Федеральный исследовательский центр
«Карельский научный центр Российской академии наук»

ТРУДЫ

КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РОССИЙСКОЙ АКАДЕМИИ НАУК

№ 4, 2025

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Петрозаводск
2025

Научный журнал
Труды Карельского научного центра Российской академии наук
№ 4, 2025
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ISSN 1997-3217 (печатная версия)
ISSN 2312-4504 (онлайн-версия)

Издается с января 2009 г.

Адрес редакции: 185910, Петрозаводск, ул. Пушкинская, 11
Тел. (8142)762018; факс (8142)769600
E-mail: trudy@krc.karelia.ru
Электронная полнотекстовая версия: http://transactions.krc.karelia.ru; http://journals.krc.karelia.ru

Karelian Research Centre of the Russian Academy of Sciences

TRANSACTIONS

of the KARELIAN RESEARCH CENTRE of the RUSSIAN ACADEMY of SCIENCES

No. 4, 2025

MATHEMATICAL MODELING
AND INFORMATION TECHNOLOGIES

Petrozavodsk
2025

Scientific Journal
Transactions of the Karelian Research Centre of the Russian Academy of Sciences
No. 4, 2025
MATHEMATICAL MODELING AND INFORMATION TECHNOLOGIES

ISSN 1997-3217 (print)
ISSN 2312-4504 (online)

Published since January 2009

8 issues a year

Editorial Office address: 11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
Tel. (8142)762018; fax (8142)769600
E-mail: trudy@krc.karelia.ru
Full-text electronic version: http://transactions.krc.karelia.ru; http://journals.krc.karelia.ru

Труды Карельского научного центра РАН. 2025. № 4. С. 5–16
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 5–16
DOI: 10.17076/mat2101

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 004.01:006.72 (470.22)

МОДЕЛИРОВАНИЕ ВОДОРОДОПРОНИЦАЕМОСТИ СПЛАВОВ ДЛЯ МЕМБРАННОГО ГАЗОРАЗДЕЛЕНИЯ

К. В. Грудова

Отдел комплексных научных исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Определение предельной растворимости водорода в гидридообразующих металлах и сплавах играет важную роль в обеспечении безопасности и надежности конструкций, изготовленных из этих материалов. Проведение комплексного исследования этого параметра позволяет разрабатывать более эффективные методы контроля за состоянием материалов и предотвращения разрушений, вызванных действием водорода. В связи с этим востребованными являются математические модели и алгоритмы обработки экспериментальных данных для оценки водородопроницаемости конструкционных материалов. В данной работе рассматривается краевая задача модели водородопроницаемости, которая в предположении квазистационарного равновесия сводится к системе обыкновенных дифференциальных уравнений. Также предлагается алгоритм локализации начальных данных полученной системы, что позволяет найти ее решение и аппроксимировать экспериментальные графики давления водорода.

Ключевые слова: водородопроницаемость; нелинейные краевые задачи; численное моделирование; параметрическая идентификация

Для цитирования: Грудова К. В. Моделирование водородопроницаемости сплавов для мембранного газоразделения // Труды Карельского научного центра РАН. 2025. № 4. С. 5–16. doi: 10.17076/mat2101

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Отдел комплексных научных исследований КарНЦ РАН).

K. V. Grudova. MODELING OF HYDROGEN PERMEABILITY OF ALLOYS FOR MEMBRANE GAS SEPARATION

Department of Multidisciplinary Scientific Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

Determination of the ultimate solubility of hydrogen in hydride-forming metals and alloys plays an important role in ensuring the safety and reliability of structures made of these materials. Conducting a comprehensive study of this parameter makes it possible to develop more effective methods for monitoring the condition of materials and preventing damage caused by the action of hydrogen. In this regard, there is demand for mathematical models and algorithms that process experimental data for assessing the hydrogen permeability of structural materials. In this paper, we consider the boundary-value problem of the hydrogen permeability model, which, assuming quasi-stationary equilibrium, is reduced to a system of ordinary differential equations. An algorithm for localization of the input data of this system is also proposed, which makes it possible to find its solution and approximate the experimental graphs of hydrogen pressure.

Keywords: hydrogen permeability; nonlinear boundary-value problems; numerical simulation; parametric identification

For citation: Grudova K. V. Modeling of hydrogen permeability of alloys for membrane gas separation. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 5–16. doi: 10.17076/mat2101

Funding. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Department for Multidisciplinary Research KarRC RAS).

Введение

Для успешного прогнозирования снижения свойств пластичности, ударной вязкости, трещиностойкости гидридообразующих металлов и сплавов при использовании их в водородосодержащей среде необходимо определить предельную растворимость водорода в рассматриваемых материалах. Данный параметр имеет решающее значение для оценки стойкости конструкционных сплавов к воздействию водорода. Подобные исследования позволяют разработать методы для предотвращения разрушения конструкций, изготовленных из таких материалов.

Предельная растворимость водорода в металлах и сплавах определяется множеством факторов, включая температуру, давление водорода, микроструктуру материала, наличие дефектов и примесей, а также скорость диффузии водорода. Исследование таких показателей позволяет более точно определить условия, при которых происходит насыщение материала водородом и возникают опасные явления (образование внутренних трещин, ускорение коррозии, изменение механических свойств и др.).

