla \nabla E_{\mathbf{w}} = H_D + H_{\mathbf{w}}, \end{aligned}$$

где H_D зависит от β , а $H_{\mathbf{w}}$ зависит от A . Так как $\nabla \nabla E_{w_i} = (\partial^2 / \partial w_i) ((1/2) \alpha_i (w_i - w_{0i})^2) = \alpha_i$, то часть гессиана $H_{\mathbf{w}}$ диагональна. Покажем, что при некоторых допущениях H_D также будет диагональной матрицей. Для этого рассмотрим два случая:

- (1) если все признаки независимы, то матрица H_D будет диагональной, так как недиагональные элементы матрицы Гессе отражают степень зависимости измеряемых величин;
- (2) при наличии в выборке шумовых или коррелирующих признаков будет наблюдаться возрастание диагональных элементов матрицы (дисперсий признаков), в сравнении с которыми недиагональными элементами можно пренебречь; т. е. и в этом случае матрицу H_D можно считать диагональной (на диагонали — собственные числа).

Итак, представим H_D в следующем виде: $H_D = \text{diag}(h_1, \dots, h_n)$. Для выявления связи между параметрами и гиперпараметрами модели рассмотрим выражение (16). Воспользуемся необходимым условием минимума и приравняем к нулю первые производные выражения (16) по α_i :

$$\frac{1}{\alpha_i} - (w_i - w_0)^2 - \frac{1}{\beta h_i + \alpha_i} = 0.$$

Данное уравнение имеет два корня. Однако один из них не имеет смысла, так как A^{-1} — диагональная ковариационная (положительно определенная) матрица, следовательно, по критерию Сильвестра (симметричная квадратная матрица является положительно определенной тогда и только тогда, когда все ее главные миноры положительны) не имеет отрицательных компонент:

$$\alpha_i = \frac{1}{2} \lambda_i \left(\sqrt{1 + \frac{4}{(w_i - w_0)^2 \lambda_i}} - 1 \right), \quad (17)$$

где $\lambda_i = \beta h_i$.

Приравняв производную по β выражения (16), найдем оптимальное значение β :

$$\frac{m}{2\beta} - E_D - \frac{1}{2\beta} \gamma = 0,$$

где

$$\gamma = \sum_{j=1}^W \frac{\lambda_j}{\lambda_j + \alpha_j}.$$

Таким образом,

$$\beta = \frac{m - \gamma}{2E_D}. \quad (18)$$

Выражения (17) и (18) не позволяют явно вычислить значения α и β . Поэтому итерационный процесс организуется следующим образом. На каждом шаге вычисляем \mathbf{w} (минимизируя функцию ошибки из выражения (8)), далее, используя полученное приближение, находим вектор гиперпараметров α , затем значение гиперпараметра β . Процедура продолжается до сходимости как параметров, так и гиперпараметров, т. е. до сходимости функции правдоподобия модели $p(D|A, \beta, \mathbf{w})$.

При появлении шумовых или коррелирующих признаков происходит возрастание диагональных элементов (большое значение дисперсии свидетельствует о неинформативности признака). Вследствие этого недиагональные элементы становятся настолько малы, что можно считать матрицу H_D диагональной. Поэтому необходимо принудительно занижать возрастающие диагональные элементы, тем самым производя отсеивание шумовых и коррелирующих признаков.

Ниже приведен псевдокод алгоритма оценки гиперпараметров регрессионной модели.

Вход: вектор зависимой переменной y , модель $\text{mdl}(\mathbf{w}, X)$
 $\mathbf{w}_0 = 0$;
 $\mathbf{w} = 0$;
 $A = \text{diag}(n, 1)$;
 $\beta = 1$;
для $k = 2, \dots, \text{MaxIterations}$
 вычислить A, β, \mathbf{w} :
 $\mathbf{w} = \text{FindParameters}(S(\mathbf{w}), A, \beta, \mathbf{w}, \mathbf{w}_0, y)$;
 для $j = 2, \dots, \text{MaxIterations}$
 добиться сходимости A и β **при данном векторе** \mathbf{w} :
 $H = \text{CalcHessian}(S(\mathbf{w}), A, \beta, \mathbf{w}, \mathbf{w}_0, y)$;
 если $\max(H)/\min(H) > 10^6$ **то**
 $\text{idx} = \text{find}(\max(H))$; {индекс строки/столбца (диагональный элемент) с max элементом}
 занулить строку и столбец Гесссиана, **содержащие максимальный элемент**;
 выход;
 $\lambda = \beta * \text{diag}(H)$;
 $A = \frac{1}{2} \lambda \left(\sqrt{1 + \frac{4}{(\mathbf{w} - \mathbf{w}_0)^2 \lambda}} - 1 \right)$;
 если $\text{idx} \neq 0$ **то**
 занулить соответствующие диагональные элементы матрицы A (необходимо для сходимости гиперпараметра α);
 выход;
 $\gamma = \sum \frac{\lambda_j}{\lambda_j + \alpha_j}$;
 $\mathbf{f} = \text{mdl}(\mathbf{w}, X)$;
 $E_D = \frac{1}{2} (\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f})$;
 $\beta = \frac{(m - \gamma)}{2E_D}$;
 если $\sum (\alpha_k - \alpha_{k-1})^2 < \text{Convergency}$ и $(\beta_k - \beta_{k-1})^2 < \text{Convergency}$; **то**
 закончить выполнение цикла на текущей итерации;
 выход;
 если $j = \text{MaxIterations}$ **то**
 вывести сообщение о величине ошибки и закончить выполнение программы;
 выход;
 если $\sum (w_k - w_{k-1})^2 < \text{Convergency}$ **то**
 закончить выполнение программы;
 конец

ПРОЦЕДУРА $\text{FindParameters}((S(\mathbf{w}), A, \beta, \mathbf{w}, \mathbf{w}_0, y)$
пока не найден минимум функции $S(\mathbf{w})$ **по** \mathbf{w}
 $\mathbf{f} = \text{mdl}(\mathbf{w}, X)$;

$$S(\mathbf{w}) = (1/2)(\mathbf{w} - \mathbf{w}_0)^T A (\mathbf{w} - \mathbf{w}_0) + (1/2)(\mathbf{y} - \mathbf{f})^T \beta I (\mathbf{y} - \mathbf{f});$$

выход;

вернуть \mathbf{w} ;

ПРОЦЕДУРА CalcHessian($S(\mathbf{w}), A, \beta, \mathbf{w}, \mathbf{w}_0, \mathbf{y}$)

$h = 10^{-6}$; {шаг разностной схемы}

для $i = 1, \dots, l$

для $j = 1, \dots, l$

посчитать элемент матрицы Гессе:

$\mathbf{e}_i = 0$; {вектор приращения}

$e_i(i) = 1$;

$\mathbf{e}_j = 0$;

$e_j(j) = 1$;

$H(i, j) = (S(\mathbf{w} + (\mathbf{e}_i + \mathbf{e}_j)h) - S(\mathbf{w} + \mathbf{e}_i h) - S(\mathbf{w} + \mathbf{e}_j h) + S(\mathbf{w})) / h^2$;

выход;

выход;

вернуть H ;

Для решения проблемы мультиколлинеарности к функционалу Q добавляют регуляризатор, штрафующий большие значения нормы вектора \mathbf{w} : $Q_\tau = \|X\mathbf{w} - \mathbf{y}\|^2 + \tau\|\mathbf{w}\|^2$. Решением полученной задачи является вектор

$$\mathbf{w}^* = (X^T X + \tau I_m)^{-1} X^T \mathbf{y}.$$

Увеличение τ приводит к уменьшению нормы вектора \mathbf{w} , однако при этом ни один из параметров не обращается в ноль. Это означает, что, повышая устойчивость модели, гребневая регрессия не производит отбор признаков.

Лассо Тибширани. В данном методе вместо добавления штрафного слагаемого к функционалу качества вводится ограничение-неравенство, запрещающее большие абсолютные значения коэффициентов:

$$Q(\mathbf{w}) = \|X\mathbf{w} - \mathbf{y}\|^2 \rightarrow \min_{\mathbf{w} \in \mathbb{W}};$$

$$\sum_{j=1}^W |w_j| < \theta.$$

6 Алгоритмы отбора признаков

Для того чтобы подчеркнуть особенности описанного в работе алгоритма, приведем примеры ранее предложенных методов регуляризации, приводящих к повышению устойчивости решения и отбору признаков в задаче линейной регрессии [5, 8].

Гребневая регрессия. Запишем функцию ошибки для линейной модели вида (1):

$$Q(\mathbf{w}) = \|X\mathbf{w} - \mathbf{y}\|^2.$$

Для минимизации функции воспользуемся необходимым условием минимума

$$\frac{\partial Q}{\partial \mathbf{w}} = 2X^T(X\mathbf{w} - \mathbf{y}) = 0,$$

откуда следует, что $X^T X \mathbf{w} = X^T \mathbf{y}$. Если матрица $X^T X$ не вырождена, то решением системы является вектор

$$\mathbf{w}^* = (X^T X)^{-1} X^T \mathbf{y}.$$

Если ковариационная матрица $X^T X$ имеет неполный ранг, то ее обращение невозможно. Также выделяют случай мультиколлинеарности: матрица $X^T X$ имеет полный ранг, но близка к некоторой матрице неполного ранга. В этом случае увеличивается разброс коэффициентов \mathbf{w}^* , появляются большие по абсолютной величине коэффициенты. Решение становится неустойчивым (небольшие изменения матрицы X ведут к большим изменениям величины \mathbf{w}^*).

Чем меньше значение θ , тем больше коэффициенты w_j обнуляется, таким образом происходит исключение j -го признака. Недостатком этого метода относительно алгоритма, представленного в работе, является необходимость в разделении выборки на две части: для обучения и контроля.

Также при использовании методов регуляризации возникает проблема выбора константы регуляризации. Для ее вычисления обычно используют скользящий контроль, что значительно повышает трудоемкость всей задачи в целом.

7 Вычислительный эксперимент

Результатом вычислительного эксперимента является отбор шумовых и коррелирующих признаков. Тестирование алгоритма производится на временном ряде продаж нарезного хлеба в зависимости от времени. Ряд содержит 195 записей. Модель, аппроксимирующая ряд: $\mathbf{y} = 0,2256 + 0,1996\xi + 0,0496 \sin(10\xi)$, где $\xi \in \mathbb{R}^n$ — регрессионная выборка. Введем следующие обозначения: ξ^0, ξ^1 — значение каждого элемента выборки в нулевой и первой степени соответственно, $\sin(10\xi)$ — поэлементное применение элементарной функции к вектору ξ . На рис. 1 представлены выборка и аппроксимирующая ее модель.

Пусть матрица плана X представлена в следующем виде $X = [\chi_1, \dots, \chi_n]$, где $\chi \in \mathbb{R}^m$. В данном случае она состоит из трех столбцов: $\xi^0, \xi^1, \sin(10\xi)$.

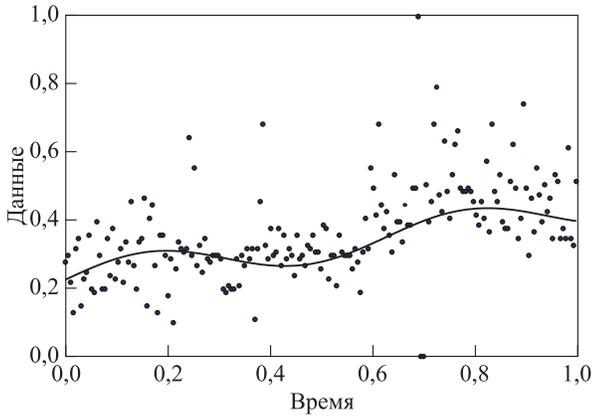


Рис. 1 Данные (точки) и аппроксимирующая модель (кривая)

Отбор шумовых признаков. Шумовая выборка сформирована при помощи добавления столбца случайных чисел с нормальным распределением. Модель, аппроксимирующая данные в эксперименте: $y = w_1\chi_1 + w_2\chi_2 + w_3\chi_3 + w_4\chi_4$, где $\chi_1 = \xi^0$, $\chi_2 \sim \mathcal{N}(0, 2)$, $\chi_3 = \xi^1$, $\chi_4 = \sin(10\xi)$. При наличии в выборке шумового элемента процедура сходится за восемь итераций. На рис. 2 проиллюстрирова-

ны изменения матрицы Гессе H на каждом шаге процедуры.

На 2-й итерации наблюдается резкое отличие диагонального элемента (2, 2). В течение итераций 2 и 3 он продолжает возрастать, пока не достигает критической относительной величины (принята эмпирическая оценка отношения максимального элемента матрицы к минимальному 10^6). Далее на 4-й итерации выполняется его зануление. Таким образом происходит выявление шумового признака.

На рис. 3 и 4 представлены диагональные элементы матрицы A . Рисунок 3 иллюстрирует изменения второго диагонального элемента α_2 , который соответствует шумовому параметру модели. Резкий скачок объясняется тем, что на данной итерации алгоритм находится вблизи локального минимума w_0 и, несмотря на возрастание диагональных элементов матрицы H , знаменатель формулы (17) мал. Далее происходит зануление элементов матрицы Гессе и соответствующий гиперпараметр α становится равным нулю.

На рис. 5 и 6 представлены скалярный гиперпараметр β и процесс изменения параметров модели w_i соответственно.

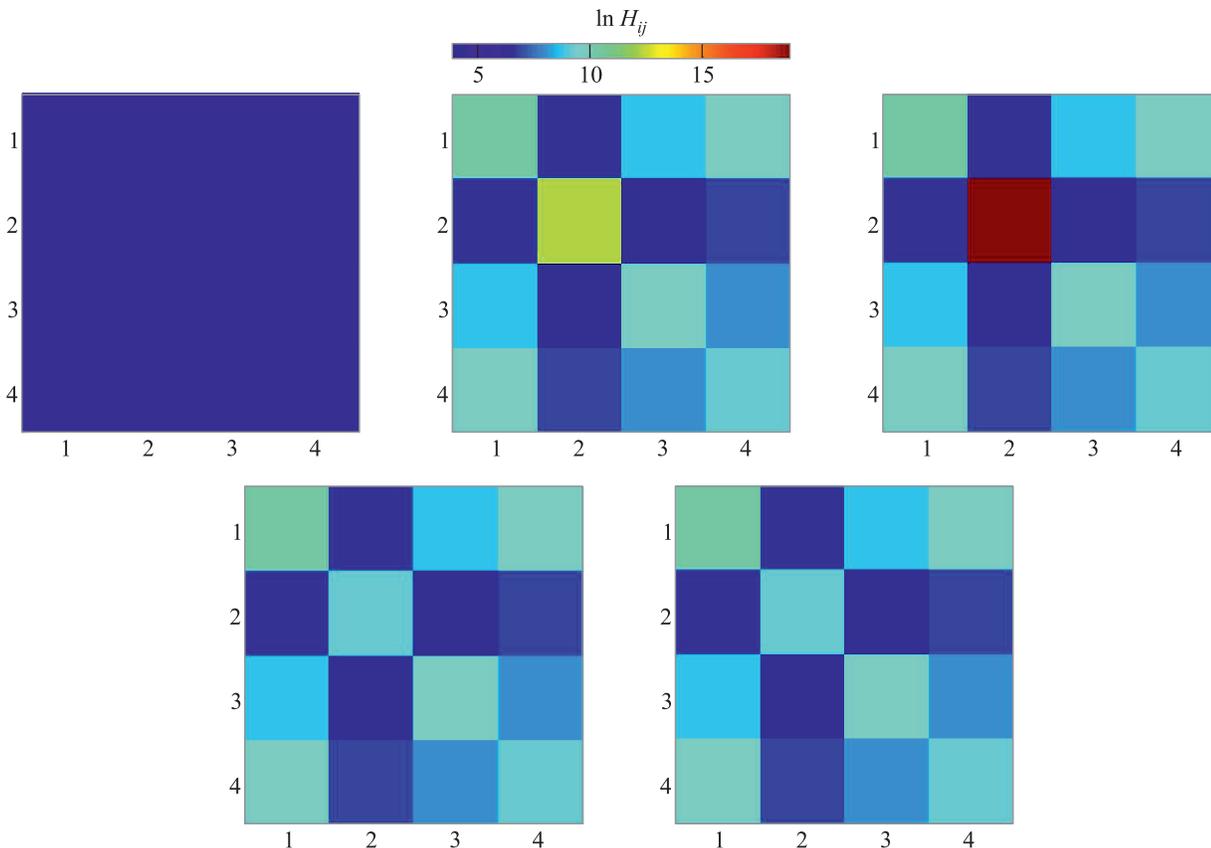


Рис. 2 Итерационный процесс для матрицы Гессе (случай шумового параметра)

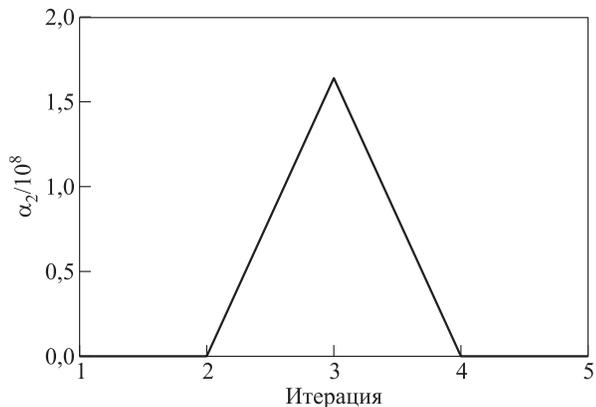


Рис. 3 Элемент матрицы A , соответствующий шумовому параметру модели

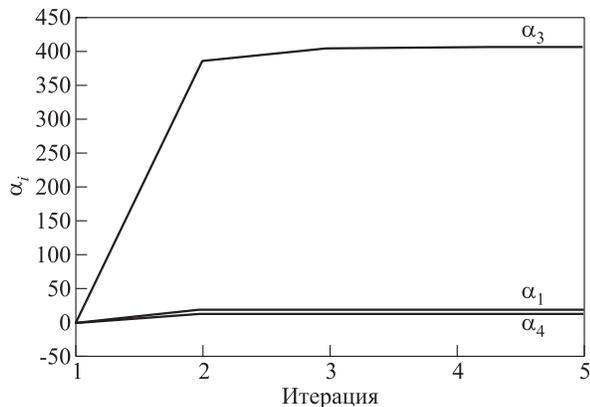


Рис. 4 Элементы матрицы A , соответствующие нешумовым параметрам модели

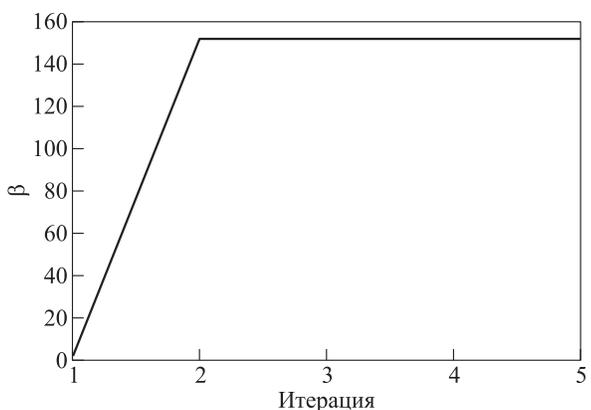


Рис. 5 Скалярный гиперпараметр β (случай шумового параметра)

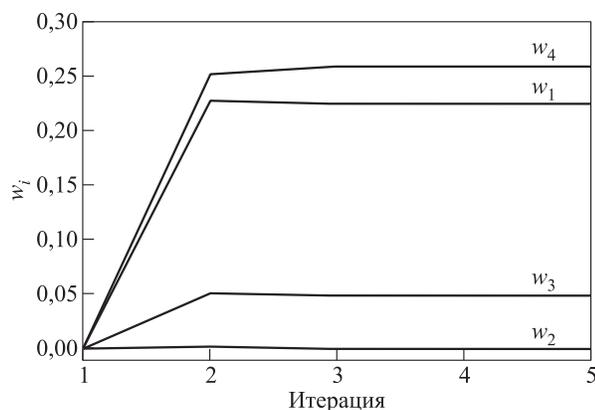


Рис. 6 Параметры модели w (случай шумового параметра)

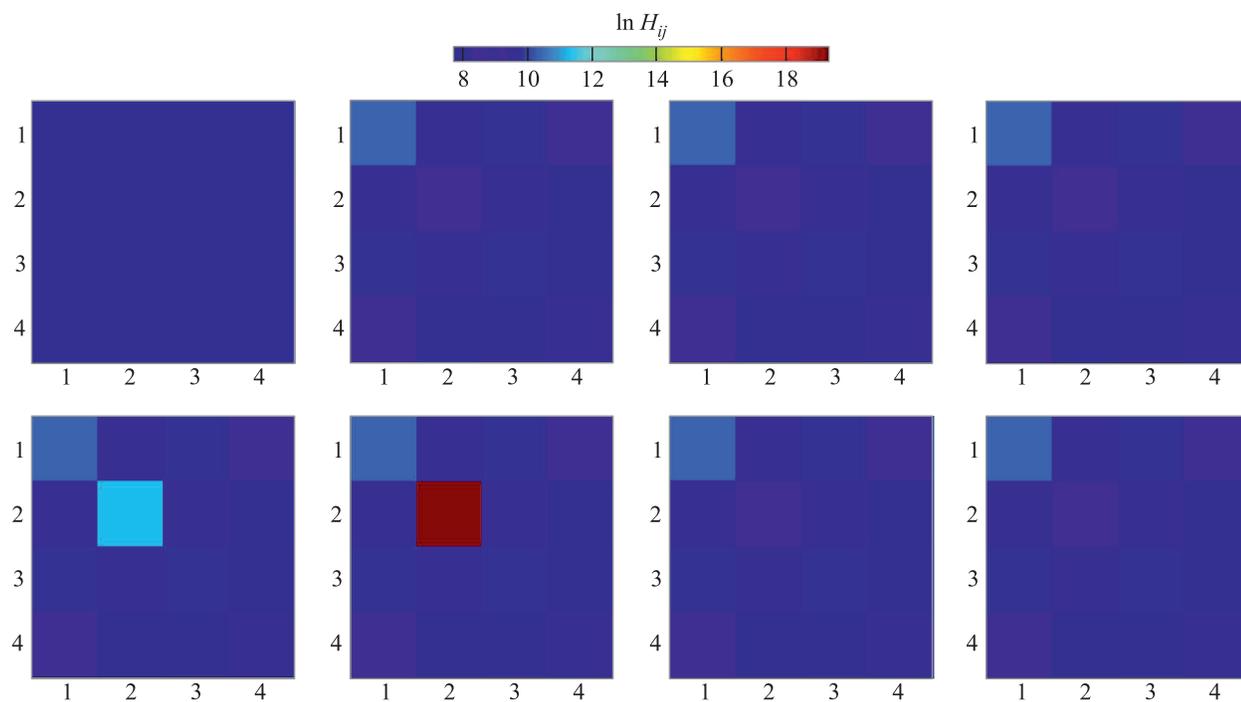


Рис. 7 Итерационный процесс для матрицы Гессе (случай коррелирующих параметров модели)

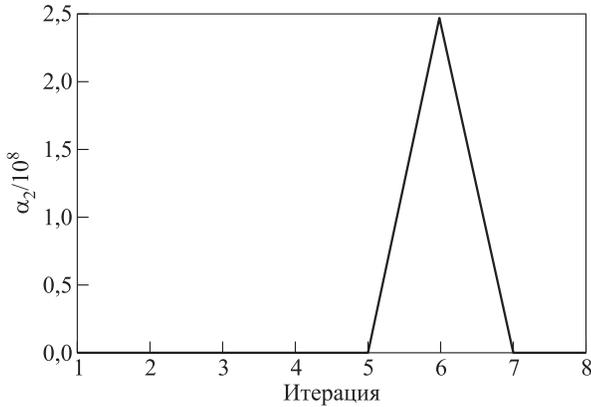


Рис. 8 Элементы матрицы A , соответствующие независимым параметрам модели

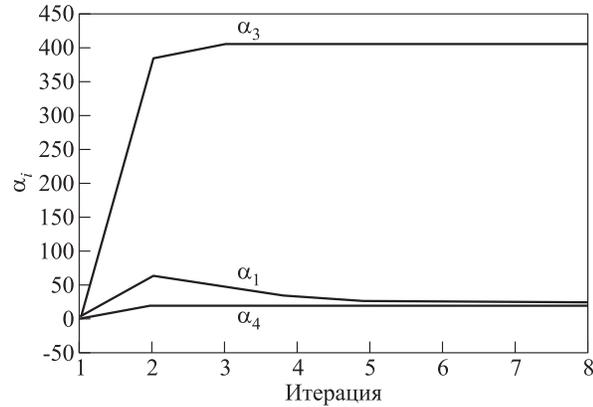


Рис. 9 Элемент матрицы A , соответствующий коррелирующему параметру модели

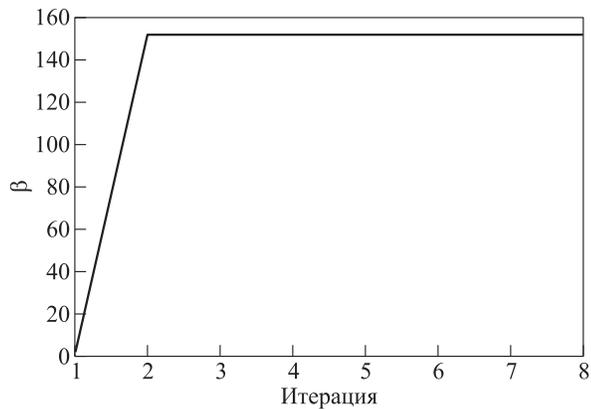


Рис. 10 Скалярный гиперпараметр β (случай зависимых параметров)

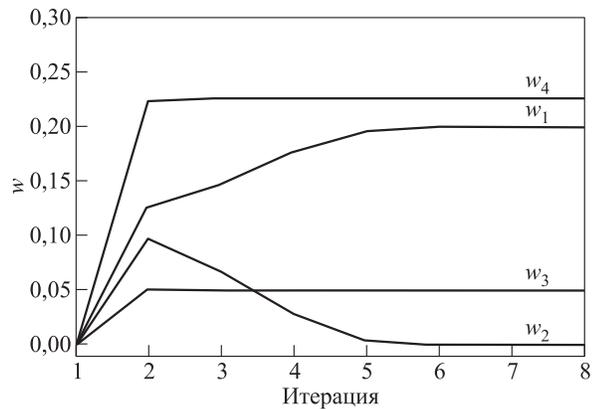


Рис. 11 Вектор параметров модели w (случай зависимых параметров)

Отбор коррелирующих признаков. Выборка с коррелирующими признаками сформирована при помощи добавления в матрицу плана столбца $1,3\chi_2$. Таким образом, модель, аппроксимирующая данные в эксперименте: $y = w_1\chi_1 + w_2\chi_2 + w_3\chi_3 + w_4\chi_4$, где $\chi_1 = \xi^0$, $\chi_2 = \xi^1$, $\chi_3 = 1,3\xi^1$, $\chi_4 = \sin(10\xi)$. На рис. 7 поэлементно проиллюстрирована матрица Гессе H .