Для определения предельной растворимости водорода в гидридообразующих материалах применяются различные методы, включая измерение изменения объема образца под действием водорода, анализ изменения электрических свойств материала, наблюдение за изменением структуры при воздействии водорода и другие. Эти методы позволяют получить данные о количестве водорода, которое может быть адсорбировано материалом без нарушения его свойств.

Одним из основных преимущественных методов исследования параметров взаимодействия водорода с материалами является метод водородопроницаемости. Это преимущество заключается в возможности измерения в одном эксперименте коэффициента диффузии, констант растворимости и проницаемости [2]. Более подробное описание метода рассмотрим в следующем разделе.

С целью повышения эффективности экспериментальных исследований, решения прикладных задач и обобщений возникает потребность в построении математической модели взаимодействия изотопов водорода с конструкционными материалами и методов их идентификации. Моментальное определение значений параметров, коэффициентов, необходимых для дальнейших исследований, во время проведения эксперимента водородопроницаемости представляет собой практически нерешаемую проблему. Составление математических моделей для процессов, происходящих в ходе эксперимента, позволяет решить этот вопрос.

На основе экспериментального опыта можно сделать вывод, что в технологических задачах, как правило, диффузионные процессы и физико-химические явления на поверхности являются лимитирующими. Технологические особенности получения партии материала влияют на параметры переноса водорода, а это значит, что нужны эффективные алгоритмы обработки экспериментальных кривых. Вместе с тем возникает потребность автоматизировать процесс определения параметров переноса, используя современные вычислительные технологии. Основой для данной работы послужила статья [4].

Таким образом, в рамках текущего исследования была поставлена цель – разработать алгоритм и его программную реализацию для оценки параметров в краевой задаче переноса водорода в сплавах для мембранного газоразделения.

Модель водородопроницаемости

В данной работе рассматривается экспериментальное исследование модифицированного метода водородопроницаемости. Эксперимент проводился с использованием вакуумной камеры, разделенной на две емкости перегородкой из исследуемого материала. Мембрана представляет собой разогретую до фиксированной температуры пластину толщиной ℓ и площадью поверхностей S (торцами можно пренебречь, поскольку $\ell \ll 1$ мм). Предварительно проводится полная дегазация системы для исключения влияния газовых примесей на результаты эксперимента.

В начале эксперимента на входе в вакуумную камеру создается скачкообразное давление молекулярного водорода путем порционного впрыскивания газа. Затем измеряется падающее давление на входе и растущее давление в выходной емкости камеры, которая изолирована. Далее фиксируются полученные кривые давлений (см. рис. 1).

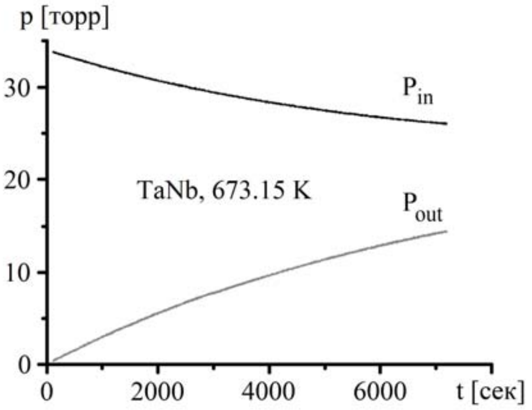


Рис. 1. Экспериментальные данные
Fig. 1. Experimental data

Информативность эксперимента ограничена, поэтому в модели водородопроницаемости учитываются только основные факторы для задачи фильтрации.

Уравнение диффузии с обратимым захватом

Рассмотрим процесс переноса водорода сквозь образец. В течение одного эксперимента температура T постоянна. Концентрация растворенного водорода в атомарном состоянии относительно мала, и можно считать, что диффузионный поток пропорционален градиенту концентрации.

Образец может содержать в себе микродефекты различной природы (например, микрополости), которые могут удерживать водород. Очевидно, что некоторая часть атомов взаимодействует с ними. Ориентируясь на прикладной смысл задачи и возможности метода проницаемости, в качестве модели диффузии с ограниченным захватом в объеме принимается нелинейная система уравнений:

$$
\begin{aligned}
&\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - f(T,z,c),\\
&\frac{\partial z}{\partial t} = f(T,z,c),\\
&f(T,z,c) \equiv a(T)\Big[1-\frac{z}{z_{\max}}\Big]c - a_{\mathrm{out}}(T)z,
\end{aligned}
$$

где $c \equiv c(t,x)$ – концентрация диффундирующего водорода (атомарного); $z \equiv z(t,x)$ – концентрация захваченного диффузанта; $D \equiv D(T)$ – коэффициент диффузии; $a \equiv a_{\mathrm{in}}$ и a_{out} – коэффициенты поглощения и высвобождения атомов H ловушками $\big([a] = 1/\text{с}\big)$; $f(T,z,c)$ – функция стока.