При наличии коррелирующих признаков также наблюдается возрастание диагональных элементов. Это происходит из-за того, что алгоритм выбирает ближайший вектор χ к вектору y (в пространстве векторов матрицы X), а коррелирующий с ним считает шумовым.

На рис. 8 и 9 представлены диагональные элементы матрицы A .

На рис. 10 представлены изменения скалярного гиперпараметра β . На рис. 11 представлены изменения параметров модели w_i в течение итерационного процесса. Коррелирующий параметр w_2

сначала возрастает, а затем стремится к нулю. Это происходит из-за того, что пространство параметров модели многоэкстремально.

8 Заключение

В работе предложен способ отсеивания шумовых и коррелирующих признаков, а также алгоритм оценки ковариационной матрицы параметров модели. Данный алгоритм имеет следующие преимущества перед методами, описанными во введении: (1) нет необходимости деления данных на обучающую и контрольную выборку; (2) алгоритм не содержит никаких параметров, которые необходимо оценивать или задавать дополнительно (как, например, в методах регуляризации); (3) добиваясь сходимости как параметров, так и гиперпараметров, предложенный алгоритм повышает устойчивость выбранной регрессионной модели.

Литература

1. Стрижов В. В. Поиск параметрической регрессионной модели в индуктивно заданном множестве // Вычислительные технологии, 2007. Т. 1. С. 93–102.
2. Strijov V. V., Weber G.-W. Nonlinear regression model generation using hyperparameter optimization // Computers Math. Appl., 2010. Vol. 60. No. 4. P. 981–988.
3. Efrogmson M. A. Multiple regression analysis // Mathematical methods for digital computers. Vol. 1 / Eds. A. Ralston, H. S. Wilf. — New York: John Wiley and Sons, 1960. P. 191.
4. Efron B., Hastie T., Johnstone J., Tibshirani R. Least angle regression // Ann. Stat., 2004. Vol. 32. No. 3. P. 407–499.
5. Tibshirani R. Regression shrinkage and Selection via the Lasso // J. R. Stat. Soc., 1996. Vol. 32. No. 1. P. 267–288.
6. Ильин В. А. О работах А. Н. Тихонова по методам решения некорректно поставленных задач // Успехи математических наук, 1997. Т. 1. С. 168–175.
7. Тихонов А. Н. Решение некорректно поставленных задач и метод регуляризации // Докл. АН СССР, 1963. Т. 151. С. 501–504.
8. Hoerl A. E., Kennard R. W. Ridge regression: Biased estimation for nonorthogonal problems // Technometrics, 1970. Vol. 3. No. 12. P. 55–67.
9. Bjorkstrom A. Ridge regression and inverse problems. Technical Report. — Stockholm: Stockholm University, 2001.
10. Belsley D. A. Conditioning diagnostics: Collinearity and weak data in regression. — New York: John Wiley and Sons, 1991.
11. Marquardt D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation // Technometrics, 1996. Vol. 12. No. 3. P. 605–607.
12. Nabney I. Bayesian techniques // Netlab: Algorithms for pattern recognition. — New York: Springer, 2002. P. 325–366.
13. MacKay D. Laplace's method // Information theory, inference, and learning algorithms. — Cambridge: Cambridge University Press, 2005. P. 341–351.
14. Стрижов В. В. Методы выбора регрессионных моделей. — М.: ВЦ РАН, 2010.
15. Bishop C. M. Linear models for classification // Pattern recognition and machine learning / Eds. M. Jordan, J. Kleinberg, B. Scholkopf. — New York: Springer Science + Business Media, 1960. P. 213–216.

HOLOGRAPHIC CODING BY WALSH–HADAMARD TRANSFORMATION OF RANDOMIZED AND PERMUTED DATA*

S. Dolev¹, S. Frenkel², and A. Cohen³

Abstract: Holographic coding has the very appealing property of obtaining partial information on the data from any part of the coded information. In the paper, holographic coding schemes based on the Walsh–Hadamard orthogonal codes are studied. It is proposed to randomize the data so that the coefficient of the Walsh–Hadamard code would be approximately uniform and thus ensure, with high probability, a monotonic gain of information. The data are xored with randomly chosen bits from random data that have been stored during a preprocessing stage or pseudorandom data produced by a pseudorandom generator. Statistical properties of the truncated sums of the Inverse Walsh–Hadamard Transformation (IWHT), taking into account the “white-noise nature” and the mentioned above holographic, is considered. Furthermore, an enhancement of the algorithm, based on random permutation and block coding is suggested. The results are compared to the Rate Distortion function and jpeg compression.

Keywords: holographic coding; Walsh–Hadamard transformation; Shannon bound

1 Introduction

Holographic coding has the very appealing property of obtaining information on the entire data, from any part of the coded information [1]. Thus, intuitively, any portion of holographic coded information (which is any subset of the corresponding codewords) represents a blurred image of the entire data. Traditionally, the blurriness is considered as a *graceful degradation* of a picture, just like the elimination of some Fourier components of high frequency from a picture spectrum. For a binary input vector, the error measure that is considered is the fraction of erroneous bits in the reconstructed file. In this case, the blurriness is defined as a uniform distribution of the errors, namely, errors in the reconstructed input vector are uniformly distributed over the entries.

The goal of this work is to consider the use of the holographic property to provide the robustness of transmission to losses and erasures and to facilitate compression when possible. The possibility of using data randomization and Walsh–Hadamard coding is investigated. Walsh–Hadamard transformation and WHT-based codes of digital random sequences are widely used in many computer and data transmission tasks, for example, for image data transmission [2]. Recently, the present authors proposed a method which combines the WHT with a randomization process that xors the data with bits of random data (organized, say, as a pseudoran-

dom vector) that have been stored during a preprocessing stage [3]. Namely, each bit of the output vector is the exclusive or (xor) of the corresponding bit of the original binary vector and a bit of the pseudorandom data sequence.

As a result, an uncorrelated random sequence with almost uniform spectrum (with properties of a discrete white noise) was obtained. The decoding was performed in a reverse manner, namely, inverse the WHT in order to get the xored sequence and then xoring with the (same) pseudorandom sequence to reveal the original data file.

As it was shown, this coding can be interpreted as a holographic code. It is holographic in the sense that any portion of the coded information (which is any subset of corresponding codewords) represents a blurred image of the entire data. The Hamming distance between the original and the reconstructed binary files is considered as a *blurriness* measure.

In this paper, statistical properties of the truncated sums of the IWHT are presented, taking into account the white-noise nature of the xored state output and the mentioned above holographic properties of this encoding method and considering the performance of the method based on the theoretic Rate Distortion bound. Using this performance measure, an improvement of the authors’ previous WHT-based holographic encoding method is suggested.

*Partially supported by Rita Altura trust chair in computer science and by the Russian Foundation for Basic Research (grant RFBR 12-07-00109).

¹Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva, Israel, dolev@cs.bgu.ac.il

²IPI RAN; Moscow Institute of Radio, Electronics, and Automation (MIREA), fsergei@mail.ru

³Department of Communication Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel, coasaf@cse.bgu.ac.il

2 Holographic Encoding

In Fig. 1, file F is xored with a pseudorandom binary matrix PN , then WHT is applied to the xored output to get the l first largest coefficients together with their indexes, then CC Block either sends the outcome to a channel or writes the outcome to a memory device (depending on the purpose of the encoding) a total l of n WHT coefficients, where l is predefined for the encoding of the given binary file F . Rxored is the reconstructed xored vector. Each of the components of the file Rxored is rounded to 0 or 1 and then xored with the PN to obtain a reconstructed blurred version R file of the original file.

Thus, the scheme is represented by the following encoding algorithm:

- (a) compute the *xored randomized data* by xoring each bit of the original file with a pseudorandom sequence PN ;
- (b) compute the WHT coefficients c_i ;
- (c) choose the $l < n$ first largest WHT coefficients for a predefined l ;
- (d) transmit the pairs (c_i, i) of these l coefficients; and
- (e) request for additional WHT coefficients if the reconstructed file resolution (estimating the values of all missing coefficients to be 0) is insufficient for the specific application requirements.

The number of coefficients l that are chosen depends on the application requirements.

The decoding algorithm (Fig. 1b) is based on the IWHT and a threshold rule for deciding on the binary value of the reconstructed xored random data which should be reconstructed by l coefficients of n (see below).

This method has the following features: (i) Walsh–Hadamard based coding that uses only additions and shifts, implying low processing complexity; (ii) input randomization for obtaining *blurred*, information revealing of the data at a positive rate, starting with more than half of the data correctly revealed after the arrival of a very few coefficients; and (iii) encryption of information transmitted through the channel.

3 Reconstruction by a Subset of Wash–Hadamard Coefficients

The WHT is based on a complete set of orthogonal functions. That is, if $b = (b_1, b_2, \dots, b_m)$ is a binary file (modeled as a binary vector), then n -character encoding of b can be represented as $c = bW$, where $c = (c_1, \dots, c_n)$, $n = 2^k$, for an integer k , are the Walsh–Hadamard coefficients. Table 1 depicts the eight (0–7) first WHT entries. Note that the sum of the entries in each row (and column), except the first one, is zero.

In more details, the spectral coefficients of WHT are $c_h = (1/n) \sum_{i=0}^{n-1} b_i W(h, i)$, and the inverse transform is $b_i = \sum_{h=0}^{n-1} c_h W(h, i)$.

Let $b = b_0, b_2, \dots, b_{n-1}$ be a random white-noise maximal entropy sequence of n bits that is obtained by xoring the original binary vector with a predefined pseudorandom vector, $\text{Prob}(b_i = 1) = \text{Prob}(b_i = 0) = 1/2$.

For compression, only the $l \ll n$ largest WHT coefficients c_1, \dots, c_l are used for the original sequence reconstruction. Correspondingly, each bit b_i of the randomized sequence b is estimated by the WHT mentioned above as $\tilde{b}_i = \sum_{j=0}^l c_j W(j, i)$, j is the ordered index of the first l largest coefficients, and $b_i = \tilde{b}_i + e_i(l)$, where $e_i(l) = \sum_{q=l+1}^{n-1} c_q W(q, i)$.

The goal is to compute a metric that captures the difference of the bits b_i and \tilde{b}_i . The result may depend on the coefficients chosen for reconstruction and the requirements of the application in hand. Each coefficient c_i is transmitted/stored with its index i in the WHT matrix, namely, the pairs (c_i, i) are stored as the representation of the data.

Due to paper size limit, an intuitive explanation for choosing the l largest coefficients for reconstruction will

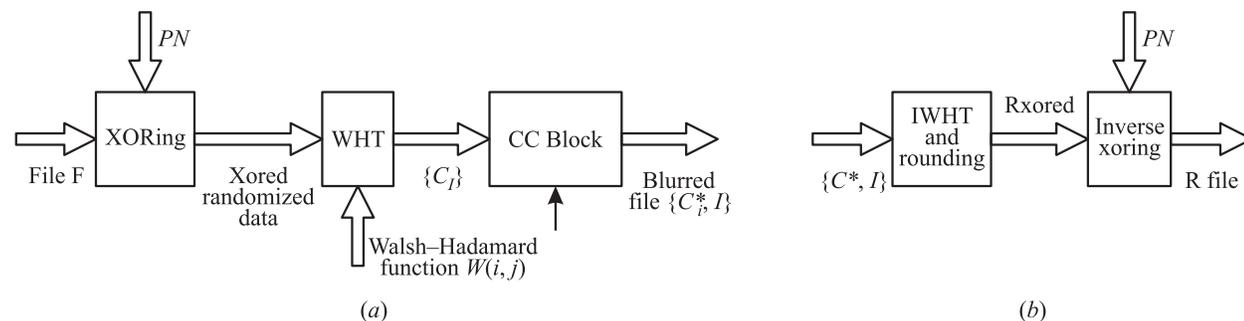


Figure 1 Walsh–Hadamard transformation based holographic encoding (a) and decoding (b)

Table 1 Walsh–Hadamard matrix

$W(0,*)=$	1	1	1	1	1	1	1	1
$W(1,*)=$	1	-1	1	-1	1	-1	1	-1
$W(2,*)=$	1	1	-1	-1	1	1	-1	-1
$W(3,*)=$	1	-1	-1	1	1	-1	-1	1
$W(4,*)=$	1	1	1	1	-1	-1	-1	-1
$W(5,*)=$	1	-1	1	-1	-1	1	-1	1
$W(6,*)=$	1	1	-1	-1	-1	-1	1	1
$W(7,*)=$	1	-1	-1	1	-1	1	1	-1

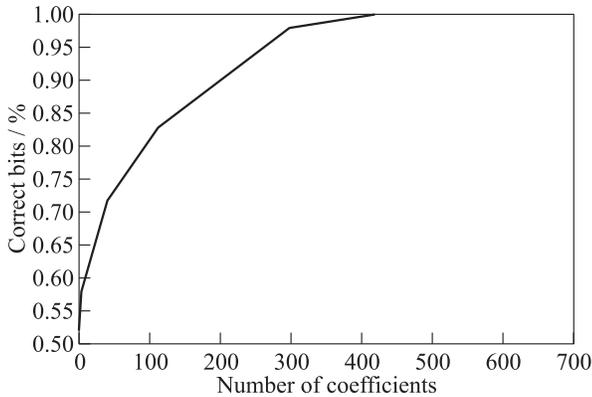


Figure 2 Fraction of correctly reconstructed bits of pseudo-random binary vector vs. number of l first largest coefficients of a 1K bits pseudorandom vectors

be just sketched. In a few words, the choice of the l largest coefficient is a consequence of the Parseval Identity:

$$n \sum_{i=0}^{n-1} b_i^2 = \sum_{i=0}^{n-1} c_i^2.$$

That is, as the l coefficients in the right side are bigger, then faster is the estimations \tilde{b}_i convergence to the accurate (original) values b_i . This is a common technique in data compression.

Figure 2 illustrates the accuracy of the xored random vector reconstruction using only the first largest coefficients (partial sums) of the IWHT.

As it can be seen in Fig. 2, a fairly small portion of the coefficients can provide a rather large fraction of correct bits. Therefore, it would be interesting to have a probabilistic characterization for the relationship between the l largest coefficients and the reconstruction error probability.

The estimation of a bit $b_i = \text{round}(\tilde{b}_i)$, where \tilde{b}_i is the estimation of the i th value before rounding (see Fig. 1b), computed by a partial sum of IWHT, is determined by the following random events:

$e_0 : (b_i = 0), e_1 : (b_i = 1)$, that is, the bit b_i of randomized file b is 0 (event e_0) or 1 (event e_1), $v_{i0} : \tilde{b}_i \leq 1/2, v_{i1} : \tilde{b}_i \geq x_0$ (defined on the space of the rational values \tilde{b}_i), where x_0 is the threshold used for the

rounding. For example, $x_0 = 1/2$ if standard rounding function is used.

Let $\text{Pr}_{\text{err}=0}(i)$ be the probability that the actually zero bit b_i was erroneously reconstructed as $b_i = 1$, and $\text{Pr}_{\text{err}=1}(i)$ be the probability that the bit $b_i = 1$ was erroneously reconstructed as $b_i = 0$.

Both probabilities $\text{Prob}(v_{i0})$ and $\text{Prob}(v_{i1})$ are the probabilities that the partial sums mentioned above have values that exceed/not exceed $1/2$.

Formally, in order to estimate error of the sequence reconstruction by truncated number of coefficients, one should know the joint distributions of the sum of l terms of the WHT $S_l = \sum_{j=0}^l c_j W(j, i)$ and the sum

of the residue $S_R = \sum_{j=n-l+1}^N c_j W(j, i)$. Then, taking into account that the sum $S_l + S_R$ is the exact value $b_i = 0$ or 1 , the conditional error probability can be characterized by $\text{Prob}(S_l \geq \text{Tr}/S_l + S_R = 0)$, $\text{Prob}(S_l \leq \text{Tr}/S_l + S_R = 1)$, where Tr is the threshold value.

In accordance with Theorem 6.4 in [4], the WHT coefficients are distributed (asymptotically) as some independent normal random values with zero mean and dispersion of $n \times f(i)$, where i is the WHT coefficient index and $f(i)$ is the (dyadic) spectral density of b .

However, as the largest coefficients are used in the partial sums mentioned above, this Gaussian asymptotic cannot be used directly for relatively small number of coefficients, as the joint distribution of maximal values of an ordered random sequence is not normal.

In order to get, nevertheless, an idea about the distributions, histograms of the partial sums S_l of the l first largest coefficients for different values of l have been computed (Fig. 3).

4 Probabilistic Features of the Walsh–Hadamard Spectral Coefficients and Partial Sums

Figure 3 shows the histograms of the partial sums (truncated IWHT) for the different number of coefficients used.

Apparently, the distributions of sums yielded by small l 's (out of the n possible) coefficients are far from normal, which, in turn, corresponds to the fact that the first largest coefficients have non-Gaussian distribution. Moreover, as the fractions l/n increase, the distributions convergence to a mixture of two normal distributions with meanings 0 and 1, respectively.

Next, an analytical model is presented to explain these results. Obviously, when l is equal to n , the value S_n is the original xored binary random value \tilde{b}_i with

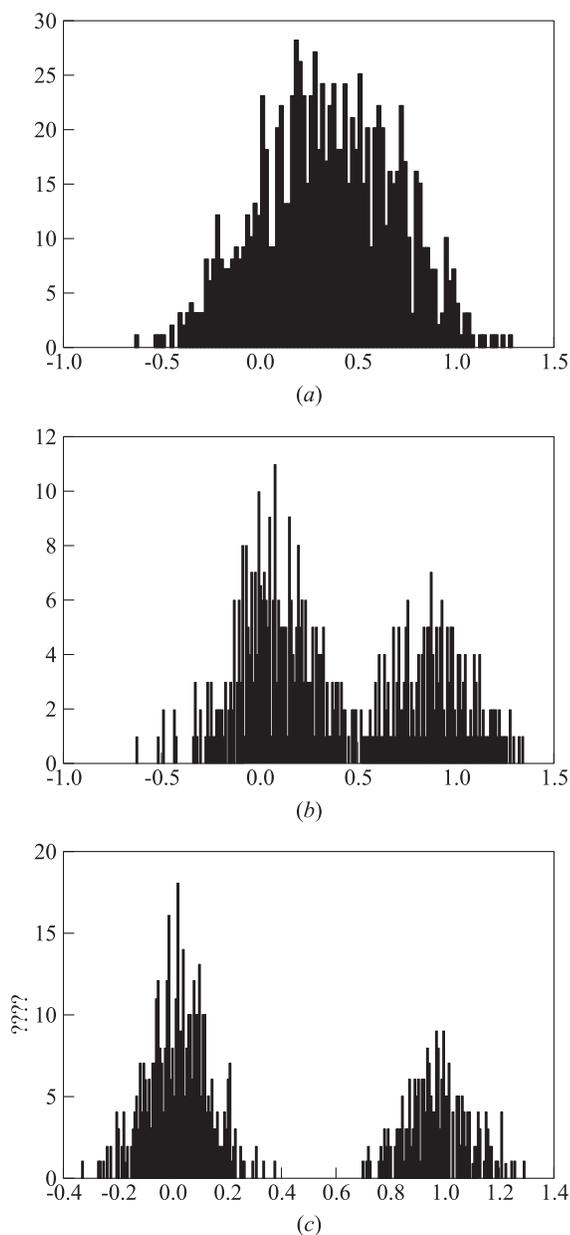


Figure 3 Histograms of partial sums with 80 (a), 280 (b), and 500 (c) first largest coefficients of 1K vector

$\text{Prob}(1) = \text{Prob}(0) = 1/2$. That is, any random value of the partial sums S_l with $l < n$ such that all n bits of the original xored vector are reconstructed correctly after rounding is a bounded (clipped) random value that forms a nonlinear function:

$$y(x) = \begin{cases} 1 & x \geq x_0; \\ 0 & x < x_0. \end{cases}$$

For the standard round function, x_0 is equal to $1/2$. That is, the partial sums behavior looks like a clipping

to the levels 0 or 1 of a random value (signal, random process) with the values from domain outranking $(0,1)$ (see Fig. 3c).

The probability density of the random variable y is

$$W(y) = [1 - F(x_0)]\delta(x - 1) + F(x_0)\delta(x)$$

where $F(\cdot)$ is the cumulative probability function of S_l that is, $\text{Prob}(S_l < x)$.

The proximity of the mixture components to the normal distribution can follow from [5], where it was proved that the sum of the largest $n - k$ out of n normal random variables is normal asymptotically under some conditions for the ratio k/n , $n \rightarrow \infty$ with k remaining fixed. Comparisons are made as well with the asymptotic distribution as both n and $k \rightarrow \infty$ with $k/n \rightarrow \alpha > 0$ (these issues are not considered in this paper).

Further, as the shape of the partial sums distribution with rather small l/n does not look like a mixture of distributions, it would be interesting to consider a possibility of using the normal approximation of these sums for the error reconstruction prediction.

Assume that it is possible to estimate the probability distribution functions of the S_l sums. Then, in order to compute the error probabilities $\text{Prob}(v_{i0})$ and $\text{Prob}(v_{i1})$, one should know the conditional probabilities of the sums (that is, the bit values estimations \tilde{b}_i) given the random value b_i . But, obviously, the random values $S_l = \tilde{b}_i$ cannot influence b_i as the bits are produced independently by a predefined pseudorandom sequence, and *vice versa*. Thus, it is enough to estimate only the marginal distribution of the partial sums, that is, the bits estimations \tilde{b}_i .

Let use the normal approximation of the IWHT sums. Then, one gets that the error probability can be expressed as the probability of a normal distributed random variable v , taking the values in accordance with the events v_{i0} and v_{i1} that falls into an interval $[a, b]$, that is (see Fig. 3c):

$$\begin{aligned} \text{Prob}(a \leq v \leq b) &= \Phi\left(\frac{b - E(v)}{s_v}\right) \\ &\quad - \Phi\left(\frac{a - E(v)}{s_v}\right) \end{aligned} \quad (1)$$

where $s_v = \text{sqrt}(\text{Var}(v))$;

$$\text{Prob}(v_{i0}) = \text{Prob}\left(b_{\min} \leq v \leq \frac{1}{2}\right); \quad (2)$$

$$\text{Prob}(v_{i1}) = \text{Prob}\left(\frac{1}{2} \leq v \leq b_{\max}\right) \quad (3)$$

where b_{\min} and b_{\max} are lower and upper bounds on v .

Note that in this computation, the sampled variance s_v used as an analytical computation of the variance is difficult to compute, seems impossible, as the first

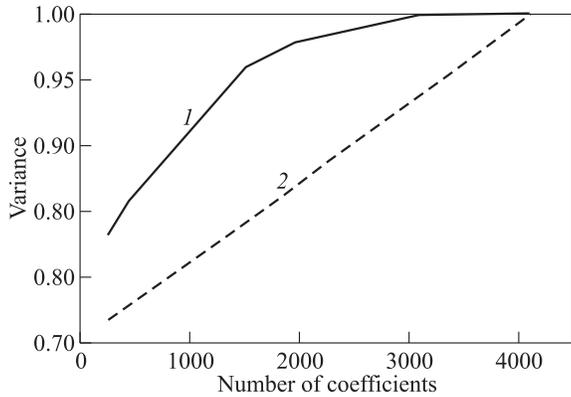


Figure 4 Computed variance of the partial sums (*l*) and the theoretical variance (*2*) (under independence assumption) vs. number of coefficients for a file of size 4K

largest *l* components of the ordering statistics corresponding to the set of all *n* coefficients are dependent. Indeed, Fig. 4 depicts that the curves of the variance values computed via simulation and the variance analytically

$$\begin{aligned} \text{Var}(S_l) &= \text{Var}(\tilde{b}_i) = \text{Var}\left(\sum_{j=0}^l c_j W(j, i)\right) \\ &= \text{Var}\left(\sum_{j=0}^l \left(\frac{1}{n}\right) \sum_{i=0}^{n-1} b_i W(h, i) W(j, i)\right) = l/(4n) \end{aligned}$$

computed assuming independence of the first large coefficients and taking into account that $\text{Var}(b_i) = 1/4$ (Bernoulli trial with parameter $p = 1/2$), are far from each other unless *l* is close to *n*, that is, the first largest coefficients are strongly dependent.

A divergence between the formulas (1)–(3) and the experiments (Fig. 5) in the vicinity of 100 percent accuracy is a consequence of the fact that the partial sums after achieving the correct (0,1) values that takes place by $l/n < 0.3\text{--}0.4$ are not dependent on the values *l* and *n*, whereas the authors’ formulae considers the partial sums as a normal distributed random values.

Standard intervals based on $\pm\sigma$ can be used for b_{\min} and b_{\max} . Then, the number of correct reconstructed bits as $n \left(\sum_{i=0}^n \text{Pr}_{\text{err}}(i)\right)$ is computed. Thus, using Eqs. (1) to (3), one may compute the probability of erroneous reconstruction, which depends on the number of transmitted coefficients *l*, the relationship between *l* and the length of the vector *n*, and the coefficients values (i. e., the value of the sum $\sum_{j=0}^l |c_j|$).

Thereby, when the number of coefficients is about 25%–30% of *n*, as in Fig. 1a, the histograms of the truncated IWHT sums disintegrate into a mixture of Gaussian distributions that is the result of the clipping effect mentioned above.

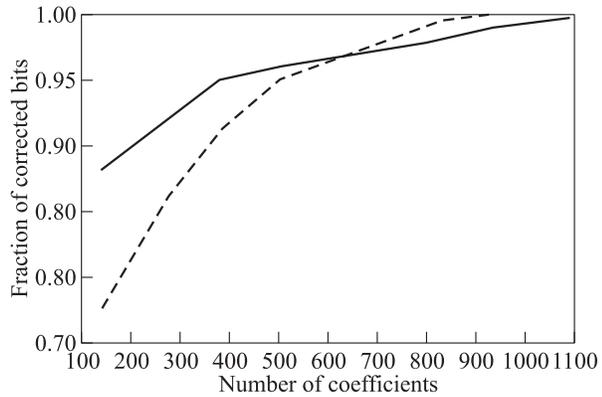


Figure 5 Fractions of correctly reconstructed bits (for a 2K bit file) vs. number of coefficients: solid curve — theoretical (Eqs. (1)–(3)) and dashed curve — experimental results

5 Comparison with Rate Distortion Curve

In order to estimate a general measure (criterion) of the performance of the method, the theoretic Rate-Distortion curve is considered:

$$R = 1 - h_b(D)$$

where *R* is the rate; $h_b(D) = -D \log_2(D) - (1 - D) \times \log_2(1 - D)$; and *D* is the fraction of erroneous bits.

The value *R* is the minimal rate of transmission which provides transmission of a message of size *n* bits. Thus, this equation is considered as a performance measure for the data transmission (coding) systems, based on reduction of spectral coefficients number transmitted to a receiver (or stored in memory) as this equation, in fact, connects

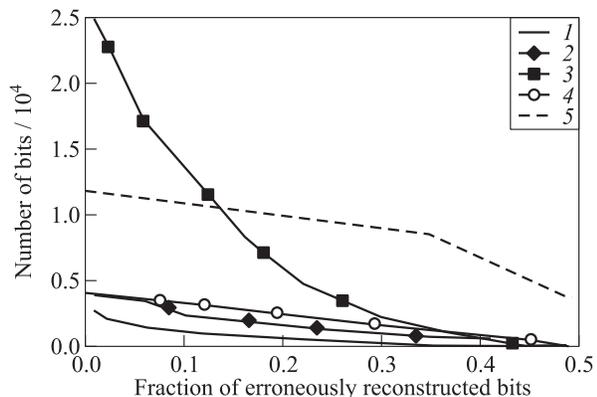


Figure 6 Comparison to the rate distortion curve (4K file): 1 — Shannon; 2 — largest coefficients, then 128, 512, 1024, 2048, 3500 blocks; 3 — largest coefficients only; 4 — blocks only; and 5 — JPEG

the number of correctly reconstructed bits and resources needed for receiving appropriate information by a user, which, in fact, allows to integrate the reconstruction accuracy metric and size of code (number of the code bits).

As it can be seen in Fig. 6, the usage of the l largest coefficients (the curve “largest coefficients only”) is not effective for rather suitable reconstruction error probability. Below, it will be shown how to improve the performance by adding some additional block (portions) of the xored randomly permuted file.

6 Enhancing Performance of Holographic Coding

In order to improve the previously considered algorithm which deals with the l first largest Walsh–Hadamard coefficients of the xored original file, the present authors suggest an algorithm based on both transmission of the first large coefficients and sending actual blocks of the xored and randomly permuted (original input) file. A block of xored and permuted bits of the original file allows recomputing (the conditional, in fact) the probability to improve the correct reconstruction of randomly chosen bit of the file. At the sender, the l largest coefficients are chosen, and both the probability of correct reconstruction $1 - D$ and their rate (number of bits in the coefficient transmitted) are estimated (computed) taking into account that the coefficients indexes must be known in order to compute the \tilde{b}_i by the IWHT. In this case, the number of R bits is the size of $(C_l) + \text{index}(C_l)$, which is the sum of bits spent to represent the coefficients and their indexes.

The usage of the largest coefficients implies a certain fraction of correctly reconstructed bits $1 - D$. Then, if the receiver receives $k < n$ actual bits (the sender transmit a k -block of original file of size n bits), the probability to reconstruct correctly the whole file is $(n - (n - k)D)/n$. Then, the Walsh–Hadamard coefficients may be recomputed (at the receiver part of the channel) for the left block (of size $n - k$ bits) as the bits of the sent portion can be extracted from the Walsh–Hadamard coefficients equation (see section 3). Note that thanks to the permutation, the bits of the sent block from the beginning of the (xored and permuted) file may correspond to arbitrary indexes of original file, that is, the independence of the reconstruction process of indexes is not violated and the blurriness property of “holographic” reconstruction is obtained.

Proposition 1. Let $C = (C_1, C_2, \dots, C_n)$, where n is the size of the original file, be the set of all n WHT coefficients of the binary vector, modeling xor-permuted file obtained from the original file (as mentioned above).

Eliminate from the whole vector a subset of the components P of size k bits. Then, one can reconstruct the bits of the rest of the vector using the following modification of Walsh–Hadamard coefficients recomputed as

$$C_i^P = C_i - \sum_{j \in P} b_j W(j, i)$$

and the rest of the bits of the vector (rest part of the file modeled by this vector) $b \setminus P$, where \setminus is the set difference, may be reconstructed as:

$$\tilde{b}_i = C_i - \sum_{j \in I/P} b_j W(j, i)$$

where I is set of the b vector components (bits) indexes, $I = \overline{1, n}$. Indeed, if the bits of P are replaced by zeros, one gets $n = 2^k$ -bit vector for which a WHT may be defined, but it is obvious that

$$C_i^P = C_i - \sum_{j \in P} b_j W(j, i) = \sum_{j \in I/P} b_j W(j, i)$$

if all items corresponding to indexes of P are zeros.

Proposition 2. Let the sender transmit a k -block ($k < n$) of the original input file and D is the previous error. Then, the probability to reconstruct correctly the entire file is $(n - (n - k)D)/n$.

The receiver does the following steps, upon receiving the identities of the (already stored and mutually agreed) chosen pseudorandom sequence:

- (a) follows receiving the first coefficient and reconstructing the entire xored permuted file, by estimating the values of all missing coefficients to be 0;
- (b) the xored-permuted file is reversed permuted to obtain the xored-only bit vector and then, in the last step, the receiver xors this vector with the original pseudorandom vector PN obtaining the original vector (see Fig. 1). Later upon receiving a signal for stopping coefficient usage and starting blocks of k bits, the receiver reconstruct the file by the coefficient and replace the first k bits and updates the already received coefficient by eliminating the contribution of the received blocks to these coefficients; and
- (c) continue according to the procedure and flags sent by the sender.

The process may stop by either the sender or the receiver upon reaching a quality criteria, for example, appropriate value of R , or/and, for example, an appropriate degree of blurriness of a picture transmitted. Note that since there are used the already sent coefficients, updating them to be a function of only the rest of the file, they become more and more accurate.

Let illustrate this approach by a simple example. Assume the original file is of 8 bits: 01001110. Let perform the following steps: (a) choose a pseudorandom number of 8 bits, say, 10010110; (b) then xor the original file bits with the pseudorandom number to get 01100101; (c) choose a random permutation, say, 8, 4, 2, 7, 5, 3, 6, 1, and permute the result of the previous step accordingly to get 10100110; (d) compute the Walsh–Hadamard coefficients which are (see Eq. (1)): $(4/8, 2/8, 0, 2/8, 0, 2/8, 0, 2/8)$; and (e) send the largest coefficient $C_1 = 4$ and check the result of reconstruction comparing with the desired 10100110. In this case, in accordance with the equation above, one gets all components $\tilde{b}_i = 4/8$, $i = (1, 2, \dots, 8)$, that is, the reconstructed vector is $(1, 1, 1, 1, 1, 1, 1, 1)$, that is, error rate $1 - D = 1/2$ (note that this rate can be achieved just with xor-permuted random file). Therefore, one should continue by sending the next largest coefficient or sending an actual block of 10100110. The next largest coefficient $2/8$ provides \tilde{b}_i vector $(6/8, 2/8, 6/8, 2/8, 6/8, 2/8, 6/8, 2/8)$, that is, the reconstructed file is $(1, 0, 1, 0, 1, 0, 1, 0)$ for which the Hamming distance from the WHT-file is 2, and the error rate is $1 - D = 1 - 6/8 = 1/4$.

Let consider how the reconstruction accuracy will change if a block consisting of the first 2 bits of the xored-permuted file is sent after sending the coefficient C_1 and the first largest coefficient C_1^P is recomputed by the formula of Proposition 1 for the rest of file, that is, $P = (01)$. In this case, the rest six reconstructed bits (in accordance with formula 8) $\tilde{b}_i = 1$, and as the first two bits of original xor-permuted vector are known, one obtains its reconstruction as 10111111. In this case, $1 - D = 5/8$. It is worse than in the case of using the 2 first largest coefficients, but, on the other hand, it requires essentially less bits to send the data, that is, 8 bits (2 bit for the 2-bit portion and $3 + 3$ bits for the coefficient C_1^P and its index) instead of 12 bits in the previous case. Therefore, it is necessary to define the trade-off between the factor of code rate and the accuracy as a performance measure.

Having the Shannon performance measure mentioned above, it is possible to guide the process of the coefficients/blocks sending by examining which of the two implies better revealing probability and act accordingly. Assume the block (portion of bits P) is found to be more efficient, then a flag of changing transmission to blocks is sent to the receiver and then the first block (say, 10) is sent, then continue as if the xored permuted file is, in fact, the rest of the file 100110 rather than 10100110 and the values of the coefficient(s) sent so far are recalculated (subtracting the related information sent in blocks according to the formula mentioned above).

As it can be seen in Fig. 6, the usage of the l largest coefficients (curve 3) is not effective for rather suitable reconstruction error probability, but adding some

additional block portions of the xored randomly permuted file can reduce the error probability dramatically, what shows the gap between the curves 3 and 2. The point of the curves divergence corresponds to starting of blocks sending. Curve 4 shows that up to very high values of accuracy (about 85%), the present approach is more effective than JPEG. Note that JPEG does not limit the operations to addition and does not achieve a *blurred* image during sequential transmission, providing the suitable quality of the information requested by different devices. JPEG is not holographic coding in the sense that a lost part of the JPEG file can lead to destruction of the decoded original file rather than blurriness. Note that the method suggested herein sums up to diverging from the distortion curve of the “largest coefficients first” compression method and continues in a straight line representing uncoded transmission. As the original curve is convex, the optimal point to diverge is simply the point where the resulting tangent is the lower.

7 Discussion and Concluding Remarks

The probabilistic features of the recently suggested holographic WHT-based encoding of binary sequences [2, 6], generated from an original binary data set has been studied. These sequences (vectors) are random ones thanks to a pseudorandom transformation of the original data by bitwise xoring with a predefined pseudorandom sequence, and this generated pseudorandom with almost uniform Walsh–Hadamard spectrum, used to ensure symmetric encoding of the original data. The goal of this research was to study the possibility of error in the data reconstruction using the IWHT truncated by the first l largest coefficients. It is important that the result of the computations of the truncated IWHT are performed in the real domain whereas the estimation metric is formed for the Hamming distance of $d(x; y) = |b - \hat{b}|$ that is an object from $\text{GF}(2)$.

It was shown that this effect can be modeled by the clipping-like nonlinear transformation of the partial (truncated) inverse Walsh–Hadamard sums. It was shown that for the computation of rather accurate approximation of the probability, it is enough to estimate only marginal distribution of the partial sums, that is, the bits estimations \tilde{b}_i . Still the estimation of the marginal distribution for rather small values of l/n remains very actual for future studies. Thanks to the symmetry property of the spectral Walsh–Hadamard coefficients, they have actual holographic properties, namely, any portion of this set defines a *blurred* image of the original file. *Blurred* information is received which becomes less and less *blurred* and more definite during a sequential trans-

mission of the encoded data. Therefore, the definition of a measure for the blurriness is of independent importance. Along with this characteristics, the authors consider the performance of this method using the shape of Rate Distortion Curve (Shannon bound). Using this performance measure, an enhancement for the authors' previous WHT-based holographic encoding method is suggested. This enhancement is provided by adding both a permutation of the xored file and a recomputation of the Walsh–Hadamard coefficients during the coefficients transmission.

Note that there can be an ambiguity in the definition of the choice of l largest over absolute values coefficients if there exist a pair of coefficients c_i, c_j , such that $\text{abs}(c_i) = \text{abs}(c_j)$. Indeed, there is a question in this case, which of the two coefficients should be included in the l -set. It is possible to use an identification of all WHT coefficients indexes that contribute significantly to the binary sequences energy. Following [7], there is a polynomial-complex algorithm for choosing the coefficients.

Thus, in both ways, the coefficients are kept in accordance with the mean square (energy) criterion, but an L_1 metric is used in order to approve them. Further, if the ordered WHT coefficients are considered as an order statistic, the problem of the original file reconstruction by $l \ll n$ coefficients can be reduced to a known method of a linear combination distribution estimation [8].

References

1. *Bruckstein A. M., Holt R. J., Netravali A. N.* Holographic representations of images // *IEEE Trans. Image Processing*, 1998. Vol. 7 (May–June). P. 1583–1597.
2. *Jung H. Y., Prost R., Choi T. Y.* A unified mathematical form of the Walsh-Hadamard transform for lossless image data compression // *Signal Processing*, 1997. Vol. 63. P. 35–43.
3. *Dolev S., Frenkel S.* A way of coding and decoding of digital data based on digital holography principles. Patent of Russian Federation No. 2010145892/08(066164) of 11.11.2010.
4. *Morettin P. A.* Walsh spectral analysis // *SIAM Review*, 1981. Vol. 23. P. 277–291.
5. *Wiens, D. P., Beaulieu N. C., Loskot P.* On the exact distribution of the sum of the largest $n - k$ out of n normal random variables with differing mean values // *Statistics*, 2006. Vol. 40, No. 2. P. 165–173.
6. *Dolev S., Frenkel S.* Multiplication free holographic coding // *IEEE 26th Convention of Electrical and Electronics Engineers Proceedings*. — Eilat, Israel, 2010.
7. *Gilbert A. C., Guhay S., Indykz P., Muthukrishnan S., Strauss M.* Near optimal sparse Fourier representations via sampling // *STOC 2002*. — Montreal, Quebec, Canada, 2002.
8. *Arellano-Valle R., Genton A.* On the exact distribution of linear combinations of order statistics from dependent random variables // *J. Multivariate Anal.*, 2007. Vol. 98. P. 1876–1894.

ГОЛОГРАФИЧЕСКОЕ КОДИРОВАНИЕ НА ОСНОВЕ ПРЕОБРАЗОВАНИЯ УОЛША–АДАМАРА РАНДОМИЗИРОВАННЫХ И ПЕРЕМЕШАННЫХ ДАННЫХ

Ш. Долев¹, С. Френкель², А. Коен³

¹Университет им. Бен-Гуриона в Негеве, Беэр-Шева, Израиль, dolev@cs.bgu.ac.il

²ИПИ РАН; Московский государственный технический университет радиотехники, электроники и автоматики, fdergei@mail.ru

³Университет им. Бен-Гуриона в Негеве, Беэр-Шева, Израиль, coasaf@cse.bgu.ac.il

Аннотация: В статье содержатся результаты статистического и вероятностного анализа способа кодирования и сжатия данных, основанного на перемешивании и рандомизации двоичных данных, с последующим преобразованием Уолша–Адамара, и на основе выполненного анализа предлагается метод улучшения производительности данного подхода.

Ключевые слова: голографическое кодирование; преобразование Уолша-Адамара; граница Шеннона

MATHEMATICAL FOUNDATION, APPLICATION, AND COMPARISON OF GENERAL DATA ASSIMILATION METHOD BASED ON DIFFUSION APPROXIMATION WITH OTHER DATA ASSIMILATION SCHEMES

K. P. Belyaev¹, C. A. S. Tanajura², and N. P. Tuchkova³

Аннотация: Data assimilation methods commonly used in numerical ocean and atmospheric circulation models for weather and climate prediction produce approximations of state variables in terms of stochastic processes. This approximation consists of random sequences of Markov chains, which converge to a diffusion-type process. The conditions for this convergence are investigated. The optimization problem associated with the search of the best possible approximation of the state variable and the results of a numerical experiment are discussed. It is shown that the data assimilation method can be used in practical applications in meteorology and oceanography. Several applications of the methods as an example of the modern operational data processing system with the ocean circulation model HYCOM and data from ARGO drifters are performed and the results as well as comparisons with other assimilation schemes are presented.

Ключевые слова: sequence of Markov chains; diffusion stochastic process; data assimilation methods; HYCOM; ARGO drifters

1 Introduction

In numerical modeling of geophysical systems, such as an ocean, an atmosphere, or a climate system, the data assimilation is a common and popular approach. It produces the initial conditions for weather and climate models and complements monitoring by correcting model variables in the direction of observations. Therefore, data assimilation methods have a substantial impact in weather forecast modeling and, consequently, in several human activities that rely on weather and climate forecasts, such as agriculture, water resources, and others.

Actually, the observed data such as the sea level, the ocean temperature, the salinity, and other tracers are collected from different sources including satellite measurements, merchant ships reports, specially designed ocean moorings and drifters, scientific expeditions reports, and so on. The processing of such enormous number of observations in the on-line regime requires to develop both the technology of data transfer and methods of their analysis. It implies a further advance in information technologies and their applications, parallel computations, mathematical methods of data analysis, etc.

In practice, the data inflow from all sources are subject to the data quality control which filters out poorly known and/or wrong observations, their preprocess analysis, further transfer to computers, and then their parallel calculations. This complex information system requires reliable and fast data assimilation methods in conjunction with circulation models.

The correction of the model output by observational data is generally based on a scheme of the following type: a system of partial differential equations usually represented in the finite-difference or finite-element form is considered within a time interval (t_0, T) . In general, t_0 can be associated with 0, while T can be considered as infinity provided it is large enough. The interval (t_0, T) can be divided into intervals $(t_0, t_1), (t_1, t_2), \dots, (t_n, t_n + 1)$. Within any time interval $(t_n, t_n + 1)$, the model starts at the moment t_n with the initial state-vector Θ_n , integrates forward until the moment t_{n+1} , when it produces (predicts) the state-vector Θ_{n+1}^m . Here and further, the superscript m indicates that the system state has been obtained only by the model integration without any other source of information. During the time interval $(t_n, t_n + 1)$, a series of observations enters, represented by the vectors ξ_1^n, \dots, ξ_k^n independent of the model, where the index k denotes the vector within the n th time

* Работа выполнена при частичной поддержке РФФИ (гранты 10-07-00487-а и 12-07-00108) и Рособразования (проект 020619-1-174).

¹Shirshov Institute of Oceanology, Russian Academy of Science, Moscow, Russia, kb@sail.msk.ru

²Federal University of Bahia, Salvador, Brazil, cast@ufba.br

³Institution of the Russian Academy of Sciences Dorodnicyn Computing Center of RAS Moscow, Russia, tuchkova@ccas.ru

interval. Usually, the vector ξ_k^n is a subset of the state vector Θ_n , since only a part of the model state can be observed. Then, the model output Θ_{n+1}^m is corrected by observations according to

$$\Theta_{n+1} = \Theta_{n+1}^m + \sum_{n=1}^k \alpha_i (\xi_i^n - \xi_i^m).$$

Here, ξ_i^m denotes the model variables, corresponding to observations, calculated at the observation time frame. The weight functions α_i also depend on time and may be either known *a priori* or determined by some appropriate algorithm. The corrected state vector Θ_{n+1} , the so-called objective analysis, is taken as the new initial condition, and the model integration resumes.

Under the name of data assimilation methods, various versions of this scheme are commonly used in geophysics. In particular, the Kalman filter approach determines the optimal weight-coefficient utilizing the statistical properties of observational data [1, 2]. Alternatively, the observations may be not considered as random. In this case, the optimal weights can be determined according to the variational or adjoint data assimilation technique [3]. The integration intervals can be considered as given, e. g., the model produces forecasts every 24 h or at random. For instance, in modern coupled ocean-atmosphere models the correction may be applied as soon as the model temperature difference between the ocean and the atmosphere exceeds a certain *a priori* chosen limit.

Despite these and other differences, all these methods in essence follow the scheme outlined in the above. Alternatively, it could be interesting to review these techniques from a different point of view. At the “moment of assimilations,” the time series of variables Θ_n undergoes a jump of its trajectories. What will happen if the interval between consecutive assimilations approaches to zero along with the values of the jump? How does the limiting behavior of the trajectories depend on the number of state variables and their distributions? Under which conditions does the limiting distribution of Θ_n exist as T goes to infinity and what is it? If it exists, it is called a stationary distribution. These and similar questions attract interest not only from the theoretical point of view, but also for quite practical reasons. For instance, knowing the limiting behavior of the time series for $\Delta t_k = t_{k+1} - t_k \rightarrow 0$, it becomes easy to calculate various parameters needed for the weather forecast, while the knowledge of stationary distribution enhances the reliability of a climate prediction. In addition, the knowledge of this limit can simplify the optimization problem for weight coefficients, which generally are the extremum of a given function, e. g., the error variance.

One of the aims of the present paper is to prove that under appropriate conditions, the trajectories of

the objective analysis Θ_n as a function of time converge to the trajectories of the stochastic diffusion process. These are continuous functions satisfying the Fokker–Planck (FP) equation. Their characteristics provide a tool for the determination of properties of the limiting trajectories, such as their maxima. Furthermore, the optimization problem of the best weight coefficients satisfying the unbiased and minimum variance estimator is solved. Finally, to illustrate the feasibility and usefulness of this method, several numerical experiments are performed. The main idea of this study is based on the classical theorem of convergence of Markov chains to diffusion process [4], and its recent generalizations [5]. It is a continuation of the works by the authors [6] and it generalizes some results previously obtained in [7].

2 Main Definitions and Notations

Let the system of equations

$$\frac{\partial \Theta(t)}{\partial t} = \Lambda(\Theta, t) \quad (1)$$

be considered on the time interval (t_0, T) . Without loss of generality, t_0 in further references will be associated with 0 while T may be both finite and infinite. In (1), $\Theta(t)$ represents the random state-vector of dimension r defined on a given probability space, $\Lambda(x, t)$ denotes nonrandom generally nonlinear operator acting in R^r , which does not explicitly involve temporal derivatives, and η denotes the random variable with zero average and finite covariance function. Symbol $\{ ' \}$ denotes a transpose of a corresponding vector and/or matrix, the symbols $| |$ or $|| |$ represent a vector or a matrix norm, respectively. A sequence of time series is considered such that in each series the interval $(0, T)$ is divided by time points $(0 = t_{1,n}, t_{2,n}, \dots, t_{k,n}, \dots)$, where the first index denotes the order number of the corresponding point while the second index refers to the time series.