Знак тождества используется как равенство по определению. Остановимся более подробно на функции стока. Скорость поглощения атомов водорода увеличивается с ростом концентрации диффузанта и уменьшается по мере заполнения ловушек. Величину $z_{\max}$ в физически разумном диапазоне можно считать независимым параметром модели. Полагается, что она достаточно мала, так как в эксперименте рассматривается металлический сплав с высокой водородопроницаемостью. Учитывая, что материал для мембранного газоразделения достаточно однороден и захват носит характер малой поправки, принимается пропорция $z_{\max} = \sigma\bar{c}$, где $\bar{c}$ – равновесная концентрация растворенного атомарного диффузионно подвижного водорода. Общая растворимость определяется концентрацией $\bar{c} + z_{\max}$. Эффект захвата учтен без дополнительных параметров в соответствии с ограниченной информативностью метода проницаемости. Температура лежит в диапазоне $T \in [400, 900]$ K. При таких температурах слагаемое $a_{\mathrm{out}}z$ несущественно. Ловушки в тонких мембранах быстро насыщаются и не оказывают значимого влияния на проникающий поток. Поэтому можно считать, что $a_{\mathrm{out}} = 0$, и в качестве функции стока рассматривать: $f(T,z,c) = a\big[1 - zz_{\max}^{-1}\big]c$. Задание значения $a_{\mathrm{out}} > 0$, в случае, когда коэффициент оказывает существенное влияние, технически не усложняет численное решение краевой задачи. Величины D, a зависят от температуры T образца по закону Аррениуса с предэкспоненциальными множителями D_0, a_0 и энергиями активации E_D, E_a (R – универсальная газовая постоянная): $D = D_0 \exp\{-E_D/[RT(t)]\}$, $a = a_0 \exp\{-E_a/[RT(t)]\}$.

Предварительно образец был дегазирован, поэтому начальные данные примут следующий вид: $c(0,x) = 0$, $z(0,x) = 0$, $x \in [0,\ell]$.

Динамические граничные условия

Из баланса потоков получаются следующие нелинейные граничные условия:

$$-\frac{dQ_{\rm in}}{dt} = \big[J_{p_0} - J_0\big]S = -SD\frac{\partial c}{\partial x}\Big|_{x=0}, \qquad (1)$$

$$-\frac{dQ_{\rm out}}{dt} = \big[J_{p_\ell} - J_\ell\big]S = SD\frac{\partial c}{\partial x}\Big|_{x=\ell}, \qquad (2)$$

где $Q_{\rm in}(t)$, $Q_{\rm out}(t)$ – количество атомов водорода во входной емкости (объем $V_{\rm in}$) и выходной емкости (объем $V_{\rm out}$), $c_0(t) \equiv c(t,0)$, $c_\ell(t) \equiv c(t,\ell)$, $J_{p_{0,\ell}}$ – плотности потоков абсорбции (потоки частиц, падающих на поверхности мембраны), $J_{0,\ell}$ – плотности потоков десорбции (выхода) из образца.

Сквозь пластину сплава проходит атомарный водород, поэтому расчет водорода ведется в атомах. Для определения связи между плотностью потока J_p и давлением p обратимся к формуле Герца – Кнудсена: $J_p = p/\sqrt{2\pi mkT}$, где k – постоянная Больцмана; m – масса молекулы водорода.

Основываясь на экспериментальных данных, в качестве единиц измерения примем $[\ell]$ = см, $[p]$ = торр. Следовательно, формулу Герца – Кнудсена можно представить в виде $J_p = \mu p$, где $\mu \equiv \mu(T) \approx 2{,}474 \cdot 10^{22}/\sqrt{T}$ $\big([\mu] = 1_{\rm H_2}/(\text{торр}\,\text{см}^2\,\text{с}),\ [\mathrm{T}] = \mathrm{K}\big)$.

Под воздействием атомов водорода на поверхности мембраны происходят различные процессы: физическая адсорбция, хемосорбция (поглощение водорода, сопровождающееся образованием химического соединения), диссоциация молекул на атомы, растворение и др. [6]. При этом большая часть атомов водорода либо отражается от поверхности пластины, либо прикрепляется к ней, но не проникает внутрь. Таким образом, из всего поступающего потока атомов только малая доля абсорбируется в объем мембраны. В математической модели этот момент отразим множителем s (коэффициент абсорбции). Тогда в качестве результирующего потока абсорбции атомов водорода сквозь поверхность принимается $J_{p_{0,\ell}} = \mu s p_{0,\ell}(t)$.

Теперь рассмотрим плотности потоков десорбции ($J_{0,\ell}$), которые можно записать как $J_{0,\ell} = bc^2_{0,\ell}$, где b – коэффициент десорбции (отклонение от квадратичности существенно лишь при экстремальных температурах). Коэффициенты s и b имеют аррениусовскую зависимость от температуры.

В условиях постоянного давления насыщения $\bar{p}$ молекулярного водорода с обеих сторон мембраны и постоянной температуры устанавливается равновесная концентрация $\bar{c}$ растворенного атомарного диффузионно подвижного водорода. Тогда, если в модели (1), (2) приравнять производные к нулю, получим $\mu s\bar{p} - b\bar{c}^2 = 0 \Rightarrow \bar{c} \propto \sqrt{\bar{p}} : \bar{c} = \Gamma\sqrt{\bar{p}}$, где $\Gamma \equiv \sqrt{\mu s/b}$.

Предложенная модель имеет упрощенный вид, и при необходимости ее можно детализировать [3]. Однако более подробное описание (учет емкости поверхности, образования слоя другого материала, деформации пластины и т. д.) требует большего количества параметров с неизвестными значениями. Исходная задача связана с улучшением технологий газоразделения, для этого достаточно оценить совокупность показателей, фиксируемых в экспериментах проницаемости. В то же время из-за весьма подвижной фазы внедрения водорода измерения с высокой точностью затруднительны. Поэтому в данной работе остановимся на принятой модели, учитывающей лишь лимитирующие факторы.