It is supposed that in each series within the interval $\Delta t_{k,n} = t_{k+1,n} - t_{k,n}$, a number of random vectors $(\xi_1^n, \dots, \xi_l^n)$, $l = \overline{1, \nu^n}$, with dimension q , $q \leq r$, are observed, where ν^n is also a random integer variable with given distribution $p_l^n = P(\nu^n = l)$ independent of the vectors ξ_l^n . Knowing the solution of the system (1), $\Theta_k^{n,m}(t)$ for initial vector $\Theta_k^n(t)$ within entire interval $\Delta t_{k,n}$, the observed variables ξ_l^n , as well as a realization of the random index ν^n , the newly constrained variables are introduced by the formula:

$$\zeta_k^n = \sum_{l=0}^{\nu^n} \alpha_{l,n} (\xi_l^n - \Theta_l^{n,m}). \quad (2)$$

In (2), the matrices $\alpha_{l,n}$ with dimension $r \times r$ referred to as weight coefficients are supposed to be known and depend on the time series.

Remark. The matrices $\alpha_{l,n}$ in (2) are supposed to be known, but arbitrary enough. Their specific determination with respect to some criteria is a matter of another problem.

For the consistency of (2), it is necessary that in (2) the vectors ξ_l^n and $\Theta_l^{n,m}(t)$ have the same dimension q which may differ from the dimension of the vector $\Theta(t)$ of the dimension r . However, without loss of a generality, it will be assumed that ζ_k^n has the same dimension as Θ_k^n setting the “dummy” components of vectors ξ_l^n and $\Theta_l^{n,m}(t)$ to zero. Ultimately, the new state variables Θ_n^{k+1} are defined by formula

$$\Theta_n^{k+1} = \Theta_n^k + \int_{t_k}^{t_{k+1}} \Lambda(\Theta_n^k, \tau) d\tau + \zeta_k^n \quad (3)$$

and Θ_n^{k+1} is taken as the initial conditions in (1) for the continuation of the integrations.

In this manner, it becomes possible to obtain the sequence of trajectories $\Theta^n(t)$ defined over the entire interval $(0, T)$. Starting from some known random vector $\Theta_0^n(t)$, the solution of (1) for each interval $\Delta t_{k,n} = t_{k+1,n} - t_{k,n}$ with breaks at moments $t_{k,n}$ can be evaluated. The goal of the present paper is to determine the limiting behavior of the solution (3) when $n \rightarrow \infty$.

3 Formulations

The following notions are introduced below.

A0. The time lattice $\Delta t_{k,n} = t_{k+1,n} - t_{k,n}$ is considered as nonequidistant, with real values in each series.

A1. The intervals $\Delta t_{k,n} \rightarrow 0$ approach zero uniformly with respect to k , i. e., $\max_k \Delta t_{k,n} \rightarrow 0, n \rightarrow \infty$.

A2. The probability distribution for random variables ν^n satisfies the conditions: $p_{l,n} = P(\nu^n = l) = p_l \Delta t_{k,n} + o(\Delta t_{k,n}), l > 0$, for any k , and $\mu = \sum_{l=1}^{\infty} l p_l < \infty$.

A3. Random vectors $\xi_1^n, \dots, \xi_l^n, \dots$ have $2 + \delta$ moments for positive δ for each n , i. e., $E \xi_i^n = \lambda_{i1}^n, E(\xi_i^n \xi_j^n) = \gamma_{ij}^n, E|\xi_i^n|^{2+\delta} < \infty$, and these variables are uniformly bounded with respect to n , i. e., $\overline{\lim} |\lambda_{i1}^n| < \infty, \overline{\lim} |\gamma_{ij}^n| < \infty, \gamma_{ii}^n = \gamma_{ii}^n, i, j = 1, 2, \dots$

A4. The operator $\Lambda(x, t)$ is a continuous function of its arguments.

A5. The set of weight coefficients $\alpha_{l,n}$ is uniformly bounded with respect to n , i. e., $\overline{\lim} \|\alpha_{l,n}\| < \infty, l = 1, 2, \dots$

A6. The sequence of distributions of random variables θ_0^n converges to the distribution of some random variable θ_0 , i. e., $P(\theta_0^n < x) \rightarrow P(\theta_0 < x), n \rightarrow \infty$, for each x .

Without loss of generality, the limit values of variables $\lambda_{i1}^n, \gamma_{ij}^n$, and $\alpha_{l,n}$ when n tends to infinity are supposed to exist and to be equal to $\lambda_{i1}, \gamma_{ij}$, and α_l , respectively. Otherwise, the corresponding subsequence can be chosen to provide the convergence to the limit points, which exist as a consequence of the conditions A3 and A5.

Theorem 1. *Let the conditions A0–A6 hold. Then, the sequence of finite-dimensional distributions of random processes $\theta^n(t)$ converges to the stochastic process, which will be a solution of the stochastic differential equation*

$$\theta(t) = \theta_0 + \int_0^t a(s, \theta(s)) ds + \int_0^t b(s, \theta(s)) dw(s)$$

where

$$a(t, x) = \sum_{l=1}^{\infty} p_l \left[\sum_{j=1}^l \alpha_j (\lambda_{j1} - x) \right] + \Lambda(x, t); \quad (4)$$

$$b^2(t, x) = \sum_{m=1}^{\infty} p_m \left\{ \sum_{i,j=1}^{m(m+1)} \alpha_i [\gamma_{ij} - (x \lambda'_{1,i} + \lambda_{1,j} x') + x x'] \alpha'_j \right\}. \quad (5)$$

The Wiener process $w(t)$ is defined on interval $(0, T)$ and it is independent of the random variable θ_0 .

Proof. The sequence of random variables $\theta_k^n, k = 0, 1, \dots$, forms the Markov chain for each series n . Hence, the general statements about the convergence of Markov chains can be applied. The proof of theorem is based on the results [4], which have been generalized in [5]. To examine the conditions of convergence given in [4, 5], the following statements have to be proved:

I There are vectors $a_k(x)$ and matrices $B_k(x)$ such that

$$\left| \frac{1}{\Delta t_{k,n}} \int_{-\infty}^{\infty} (y - x) dp_{k,n} \left(\frac{y}{x} \right) - a_k(x) \right| \rightarrow 0;$$

$$\left\| \frac{1}{\Delta t_{k,n}} \int_{-\infty}^{\infty} (y - x)(y - x)' dp_{k,n} \left(\frac{y}{x} \right) - B_k(x) \right\| \rightarrow 0$$

for each k , when $n \rightarrow \infty$.

Here and further, $p_{k,n}(y/x)$ denotes the transitional probabilities of Markov chains in each time series, i. e., $P(\theta_{k+1}^n = y / \theta_k^n = x) = p_{k,n}(y/x)$.

This equality is correct for both discrete and absolutely continuous random variables ξ_l^n and $\theta_k^n(t)$. In the latter case, this equality applies for the corresponding probability densities.

II The convergence of sequences

$$E(\theta_{k+1}^n - \theta_k^n)^{2+\delta} \xrightarrow{n \rightarrow \infty} 0;$$

$$P\left(\frac{|\xi_k^n - \xi_{k+p}^n|^{2+\delta}}{\xi_k^n}\right) \rightarrow 0$$

for any k, p , when $n \rightarrow \infty$, has to be shown.

III The sequence $\{|a_k(x)|/|B_k(x)|\}$ is uniformly bounded.

As in [4, 5], these conditions provide the basis of the proof of the theorem. The transitional probabilities $p_{k,n}(y/x)$ are directly calculated as

$$P(\theta_{k+1}^n = y/\theta_k^n = x) = p_n\left(\frac{y}{x}\right)$$

$$= P\left(\theta_k^n + \int_{t_k}^{t_{k+1}} \Lambda(\theta_k^n, \tau) d\tau + \zeta_k^n = y/\theta_k^n = x\right)$$

$$= P\left(\zeta_k^n + x + \int_{t_k}^{t_{k+1}} \Lambda(x, \tau) d\tau = y\right)$$

$$= P\left(\sum_{l=0}^{\nu^n} \alpha_{l,n}(\xi_l^n - \theta_l^{n,m}) = y - x - \int_{t_k}^{t_{k+1}} \Lambda(x, \tau) d\tau\right).$$

Then, this equality may be extended to

$$p_n(y/x) = p\left(\sum_{l=0}^{\nu^n} \alpha_{l,n}\left(\xi_l^n - x - \int_{t_k}^{s_l} \Lambda(x, \tau) d\tau\right)\right)$$

$$= y - x - \int_{t_k}^{t_{k+1}} \Lambda(x, \tau) d\tau \quad (6)$$

where $s_l, s_1 < s_2 < \dots < s_l, l = 0, 1, \dots, \nu^n$, are chronologically ordered moments within the interval (t_k, t_{k+1}) . Using the continuity condition A4, the integrals in (6) can be represented as $\int_{t_k}^{s_l} \Lambda(x, \tau) d\tau = \Lambda(x, \tau_*)(s_l - t_k)$. On account of this relation, one obtains from the conditional probabilities (6):

$$p\left(\sum_{l=0}^{\nu^n} \alpha_{l,n}(\xi_l^n - x)\right)$$

$$= y - x - \Delta t_{k,n} \Lambda(x, \tau_*) + \sum_{l=0}^{\nu^n} (s_l - t_{k,n}) \Lambda(x, \tau_{*,k})$$

where $\tau_* \in (t_k, s_l)$. Using this relation, the conditional average $E(y - x)p(y/x)$ may be written as

$$E(y - x)p_n(y/x) = \int_{-\infty}^{\infty} (y - x) d_y p$$

$$\times \left(\sum_{l=0}^{\nu^n} \alpha_{l,n}(\xi_l^n - x) = y - x - \Delta t_{k,n} \Lambda(x, \tau_*)\right.$$

$$\left. + \sum_{l=0}^{\nu^n} (s_l - t_{k,n}) \Lambda(x, \tau_{*,k})\right)$$

$$= \int_{-\infty}^{\infty} \left(z + \Delta t_{k,n} \Lambda(x, \tau_*) - \sum_{l=0}^{\nu^n} (s_l - t_{k,n}) \Lambda(x, \tau_{*,k})\right)$$

$$\times d_z p\left(\sum_{l=0}^{\nu^n} \alpha_{l,n}(\xi_l^n - x) = z\right)$$

$$= \int_{-\infty}^{\infty} z d_z p\left(\sum_{l=0}^{\nu^n} \alpha_{l,n}(\xi_l^n - x) = z\right) + \Delta t_{k,n} \Lambda(x, \tau_*)$$

$$- \sum_{l=0}^{\nu^n} (s_l - t_{k,n}) \Lambda(x, \tau_{*,k}) \quad (7)$$

Since the last two terms are independent of z , they can be taken out of the integral. The first term on the right-hand side of Eq. (7) is the expectation value of a series of random terms. It may be expanded as

$$E\left(\sum_{l=0}^{\nu^n} \alpha_{l,n}(\xi_l^n - x)\right) = \sum_{m=0}^{\infty} p_{m,n} \sum_{l=0}^m \alpha_{l,n}(\xi_l^n - x)$$

$$= \sum_{m=1}^{\infty} p_{m,n} \sum_{l=1}^m \alpha_{l,n}(\xi_l^n - x),$$

because the term with $m = 0$ vanishes. Inserting this expression into Eq. (7), the equality is obtained:

$$E(y - x)p_n(y/x)$$

$$= \sum_{m=1}^{\infty} p_{m,n} \sum_{l=1}^m \alpha_{l,n}(\xi_l^n - x) + \Delta t_{k,n} \Lambda(x, \tau_*)$$

$$- \sum_{m=1}^{\infty} p_{m,n} \sum_{l=1}^m (s_l - t_{k,n}) \Lambda(x, \tau_{*,k}). \quad (8)$$

Similarly, for conditional variance $E(y - x)(y - x)'p_n(y/x)$, one arrives at the relation (omitting intermediate calculations):

$$E(y - x)(y - x)'p_n(y/x) = \sum_{m=1}^{\infty} p_{m,n}$$

$$\times \left\{ \sum_{i,j=1}^{m(m+1)} \alpha_{i,n} \left[\gamma_{ij}^n - x \lambda_{1,i}^{n'} - \lambda_{1,j}^n x' + x x' \right] \alpha_{j,n}' \right\}$$

$$\begin{aligned}
 &+2 \left(\Delta t_{k,n} \Lambda(x, \tau_{\bullet}) - \sum_{m=1}^{\infty} p_{m,l} \sum_{l=1}^m (s_l - t_{k,n}) \Lambda(x, \tau_{*,k}) \right) \\
 &\quad \times E(y - x) p_n(y/x) + \left(\Delta t_{k,n} \Lambda(x, \tau_{\bullet}) \right. \\
 &\quad \left. - \sum_{m=1}^{\infty} p_{m,l} \sum_{l=1}^m (s_l - t_{k,n}) \Lambda(x, \tau_{*,k}) \right)^2. \quad (9)
 \end{aligned}$$

Now, the last term in right-hand side of (8) can be estimated as

$$\begin{aligned}
 &\sum_{m=1}^{\infty} p_{m,n} \sum_{l=1}^m (s_l - t_{k,n}) \Lambda(x, \tau_{*,k}) \\
 &\leq \sum_{m=1}^{\infty} p_{m,n} \max_{t_k \leq \tau \leq t_{k+1}} |\Lambda(x, \tau)| m (s_m - t_k) \\
 &\leq \Delta t_{k,n} \max_{t_k \leq \tau \leq t_{k+1}} |\Lambda(x, \tau)| (\mu \Delta t_{k,n} + o(\Delta t_{k,n})).
 \end{aligned}$$

Similar, in (9),

$$\begin{aligned}
 &(\Delta t_{k,n} \Lambda(x, \tau_{\bullet}) - \sum_{m=1}^{\infty} p_{m,n} \sum_{l=1}^m (s_l - t_{k,n}) \Lambda(x, \tau_{*,k}))^2 \\
 &\leq (\Delta t_{k,n} \max_{t_k \leq \tau \leq t_{k+1}} |\Lambda(x, \tau)| \mu \Delta t_{k,n})^2.
 \end{aligned}$$

From (8), (9), conditions A2, A3, A4, and the last inequalities, one finally yields when $n \rightarrow \infty$ that

$$\begin{aligned}
 &(\Delta t_{k,n})^{-1} E(y - x) p_n(y/x) \\
 &\quad \rightarrow \sum_{m=1}^{\infty} p_m \sum_{l=1}^m \alpha_l (\xi_l - x) + \Lambda(x, t); \\
 &(\Delta t_{k,n})^{-1} E(y - x)(y - x)' p_n(y/x) \\
 &\quad - \Delta t_{k,n} (E(y - x) p_n(y/x))^2 \\
 &\rightarrow \sum_{m=1}^{\infty} p_m \left\{ \sum_{i,j=1}^{m(m+1)} \alpha_i [\gamma_{ij} - x' \lambda_{1,i} - \lambda_{1,j} x' + x x'] \alpha_j' \right\}.
 \end{aligned}$$

Hence, condition I is satisfied.

For the proof of the entire theorem, it is necessary to verify the convergence conditions II and III. However, it is readily seen that the condition II is satisfied as a consequence of the continuity of the operator and conditions A2 and A3. Finally, the condition III follows from the existence of a uniform bound of weight coefficients (A5) and the nonzero variance, if the distribution of the random index ν is not zero with probability 1. This proves the theorem completely.

Remark. All parameters p , α , λ , and γ depend, in general, on both t and x . To avoid the overloading of the text and to simplify the notation, this dependence is not explicitly shown.

This theorem may be generalized on the case when a random index ν is a vector with different distributions for each component, i. e.,

$$p(\nu = l) = p(\nu_1 = l_1, \dots, \nu_r = l_r).$$

Corollary 1. If no observations are assimilated, $a(t, x) = \Lambda(t, x)$ and $b(t, x) = 0$. Thus, the limiting diffusion process coincides with the initial model state, which should be expected.

Corollary 2. The probability distribution of the trajectory is determined by the Fokker–Planck equation (Kolmogorov’s second equation)

$$\begin{aligned}
 &\frac{\partial p(t, x)}{\partial t} \\
 &= - \frac{\partial(a(t, x)p(t, x))}{\partial x} + \frac{1}{2} \frac{\partial^2(b^2(t, x)p(t, x))}{\partial x^2} \quad (10)
 \end{aligned}$$

with the initial conditions $p(0, x) = p(\theta_0 = x)$ and boundary conditions $p(t, \pm\infty) = 0$.

In Eq. (10), the drift vector $a(t, x)$ and the diffusion matrix $b^2(t, x)$ are given by (4) and (5).

4 Optimization Problem

Above, the convergence of a sequence of random variables was considered with all parameters fixed. However, in practical applications, some parameters may be unknown, but rather be sought according to given criteria. The most relevant physical problem is the determination of optimal weight coefficients α_l . By Theorem 1, the limiting trajectories will be those of a diffusion process with the drift vector $a(t, x)$ and the diffusion matrix $b^2(t, x)$, which determine the mean and its variance. More precisely, the following equalities are valid:

$$\begin{aligned}
 &\frac{\partial}{\partial t} E \left(\frac{\theta(t)}{\theta_0} \right) = \int_{-\infty}^{\infty} a(t, x) d_x p(t, x); \\
 &\frac{\partial}{\partial t} \text{var} \left(\frac{\theta(t)}{\theta_0} \right) = \int_{-\infty}^{\infty} b^2(t, x) d_x p(t, x).
 \end{aligned}$$

These relations render it sensible to investigate the optimization problem under the following constrains.

Let the unknown “true” field θ_t be estimated by observations and by the model from the aforementioned scheme. It is supposed that the unknown field is governed by the equation $d\theta_t = C dt + \eta dW$ with the initial value θ_0 being known. The problem is to define the coefficients so that the variance of the estimator is minimized while the average remains unchanged. This is the well-known problem of constraining the unbiased estimator with minimum variance.

Theorem 2. *Let the conditions of theorem 1 A1–A6 hold. Also, let $\alpha_{l,n}^*$ be the coefficients that provide the minimum norm of the conditional variance $(\Delta t_{k,n})^{-1} \|E[(\theta_k^n) - E\theta_k^n](\theta_k^n - E\theta_k^n)' / \theta_0\|$ while the conditional average $(\Delta t_{k,n})^{-1} E[(\theta_k^n) / \theta_0]$ remains fixed and equals to C in each series. The numerical sequence $\alpha_{l,n}^*$ will then converge to coefficients α_l^* , satisfying the system of equations with the unknown vector $\Phi = (\varphi_1, \dots, \varphi_r)$:*

$$\sum_{j=1}^{\infty} p_j \left(\sum_{l=1}^j \alpha_l^* g_{lj} + \Phi \lambda'_{1,i} - E\xi_j \eta dW \right) = 0 \quad \text{for each } i = 1, \dots, \quad (11)$$

where $g_{ij} = \gamma_{ij} - (x \lambda'_{1,i} - \lambda_{1,j} x') + x x'$, and

$$\sum_{l=1}^{\infty} p_l \left[\sum_{j=1}^l \alpha_j^* (\lambda_{j-1} - x) \right] + \Lambda(x, t) = C. \quad (12)$$

Equation (11) is a matrix equation with unknown matrices α_i , the equality is valid for each element of the corresponding matrices. At any given constant C , the system (11), (12) has a unique solution for α_l^* and Φ if and only if the observations as random vectors are linearly independent of each other, i. e., no observation can be represented as a linear combination of the others.

Remark. Statistically, the constant C measures the model bias with respect to the observations. This bias can be eliminated from the algorithm if it is known and/or previously determined.

Proof. The expression for the extremes is obtained by utilizing the method of Lagrange multipliers. For this, it is necessary to determine the minimum of the following function of α and Φ :

$$F(\alpha, \Phi) = \|E(\theta - E\theta - (\theta_t - E\theta_t))(\theta - E\theta - \theta_t - E\theta_t)'\| + \Phi(E\theta - E\theta_t) \quad (13)$$

where Φ is a r -dimensional unknown vector. The minimum of this function requires a minimum of the sum of the elements in each row of matrices α_i for all $i = 1, \dots$

The scheme of the proof is as follows: (i) inserting into Eq. (13) the explicit expression for θ and θ_t according to (4) and (5); (ii) taking the derivatives of Eq. (13) with respect to α_i and Φ and setting them to zero; (iii) substituting of matrices g_{ij} and g_{ji} because they are identical; and (iv) performing ordinary calculations which lead to Eqs. (11) and (12). The convergence of the extremes is proven along the same line of arguments as in the proof of Theorem 1. Actually, it is only necessary to prove the existence and uniqueness of the solution of (11), (12) if all matrices g_{ij} are linearly independent.

Indeed, system (11), (12) is a linear one. For large m , the residual of the sums $\sum_{i=m}^{\infty} p_i$ go to zero. Thus, it is sufficient to consider only a finite number of equations in (11), (12). Let l equations be considered. The matrix in system (11), (12) has the rang $r^2 l + r$. Namely, it consists of matrices $g_{kn} = g_{nk}$, $k, n = 1, \dots, l$, with symmetrically attached last row and column:

$$A = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1l} & f_1 \\ g_{21} & g_{22} & \cdots & g_{2l} & f_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{l1} & g_{l2} & \cdots & g_{ll} & f_l \\ f_1 & f_2 & \cdots & f_l & 0 \end{pmatrix}.$$

Therefore, the determinant of A does not vanish, if rows or columns of matrix A are linearly independent. The variances form the diagonal of this matrix while covariances are placed outside of the diagonal. Because of linear independence of observations, all rows are linearly independent. This completes the proof of Theorem 2.

Theorem 2 states that at each small time interval, once the initial conditions and observation statistics are known, the optimal weight coefficients are to be determined by formulae (11) and (12). However, in practical applications, it is desirable to define these coefficients once and forever, knowing only the initial values and the temporal properties of observations, such as the forecast of their mean values and covariance. The most relevant physical problem in this case is to minimize the functional T^{-1}

$$\int_0^T \left\| E \left[\frac{(\theta(t) - E\theta)(\theta(t) - E\theta)'}{\theta_0} \right] \right\| dt \quad (14)$$

with conditions $T^{-1} \int_0^T E [(\theta(t)) / \theta_0] dt = C$.

Theorem 3. *Let the same scheme of constraining the trajectories be considered and all the conditions A0–A6 hold. If the limiting average vectors $\lambda_{1,i}$ and covariance matrices γ_{ij} are continuous in time, then, starting from the known vector θ_0 , the optimal trajectories with respect to criteria of Theorem 2 will be given by formulae (11) and (12).*

Proof. After passing the limit under integral, the functional (14) takes the form of (13) where all parameters will be considered as functions of time t . The classical Euler equation for optimal trajectory of functional (14) with respect to $\alpha_i(t)$ coincides with the system (11), (12). The limiting transform under the integral is proven by the continuity of limiting functions and upper bound conditions of functions α_i . Ultimately, the initial conditions $\theta(0) = \theta_0$ simply $\alpha_i(0) = 0$ and, hence, no constants should be added to obtain system (13). This proves Theorem 3.

5 Applications to the Ocean Circulation Model HYCOM and ARGO Drifter Data

In this section, several numerical experiments are conducted as illustrations of the applicability of Theorems 2 and 3. To realize the assimilation scheme, the Hybrid Coordinate Ocean Model (HYCOM) was used. HYCOM solves five prognostic equations: two for the horizontal motion, one for the continuity equation, and two for the thermodynamic conservation that can be the salinity and the potential temperatures [8, 9]. The model is structured with a hybrid vertical coordinate system to solve the prognostic equations associated with the shallow water physics. It uses isopycnic coordinates for the open stratified ocean, which reverts to terrain-following coordinates in shallow coastal regions, and z -level coordinates in the mixed layer and over unstratified ocean regions.

The data from ARGO drifters available from the site <http://www.coriolis.eu.org> were assimilated according to the aforementioned scheme. The data selected for the experiments were those daily assimilated during January 2008 in the Atlantic Ocean. The assimilation technique based on the covariance evolution is rather applicable to each model vertical layer than in a fixed depth. It is consistent to suggest that the real link between the physical variables is propagated mostly along isopycnic layers. This allows exploiting the similarity of physics within the same layer due to the concept of identity of water masses. Therefore, the covariance among physical variables in the same water mass depends mostly on the properties of this water mass and on the distance in isopycnic layer between two considered points. This makes the physical justification of the mathematical formalism described above.

The computations were performed on the IBM cluster "Neptune" by using multiparallel computation scheme. The entire Atlantic was divided into 64 subdomains, and all computations were carried out independently in each domain with a parallel technology from MPI library with the exchange of the current information among processors.

Two functions have to be determined prior to solving the system. These are the bias C and the covariance $g_{ij} = \gamma_{ij} - (x_j E \xi'_i - E \xi_j x'_i) + x_i x'_j$ or simply the matrix G of the covariance of the error between a pair of observed vectors $\xi_i - H_i \theta$ and $\xi_j - H_j \theta$ where $H_i \theta$ and $H_j \theta$ are the projections of the prognostic model state θ onto the observational points $i, j = 1, \dots, N$. The number of observations N has been given at the day of assimilation so that the probability distribution of the random index ν was set up as $P(\nu(t) = N) = 1$, $P(\nu(t) = L \neq N) = 0$. Equations (11) and (12) are

similar to the standard Kalman filter theory, but include explicitly the model bias. If the bias is zero, the covariance g_{ij} will be the error covariance matrix widely used in the Kalman filter theory. Additionally, it is necessary to set up the covariance $E \xi_i \eta dW$ between an arbitrary grid point and observations at the point i .

Three different approaches were used to constrain the matrix of covariance G and the covariance $E \xi_i \eta dW$.

(A) A common way to define this matrix is through the extended Kalman filter approach (EKF), which is a version of the Monte-Carlo scheme [2]. The observations are supposed to follow Eq. (2) with zero average and the variance of the noise part being known. Ensemble experiments are conducted with different initial conditions and the error covariance matrix is estimated from the series of model outputs. This approach requires substantial computational costs and its accuracy increases with the square root of M where M is the number of ensemble members. However, the generation of an adequate ensemble set is a challenge, since for practical applications, only a relatively small number of members may be used, and this may not lead to a good estimation.

In the present work, 10 ensemble members were set to realize the EKF scheme to constrain the covariance. The method of their contraction was the following: the model was initialized with the World Ocean Atlas temperature and salinity fields at rest. It was integrated for 40 years with atmospheric forcing from the Comprehensive Ocean-Atmosphere Data Set (COADS) climatological monthly mean fields available at <http://coads.noaa.gov> as the spin-up run. After that, each January 1 of the last 10 years of the spin-up was used as initial condition to create a 10-member ensemble. Ten members were forced by 6-hour reanalysis data from the U.S. National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction (NOAA/NCEP) with 1 degree horizontal resolution during 13 months, from January 1, 2007 until January 31, 2008.

Knowing the model output for all grid points after 10 independent runs, it is easily to estimate the covariance matrix simply as

$$g_{ij} = \frac{1}{10} \sum_{l=1}^{10} (\theta_i^l - E\theta_i) (\theta_j^l - E\theta_j)'; \quad E\theta_i = \frac{1}{10} \sum_{l=1}^{10} \theta_i^l$$

where θ_i^l is the value of the l th model output at the grid point i , $i = 1, L$, $l = 1, \dots, 10$, and L is the total number of grid points. Since the model temperature and salinity are considered, $r = 2$. The numerical consumption regarding this scheme is very high due to the huge size of this matrix, however this scheme can be realized. The issue of the numerical realization and its cost is out of the scope of this paper. The bias value C was zero for this scheme.

(B) The scheme of the creation of the matrix G stems from the limit approximation as diffusion-type process of the model output θ and, hence, the representation of this matrix through the solution of the FP equation

$$\frac{\partial p(t, u)}{\partial t} = -\frac{\partial(Ap)}{\partial u} + \frac{1}{2} \frac{\partial^2(Bp)}{\partial u^2}; \quad (15)$$

$$g_{ij} = \int_{-\infty}^{\infty} (u_i - E\theta_i)(u_j - E\theta_j)p(t, u_i, u_j) du_i du_j$$

where $p(t, u) = p(t, u_i, u_j)$ is the joint probability density of the error between points i, j , $i, j = 1, \dots, N$; t is the time; and A and B are the drift and diffusion coefficients, respectively. All pairs for temperature–temperature, temperature–salinity, and salinity–salinity are considered. These coefficients are defined through the model output and data as it is done in [6]. Briefly, their calculation scheme is the following: at time moment t , all grid points where temperature and/or salinity values are equaled to $u = (u_i, u_j)$ are marked. Let this value be $L(u)$. Then, at the time moment $t + dt$, among the marked grid points, take all grid points in which the considered variable is equal to $v = (v_i, v_j)$ and let this number be $L(v)$. Then, the ratio $p(v/u) = L(v)/L(u)$ is taken as the estimation of the conditional probability $p(t, v/u) = p(\theta(t + dt) = v/\theta(t) = u)$. Following [4], the drift coefficient is $A(t, u) = (dt)^{-1} \int_{-\infty}^{\infty} (v - u)p(v/u) dv$ and

diffusion matrix is $B(t, u) = (dt)^{-1} \int_{-\infty}^{\infty} (v - u)(v - u)'p(v/u) dv - dtA^2(t, u)$. In the current work, the diffusion coefficient was prescribed and equaled to $B = (1/(N - 1)) \sum_{l,j=1}^N (\xi_l - H_l\theta)(\xi_j - H_j\theta)'$. Once the coefficients are determined, Eq. (15) are solved with the boundary conditions $p(t, u) = 0, u \rightarrow \pm\infty$ and the known initial conditions, $p(0, u) = \delta(u - u^0)$, $\delta(u)$ is the Dirac delta-function. The latter condition means that at the initial moment, the error was known and equaled to u^0 . The bias C at an arbitrary grid point was set up as

$C = (1/\bar{N}) \sum_{l=1}^{\bar{N}} (\xi_l - H_l\theta)$ if the distance between the location of ξ_l and the considered grid point did not exceed the selected cut-off radius and \bar{N} denotes the quantity of observed points within this circle. For remote grid points where the locations of observed points exceeded the cut-off radius, the bias was set to zero.

(C) The matrix G was prescribed and set up as

$$g_{ij} = \sigma^2 \exp(-\lambda d_{ij})$$

where d_{ij} denotes the grid distance between points i and j , i.e., all neighboring points in any horizontal

direction have the distance equaled to 1, and λ is a known dimensionless factor (normally varying between 0.1 and 0.3). The bias was set up as in scheme (B).

6 Results of Experiments and the Comparison of the Data Assimilation Scheme

For illustration of the feasibility of the considered assimilation schemes, Fig. 1 presents the results of the modeling with and without assimilation. This figure shows the snapshot of the sea surface temperature (SST) taken on January 30, 2008 for all four assimilation schemes and for the control run, i.e., a model run without assimilation. It also contains the Reynolds SST analysis from remote sensed and *in situ* data on the same day. Data were retrieved from the site <http://www.nhc.noaa.gov/aboutsst.shtml>. The control run overestimated the temperature in the tropics and in the North Atlantic mid-latitude. This can be seen by the area covered by the 27, 24, and the 21 °C isotherms. Also, the 21 °C isotherm to the west of Southern Africa associated with the Benguela Current does not match observations. All assimilation runs produced substantial reduction of temperature in the tropics and cooling to the west of southern place of Africa around 30°S. This correction makes all analysis closer to the Reynolds SST in many regions. However, the correction in the equatorial western Atlantic and in the Caribbean Sea was intense, and the analyses produced cooler values than Reynolds SST in these regions. The EKF produced the strongest cooling. The FP result is in between the EKF and OI run.

In order to numerically compare the skill of all methods, the time behavior of the variance for each method is

presented in Fig. 2. Let $\sigma = \sqrt{\sum_{i=1}^N (\xi_i - H_i\theta)^2 / (N - 1)}$

be the root mean squared error (RMS) averaged over all observations. Along with the variable σ , two other variables, σ_b and σ_a , are considered where σ_b is the RMS for the one-day forecast error, i.e., the deviation is considered between the one-day forecast run made after correction and the data at this moment; and σ_a is the RMS for the analysis error, i.e., the deviation is considered after correction at the instant of the correction. These two variables are calculated for all assimilation methods independently and for each day from January 2 until January 30, 2008.

Figures 2a–2c show the time behavior of σ , σ_b , and σ_a for the temperature according to the EKF, the OI, and the FP schemes, respectively. The EKF forecast error in the beginning of the assimilation run is larger

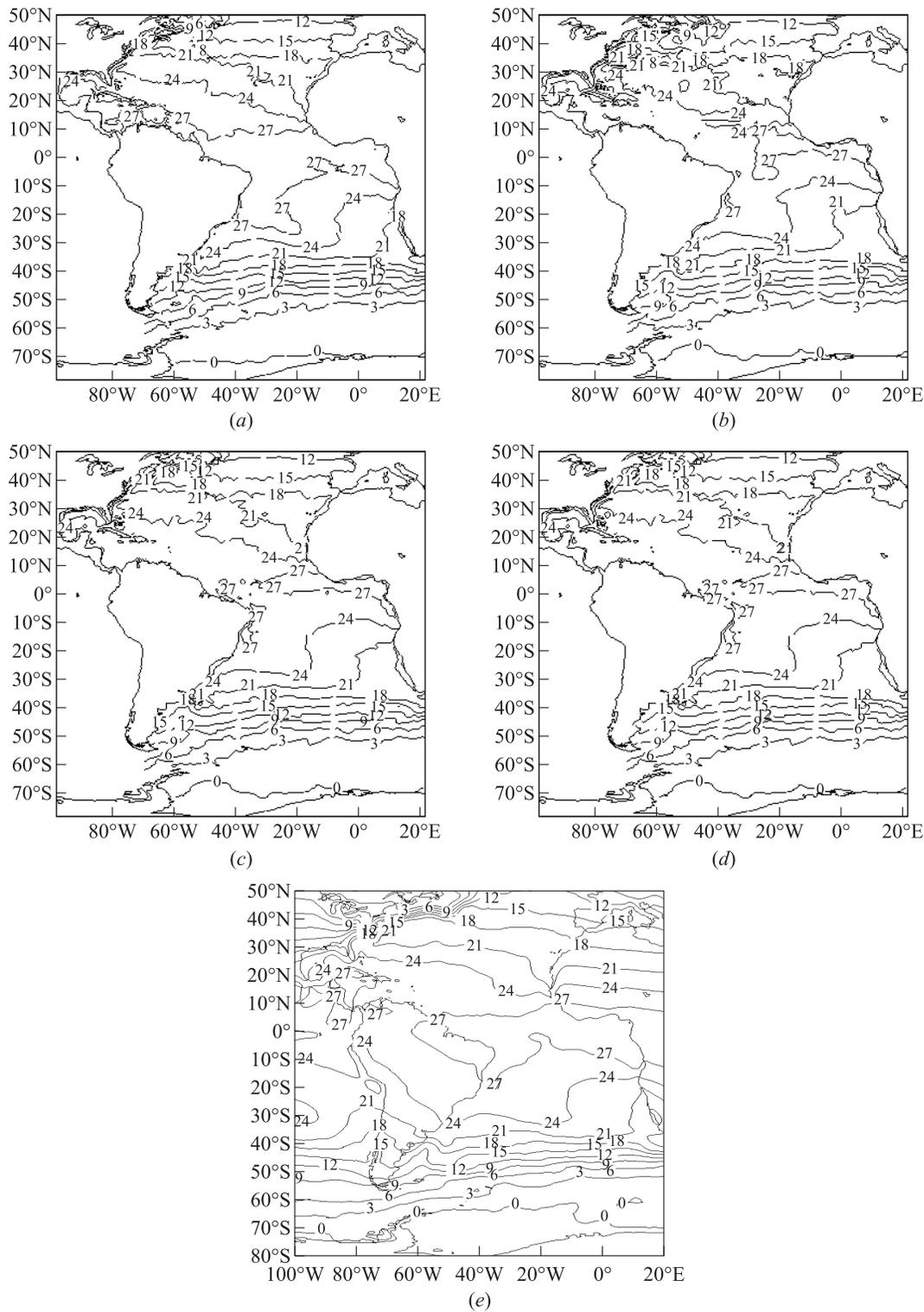


Figure 1 Sea surface temperature (in °C) modeled by HYCOM with and without correction on January 30, 2008 and SST observed by satellite on the same day: (a) control; (b) OI; (c) EKF; (d) FP; and (e) Reynolds SST

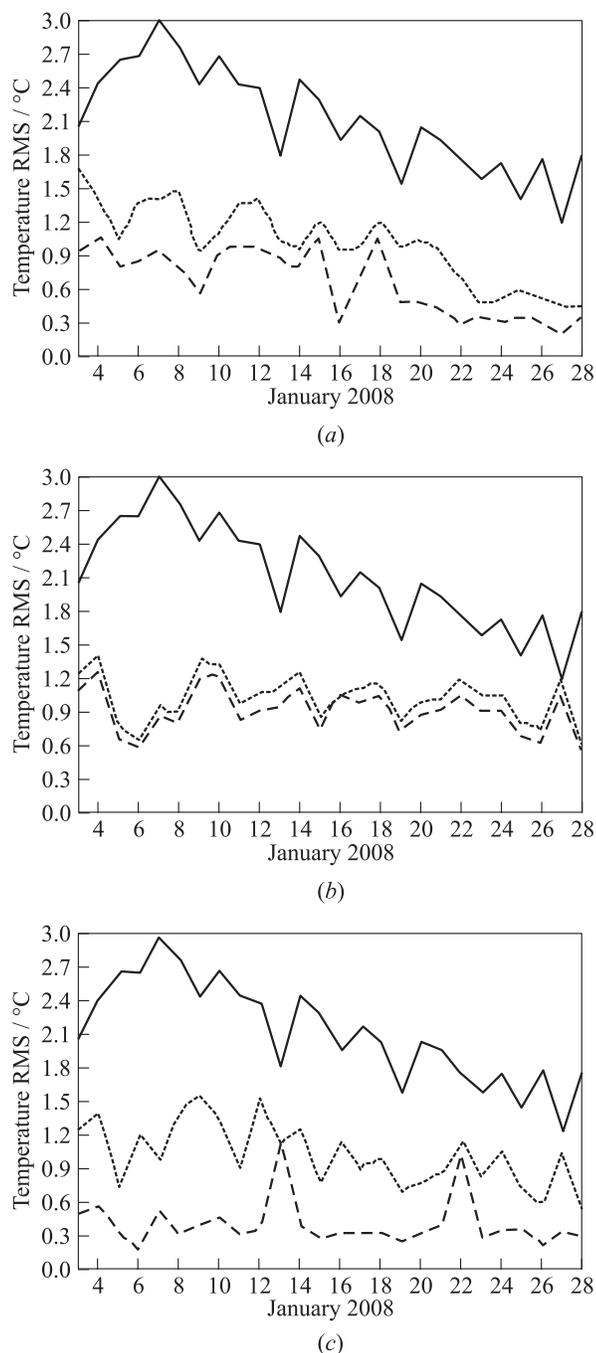


Figure 2 Root mean squared error of temperature over the period January 2–30, 2008 for all assimilation schemes: solid curves — control run; dotted curves — the one-day-forecast error, and dashed curves — the analysis error: (a) OI; (b) EKF; and (c) FP

than the error of the control run. In the second day, the one-day forecast error is already smaller than the control run error, and there is a substantial reduction of the forecast error in time from around 1.5°C in the first day to 0.8°C in the end of the month. The EKF analysis errors have similar behavior, starting around

1.0°C and reaching a minimum smaller than 0.3°C few days before the end of the experiment. The OI forecast errors (see Fig. 2b) oscillated in the beginning of the experiment between 0.6 and 1.4°C . However, it does not decrease in time, so that the forecast error by the end of the experiment is of the same magnitude as in the beginning. The OI analysis error has almost exactly the same variability of its forecast errors, but with values about 0.1°C lower along the whole period. The FP forecast errors (see Fig. 2c) oscillated from 0.7 to 1.5°C in the first 15 days and then it went down to lower values by the end of the experiment. The FP analysis error is the smallest among the three assimilation runs, despite the three pronounced spikes in the days 13, 22, and 29. Without considering these spikes, it oscillated around 0.3°C .

7 Discussion and Concluding Remarks

The presented results are of both theoretical and practical interest. From the theoretical point of view, it is important to know when and under which conditions the solution of differential equations (1) along with the correction (2) may be approximated by a continuous function and how the probability of limiting function could be calculated. In practical applications, the scheme can be utilized to determine a variety of relevant parameters such as the probability of extremes, the probability for crossing some levels, and similar issues. Also, the optimization scheme generalizes the conventional schemes in the assimilation problem; besides, it includes the bias and random quantity of observations.

Nevertheless, several issues still have to be addressed specifically. The system of Eqs. (10) solves the assimilation problem ultimately and uniquely, once the observational statistics, averaged values, covariance, and their time evolution are known. However, in practice, it is not an easy task. Data insufficiency, irregularity of measurements in space and time as well as their inhomogeneous spatial-temporal distribution often make corresponding estimations highly unreliable.

The application of proposed methods was presented in [10] in a climate research with a couple model EGMAM, and this application demonstrated the ability and fruitfulness of the used approach.

Acknowledgments

This work was financially supported by PETROBRAS and Agencia Nacional do Petroleo, Gas Natural e Biocombustiveis (ANP), Brazil, via the Oceanographic

Modeling and Observation Research Network (REMO); Russian Foundation for Basic Research (project 11-07-00161-a).

References

1. Gill M., Malanotte-Rizzoli P. Data assimilation in meteorology and oceanography // Adv. Geophys., 1991. Vol. 33. P. 141–266.
2. Evensen G. Sequential data assimilation with a non-linear quasi geostrophic model using Monte-Carlo methods to forecast error statistics // J. Geophys. Res., 1994. Vol. 6. P. 10143–11062.
3. Cohn S. An introduction to estimation theory // J. Meteor. Soc. Japan, 1997. Vol. 75. P. 257–288.
4. Gikhman I. I., Skorokhod A. S. Introduction to the theory of random processes. — Dover Publications, 1996.
5. Strook D., Varadhan S. R. S. Multidimensional random processes. — Berlin: Springer-Verlag, 1995.
6. Belyaev K., Tanajura C. A. S., O'Brien J. J. A data assimilation technique with an ocean circulation model and its application to the tropical Atlantic // Appl. Math. Model., 2001. Vol. 25. P. 655–670.
7. Tanajura C. A. S., Belyaev K. A sequential data assimilation method based on the properties of diffusion-type process // Appl. Math. Model., 2009. Vol. 33. P. 2165–2174.
8. Bleck R., Boudra D. B. Initial testing of a numerical ocean circulation model using a hybrid quasi-isopycnal vertical coordinate // J. Phys. Oceanogr., 1981. Vol. 11. P. 750–770.
9. Bleck R. An oceanic general circulation model framed in hybrid isopycnal Cartesian coordinates // Ocean Model., 2002. Vol. 4. P. 55–88.
10. Belyaev K. P., Tuchkova N. P., Cubasch U. Response of a coupled ocean-ice-atmosphere model to data assimilation in the tropical zone of the Pacific Ocean // J. Oceanology, 2010. Vol. 50. No. 3. P. 306–316.

МАТЕМАТИЧЕСКОЕ ОБОСНОВАНИЕ, ПРИМЕНЕНИЕ И СРАВНЕНИЕ ОБОБЩЕННОГО МЕТОДА УСВОЕНИЯ ДАННЫХ НАБЛЮДЕНИЙ, ОСНОВАННОГО НА МЕТОДАХ ДИФФУЗИОННОЙ АППРОКСИМАЦИИ, С ДРУГИМИ МЕТОДАМИ УСВОЕНИЯ ДАННЫХ

К. П. Беляев¹, К. А. С. Танажура², Н. П. Тучкова³

¹Институт океанологии им. П.П. Ширшова Российской академии наук, kb@sail.msk.ru

²Федеральный университет штата Баийя, Бразилия, cast@ufba.br

³Вычислительный центр им. А. А. Дородницына Российской академии наук, tuchkova@ccas.ru

Аннотация: Многие методы усвоения данных, применяемые в численных океанских и атмосферных моделях, базируются на теории случайных процессов. Предложен метод усвоения, основанный на построении специальной последовательности цепей Маркова, с помощью которой строится сходимость к состоянию модели. Исследуются условия этой сходимости. Решается проблема оптимизации параметров этой цепи для наилучшего приближения, и обсуждаются результаты численных экспериментов. Показано, что предложенный метод усвоения данных может использоваться в практическом применении в метеорологии и океанографии. В данном исследовании метод применялся для океанской модели НУСОМ и данных наблюдений с дрейфтеров АРГО. В работе также выполнялись эксперименты с другими методами усвоения. Представлены результаты сравнения и анализа.

Ключевые слова: последовательность цепей Маркова; диффузионный случайный процесс; методы усвоения данных наблюдений; НУСОМ (Гибридная модель циркуляции океана); дрейфтеры АРГО

COMPLETE CONVERGENCE FOR ARRAYS OF NEGATIVELY DEPENDENT RANDOM VARIABLES

S. H. Sung¹, K. Budsaba², and A. Volodin³

Abstract: A general result establishing complete convergence for the row sums of an array of rowwise negatively dependent random variables is presented. From this result, a number of complete convergence results have been obtained for weighted sums of negatively dependent random variables.

Keywords: complete convergence; negatively dependent; weighted sums; arrays

1 Introduction

The concept of complete convergence of a sequence of random variables was introduced by Hsu and Robbins [1] as follows. A sequence $\{U_n, n \geq 1\}$ of random variables converges completely to the constant θ if

$$\sum_{n=1}^{\infty} P(|U_n - \theta| > \varepsilon) < \infty \quad \text{for all } \varepsilon > 0.$$

In view of the Borel–Cantelli lemma, this implies that $U_n \rightarrow \theta$ almost surely. The converse is true if $\{U_n, n \geq 1\}$ are independent random variables. Hsu and Robbins [1] and Katz [2] ($p = 1$ and $1 < p < 2$, respectively) proved that if $\{X_n, n \geq 1\}$ is a sequence of independent and identically distributed random variables with mean zero and $E|X_1|^{2p} < \infty$, then $\sum_{i=1}^n X_i/n^{1/p}$ converges completely to zero.

The paper [1] initiated numerous explorations of the complete convergence of sums of independent random variables. The research was continued by Erdős [3, 4], Spitzer [5], Baum and Katz [6], and Gut [7]. This subject is actively discussed in scientific press during the last few decades. For example, Hu *et al.* [8] extended the result of Hsu–Robbins–Katz to the case where $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ is an array of rowwise independent random variables which are stochastically dominated by a random variable X satisfying $E|X|^{2p} < \infty$ for some $1 \leq p < 2$.

The papers [9, 10] contain, up to the authors' knowledge, the most general theorems that provide sufficient conditions for complete convergence for sums of arrays of rowwise independent random variables.

In the following, let $\{k_n, n \geq 1\}$ be a sequence of positive integers. In general, the case $k_n = \infty$ is not

precluded. When $k_n = \infty$, it will be assumed that $\sum_{i=1}^{\infty} X_{ni}$ converges almost surely. Recall that an array $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ of random variables is said to be *stochastically dominated* by a random variable X if there exists a positive constant $C > 0$ such that

$$P\{|X_{ni}| > x\} \leq CP\{|X| > x\} \\ \text{for all } x > 0, 1 \leq i \leq k_n, \text{ and } n \geq 1.$$

Recently, some complete convergence theorems for negatively dependent random variables have been obtained by many authors (see, for example, [11, 12] and references in these papers). Taylor *et al.* [11] extended the result of Hu *et al.* [8] to the array of rowwise negatively dependent random variables. Giuliano *et al.* [12] considered so-called acceptable random variable, which is more general notion than negative dependency.

The finite set of random variables X_1, \dots, X_n is said to be *negatively dependent* if

$$P\{X_1 \leq x_1, \dots, X_n \leq x_n\} \\ \leq P\{X_1 \leq x_1\} \cdots P\{X_n \leq x_n\};$$

$$P\{X_1 > x_1, \dots, X_n > x_n\} \\ \leq P\{X_1 > x_1\} \cdots P\{X_n > x_n\}$$

for all real x_1, \dots, x_n . An infinite sequence $\{X_n, n \geq 1\}$ is said to be negatively dependent if every finite subset of the sequence $\{X_1, \dots, X_n\}$ is negatively dependent.

In this paper, a general result establishing complete convergence for the row sums of an array of rowwise negatively dependent random variables is presented. It also specifies the corresponding rate of convergence. From this result, a number of complete convergence

¹Department of Applied Mathematics, Pai Chai University, Taejon, South Korea, sungsh@pcu.ac.kr

²Center of Excellence in Mathematics, CHE, Bangkok, Thailand; Department of Mathematics and Statistics, Thammasat University Rangsit Center, Pathumthani, Thailand, kamon@mathstat.sci.tu.ac.th

³School of Mathematics and Statistics, University of Western Australia, Crawley, Australia; University of Regina, Canada, Andrei.Volodin@uregina.ca

results for negatively dependent random variables have been obtained. As a corollary, the result of Taylor *et al.* [11] is obtained.

Throughout this paper, C denotes a positive constant which may be different in various places, and it is convenient to define $\log x = \max\{1, \ln x\}$.

2 Preliminary Lemmas

To prove the main result, the following lemmas are necessary. The first two lemmas are well known and can be found, for example, in [11].

Lemma 1. *Let $\{X_n, n \geq 1\}$ be a sequence of negatively dependent random variables and $\{f_n, n \geq 1\}$ be a sequence of Borel functions all of which are monotone increasing (or monotone decreasing), then $\{f_n(X_n), n \geq 1\}$ is a sequence of negatively dependent random variables.*

The second lemma mainly states that negatively dependent random variables are negatively correlated.

Lemma 2. *Let X_1, \dots, X_n be nonnegative negatively dependent integrable random variables. Then*

$$\mathbb{E} \prod_{i=1}^n X_i \leq \prod_{i=1}^n \mathbb{E} X_i.$$

The following lemma plays an essential role in the main result. Of course, this lemma is of interest only if positive constants d_i , and, hence, second moments $\mathbb{E} X_i^2, 1 \leq i \leq n$, are close to zero (at least less than one). Otherwise, there is an alternative so-called subgaussian estimations (see, for example, [12]).