Уточним условия эксперимента. Объемы входной и выходной емкостей $V_{\rm in,out}$ несколько литров, толщина мембраны ℓ меньше 1 мм, площадь S около 1 см^2, давление напуска $p_0(0)$ несколько десятков торр. Диапазон давлений $[p_{\min}, p_{\max}]$ невелик, для концентрации захваченного диффузанта ограничимся $z_{\max} = \sigma\bar{c}$, где $\sigma \leqslant 0{,}1$. При этом будет выполняться закон Сивертса [7] $\big(\bar{c} + z_{\max} \propto \sqrt{\bar{p}}\big)$, а точнее в пределах экспериментальной точности: $\bar{c} + z_{\max} \approx \bar{c} = \Gamma\sqrt{\bar{p}}$. Известно, что в масштабе времени переноса газ находится в термодинамическом квазиравновесии с поверхностью, поэтому для определения $Q_{\rm in}$, $Q_{\rm out}$ воспользуемся формулой $N = pV/(kT)$, где N – количество частиц газа, занимаемого объемом V при температуре T и давлении p. При необходимости, в зависимости от конкретных условий, можно уточнить $T = T_V$. Единицы измерения в системе СИ $[p]$ = Па, $[V]$ = м^3, $[k]$ = Дж/К) [8]. Учитывая торр = 133,322 Па, Па = Дж/м^3 (формально), можем получить для соответствующих давлений и объемов в граничных условиях (1), (2) следующие выражения $Q = 2N = \alpha pV/T$, $\alpha \approx 1{,}931 \cdot 10^{19}$, где p, V, T – численные значения в выбранных единицах измерения (торр, см^3, К).

Безразмерная форма краевой задачи

Учитывая вышесказанное, мы получим следующую модель:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - a(T)\Big[1 - \frac{z}{z_{\max}}\Big]c,$$
$$\frac{\partial z}{\partial t} = a(T)\Big[1 - \frac{z}{z_{\max}}\Big]c,$$

где начальные данные и граничные условия запишутся в виде:

$$c(0,x) = 0,\ z(0,x) = 0,\ x \in [0,\ell],$$
$$\mu s p_{0,\ell}(t) - b c_{0,\ell}^2 = \mp D\frac{\partial c}{\partial x}\Big|_{x=0,\ell},$$
$$\frac{dQ_{\text{in,out}}}{dt} = -\big[\mu s p_{0,\ell}(t) - b c_{0,\ell}^2\big]S,$$
$$Q_{\text{in,out}} = \alpha p_{0,\ell}(t) V_{\text{in,out}} T^{-1}.$$

Перейдем к безразмерной координате $y = x/\ell$ с концентрациями $u = c/\bar{c}$, $v = z/\bar{c}$, $v_{\max} = z_{\max}/\bar{c}$ ($z_{\max} = \sigma\bar{c} \ll \bar{c}$) и временем τ: $t = D^{-1}\ell^2\tau$ ($[D] = \text{см}^2/\text{с}$). Температура T фиксирована. Равновесную концентрацию диффундирующего водорода $\bar{c} = \Gamma\sqrt{\bar{p}_0}$ и количества атомов $\overline{Q}_{\text{in}} = \alpha\bar{p}_0 V_{\text{in}}/T$, $\overline{Q}_{\text{out}} = \alpha\bar{p}_0 V_{\text{out}}/T$ определяют по соответствующему максимальному давлению $\bar{p}_0 = p_0(0)$. Таким образом, получаем следующую краевую задачу:

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial y^2} - \tilde{a}\Big[1 - \frac{v}{v_{\max}}\Big]u,\ \frac{\partial v}{\partial \tau} = \tilde{a}\Big[1 - \frac{v}{v_{\max}}\Big]u,$$
$$u, v \in [0,1],\ y \in (0,1),\ \tau > 0;$$

$$u(0,y) = 0,\ v(0,y) = 0,\ y \in [0,1],$$
$$W\big[\widetilde{p}_0 - u_0^2\big] = -\frac{\partial u}{\partial y}\Big|_{y=0},\ W\big[\widetilde{p}_1 - u_1^2\big] = \frac{\partial u}{\partial y}\Big|_{y=1}$$
$$\overline{Q}_{\text{in}}\frac{d\widetilde{p}_0}{d\tau} = -\overline{Q}_m W\big[\widetilde{p}_0 - u_0^2\big],\quad \widetilde{p}_0(0) = 1,$$
$$\overline{Q}_{\text{out}}\frac{d\widetilde{p}_1}{d\tau} = -\overline{Q}_m W\big[\widetilde{p}_1 - u_1^2\big],\quad \widetilde{p}_1(0) = 0,$$

где используются следующие обозначения:

$$\tilde{a} \equiv a(T)\ell^2 D^{-1},\quad u_{0,1} \equiv u(\tau,y)|_{y=0,1},$$
$$W \equiv b\bar{c}\ell D^{-1} = \overline{P}_0\ell[D\bar{c}]^{-1},\quad \overline{P}_0 \equiv \mu s\bar{p}_0,$$
$$\overline{Q}_{\text{in}} \equiv \alpha V_{\text{in}}\,\bar{p}_0 T^{-1},\quad \overline{Q}_{\text{out}} \equiv \alpha V_{\text{out}}\,\bar{p}_0 T^{-1}$$
$$\widetilde{p}_{0,1} \equiv p_{0,\ell}(t)\overline{p}_0^{\,-1},\quad \overline{Q}_m \equiv S\ell\bar{c} \equiv V_m\bar{c}.$$

Величина $\overline{Q}_m$ обозначает количество атомов диффузионно подвижного водорода в образце в режиме равновесного насыщения ($\bar{p}_0 = p_0(0)$, температура T постоянна). Транспортный параметр W немаловажен при исследовании метода прорыва (такой вариант метода проницаемости, когда на выходе производится постоянная откачка водорода).

Результаты численного моделирования

Результат аппроксимации экспериментальных данных входного и выходного давления молекулярного водорода по сплаву $Ta_{77}Nb_{23}$ [9] (рис. 1) модельными кривыми представлен на рис. 2. Условия эксперимента: $T = 673{,}15\,\text{К}$, $\ell = 0{,}014\,\text{см}$, $S = 0{,}785\,\text{см}^2$, $V_{\text{in}} = 3000\,\text{см}^3$, $V_{\text{out}} = 1750\,\text{см}^3$.

Как было ранее сказано, в условиях насыщения ($p = \text{const}$, $T = \text{const}$) равновесная концентрация $\bar{c} = \Gamma\sqrt{p}$, где $\Gamma \equiv \sqrt{\mu s/b}$ – коэффициент растворимости. Однако в этом случае учитывается только диффузионно подвижный растворенный атомарный водород. При полной дегазации $\bar{c} + z_{\max} = \Gamma_{\max}\sqrt{p}$, где в условиях сохранения закона Сивертса $\Gamma_{\max} = \Gamma[1+\sigma]$. В таком случае аппроксимация кривых давления примет вид как на рис. 3.

Можно заметить, что на рис. 3 с гипотетическим значением $z_{\max} = 10\,\bar{c}$ (в отличие от рис. 2) общая растворимость на порядок больше. Если в эксперименте прорыва ($p_0(t) = p = \text{const}$, на выходе – вакуум) рассматривать установившийся поток водородопроницаемости $J = -D\partial_x c$, то при условии $c_0 = \bar{c}_0 = \Gamma\sqrt{p}$, $c_\ell = 0$ получим $J = \Phi\sqrt{p}/\ell$, $\Phi \equiv D\Gamma = D\sqrt{\mu s/b}$. Исходя из формулы, можно провести серию экспериментов при $T = \text{const}$, варьируя p и фиксируя пропорциональность $J \propto \sqrt{p}$. Далее по полученным значениям p, J определить коэффициент водородопроницаемости Φ, который характеризуется диффундирующим водородом (в модели $\Phi = D\Gamma$).

Если же Φ вычислять напрямую, по имеющимся независимым значениям коэффициентов диффузии и растворимости (эксперименты различны и растворимость общая, с учетом захваченного водорода), то можем получить $D\Gamma_{\max} \gg D\Gamma$. Возможно, это является одной из причин расхождения литературных данных по растворимости и проницаемости. Проблема заключается в том, что для режима проницаемости в тонких мембранах сложно определить количество водорода, захваченного в ловушки, даже при большой их емкости, хотя они оказывают значительное влияние на общую растворимость в материале (при пересчете на см^3 и м^3 в условиях $\ell, S \ll 1$).

Рассмотрим динамику объемных концентраций в эксперименте. Проанализируем приповерхностные концентрации ($x = 0, \ell$), изображенные на рис. 4, 5.

Ловушки заполняются в течение минуты $z_0 \approx z_\ell \approx z_{\max}$, и концентрации $c_0(t) = c(t, 0)$, $c_\ell(t) = c(t, \ell)$ быстро приходят в состояние равновесия. Однако в масштабе двух часов заметна смена переходного режима всплеска на медленный тренд. За счет малой толщины мембраны ($\ell \ll 1$ мм) и быстрой водородопроницаемости сплава время стабилизации c_ℓ сравнимо по отношению к c_0. Диффузия ℓ^2/D ($D \sim 10^{-5}$) происходит в пределах десяти секунд, поэтому переходные процессы происходят с достаточно высокой скоростью. Общая концентрация задается суммой $c + z$.

Из соотношений $\mu s p_{0,\ell} = b c_{0,\ell}^2$ можно определить квазиравновесные концентрации диффундирующего водорода $\bar{c}_{0,\ell}(t) = \Gamma\sqrt{p_{0,\ell}(t)}$. С учетом захвата водорода ловушками в качестве концентраций будем рассматривать их суммы $\bar{c}_{0,\ell}(t) + z_{\max}$ и $c_{0,\ell}(t) + z_{0,\ell}(t)$.

Ориентируясь на рис. 5, можно оценить разницу между рассогласованием концентраций на выходе ($c_\ell - \bar{c}_\ell$) и входе ($\bar{c}_0 - c_0$). Заметно, что первое больше последнего. Однако по мере приближения к равновесию происходит сближение этих величин.

В текущей задаче газоразделения нас интересует проникающий поток. Но переход к квазиравновесным оценкам ухудшает применение системы разностных схем для аппроксимации градиента $\partial_x c$ как на входе ($\bar{c}_0 > c_0$), так и на выходе ($\bar{c}_\ell < c_\ell$).