Lemma 3. *Let X_1, \dots, X_n be negatively dependent mean zero random variables such that*

$$|X_i| \leq d_i, \quad 1 \leq i \leq n,$$

for a sequence of positive constants d_1, \dots, d_n . Then, for any $t > 0$,

$$\mathbb{E} \exp \left\{ t \sum_{i=1}^n X_i \right\} \leq \exp \left\{ \frac{t^2}{2} \sum_{i=1}^n e^{td_i} \mathbb{E} X_i^2 \right\}.$$

Proof. From the inequality $e^x \leq 1 + x + (x^2/2)e^{|x|}$, which is true for all x , one has

$$\begin{aligned} \mathbb{E} e^{tX_i} &\leq 1 + t\mathbb{E} X_i + \frac{t^2}{2} \mathbb{E} \left(X_i^2 e^{t|X_i|} \right) \\ &= 1 + \frac{t^2}{2} \mathbb{E} \left(X_i^2 e^{t|X_i|} \right) \quad (\text{since } X_i \text{ have mean zero}) \\ &\leq 1 + \frac{t^2}{2} e^{td_i} \mathbb{E} X_i^2 \leq \exp \left\{ \frac{t^2}{2} e^{td_i} \mathbb{E} X_i^2 \right\}, \end{aligned}$$

since $1 + x \leq e^x$ for all x . It follows from Lemmas 1 and 2 that

$$\begin{aligned} \mathbb{E} \exp \left\{ t \sum_{i=1}^n X_i \right\} &\leq \prod_{i=1}^n \mathbb{E} e^{tX_i} \\ &\leq \prod_{i=1}^n \exp \left\{ \frac{t^2}{2} e^{td_i} \mathbb{E} X_i^2 \right\} = \exp \left\{ \frac{t^2}{2} \sum_{i=1}^n e^{td_i} \mathbb{E} X_i^2 \right\}. \quad \square \end{aligned}$$

3 Main Result

With the preliminary lemmas, the main result may now be stated and proved.

Theorem. *Let $\{X_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of rowwise negatively dependent random variables, $\{a_n, n \geq 1\}$ be a sequence of positive constants, and $\{b_n, n \geq 1\}$ be a sequence of positive constants such that $\lim_{n \rightarrow \infty} b_n = \infty$. Suppose that*

- (i) $\sum_{n=1}^{\infty} a_n \sum_{i=1}^{k_n} \mathbb{P}\{|X_{ni}| > \varepsilon\} < \infty$ for all $\varepsilon > 0$;
- (ii) $\sum_{n=1}^{\infty} a_n \left(\sum_{i=1}^{k_n} \mathbb{P}\{|X_{ni}| > 1/b_n\} \right)^{N_1} < \infty$ for some $N_1 > 0$;
- (iii) $b_n \sum_{i=1}^{k_n} \mathbb{E} X_{ni}^2 I\{|X_{ni}| \leq 1/b_n\} \rightarrow 0$ as $n \rightarrow \infty$;
and
- (iv) $\sum_{n=1}^{\infty} a_n \exp\{-N_2 b_n\} < \infty$ for some $N_2 > 0$.

Then

$$\sum_{n=1}^{\infty} a_n \mathbb{P} \left\{ \left| \sum_{i=1}^{k_n} X_{ni} - \mathbb{E} X_{ni} I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} \right| > \varepsilon \right\} < \infty$$

for all $\varepsilon > 0$.

Proof. The set of all natural numbers is partitioned into two subsets:

$$\begin{aligned} A' &= \left\{ n : \sum_{i=1}^{k_n} \mathbb{P} \left\{ |X_{ni}| > \frac{1}{b_n} \right\} \leq 1 \right\}; \\ A'' &= \left\{ n : \sum_{i=1}^{k_n} \mathbb{P} \left\{ |X_{ni}| > \frac{1}{b_n} \right\} > 1 \right\}. \end{aligned}$$

Applying (ii), one obtains

$$\begin{aligned} &\sum_{n \in A''} a_n \mathbb{P} \left\{ \left| \sum_{i=1}^{k_n} X_{ni} - \mathbb{E} X_{ni} I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} \right| > \varepsilon \right\} \\ &\leq \sum_{n \in A''} a_n \leq \sum_{n \in A''} a_n \left(\sum_{i=1}^{k_n} \mathbb{P} \left\{ |X_{ni}| > \frac{1}{b_n} \right\} \right)^{N_1} < \infty. \end{aligned}$$

Hence, it is enough to show that

$$\sum_{n \in A'} a_n \mathbb{P} \left\{ \left| \sum_{i=1}^{k_n} X_{ni} - \mathbb{E} X_{ni} I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} \right| > \varepsilon \right\} < \infty \text{ for all } \varepsilon > 0.$$

For $1 \leq i \leq k_n$ and $n \geq 1$, define

$$\begin{aligned} Y_{ni} &= X_{ni} I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} + \frac{1}{b_n} I \left\{ X_{ni} > \frac{1}{b_n} \right\} \\ &\quad - \frac{1}{b_n} I \left\{ X_{ni} < -\frac{1}{b_n} \right\}; \\ U_{ni} &= \frac{1}{b_n} \left(I \left\{ X_{ni} < -\frac{1}{b_n} \right\} - \mathbb{P} \left\{ X_{ni} < -\frac{1}{b_n} \right\} \right); \\ V_{ni} &= -\frac{1}{b_n} \left(I \left\{ X_{ni} > \frac{1}{b_n} \right\} - \mathbb{P} \left\{ X_{ni} > \frac{1}{b_n} \right\} \right); \\ Z_{ni} &= X_{ni} I \left\{ \frac{1}{b_n} < |X_{ni}| \leq \frac{\varepsilon}{4[N_1 + 1]} \right\}. \end{aligned}$$

Then, $\{Y_{ni} - \mathbb{E}Y_{ni}, 1 \leq i \leq k_n, n \geq 1\}$, $\{U_{ni}, 1 \leq i \leq k_n, n \geq 1\}$, and $\{V_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ are the arrays of rowwise negatively dependent random variables by Lemma 1.

Note that if one defines

$$W_{ni} = \frac{1}{b_n} \left(I \left\{ |X_{ni}| > \frac{1}{b_n} \right\} - \mathbb{P} \left\{ |X_{ni}| > \frac{1}{b_n} \right\} \right),$$

then it cannot be stated that $\{W_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is an array of negatively dependent random variables. This is a sort of the main disadvantage when one is dealing with negatively dependent random variables.

Since $\lim_{n \rightarrow \infty} b_n = \infty$, there exists a positive integer M such that

$$\frac{\varepsilon}{4[N_1 + 1]} > \frac{1}{b_n}$$

for all $n > M$. For $n > M$, one can write that

$$\begin{aligned} &\sum_{i=1}^{k_n} X_{ni} - \mathbb{E} X_{ni} I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} \\ &= \sum_{i=1}^{k_n} (Y_{ni} - \mathbb{E}Y_{ni}) + \sum_{i=1}^{k_n} U_{ni} + \sum_{i=1}^{k_n} V_{ni} + \sum_{i=1}^{k_n} Z_{ni} \\ &\quad + \sum_{i=1}^{k_n} X_{ni} I \left\{ |X_{ni}| > \frac{\varepsilon}{4[N_1 + 1]} \right\}. \end{aligned}$$

It follows that

$$\begin{aligned} &\sum_{\substack{n > M, \\ n \in A'}} a_n \mathbb{P} \left\{ \sum_{i=1}^{k_n} X_{ni} - \mathbb{E} X_{ni} I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} > \varepsilon \right\} \\ &\leq \sum_{n > M, n \in A'} a_n \mathbb{P} \left\{ \sum_{i=1}^{k_n} Y_{ni} - \mathbb{E}Y_{ni} > \frac{\varepsilon}{4} \right\} \\ &\quad + \sum_{n > M, n \in A'} a_n \mathbb{P} \left\{ \sum_{i=1}^{k_n} U_{ni} > \frac{\varepsilon}{4} \right\} \\ &\quad + \sum_{n > M, n \in A'} a_n \mathbb{P} \left\{ \sum_{i=1}^{k_n} V_{ni} > \frac{\varepsilon}{4} \right\} \\ &\quad + \sum_{n > M, n \in A'} a_n \mathbb{P} \left\{ \sum_{i=1}^{k_n} Z_{ni} > \frac{\varepsilon}{4} \right\} \\ &+ \sum_{\substack{n > M, \\ n \in A'}} a_n \mathbb{P} \left\{ |X_{ni}| > \frac{\varepsilon}{4[N_1 + 1]} \text{ for some } 1 \leq i \leq k_n \right\} \\ &=: I_1 + I_2 + I_3 + I_4 + I_5. \end{aligned}$$

Now, let estimate each sum separately.

For I_1 , note that $|Y_{ni}| \leq 1/b_n$ and

$$Y_{ni}^2 = X_{ni}^2 I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} + \left(\frac{1}{b_n} \right)^2 I \left\{ |X_{ni}| > \frac{1}{b_n} \right\}.$$

Moreover, one has that

$$\frac{1}{b_n} \sum_{i=1}^{k_n} \mathbb{P} \left\{ |X_{ni}| > \frac{1}{b_n} \right\} = o(1) \text{ for } n \in A'.$$

By Lemma 3 with $t = 4(N_2 + 1)b_n/\varepsilon$, one obtains that for $n \in A'$,

$$\begin{aligned} &\mathbb{P} \left\{ \sum_{i=1}^{k_n} (Y_{ni} - \mathbb{E}Y_{ni}) > \frac{\varepsilon}{4} \right\} \\ &\leq \exp \left\{ -\frac{t\varepsilon}{4} \right\} \mathbb{E} \exp \left\{ t \sum_{i=1}^{k_n} Y_{ni} - \mathbb{E}Y_{ni} \right\} \\ &\leq \exp \left\{ -\frac{t\varepsilon}{4} \right\} \exp \left\{ \frac{t^2}{2} e^{2t/b_n} \sum_{i=1}^{k_n} \mathbb{E}(Y_{ni} - \mathbb{E}Y_{ni})^2 \right\} \\ &\leq \exp \left\{ -\frac{t\varepsilon}{4} \right\} \exp \left\{ \frac{t^2}{2} e^{2t/b_n} \sum_{i=1}^{k_n} \mathbb{E}Y_{ni}^2 \right\} = \exp \left\{ -\frac{t\varepsilon}{4} \right\} \\ &\quad \times \exp \left\{ \frac{t^2}{2} e^{2t/b_n} \sum_{i=1}^{k_n} \mathbb{E}X_{ni}^2 I \left\{ |X_{ni}| \leq \frac{1}{b_n} \right\} \right\} \\ &\quad + \frac{1}{b_n^2} \mathbb{P} \left\{ |X_{ni}| > \frac{1}{b_n} \right\} \leq \exp \left\{ -(N_2 + 1)b_n \right\} \\ &\quad + 8(N_2 + 1)^2 e^{8(N_2 + 1)/\varepsilon} \varepsilon^{-2} o(1)b_n \quad (\text{by (iii)}) \\ &= \exp \left\{ -(N_2 + 1 - o(1))b_n \right\} \leq \exp \left\{ -N_2 b_n \right\} \end{aligned}$$

for all large n . Thus, $I_1 < \infty$ by (iv).

For I_2 , it can be observed that $|U_{ni}| \leq 1/b_n$ and $EU_{ni}^2 \leq P(|X_{ni}| > 1/b_n)/b_n^2$. Hence,

$$\sum_{i=1}^{k_n} EU_{ni}^2 \leq \frac{1}{b_n^2} \sum_{i=1}^{k_n} P\left\{|X_{ni}| > \frac{1}{b_n}\right\} = \frac{1}{b_n} o(1)$$

for $n \in A'$.

By Lemma 3 with $t = 4(N_2 + 1)b_n/\varepsilon$, one obtains that for $n \in A'$,

$$\begin{aligned} P\left\{\sum_{i=1}^{k_n} U_{ni} > \frac{\varepsilon}{4}\right\} &\leq \exp\left\{-\frac{t\varepsilon}{4}\right\} E \exp\left\{t \sum_{i=1}^{k_n} U_{ni}\right\} \\ &\leq \exp\left(-\frac{t\varepsilon}{4}\right) \exp\left\{\frac{t^2}{2} e^{t/b_n} \sum_{i=1}^{k_n} EU_{ni}^2\right\} \\ &\leq \exp\left\{-(N_2 + 1)b_n\right\} \\ &+ 8(N_2 + 1)^2 e^{4(N_2+1)/\varepsilon} \varepsilon^{-2} o(1)b_n \leq \exp\{-N_2 b_n\} \end{aligned}$$

for all large n . Thus, $I_2 < \infty$ by (iv).

Similarly to I_2 , one gets $I_3 < \infty$.

For I_4 , note that

$$P\left\{\sum_{i=1}^{k_n} Z_{ni} > \frac{\varepsilon}{4}\right\} \leq P\{\text{at least } [N_1 + 1] \text{ of } Z_{ni} \neq 0\}$$

because

$$\begin{aligned} Z_{ni} &< \frac{\varepsilon}{4[N_1 + 1]} \\ &= P\left\{\text{at least } [N_1 + 1] \text{ of } X_{ni} \text{ have the property}\right. \\ &\quad \left.\frac{1}{b_n} < |X_{ni}| \leq \frac{\varepsilon}{(4[N_1 + 1])}\right\} \\ &\leq \sum_{j_1 < \dots < j_{[N_1+1]}} P\left\{X_{n,j_1} > \frac{1}{b_n}, \dots, X_{n,j_{[N_1+1]}} > \frac{1}{b_n}\right\} \end{aligned}$$

(where the summation is taken for all $[N_1 + 1]$ – tuple $(j_1, \dots, j_{[N_1+1]})$)

with $j_1 < \dots < j_{[N_1+1]}$ and $j_i = 1, \dots, k_n$ for each i)

$$\leq \sum_{j_1 < \dots < j_{[N_1+1]}} P\left\{X_{n,j_1} > \frac{1}{b_n}\right\} \dots P\left\{X_{n,j_{[N_1+1]}} > \frac{1}{b_n}\right\}$$

(by negative dependence)

$$\begin{aligned} &= \sum_{j_1 < \dots < j_{[N_1+1]}} \prod_{k=1}^{[N_1+1]} P\left\{X_{n,j_k} > \frac{1}{b_n}\right\} \\ &\leq \sum_{j_1, \dots, j_{[N_1+1]}} \prod_{i=1}^{[N_1+1]} P\left\{X_{n,j_i} > \frac{1}{b_n}\right\} \end{aligned}$$

(where the summation is taken for all possible

$$\begin{aligned} &[N_1 + 1] \text{ – tuple } (j_1, \dots, j_{[N_1+1]}) \\ &\text{and } j_i = 1, \dots, k_n \text{ for each } i) \\ &= \left(\sum_{i=1}^{k_n} P\left\{|X_{ni}| > \frac{1}{b_n}\right\}\right)^{[N_1+1]}. \end{aligned}$$

Thus, $I_4 < \infty$ by (ii).

Obviously, $I_5 < \infty$ by (i).

Therefore, one has that

$$\sum_{\substack{n > M, \\ n \in A'}} a_n P\left(\sum_{i=1}^{k_n} \left(X_{ni} - EX_{ni} I\left(|X_{ni}| \leq \frac{1}{b_n}\right)\right) > \varepsilon\right) < \infty.$$

Since $\{-X_{ni}\}$ is also an array of rowwise negatively dependent random variables, one can replace X_{ni} by $-X_{ni}$ in the above statement. That is,

$$\sum_{\substack{n > M, \\ n \in A'}} a_n P\left(\sum_{i=1}^{k_n} \left(X_{ni} - EX_{ni} I\left(|X_{ni}| \leq \frac{1}{b_n}\right)\right) < -\varepsilon\right) < \infty. \quad \square$$

Remark 1. In view of assumption (iii), it is interesting to consider sequences $\{b_n, n \geq 1\}$ that increase to infinity as slow as possible for (iv) still be true. If the sequence $\{a_n, n \geq 1\}$ has a polynomial growth or a constant (that is, $a_n = n^t, t \geq 0$), then the good choice is $b_n = \log n, n \geq 1$, which has been explored in [10] for the case of rowwise independent arrays. But the present theorem can be applied for sequences $\{a_n, n \geq 1\}$ with a different than polynomial behavior. The main idea is that it is possible to link sequences $\{a_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$ according to assumption (iv).

4 Corollaries

The theorem presented and proved in the previous section can be applied in different situations for various choices of weights and moment conditions.

Corollary 1. Let $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ be an array of rowwise negatively dependent mean zero random variables which are stochastically dominated by a random variable X with $E|X|^{2p} < \infty$ for some $p \geq 1$. Let $\{a_{ni}, 1 \leq i \leq n, n \geq 1\}$ be an array of real numbers and $\{b_n, n \geq 1\}$ be a sequence of positive constants such that

(a) $\lim_{n \rightarrow \infty} b_n = \infty$;

(b) $b_n = O(n^q)$ for some $0 < q < 1/(2p)$;

(c) $\sum_{n=1}^{\infty} \exp\{-N_2 b_n\} < \infty$ for some $N_2 > 0$;

(d) $b_n \sum_{i=1}^n a_{ni}^2 = o(1)$ as $n \rightarrow \infty$; and

(e) $\max_{1 \leq i \leq n} |a_{ni}| = O(1/n^{1/p})$.

Then, $\sum_{i=1}^n a_{ni} X_{ni} \rightarrow 0$ completely.

Proof. Without loss of generality, one may assume that $a_{ni} \geq 0$ for $1 \leq i \leq n$ and $n \geq 1$. Otherwise, let prove the result separately for two arrays of constants $\{a_{ni}^+, 1 \leq i \leq n, n \geq 1\}$ and $\{a_{ni}^-, 1 \leq i \leq n, n \geq 1\}$, where the notations $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$ are used. Then, $\{a_{ni} X_{ni}, 1 \leq i \leq n, n \geq 1\}$ is an array of rowwise negatively dependent random variables by Lemma 1. It can be also assumed that $\max_{1 \leq i \leq n} a_{ni} \leq 1/n^{1/p}$.

Let apply the theorem with $a_n = 1, n \geq 1$, and X_{ni} replaced by $a_{ni} X_{ni}, 1 \leq i \leq n, n \geq 1$.

In order to check condition (i) of the theorem, note that by the stochastic domination hypothesis,

$$\begin{aligned} \sum_{n=1}^{\infty} \sum_{i=1}^n P\{|a_{ni} X_{ni}| > \varepsilon\} &\leq \sum_{n=1}^{\infty} \sum_{i=1}^n P\{|X_{ni}| > \varepsilon n^{1/p}\} \\ &\leq C \sum_{n=1}^{\infty} n P\{|X| > \varepsilon n^{1/p}\}. \end{aligned}$$

The sum $\sum_{n=1}^{\infty} n P\{|X|^p > n\} < \infty$ if and only if $E|X|^{2p} < \infty$. Thus, condition (i) of the theorem holds.

For condition (ii), taking $N_1 > 1/(1 - 2pq)$, one has by Markov's inequality and the stochastic domination hypothesis that

$$\begin{aligned} &\sum_{n=1}^{\infty} \left(\sum_{i=1}^n P\left\{|a_{ni} X_{ni}| > \frac{1}{b_n}\right\} \right)^{N_1} \\ &\leq \sum_{n=1}^{\infty} \left(b_n^{2p} \sum_{i=1}^n |a_{ni}|^{2p} E|X_{ni}|^{2p} \right)^{N_1} \\ &\leq \sum_{n=1}^{\infty} \left(CE|X|^{2p} \frac{b_n^{2p}}{n} \right)^{N_1} \quad \left(\text{by assumption (e)} \right) \\ &< \infty \quad \left(\text{by assumption (b) and the fact} \right. \\ &\quad \left. \text{that } N_1 > \frac{1}{1 - 2pq} \right). \end{aligned}$$

Thus, condition (ii) holds.

For condition (iii),

$$\begin{aligned} &b_n \sum_{i=1}^n E(a_{ni} X_{ni})^2 I\left(|a_{ni} X_{ni}| \leq \frac{1}{b_n}\right) \\ &\leq b_n \sum_{i=1}^n a_{ni}^2 EX_{ni}^2 \leq CE X^2 b_n \sum_{i=1}^n a_{ni}^2 \rightarrow 0 \quad (\text{by (d)}) \end{aligned}$$

Thus, condition (iii) holds.

Condition (iv) holds by the assumption (c).

By the theorem, one obtains that

$$\sum_{n=1}^{\infty} P\left\{\left|\sum_{i=1}^n a_{ni} \left(X_{ni} - EX_{ni} I\left\{|a_{ni} X_{ni}| \leq \frac{1}{b_n}\right\}\right)\right| > \varepsilon\right\} < \infty$$

for all $\varepsilon > 0$. It remains to show that

$$\sum_{i=1}^n a_{ni} EX_{ni} I\left\{|a_{ni} X_{ni}| \leq \frac{1}{b_n}\right\} \rightarrow 0.$$

Since $EX_{ni} = 0$,

$$EX_{ni} I\left\{|a_{ni} X_{ni}| \leq \frac{1}{b_n}\right\} = -EX_{ni} I\left\{|a_{ni} X_{ni}| > \frac{1}{b_n}\right\}.$$

It follows that

$$\begin{aligned} &\left|\sum_{i=1}^n a_{ni} EX_{ni} I\left\{|a_{ni} X_{ni}| \leq \frac{1}{b_n}\right\}\right| \\ &\leq \sum_{i=1}^n |a_{ni}| E|X_{ni}| I\left\{|a_{ni} X_{ni}| > \frac{1}{b_n}\right\} \\ &\leq \frac{1}{n^{1/p}} \sum_{i=1}^n E|X_{ni}| I\left\{|X_{ni}| > \frac{n^{1/p}}{b_n}\right\} \\ &\quad (\text{by assumption (e)}) \\ &\leq C n^{-1-1/p} E|X| I\left\{|X| > \frac{n^{1/p}}{b_n}\right\} \\ &\leq C n^{-1-1/p} E|X|^{2p} |X|^{1-2p} I\left\{|X| > \frac{n^{1/p}}{b_n}\right\} \\ &\leq CE|X|^{2p} n^{-1-1/p} \left(\frac{b_n}{n^{1/p}}\right)^{2p-1} \leq C n^{-1/(2p)} \rightarrow 0 \end{aligned}$$

since $b_n < C n^{1/(2p)}$ for n large enough. Thus, the proof is completed. \square

As a special case of Corollary 1, one gets the following corollary which was proved by Taylor *et al.* [11].

Corollary 2. Let $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ be an array of rowwise negatively dependent mean zero random variables which are stochastically dominated by a random variable X with $E|X|^{2p} < \infty$ for some $1 \leq p < 2$. Then, $\sum_{i=1}^n X_{ni}/n^{1/p} \rightarrow 0$ completely.

Proof. Let $a_{ni} = 1/n^{1/p}$ for $1 \leq i \leq n$ and $n \geq 1$. Then, conditions of Corollary 1 are trivially satisfied with $b_n = n^q$ for some $0 < q < \min \{1/(2p), 2/p - 1\}$. \square

Corollary 3. Let $t > -1, p > 0$, and $\beta \in \mathbb{R}$. Denote $\Delta = p(t + \beta + 1)$ and assume that $\Delta \geq 1$. Let $\{X_{ni}, i \geq 1, n \geq 1\}$ be an array of rowwise negatively dependent mean zero random variables which are stochastically dominated by a random variable X with $E|X|^\Delta < \infty$. Let $\{a_{ni}, i \geq 1, n \geq 1\}$ be a bounded array of real numbers such that

- (1) $\sum_{i=1}^\infty |a_{ni}|^q = O(n^\beta)$ for some $q < \Delta$; and
- (2) If $\Delta \geq 2$, then $\sum_{i=1}^\infty a_{ni}^2 = O(n^\gamma)$ for some $\gamma < 2/p$.

Then,

$$\sum_{n=1}^\infty n^t P \left\{ \left| \frac{\sum_{i=1}^\infty a_{ni} X_{ni}}{n^{1/p}} \right| > \varepsilon \right\} < \infty \text{ for all } \varepsilon > 0.$$

Proof. The same as in the proof of Corollary 1, without loss of generality, one may assume that $a_{ni} \geq 0$ for $i \geq 1, n \geq 1$. Then, $\{a_{ni} X_{ni}/n^{1/p}, i \geq 1, n \geq 1\}$ is an array of rowwise negatively dependent random variables by Lemma 1. Let apply the theorem with $a_n = n^t$, $n \geq 1$, and X_{ni} replaced by $a_{ni} X_{ni}/n^{1/p}, i \geq 1, n \geq 1$.

Consider the sequence $b_n = n^\alpha, n \geq 1$, where $0 < \alpha < (t + 1)/\Delta$. For the case $\Delta \geq 2$, let require additionally that $0 < \alpha < 2/p - \gamma$.