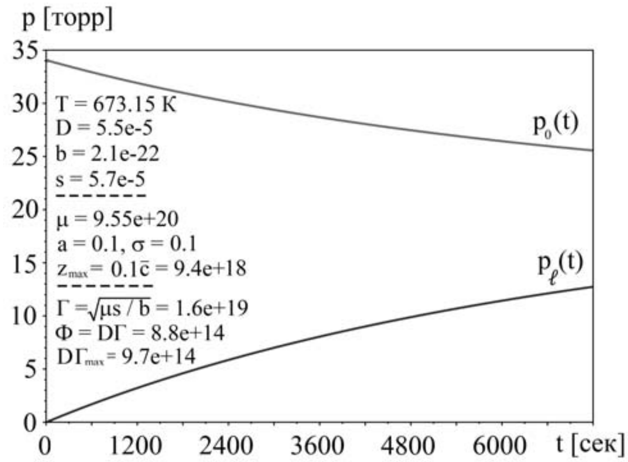


Рис. 2. Аппроксимация давлений, 673,15 K
Fig. 2. Pressure approximation, 673.15 K

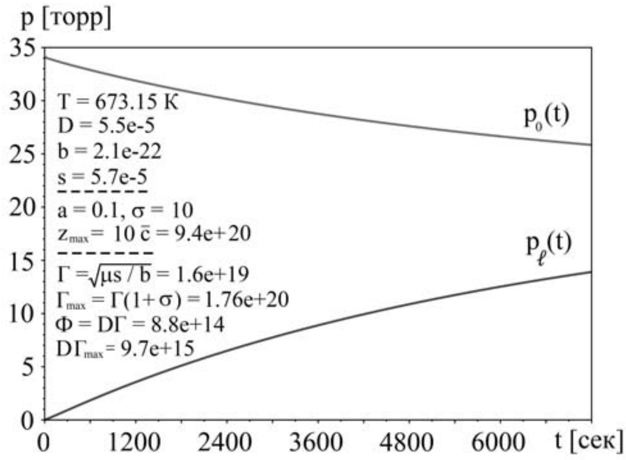


Рис. 3. Давления в случае $z_{\max} = 10\,\bar{c}$
Fig. 3. Pressure in case of $z_{\max} = 10\,\bar{c}$

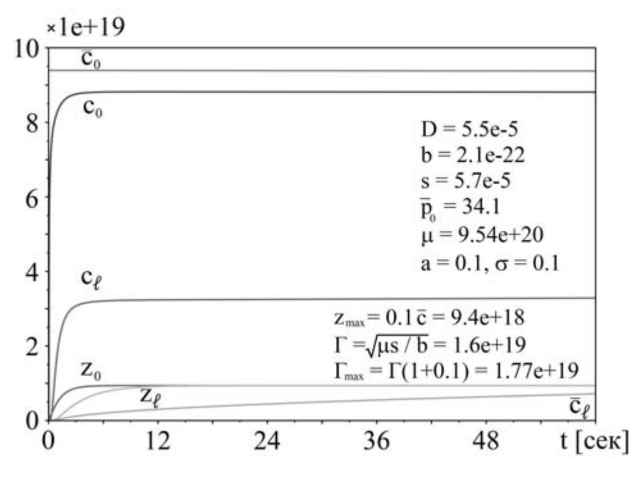


Рис. 4. Приповерхностные концентрации, 1 мин
Fig. 4. Near-surface concentrations, 1 min

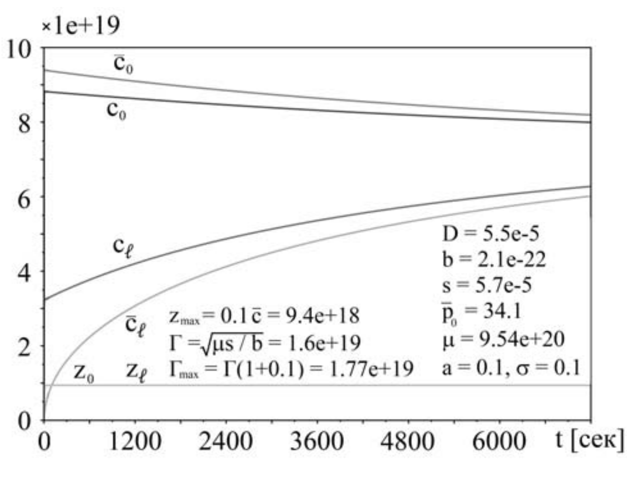


Рис. 5. Приповерхностные концентрации, 120 мин
Fig. 5. Near-surface concentrations, 120 min

Модель быстрой водородопроницаемости

Как было сказано выше, подстановка равновесных концентраций приводит к заметной погрешности при толщине мембраны $\ell \ll 1$ мм и быстрой проницаемости сплава. Поэтому рассмотрим задачу моделирования концентраций $c_{0,\ell}$ по кривым давления $p_{0,\ell}$ без квазиравновесного упрощения $c(t) = \Gamma\sqrt{p(t)}$.

В течение короткого (в масштабе эксперимента) периода t_0 ловушки насыщаются, и наступает квазистационарный режим. Применим аппроксимацию $\partial_x c = -[c_0 - c_\ell]\ell^{-1}$:

$$\dot{p}_{0,\ell}(t) = \mp\beta_{0,\ell}\big[c_0 - c_\ell\big], \tag{3}$$

$$\mu s p_{0,\ell}(t) - bc_{0,\ell}^2 = \pm D\ell^{-1}\big[c_0 - c_\ell\big], \tag{4}$$

$$\beta_{0,\ell} \equiv SD\left[\alpha V_{\mathrm{in,out}}\ell\right]^{-1}T,\ t \geqslant t_0 > 0.$$

Рассмотрим систему (4). Выполним преобразования уравнения при $x = 0$ следующим образом:

$$\mu s p_0(t) - bc_0^2 = D\big[c_0 - c_\ell\big]\ell^{-1},$$
$$\mu s p_0(t) - bc_0^2 - D\ell^{-1}c_0 = -D\ell^{-1}c_\ell,$$
$$\frac{\mu s p_0(t)}{b} - c_0^2 - \frac{D}{b\ell}c_0 = -\frac{D}{b\ell}c_\ell.$$

Введем коэффициент растворимости $\Gamma \equiv \sqrt{\mu s/b}$ в полученную систему и выделим полный квадрат относительно $c_0(t)$:

$$\Gamma^2 p_0(t) - \left[c_0 + \frac{D}{2b\ell}\right]^2 + \frac{D^2}{4b^2\ell^2} = -\frac{D}{b\ell}c_\ell,$$
$$\Gamma^2 p_0(t) + \frac{D^2}{4b^2\ell^2} + \frac{D}{b\ell}c_\ell = \left[c_0 + \frac{D}{2b\ell}\right]^2.$$

В левой части уравнения выделим выражение относительно $c_\ell(t)$ аналогично выражению, возводимому в квадрат $(c_0 + D(2b\ell)^{-1})$:

$$\Gamma^2 p_0(t) - \frac{D^2}{4b^2\ell^2} + \frac{D}{b\ell}\left[c_\ell + \frac{D}{2b\ell}\right] = \left[c_0 + \frac{D}{2b\ell}\right]^2.$$

Домножим уравнение на $(2b\ell)^2D^{-2}$:

$$4(\Gamma b\ell)^2 p_0(t)D^{-2} - 1 + 2\left[2b\ell D^{-1}c_\ell + 1\right] = \left[2b\ell D^{-1}c_0 + 1\right]^2. \tag{5}$$

Выполним аналогичные преобразования уравнения системы (4) при $x = \ell$. В результате получим уравнение вида (5), в котором величины c_0 и c_ℓ заменяются на c_ℓ и c_0 соответственно. Введем безразмерные переменные:

$$X_{0,\ell}(t) = 1 + 2\ell c_{0,\ell}(t)bD^{-1},$$
$$a_{0,\ell}(t) = 4(\ell\Gamma b)^2 p_{0,\ell}(t)D^{-2} - 1. \tag{6}$$

Таким образом, система уравнений (4) запишется в симметричном виде:

$$a_0 + 2X_\ell = X_0^2,\ a_\ell + 2X_0 = X_\ell^2. \tag{7}$$

Для переменной $X \equiv X_\ell$ получаем неполное уравнение четвертой степени

$$\big[X^2 - a_\ell\big]^2 = 4\big[2X + a_0\big], \tag{8}$$

которое решается в радикалах. Исходя из физических соображений, нас интересует положительный корень. Выведем дифференциальные уравнения для $X_{0,\ell}$, поскольку информация о динамике граничных концентраций $c_{0,\ell}$ представляет и самостоятельный интерес. Продифференцируем выражения:

$$\dot{a}_0 + 2\dot{X}_\ell = 2X_0\dot{X}_0,\ \dot{a}_\ell + 2\dot{X}_0 = 2X_\ell\dot{X}_\ell. \tag{9}$$

Решая систему (9) как систему линейных уравнений относительно производных $\dot{X}_0$ и $\dot{X}_\ell$, получим:

$$\dot{X}_0 = \frac{\dot{a}_0 X_\ell + \dot{a}_\ell}{2[X_0X_\ell - 1]}, \tag{10}$$

$$\dot{X}_\ell = \frac{\dot{a}_\ell X_0 + \dot{a}_0}{2[X_0X_\ell - 1]}. \tag{11}$$

Используя выражения (6) и (3), получаем:

$$\dot{a}_0 = 4(\ell\Gamma b)^2D^{-2}\dot{p}_0(t) = -4\beta_0[c_0 - c_\ell](\ell\Gamma b)^2D^{-2}. \tag{12}$$

Из уравнений (6) найдем:

$$X_0 - X_\ell = 2bD^{-1}\ell\big[c_0(t) - c_\ell(t)\big],$$

откуда, выразив $c_0(t) - c_\ell(t)$ и подставив в (12), получим:

$$\dot{a}_0 = -2\beta_0[X_0 - X_\ell]\ell bD^{-1}\Gamma^2.$$

После подстановки выражений

$$\beta_0 \equiv SDT(\alpha V_{\mathrm{in}}\ell)^{-1},\ \Gamma \equiv \sqrt{\mu s/b}$$

в полученное уравнение будем иметь:

$$\dot{a}_0 = -2[X_0 - X_\ell]sM_0, \quad M_0 \equiv \frac{\mu ST}{\alpha V_{\mathrm{in}}}.$$

Аналогичным образом получаем:

$$\dot{a}_\ell = 2[X_0 - X_\ell]sM_\ell, \quad M_\ell \equiv \frac{\mu ST}{\alpha V_{\mathrm{out}}}.$$