The fact that

$$\sum_{n=1}^\infty n^t \sum_{i=1}^\infty P \left(\left| a_{ni} n^{-1/p} X_{ni} \right| > \varepsilon \right) \leq C E|X|^{p(t+\beta+1)} < \infty$$

was established in many papers (see, for example, [13]) (beginning of the proof of Theorem 3.1), [14] (beginning of the proof of Theorem 3.1), and [10] (beginning of the proof of Theorem 2 and Lemma 3). Note also that the proof presented in [13] is rather complicated once it uses the Stieltjes integration technique, summation by parts lemma, and so on. The proof presented in [14] is much more elegant. Also, Hu *et al.* [13] and Ahmed *et al.* [14] are dealing with an array of constants $\{a_{ni} X_{ni}, i \geq 1, n \geq 1\}$ rather than the array $\{a_{ni} X_{ni}/n^{1/p}, i \geq 1, n \geq 1\}$ which is considered in [10] and this paper.

According to the inequality presented above, condition (i) of the theorem holds.

For (ii), taking $N_1 > (t + 1)/(t + 1 - \alpha\Delta) > 0$, one has by Markov's inequality, $|a_{ni}| = O(1)$, and (1) that

$$\sum_{n=1}^\infty n^t \left(\sum_{i=1}^\infty P \left\{ \left| a_{ni} n^{-1/p} X_{ni} \right| > \frac{1}{b_n} \right\} \right)^{N_1}$$

$$\begin{aligned} &\leq \sum_{n=1}^\infty n^t \left(b_n^\Delta n^{-(t+\beta+1)} \sum_{i=1}^\infty |a_{ni}|^\Delta E|X_{ni}|^\Delta \right)^{N_1} \\ &\leq C \sum_{n=1}^\infty n^t \left(b_n^\Delta n^{-(t+\beta+1)} \sum_{i=1}^\infty |a_{ni}|^q |a_{ni}|^{\Delta-q} \right)^{N_1} \\ &\leq C \sum_{n=1}^\infty n^{t+\alpha\Delta N_1 - (t+1)N_1} < \infty, \end{aligned}$$

since $t + \alpha\Delta N_1 - (t + 1)N_1 < -1$. Thus, condition (ii) of the theorem holds.

For condition (iii), let consider two cases. If $1 \leq \Delta < 2$, by (1), one obtains

$$\begin{aligned} &b_n \sum_{i=1}^\infty E \left(a_{ni} n^{-1/p} X_{ni} \right)^2 I \left\{ \left| a_{ni} n^{-1/p} X_{ni} \right| \leq \frac{1}{b_n} \right\} \\ &= b_n \sum_{i=1}^\infty E \left| a_{ni} n^{-1/p} X_{ni} \right|^\Delta \left| a_{ni} n^{-1/p} X_{ni} \right|^{2-\Delta} \\ &\quad \times I \left\{ \left| a_{ni} n^{-1/p} X_{ni} \right| \leq \frac{1}{b_n} \right\} \\ &\leq b_n^{\Delta-1} \sum_{i=1}^\infty E \left| a_{ni} n^{-1/p} X_{ni} \right|^\Delta I \left\{ \left| a_{ni} n^{-1/p} X_{ni} \right| \leq \frac{1}{b_n} \right\} \\ &\leq b_n^{\Delta-1} \sum_{i=1}^\infty E \left| a_{ni} n^{-1/p} X_{ni} \right|^\Delta \\ &\leq C b_n^{\Delta-1} E|X|^\Delta \sum_{i=1}^\infty \left| a_{ni} n^{-1/p} \right|^\Delta \\ &\leq C n^{\alpha\Delta - \alpha - t - 1} < C n^{-\alpha} \rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

by the choice of α .

If $\Delta \geq 2$, then by (2)

$$\begin{aligned} &b_n \sum_{i=1}^\infty E \left(a_{ni} n^{-1/p} X_{ni} \right)^2 I \left\{ \left| a_{ni} n^{-1/p} X_{ni} \right| \leq \frac{1}{b_n} \right\} \\ &\leq C b_n E X^2 \sum_{i=1}^\infty \frac{a_{ni}^2}{n^{2/p}} \leq C E X^2 n^{\alpha+\gamma-2/p} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ by the choice of α . Thus, condition (iii) of the theorem holds.

Condition (iv) holds trivially.

Hence, one gets by the theorem that

$$\begin{aligned} &\sum_{n=1}^\infty n^t P \left\{ \left| \sum_{i=1}^\infty a_{ni} n^{-1/p} (X_{ni} \right. \right. \\ &\quad \left. \left. - E X_{ni} I \left\{ \left| a_{ni} X_{ni} \right| \leq \frac{n^{1/p}}{b_n} \right\} \right) \right| > \varepsilon \right\} < \infty \end{aligned}$$

for all $\varepsilon > 0$. It remains to show that

$$\sum_{i=1}^{\infty} a_{ni} n^{-1/p} \mathbb{E} X_{ni} I \left\{ |a_{ni} X_{ni}| \leq \frac{n^{1/p}}{b_n} \right\} \rightarrow 0.$$

Since $\mathbb{E} X_{ni} = 0$,

$$\begin{aligned} \mathbb{E} X_{ni} I \left(|a_{ni} X_{ni}| \leq \frac{n^{1/p}}{b_n} \right) \\ = -\mathbb{E} X_{ni} I \left\{ |a_{ni} X_{ni}| > \frac{n^{1/p}}{b_n} \right\}. \end{aligned}$$

It follows that

$$\begin{aligned} & \left| \sum_{i=1}^{\infty} a_{ni} n^{-1/p} \mathbb{E} X_{ni} I \left\{ |a_{ni} X_{ni}| \leq \frac{n^{1/p}}{b_n} \right\} \right| \\ & \leq n^{-1/p} \sum_{i=1}^{\infty} \mathbb{E} |a_{ni} X_{ni}| I \left\{ |a_{ni} X_{ni}| > \frac{n^{1/p}}{b_n} \right\} \\ & \leq n^{-1/p} \left(\frac{b_n}{n^{1/p}} \right)^{\Delta-1} \\ & \quad \times \sum_{i=1}^{\infty} \mathbb{E} |a_{ni} X_{ni}|^{\Delta} I \left\{ |a_{ni} X_{ni}| > \frac{n^{1/p}}{b_n} \right\} \\ & \leq \frac{C(b_n)^{\Delta-1} \mathbb{E} |X|^{\Delta}}{n^{t+1}} \leq C n^{\alpha(\Delta-1)-t-1} \rightarrow 0 \end{aligned}$$

by the choice of α .

Thus, the proof is completed. \square

Remark 2. If $t < -1$, then the conclusion of Corollary 3 holds trivially. When $t \geq -1$, Sung [10] proved Corollary 3 under the stronger condition that $\{X_{ni}, i \geq 1, n \geq 1\}$ is an array of rowwise independent random variables. However, the relatively important case $t = -1$ in Corollary 3 cannot be proved by using the theorem. The present authors left as an open problem whether Corollary 3 holds for $t = -1$.

As a special case of Corollary 3, let get the following corollary.

Corollary 4. Let $t > -1$ and $1 \leq p < 2$. Let $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ be an array of rowwise negatively dependent mean zero random variables which are stochastically dominated by a random variable X with $\mathbb{E}|X|^{p(t+2)} < \infty$. Then,

$$\sum_{n=1}^{\infty} n^t \mathbb{P} \left\{ \frac{\left| \sum_{i=1}^n X_{ni} \right|}{n^{1/p}} > \varepsilon \right\} < \infty \text{ for all } \varepsilon > 0.$$

Proof. Let $a_{ni} = 1$ for $1 \leq i \leq n$ and $a_{ni} = 0$ for $i > n$. Then, for $q < p(t+2)$, $\sum_{i=1}^{\infty} |a_{ni}|^q = n$. Thus, assumption (1) of Corollary 3 holds for $\beta = 1$. Since $1 \leq p < 2$, assumption (2) holds for $\gamma = 1$. Thus, the result follows from Corollary 3. \square

Remark 3. When $t = 0$, Corollary 4 is the same as Corollary 2.

Acknowledgments

This research is partially supported by the Center of Excellence in Mathematics, the Commission on Higher Education, Thailand.

References

1. Hsu P. L., Robbins H. Complete convergence and the law of large numbers // Proc. Nat. Acad. Sci. USA, 1947. Vol. 33. P. 25–31.
2. Katz M. The probability in the tail of a distribution // Ann. Math. Stat., 1963. Vol. 34. P. 312–318.
3. Erdős P. On a theorem of Hsu and Robbins // Ann. Math. Statist., 1949. Vol. 20. P. 286–291.
4. Erdős P. Remark on my paper “On a theorem of Hsu and Robbins” // Ann. Math. Statist., 1950. Vol. 21. P. 138.
5. Spitzer F. L. A combinatorial lemma and its applications // Trans. Amer. Math. Soc., 1956. Vol. 82. P. 323–339.
6. Baum K. B., Katz M. Convergence rates in the law of large numbers // Trans. Amer. Math. Soc., 1965. Vol. 120. P. 108–123.
7. Gut A. Complete convergence for arrays // Periodica Math. Hungarica, 1992. Vol. 25. P. 51–75.
8. Hu T.-C., Móricz F., Taylor R. L. Strong laws of large numbers for arrays of rowwise independent random variables // Acta Math. Hung., 1989. Vol. 54. P. 153–162.
9. Kruglov V. M., Volodin A. I., Hu T.-C. On complete convergence for arrays // Stat. Prob. Lett., 2006. Vol. 76. P. 1631–1640.
10. Sung S. H. Complete convergence for weighted sums of random variables // Stat. Prob. Lett., 2007. Vol. 77. P. 303–311.
11. Taylor R. L., Patterson R. F., Bozorgnia A. A strong law of large numbers for arrays of rowwise negatively dependent random variables // Stochastic Anal. Appl., 2002. Vol. 20. P. 643–656.
12. Giuliano Antonini R., Kozachenko V., Volodin A. Convergence of series of dependent φ -subgaussian random variables // J. Math. Anal. Appl., 2008. Vol. 338. P. 1188–1203.
13. Hu T.-C., Li D., Rosalsky A., Volodin A. On the rate of complete convergence for weighted sums of Banach space valued random elements // Theor. Prob. Appl., 2002. Vol. 47. P. 5455–5468.
14. Ahmed S. E., Giuliano Antonini R., Volodin A. On the rate of complete convergence for weighted sums of arrays of Banach space valued random elements with application to moving average process // Statist. Prob. Lett., 2002. Vol. 58. P. 185–194.

ПОЛНАЯ СХОДИМОСТЬ СУММ В СХЕМЕ СЕРИЙ ОТРИЦАТЕЛЬНО ЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

С. Х. Санг¹, К. Будсаба², А. Володин³

¹Университет Пай Чай, Республика Корея, sungsh@pcu.ac.kr

²Университет Таммасат, Таиланд, kamon@mathstat.sci.tu.ac.th

³Университет Реджайны, Канада, Andrei.Volodin@uregina.ca

Аннотация: Приводится результат о полной сходимости для сумм в схеме серий для отрицательно зависимых случайных величин в весьма общей форме. Из этого результата следуют многие факты о полной сходимости взвешенных сумм отрицательно зависимых случайных величин.

Ключевые слова: полная сходимость; отрицательная зависимость; взвешенные суммы; схема серий

ANALYTICAL MODELING INVARIANT MEASURE DISTRIBUTIONS IN STOCHASTIC SYSTEMS WITH AUTOCORRELATED NOISES

I. N. Sinitsyn

IPI RAN, sinitsin@dol.ru

For multidimensional nonlinear normal (Gaussian) differential systems with un- and autocorrelated noises, on the basis of normal approximation, the correlational algorithms for analytical modeling of stochastic regimes with invariant measure are considered. Special software tools in MATLAB are developed. Test examples confirm practical accuracy.

Keywords: analytical modeling; autocorrelated noise; correlational algorithm; distribution with invariant measure; multidimensional nonlinear differential stochastic system; normal approximation method

ON THE ACCURACY OF SOME MATHEMATICAL MODELS OF CATASTROPHICALLY ACCUMULATED EFFECTS IN PREDICTION OF RISKS OF EXTREMAL EVENTS

I. A. Duchitskii¹, V. Yu. Korolev², and I. A. Sokolov³

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, duchik@gmail.com

²M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

³IPI RAN, isokolov@ipiran.ru

Estimates are constructed for the accuracy of approximation of the distributions of extrema of special random sums by scale mixtures of half-normal laws. The possibility of the application of these results in prediction of risks of extremal events due to catastrophically accumulated effects is discussed.

Keywords: nonhomogeneous flows of events; doubly stochastic Poisson process; negative binomial distribution; gamma-distribution; convergence rate estimate

ABOUT ADAPTIVE STRATEGIES AND THEIR EXISTENCE CONDITIONS

M. G. Kononov

IPI RAN, mkononov@ipiran.ru

The optimal control problem is considered under deficiency of *a priori* information about a controlled object. The solution of the problem is the construction of adaptive strategies on the base of in-control available observations. Some conditions of adaptive controllability are studied. Controlled random sequences are used as mathematical model.

Keywords: controlled random sequences; adaptive strategies; existence conditions

BOUNDS IN NULL ERGODIC CASE FOR SOME QUEUEING SYSTEMS

A. I. Zeifman¹, A. V. Korotysheva², Ya. Satin³, and S. Ya. Shorgin⁴¹Vologda State Pedagogical University; IPI RAN; VSCC CEMI RAS, a_zeifman@mail.ru²Vologda State Pedagogical University, a_korotysheva@mail.ru³Vologda State Pedagogical University, yacovi@mail.ru⁴IPI RAN, SShorgin@ipiran.ru

Markovian queueing models with batch arrivals and group services are considered. The bounds on the rate of convergence in null ergodic situation are obtained. Also, a class of such queueing systems is considered.

Keywords: nonstationary queueing systems with batch arrivals and group services; null ergodicity; bounds

GENERALIZED LAPLACE DISTRIBUTION AS A LIMIT LAW FOR RANDOM SUMS AND STATISTICS CONSTRUCTED FROM SAMPLES WITH RANDOM SIZES

V. Yu. Korolev¹, V. E. Bening², L. M. Zaks³, and A. I. Zeifman⁴¹M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su²Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, bening@cs.msu.su³Department of Modeling and Mathematical Statistics, Alpha-Bank, lily.zaks@gmail.com⁴Vologda State Pedagogical University; IPI RAN; VSCC CEMI RAS, a_zeifman@mail.ru

Limit theorems establishing necessary and sufficient conditions of convergence of random sums and statistics constructed from the samples with random sizes to the generalized Laplace distribution are proved.

Keywords: generalized Laplace distribution; symmetric stable distribution; one-sided stable distribution; scale mixture of normal laws; random sum; sample with random size; mixed Poisson distribution

LOWER BOUNDS FOR THE STABILITY OF NORMAL MIXTURE MODELS WITH RESPECT TO PERTURBATIONS OF MIXING DISTRIBUTION

A. Nazarov

Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University nazarov.vmik@gmail.com

The stability of normal mixture models with respect to perturbations of mixing distribution is investigated. Inequality estimating the distance between two mixing distributions through the closeness of the corresponding mixtures is presented. Existence theorem for stability estimates is proved for subclasses of scale and shift mixtures of normal distributions. For the class of shift mixtures, the estimate is obtained in an explicit form. It is shown that the presented results cannot be radically improved without additional assumptions.

Keywords: normal distribution mixtures; stability problems for stochastic models; Fourier transform; Plancherel theorem; Prokhorov's theorem; Lévy metric; lower bounds

PREPROCESSING OF TEXT RECOGNITION UNDER THE POOR QUALITY IMAGE

M. P. Krivenko

IPI RAN, mkkrivenko@ipiran.ru

The methods of preprocessing of text images including the skew correction and the line segmentation are discussed for the case where the recognizable image is of low quality being obtained with high resolution. Provided that the brightness of the pixel rows of characters differs, even slightly, from the brightness of the background pixels, the procedures for the skew correction and segmentation of the text lines are proposed and analyzed.

Keywords: text recognition; image preprocessing; skew correction; text line segmentation

RANDOM GRAPHS MODEL FOR DESCRIPTION OF INTERACTIONS IN THE NETWORK

A. Grusho¹ and E. Timonina²¹IPI RAN; Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, grusho@yandex.ru²IPI RAN, eltimon@yandex.ru

A new class of random graphs urged to simulate network functioning in time is considered. It is supposed that observations over a network are carried by means of a “window” method. To detect the anomalies, normal behavior which can be watched in “windows” of a considered model is studied. The asymptotic value of the maximum degree of vertices in graph which is generated by a “window” of certain size is analyzed.

Keywords: random graphs; simulation of wide area networks; information security; abnormal behavior

ON THE OPTIMAL CORRECT RECODING OF INTEGER DATA IN RECOGNITION

E. V. Djukova¹, A. V. Sizov², and R. M. Sotnezov³¹Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, edjukova@mail.ru²Moscow State University, box.sizov@gmail.com³Institution of the Russian Academy of Sciences Dorodnicyn Computing Center of RAS, rom.sot@gmail.com

Questions of application of logical procedure of recognition by precedents in the case of float information and high-atomicity integer information are investigated. The problem of correct reducing the data atomicity is considered. Genetic algorithms for the search of optimal correct recoding of source information are developed. Developed algorithms are tested on real data.

Keywords: pattern recognition; correct recoding; covering of the Boolean matrix

ESTIMATION OF LINEAR MODEL HYPERPARAMETERS FOR NOISE OR CORRELATED FEATURE SELECTION PROBLEM

A. A. Tokmakova¹ and V. V. Strijov²¹Moscow Institute of Physics and Technology, aleksandra-tok@yandex.ru²Computing Center RAS, strijov@ccas.ru

The problem of feature selection in linear regression models has been solved. To select the features, the authors estimate the covariance matrix of the model parameters. Dependent variable and model parameters are assumed to be normally distributed vectors. Laplace approximation is used for estimation of the covariance matrix: logarithm of the error function is approximated by the normal distribution function. The problem of noise or correlated features is also examined, since in this case, the covariance matrix of the model parameters becomes singular. An algorithm for feature selection is suggested. The results of the study for a time series are given in the computational experiment.

Keywords: feature selection; regression; coherent Bayesian inference; covariance matrix; model parameters

HOLOGRAPHIC CODING BY WALSH–HADAMARD TRANSFORMATION OF RANDOMIZED AND PERMUTED DATA

S. Dolev¹, S. Frenkel², and A. Cohen³¹Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva, Israel, dolev@cs.bgu.ac.il²IPI RAN; Moscow Institute of Radio, Electronics, and Automation (MIREA), fsergei@mail.ru³Department of Communication Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel, coasaf@cse.bgu.ac.il

Holographic coding has the very appealing property of obtaining partial information on data, from any part of the coded information. Holographic coding schemes are studied based on the Walsh–Hadamard orthogonal codes. It is proposed to randomize the data so that the coefficient of the Walsh–Hadamard code would be approximately uniform in order to ensure, with high probability, a monotonic gain of information. The data are xored with randomly chosen bits from random data that have been stored during a preprocessing stage or pseudorandom data produced by a pseudorandom generator. Statistical properties of the truncated sums of Inverse Walsh–Hadamard Transformation (WHT), taking into account the “white-noise nature” and the mentioned above holographic properties of this encoding method, and the performance of the method is considered based on the theoretic Shannon bound. Using this performance measure, an enhancement for the authors’ previous WHT-based holographic coding method is suggested. This enhancement is based on a random permutation.

Keywords: holographic coding; Walsh–Hadamard transformation; Shannon bound

MATHEMATICAL FOUNDATION, APPLICATION, AND COMPARISON OF GENERAL DATA ASSIMILATION METHOD BASED ON DIFFUSION APPROXIMATION WITH OTHER DATA ASSIMILATION SCHEMES

K. P. Belyaev¹, C. A. S. Tanajura², and N. P. Tuchkova³

¹Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia, kb@sail.msk.ru

²Federal University of Bahia, Salvador, Brazil, cast@ufba.br

³Institution of the Russian Academy of Sciences Dorodnicyn Computing Center of RAS, Moscow, Russia, tuchkova@ccas.ru

Data assimilation methods commonly used in numerical ocean and atmospheric circulation models for weather and climate prediction produce approximations of state variables in terms of stochastic processes. This approximation consists of random sequences of Markov chains, which converge to a diffusion-type process. The conditions for this convergence are investigated. The optimization problem associated with the search of the best possible approximation of the state variable and the results of a numerical experiment are discussed. It is shown that the data assimilation method can be used in practical applications in meteorology and oceanography. Several applications of the methods as an example of the modern operational data processing system with the ocean circulation model HYCOM and data from ARGO drifters are performed and the results as well as comparisons with other assimilation schemes are presented.

Keywords: sequence of Markov chains; diffusion stochastic process; data assimilation methods; HYCOM; ARGO drifters

COMPLETE CONVERGENCE FOR ARRAYS OF NEGATIVELY DEPENDENT RANDOM VARIABLES

S. H. Sung¹, K. Budsaba², and A. Volodin³

¹Department of Applied Mathematics, Pai Chai University, Taejon, South Korea, sungsh@pcu.ac.kr

²Center of Excellence in Mathematics, CHE, Bangkok, Thailand; Department of Mathematics and Statistics, Thammasat University Rangsit Center, Pathumthani, Thailand, kamon@mathstat.sci.tu.ac.th

³School of Mathematics and Statistics, University of Western Australia, Crawley, Australia; University of Regina, Canada, Andrei.Volodin@uregina.ca

A general result establishing complete convergence for the row sums of an array of row-wise negatively dependent random variables is presented. From this result, a number of complete convergence results is obtained for weighted sums of negatively dependent random variables.

Keywords: complete convergence; negatively dependent; weighted sums; arrays

Об авторах

Беляев Константин Павлович (р. 1955) — доктор физико-математических наук, ведущий научный сотрудник Института океанологии им. П. П. Ширшова Российской академии наук; научный сотрудник Федерального университета штата Баийя, Сальвадор, Бразилия

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

Будсаба Камон (р. 1963) — доктор философии по статистике, доцент университета Таммасат, Таиланд

Володин Андрей Игоревич (р. 1961) — доктор философии по статистике, профессор университета Реджайны, Канада

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

Долев Шломи (р. 1958) — профессор, доктор наук по информатике, декан факультета естественных наук Университета им. Бен-Гуриона в Негаве, Беэр-Шева, Израиль

Дучицкий Игорь Александрович (р. 1987) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Дюкова Елена Всеволодовна (р. 1945) — доктор физико-математических наук, доцент кафедры математических методов прогнозирования факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; главный научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Закс Лилия Михайловна (р. 1989) — главный специалист отдела моделирования и математической статистики Альфа-банка

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, декан факультета прикладной математики и компьютерных технологий Вологодского государственного педагогического университета; старший научный сотрудник ИПИ РАН, ведущий научный сотрудник ВНКЦ ЦЭМИ РАН

Коен Асаф (р. 1974) — доктор философии по электротехнике, преподаватель и исследователь кафедры телекоммуникаций Университета им. Бен-Гуриона в Негаве, Беэр-Шева, Израиль

Коновалов Михаил Григорьевич (р. 1950) — доктор технических наук, заведующий сектором ИПИ РАН

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; ведущий научный сотрудник ИПИ РАН

Коротышева Анна Владимировна (р. 1988) — аспирант факультета прикладной математики и компьютерных технологий Вологодского государственного педагогического университета

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник ИПИ РАН

Назаров Алексей Леонидович (р. 1989) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Санг Су Хак (р. 1960) — доктор философии по математике, профессор университета Пай Чай, Республика Корея

Сатин Яков Александрович (р. 1978) — кандидат физико-математических наук, старший преподаватель Вологодского государственного педагогического университета

Сизов Алексей Валерьевич (р. 1989) — выпускник кафедры математических методов прогнозирования факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом ИПИ РАН

Соколов Игорь Анатольевич (р. 1954) — академик (действительный член) Российской академии наук, доктор технических наук, директор ИПИ РАН

Стрижов Вадим Викторович (р. 1967) — кандидат физико-математических наук, доцент, научный сотрудник Вычислительного центра Российской академии наук

Сотнезов Роман Михайлович (р. 1985) — кандидат физико-математических наук, инженер-исследователь Вычислительного центра им. А. А. Дородницына Российской академии наук

Танажура Клементе А. С. (р. 1960) — доктор философии, профессор, Федеральный университет штата Баийя, Сальвадор, Бразилия

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник ИПИ РАН

Токмакова Александра Алексеевна (р. 1991) — студентка Московского физико-технического института

Тучкова Наталья Павловна (р. 1955) — кандидат физико-математических наук, старший научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук

Френкель Сергей Лазаревич (р. 1951) — кандидат технических наук, старший научный сотрудник ИПИ РАН; доцент Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА)

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора ИПИ РАН

About Authors

Belyaev Konstantin P. (b. 1955) — Doctor of Science in physics and mathematics, leading scientist, Shirshov Institute of Oceanology of RAS; researcher, Federal University of Bahia, Salvador, Brazil

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

Budsaba Kamon (b. 1963) — Doctor of Philosophy (PhD) in Statistics, associate professor, Thammasat University, Thailand

Cohen Asaf (b. 1974) — Doctor of Philosophy (PhD) in Electrical Engineering, lecturer and researcher, Department of Communication Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Djukova Elena V. (b. 1945) — Doctor of Science in physics and mathematics, associate professor, Department of Mathematical Methods of Forecasting, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; principal scientist, Dorodnicyn Computing Center, Russian Academy of Sciences

Dolev Shlomi (b. 1958) — professor, Doctor of Science in computer science, Dean of the Faculty of Natural

Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Duchitskii Igor A. (b. 1987) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Frenkel Sergey L. (b. 1951) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; associate professor, Moscow Institute of Radio, Electronics, and Automation “MIREA”

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

Kononov Mikhail G. (b. 1950) — Doctor of Science in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

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

Korotysheva Anna V. (b. 1988) — PhD student, Faculty of Applied Mathematics and Computer Technologies, Vologda State Pedagogical University

Krivenko Mikhail P. (b. 1946) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Nazarov Alexey L. (b. 1989) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Satin Yakov A. (b. 1978) — Candidate of Science (PhD) in physics and mathematics; senior lecturer, Vologda State Pedagogical University

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics; professor, Deputy Director, 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

Sizov Alexey V. (b. 1989) — graduate student, Department of Mathematical Methods of Forecasting, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences

Sotnezov Roman M. (b. 1985) — Candidate of Science (PhD) in physics and mathematics, engineer-researcher, Dorodnicyn Computing Center, Russian Academy of Sciences

Strijov Vadim V. (b. 1967) — Candidate of Science (PhD) in physics and mathematics; associate professor, scientist, Computing Center, Russian Academy of Sciences

Sung Soo Hak (b. 1960) — Doctor of Philosophy (PhD) in Mathematics, professor, Pai Chai University, Republic of Korea

Tanjura Clemente A. S. (b. 1960) — Doctor of Philosophy (PhD), professor, Federal University of Bahia, Salvador, Brazil

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Tokmakova Alexandra A. (b. 1991) — student, Moscow Institute of Physics and Technology

Tuchkova Natalia P. (b. 1955) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Dorodnicyn Computing Center, Russian Academy of Sciences

Volodin Andrei I. (b. 1961) — Doctor of Philosophy (PhD) in Statistics, professor, University of Regina, Canada

Zaks Lily M. (b. 1989) — principal officer, Department of Modeling and Mathematical Statistics, Alpha-Bank

Zeifman Alexander I. (b. 1954) — Doctor of Science in physics and mathematics; professor, Dean of the Faculty of Applied Mathematics and Computer Technologies, Vologda State Pedagogical University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; leading scientist, VSCC CEMI RAS

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2012 г.

	Выпуск	Стр.
Агаларов Я. М. Алгоритм вычисления характеристик модели телекоммуникационной сети с повторами передач и неполнодоступной схемой управления буферами . . .	2	2
Белых И. Н., Капустин А. И., Козлов А. В., Лоханова А. И., Матвеев Ю. Н., Пеховский Т. С., Симончик К. К., Шулипа А. К. Система идентификации дикторов по голосу для конкурса <i>NIST SRE 2010</i>	1	91
Беляев К. П., Танажура К. А. С., Тучкова Н. П. Математическое обоснование, применение и сравнение обобщенного метода усвоения данных наблюдений, основанного на методах диффузионной аппроксимации, с другими методами усвоения данных	4	84
Бендерина М. В., Борохов С. В., Будзко В. И., Степанов П. В., Сучков А. П. Управление учетными записями и правами доступа пользователей в центрах обработки данных высокой доступности	1	59
Бенинг В. Е., Закс Л. М., Королев В. Ю. Оценки скорости сходимости распределений случайных сумм к дисперсионным гамма-распределениям	3	69
Бенинг В. Е. см. Королев В. Ю.		
Боков М. В. см. Гудков В. Ю.		
Боровикова О. И. см. Загоруйко Ю. А.		
Борохов С. В. см. Бендерина М. В.		
Босов А. В. Задачи анализа и оптимизации для модели пользовательской активности. Часть 2. Оптимизация внутренних ресурсов	1	19
Босов А. В. Задачи анализа и оптимизации для модели пользовательской активности. Часть 3. Оптимизация внешних ресурсов	2	14
Будзко В. И. см. Бендерина М. В.		
Будсаба К. см. Санг С. Х.		
Визильтер Ю. В., Горбацевич В. С., Каратеев С. Л., Костромов Н. А. Обучение алгоритмов выделения кожи на цветных изображениях лиц	1	108
Володин А. см. Санг С. Х.		
Гайдамака Ю. В., Ефимушкина Т. В., Самуйлов А. К., Самуйлов К. Е. Задачи оптимального планирования межуровневого интерфейса в беспроводных сетях	3	74
Горбацевич В. С. см. Визильтер Ю. В.		
Горшенин А. К. Об устойчивости сдвиговых смесей нормальных законов по отношению к изменениям смешивающего распределения	2	22
Грушо А. А., Тимонина Е. Е. Модель случайных графов для описания взаимодействий в сети	4	57
Гудков В. Ю., Боков М. В. Быстрая обработка изображений отпечатков пальцев	1	99
Де Никола К., Хохлов Ю. С., Пагано М., Сидорова О. И. Дробное движение Леви с зависимыми приращениями и его приложение к моделированию сетевого трафика	3	59
Долев Ш., Френкель С., Коен А. Голографическое кодирование на основе преобразования Уолша–Адамара рандомизированных и перемешанных данных	4	76
Дулин С. К., Розенберг И. Н., Уманский В. И. Обработка геопространственной информации на базе репозитория геоинформационной системы	2	29
Дучицкий И. А., Королев В. Ю., Соколов И. А. О точности некоторых математических моделей катастрофически накапливающихся эффектов при прогнозировании риска экстремальных событий	4	9

	Выпуск	Стр.
Дюкова Е. В., Сизов А. В., Сотнезов Р. М. Об оптимальном корректном перекодировании целочисленных данных в распознавании	4	61
Ефимушкина Т. В. см. Гайдамака Ю. В.		
Жевнерчук Д. В., Николаев А. В. Методика моделирования нагрузки на сервер в открытых системах облачных вычислений	2	43
Желенкова О. П. Исследование радиоисточников средствами виртуальной обсерватории	3	5
Жижимов О. Л. см. Скачков Д. М.		
Загорулько Ю. А., Боровикова О. И., Кононенко И. С., Соколова Е. Г. Методологические аспекты разработки электронного русско-английского тезауруса по компьютерной лингвистике	3	22
Закс Л. М. см. Бенинг В. Е.		
Закс Л. М. см. Королев В. Ю.		
Захаров В. Н. см. Калиниченко Л. А.		
Зейфман А. И., Коротышева А. В., Сатин Я. А., Шоргин С. Я. Оценки в нуль-эргодическом случае для некоторых систем обслуживания	4	27
Зейфман А. И. см. Королев В. Ю.		
Калёнов Н. Е. Задачи и функции библиотек РАН в современных условиях	2	51
Калиниченко Л. А., Ступников С. А. Унификация языков систем на правилах для обеспечения интероперабельности декларативных программ	2	59
Калиниченко Л. А., Ступников С. А., Захаров В. Н. Развитие технологий интеграции информации для решения задач над неоднородными информационными ресурсами	1	70
Капустин А. И. см. Белых И. Н.		
Каратеев С. Л. см. Визильтер Ю. В.		
Карпов А. А. Когнитивные исследования ассистивного многомодального интерфейса для бесконтактного человеко-машинного взаимодействия	2	77
Когаловский М. Р., Паринов С. И. Классификация и использование семантических связей между информационными объектами в научных электронных библиотеках	3	32
Коен А. см. Долев Ш.		
Козлов А. В. см. Белых И. Н.		
Коновалов М. Г. Об адаптивных стратегиях и условиях их существования	4	18
Коновалов М. Г. Оптимизация работы вычислительного комплекса с помощью имитационной модели и адаптивных алгоритмов	1	37
Кононенко И. С. см. Загорулько Ю. А.		
Королев В. Ю., Бенинг В. Е., Закс Л. М., Зейфман А. И. Обобщенное распределение Лапласа как предельное для случайных сумм и статистик, построенных по выборкам случайного объема	4	34
Королев В. Ю., Соколов И. А. Скошенные распределения Стьюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации	1	3
Королев В. Ю. см. Бенинг В. Е.		
Королев В. Ю. см. Дучицкий И. А.		
Королев В. Ю. см. Соколов И. А.		
Коротышева А. В. см. Зейфман А. И.		
Костромов Н. А. см. Визильтер Ю. В.		
Кривенко М. П. Предварительная обработка при распознавании текстов по изображению низкого качества	4	49
Кузнецов В. В. см. Урмаев О. С.		
Кузнецов И. П., Сомин Н. В. Выявление имплицитной информации из текстов на естественном языке: проблемы и методы	1	49
Кузнецов И. П. см. Шарнин М. М.		
Куракин А. В. Распознавание жестов ладони в реальном времени на основе плоских и пространственных скелетных моделей	1	114

	Выпуск	Стр.
Лоханова А. И. см. Белых И. Н.		
Лукашенко О. В., Морозов Е. В. Асимптотика максимума процесса нагрузки для некоторого класса гауссовских очередей	3	81
Маркова Н. А. Логика биографических фактов	2	87
Матвеев Ю. Н. см. Белых И. Н.		
Морозов Е. В., Некрасова Р. С. Об оценивании вероятности переполнения конечного буфера в регенеративных системах обслуживания	3	90
Морозов Е. В., Румянцев А. С. Вероятностные модели многопроцессорных систем: стационарность и моментные свойства	3	99
Морозов Е. В. см. Лукашенко О. В.		
Мурашов Д. М. Комбинированный подход к локализации различий многомодальных изображений	1	122
Назаров А. Л. Нижние оценки устойчивости смесей нормальных распределений к возмущениям смешивающих распределений	4	40
Некрасова Р. С. см. Морозов Е. В.		
Николаев А. В. см. Жевнерчук Д. В.		
Павлов И. В. Расчет и оптимизация некоторых характеристик для модели вычислительного комплекса	2	97
Пагано М. см. Де Никола К.		
Паринов С. И. см. Когаловский М. Р.		
Пеховский Т. С. см. Белых И. Н.		
Печинкин А. В., Соколов И. А., Шоргин С. Я. Ограничение на суммарный объем заявок в дискретной системе Geo/G/1/∞	3	107
Попов С. В. Уточнение неравномерных оценок скорости сходимости в центральной предельной теореме при существовании моментов не выше второго	1	32
Розенберг И. Н. см. Дулин С. К.		
Рудаков К. В., Торшин И. Ю. Анализ информативности мотивов на основе критерия разрешимости в задаче распознавания вторичной структуры белка	1	79
Румянцев А. С. см. Морозов Е. В.		
Самуйлов А. К. см. Гайдамака Ю. В.		
Самуйлов К. Е. см. Гайдамака Ю. В.		
Санг С. Х., Будсаба К., Володин А. Полная сходимость сумм в схеме серий отрицательно зависимых случайных величин	4	95
Сатин Я. А. см. Зейфман А. И.		
Семенов К. К. Нечеткие переменные как способ формализации характеристик погрешности в задачах математической обработки	2	101
Сидорова О. И. см. Де Никола К.		
Сизов А. В. см. Дюкова Е. В.		
Симончик К. К. см. Белых И. Н.		
Синицын И. Н. Аналитическое моделирование распределений с инвариантной мерой в стохастических системах с автокоррелированными шумами	4	4
Синицын И. Н. Математическое обеспечение для анализа нелинейных многоканальных круговых стохастических систем, основанное на параметризации распределений	1	12
Скачков Д. М., Жижимов О. Л. Об интеграции географических метаданных посредством ретроспективного тезауруса	3	43
Соколов И. А., Королев В. Ю. Предисловие	4	3
Соколов И. А. см. Дучицкий И. А.		
Соколов И. А. см. Королев В. Ю.		
Соколов И. А. см. Печинкин А. В.		
Соколова Е. Г. см. Загоруйко Ю. А.		
Сомин Н. В. см. Кузнецов И. П.		
Сотнезов Р. М. см. Дюкова Е. В.		
Степанов П. В. см. Бендерина М. В.		

Стрижов В. В. см. Токмакова А. А.		
Ступников С. А. см. Калиниченко Л. А.		
Ступников С. А. см. Калиниченко Л. А.		
Сучков А. П. см. Бендерина М. В.		
Танажура К. А. С. см. Беляев К. П.		
Тимонина Е. Е. см. Грушо А. А.		
Токмакова А. А., Стрижов В. В. Оценивание гиперпараметров линейных регрессионных моделей при отборе шумовых и коррелирующих признаков.	4	66
Торшин И. Ю. см. Рудаков К. В.		
Тучкова Н. П. см. Беляев К. П.		
Уманский В. И. см. Дулин С. К.		
Ушаков А. В. Анализ системы обслуживания с гиперэкспоненциальным входящим потоком в условиях критической загрузки.	3	114
Ушаков А. В. О виртуальном времени ожидания в системе с относительным приоритетом и гиперэкспоненциальным входящим потоком.	1	27
Ушмаев О. С., Кузнецов В. В. Алгоритмы защищенной биометрической верификации на основе бинарного представления топологии отпечатков пальцев.	1	132
Френкель С. см. Долев Ш.		
Хохлов Ю. С. см. Де Никола К.		
Шарапов Р. В. см. Шарапова Е. В.		
Шарапова Е. В., Шарапов Р. В. Универсальная система проверки текстов на плагиат «Автор.NET».	3	52
Шарнин М. М., Кузнецов И. П. Особенности семантического поиска информационных объектов на основе технологии баз знаний.	2	113
Шестаков О. В. О скорости сходимости оценки риска пороговой обработки вейвлет-коэффициентов к нормальному закону при использовании робастных оценок дисперсии.	2	122
Шоргин С. Я. см. Зейфман А. И.		
Шоргин С. Я. см. Печинкин А. В.		
Шулипа А. К. см. Белых И. Н.		
Янушкявичене О. Л., Янушкявичюс Р. О скорости сходимости некоторой U-статистики	3	64
Янушкявичюс Р. см. Янушкявичене О. Л.		

2012 AUTHOR INDEX

	Issue	Page
Agalarov Ya. M. Algorithm for Computation Characteristics of Telecommunication Network with Resubmits and Incomplete Circuit Buffer Management Model.....	2	2
Belyaev K. P., Tanajura C. A. S., and Tuchkova N. P. Mathematical Foundation, Application, and Comparison of General Data Assimilation Method Based on Diffusion Approximation with Other Data Assimilation Schemes.....	4	84
Belykh I. N., Kapustin A. I., Kozlov A. V., Lohanova A. I., Matveev Yu. N., Pekhovsky T. S., Simonchik K. K., and Shulipa A. K. Speaker Identification System for the <i>NIST SRE 2010</i>	1	91
Benderina M. V., Borokhov S. V., Budzko V. I., Stepanov P. V., and Suchkov A. P. Identity and Access Management of the Users' Rights in High Available Data Center.....	1	59
Bening V. E., Zaks L. M., and Korolev V. Yu. Estimates of the Rate of Convergence of the Distributions of Random Sums to Variance-Gamma Distributions.....	3	69
Bening V. E. see Korolev V. Yu.		
Bokov M. V. Gudkov V. J.		
Borokhov S. V. see Benderina M. V.		
Borovikova O. I. see Zagorulko Yu. A.		
Bosov A. V. Analysis and Optimization Problems for Some Users Activity Model. Part 2. Internal Resources Optimization.....	1	19
Bosov A. V. Analysis and Optimization Problems for Some Users Activity Model. Part 3. External Resources Optimization.....	2	14
Budsaba K. see Sung S. H.		
Budzko V. I. see Benderina M. V.		
Cohen A. see Dolev S.		
De Nikola C., Khokhlov Y. S., Pagano M., and Sidorova O. I. Fractional Levy Motion with Dependent Increments and Its Application to Network Traffic Modeling.....	3	59
Djukova E. V., Sizov A. V., and Sotnezov R. M. On the Optimal Correct Recoding of Integer Data in Recognition.....	4	61
Dolev S., Frenkel S., and Cohen A. Holographic Coding by Walsh–Hadamard Transformation of Randomized and Permuted Data.....	4	76
Duchitskii I. A., Korolev V. Yu., and Sokolov I. A. On the Accuracy of Some Mathematical Models of Catastrophically Accumulated Effects in Prediction of Risks of Extremal Events.....	4	9
Dulin S. K., Rozenberg I. N., and Umansky V. I. Geospatial Information Processing on the Base of GIS Repository.....	2	29
Efimushkina T. V. see Gaidamaka Y. V.		
Frenkel S. see Dolev S.		
Gaidamaka Y. V., Efimushkina T. V., Samuylov A. K., and Samouylov K. E. Cross-Layer Optimization Planning Problems in Wireless Networks.....	3	74
Gorbatcevich V. see Vizilter Y.		
Gorshenin A. K. On Stability of Normal Location Mixtures with Respect to Variations in Mixing Distribution.....	2	22
Grusho A. and Timonina E. Random Graphs Model for Description of Interactions in the Network.....	4	57
Gudkov V. J. and Bokov M. V. Fast Processing of Fingerprint Images.....	1	99
Kalenov N. E. Russian Academy of Sciences Libraries Tasks and Functions in Modern Conditions.....	2	51

	Issue	Page
Kalinichenko L. A. and Stupnikov S. A. Unification of the Rule-Based System Languages to Provide Interoperability of Declarative Programs	2	59
Kalinichenko L. A., Stupnikov S. A., and Zakharov V. N. Extending Information Integration Technologies for Problem Solving over Heterogeneous Information Resources	1	70
Kapustin A. I. see Belykh I. N.		
Karateev S. see Vizilter Y.		
Karpov A. A. Cognitive Study of an Assistive Multimodal User Interface for Hands-Free Human–Computer Interaction	2	77
Khokhlov Y. S. see De Nikola C.		
Kogalovsky M. R. and Parinov S. I. Classification and Utilization of Semantic Linkages Between Information Objects in Scientific Digital Libraries	3	32
Kononenko I. S. see Zagorulko Yu. A.		
Konovalov M. G. About Adaptive Strategies and Their Existence Conditions	4	18
Konovalov M. G. Computer System Optimization Using Simulation Model and Adaptive Algorithms	1	37
Korolev V. Yu., Bening V. E., Zaks L. M., and Zeifman A. I. Generalized Laplace Distribution As a Limit Law for Random Sums and Statistics Constructed from Samples with Random Sizes	4	34
Korolev V. Yu. and Sokolov I. A. Skew Student Distributions, Variance-Gamma Distributions and Their Generalizations as Asymptotic Approximations	1	3
Korolev V. Yu. see Bening V. E.		
Korolev V. Yu. see Duchitskii I. A.		
Korotysheva A. V. see Zeifman A. I.		
Kostromov N. see Vizilter Y.		
Kozlov A. V. see Belykh I. N.		
Krivenko M. P. Preprocessing of Text Recognition Under the Poor Quality Image	4	49
Kurakin A. V. Real-Time Hand Gesture Recognition by Planar and Spatial Skeletal Models	1	114
Kuznetsov I. P. and Somin N. V. Extraction of Implicit Information from the Texts in Natural Language: Problems and Methods	1	49
Kuznetsov I. P. see Sharnin M. M.		
Kuznetsov V. V. see Ushmaev O. S.		
Lohanova A. I. see Belykh I. N.		
Lukashenko I. V. and Morozov A. V. Asymptotics of the Maximum Workload for a Class of Gaussian Queueing Systems	3	81
Markova N. A. A Logic of Biographical Facts	2	87
Matveev Yu. N. see Belykh I. N.		
Morozov A. V. see Lukashenko I. V.		
Morozov E. V. and Nekrasova R. S. On the Estimation of the Overflow Probability in Regenerative Finite Buffer Queueing Systems	3	90
Morozov E. V. and Rumyantsev A. S. Stochastic Models of Multiprocessor Systems: Stability and Moment Properties	3	99
Murashov D. M. Combined Approach to Localization of Differences for Multimodal Images	1	122
Nazarov A. Lower Bounds for the Stability of Normal Mixture Models with Respect to Perturbations of Mixing Distribution	4	40
Nekrasova R. S. see Morozov E. V.		
Nikolaev A. V. see Zhevnerchuk D. V.		
Pagano M. see De Nikola C.		
Parinov S. I. see Kogalovsky M. R.		
Pavlov I. V. Calculation and Optimization of Some Characteristics of the Model for Computational Complex	2	97
Pechinkin A. V., Sokolov I. A., and Shorgin S. Ya. A Restriction on the Total Volume of Demands in the Discrete-Time System $Geo/G/1/\infty$	3	107
Pekhovsky T. S. see Belykh I. N.		

	Issue	Page
Popov S. V. A Refinement of Nonuniform Estimates of the Rate of Convergence in the Central Limit Theorem Under the Existence of Moments of Order Not Higher Than the Second	1	32
Rozenberg I. N. see Dulin S. K.		
Rudakov K. V. and Torshin I. Yu. The Motif Information Analysis Based on the Solvability Criterion for the Protein Secondary Structure Recognition	1	79
Rumyantsev A. S. see Morozov E. V.		
Samouylov K. E. see Gaidamaka Y. V.		
Samuylov A. K. see Gaidamaka Y. V.		
Satin Ya. see Zeifman A. I.		
Semenov K. K. Fuzzy Variables as a Tool for Expressing Error Characteristics in Data Processing	2	101
Sharapov R. V. see Sharapova E. V.		
Sharapova E. V. and Sharapov R. V. Universal System of Plagiarism Detection "Autor.NET" ...	3	52
Sharnin M. M. and Kuznetsov I. P. Semantic Search of Natural Language Information on the Basis of Knowledge Base Technology	2	113
Shestakov O. V. On the Rate of Convergence to the Normal Law of Risk Estimate for Wavelet Coefficients Thresholding When Using Robust Variance Estimates	2	122
Shorgin S. Ya. see Pechinkin A. V.		
Shorgin S. Ya. see Zeifman A. I.		
Shulipa A. K. see Belykh I. N.		
Sidorova O. I. see De Nikola C.		
Simonchik K. K. see Belykh I. N.		
Sinitsyn I. N. Analytical Modeling Invariant Measure Distributions in Stochastic Systems with Autocorrelated Noises	4	4
Sinitsyn I. N. Mathematical Support for Nonlinear Multichannel Circular Stochastic Systems Analysis Based on Distribution Parametrization	1	12
Sizov A. V. see Djukova E. V.		
Skachkov D. M. and Zhizhimov O. L. Integration of Geographic Metadata Using Retrospective Thesaurus	3	43
Sokolov I. A. see Duchitskii I. A.		
Sokolov I. A. see Korolev V. Yu.		
Sokolov I. A. see Pechinkin A. V.		
Sokolova E. G. see Zagorulko Yu. A.		
Somin N. V. see Kuznetsov I. P.		
Sotnezov R. M. see Djukova E. V.		
Stepanov P. V. see Benderina M. V.		
Strijov V. V. see Tokmakova A. A.		
Stupnikov S. A. see Kalinichenko L. A.		
Stupnikov S. A. see Kalinichenko L. A.		
Suchkov A. P. see Benderina M. V.		
Sung S. H., Budsaba K., and Volodin A. Complete Convergence for Arrays of Negatively Dependent Random Variables	4	95
Tanajura C. A. S. see Belyaev K. P.		
Timonina E. see Grusho A.		
Tokmakova A. A. and Strijov V. V. Estimation of Linear Model Hyperparameters for Noise or Correlated Feature Selection Problem	4	66
Torshin I. Yu. see Rudakov K. V.		
Tuchkova N. P. see Belyaev K. P.		
Umansky V. I. see Dulin S. K.		
Ushakov A. V. Head-of-the-Line Priority and Hyperexponential Input Stream	1	27
Ushakov A. V. Heavy-Traffic Analysis for the Queueing System with Hyperexponential Stream	3	114
Ushmaev O. S. and Kuznetsov V. V. Secured Biometric Verification Based on Fingerprint Topology Binary Representation	1	132

	Issue	Page
Vizilter Y., Gorbatceвич V., Karateev S., and Kostromov N. Teaching of Skin Extraction Algorithms for Human Face Color Images	1	108
Volodin A. see Sung S. H.		
Yanushkevichiene O. and Yanushkevichius R. About the Rate of Convergence of One U-Statistic	3	64
Yanushkevichius R. see Yanushkevichiene O.		
Zagorulko Yu. A., Borovikova O. I., Kononenko I. S., and Sokolova E. G. Methodological Aspects of Development of the Electronic Russian–English Thesaurus on Computational Linguistics	3	22
Zakharov V. N. see Kalinichenko L. A.		
Zaks L. M. see Bening V. E.		
Zaks L. M. see Korolev V. Yu.		
Zeifman A. I., Korotysheva A. V., Satin Ya., and Shorgin S. Ya. Bounds in Null Ergodic Case for Some Queueing Systems	4	27
Zeifman A. I. see Korolev V. Yu.		
Zhelenkova O. P. Radio Sources Study with the Virtual Observatory Tools	3	5
Zhevnerchuk D. V. and Nikolaev A. V. Technique for Modeling the Server Loading in Open Cloud Computing Systems	2	43
Zhizhimov O. L. see Skachkov D. M.		

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. Корректурa статей высылается 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 L^AT_EX (.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