Подставив полученные выражения в (10), будем иметь:

$$\dot{X}_0 = -sM_0[X_0 - X_\ell]\frac{X_\ell - M_\ell M_0^{-1}}{X_0X_\ell - 1}.$$

Выражение для $\dot{X}_\ell$ выводится аналогично. В результате мы имеем следующую систему дифференциальных уравнений ($\mu \propto 1/\sqrt{T}$):

$$\dot{X}_0(t) = -sM_0\big[X_0 - X_\ell\big]\frac{X_\ell - V_{\rm in}V_{\rm out}^{-1}}{X_0X_\ell - 1}, \qquad (13)$$

$$\dot{X}_\ell(t) = sM_\ell\big[X_0 - X_\ell\big]\frac{X_0 - V_{\rm in}^{-1}V_{\rm out}}{X_0X_\ell - 1}. \qquad (14)$$

$$M_\ell \equiv \mu ST(\alpha V_{\rm out})^{-1}, \quad M_0 \equiv \mu ST(\alpha V_{\rm in})^{-1}.$$

Локализация начальных данных для системы ОДУ

Рассмотрим более подробно уравнение четвертой степени (8). При помощи стандартных математических пакетов можно получить четыре корня, из которых нужно выбрать соответствующий физическому смыслу. Возникает задача автоматизации выбора корня.

Итак, рассмотрим уравнение

$$\big[X^2 - a_\ell\big]^2 = 4\big[2X + a_0\big],$$

где $X \equiv 1 + 2\ell c_\ell(t_0)bD^{-1}$, $a_{0,\ell}(t_0) \equiv 4(\ell\Gamma b)^2 p_{0,\ell}(t_0)D^{-2} - 1$, t_0 – момент времени после быстрого переходного периода, когда квазистационар практически установился ($\bar{c}_{0,\ell}(t_0) = \Gamma\sqrt{p_{0,\ell}(t_0)}$). Так как

$$2X + a_0 = 1 + \frac{4\ell b}{D}c_\ell(t_0) + \frac{4(\ell b)^2}{D^2}\bar{c}_0^2(t_0) > 1,$$

то вместо уравнения четвертой степени (8) будем решать уравнение $X^2 - a_\ell = 2\sqrt{2X + a_0}$.

Выполняя аналогичные вычисления, получим:

$$X^2 - a_\ell = 2X + 4\ell^2 b^2 D^{-2}(c_\ell^2(t_0) - \bar{c}_\ell^2(t_0)).$$

На рис. 5 видно, что $c_\ell(t_0) > \bar{c}_\ell(t_0)$, следовательно,

$$X^2 - a_\ell > 2X > 2,$$
$$X > 1 + \sqrt{1 + a_\ell} = 1 + 2\ell\Gamma\sqrt{p_\ell(t_0)}bD^{-1}.$$

Полученное неравенство совпадает с оценкой X в (6), если вместо c_ℓ подставить $\bar{c}_\ell$. Таким образом, данная нижняя оценка X для принятой модели неулучшаема.

Для автоматизации графического решения уравнения $X^2 - a_\ell = 2\sqrt{2X + a_0}$ найдем верхнюю оценку X.

Образец исследуемого материала разбивает вакуумную камеру на два сосуда. Обозначим условно их как левый (входной) объем V_1 и правый (выходной) объем V_2. При этом символы $V_{1,2}$ будем использовать как для самих сосудов, так и для их объемов (в зависимости от контекста). Со временем давление водорода в объеме V_2 будет увеличиваться до определенного уровня $\bar{p} < \bar{p}_0(0)$, следовательно, $\max c_\ell < \Gamma\sqrt{\bar{p}_0(0)}$. Значит, можно получить верхнюю оценку: $X_{\max} < 1 + 2\ell\Gamma\sqrt{\bar{p}_0(0)}bD^{-1}$. Однако такая оценка – грубая, постараемся уточнить значение $\bar{p}$.

Перед началом эксперимента проводится дегазация, объемы V_1 и V_2 – вакуумированы. Затем в начальный момент времени ($t_0 = 0$) в V_1 скачкообразно создается давление молекулярного водорода ($p_0 = p_0(0)$). Обычно $p_0 \approx 10$–100 Торр, так что газ можно считать практически идеальным. В таком случае, воспользовавшись основным уравнением молекулярно-кинетической теории, получим $p_0V_1 = n_0kT$, где n_0 – число молекул H_2 в начальный момент времени; k – постоянная Больцмана; T – абсолютная температура газа, которая принимает постоянное значение в исследуемом эксперименте ($T = \text{const}$).

В процессе продолжения эксперимента все большая часть молекул водорода переходит в V_2. Поскольку исследуемая пластина сплава пренебрежительно мала по объему и нас интересует верхняя оценка давления в V_2, будем считать, что $n_0 = n_1 + n_2$, где n_1, n_2 – число молекул в V_1 и V_2 соответственно, когда давления в $V_{1,2}$ практически равны ($\bar{p}$).

Таким образом, для каждого сосуда получаем уравнения:

$$\bar{p}V_1 = n_1kT, \quad \bar{p}V_2 = n_2kT,$$

из которых можно выразить $\bar{p}$:

$$\bar{p}(V_1 + V_2) = n_0kT = \frac{p_0V_1}{kT}kT$$
$$= p_0V_1 \quad \Rightarrow \quad \bar{p} = \frac{p_0V_1}{V_1 + V_2}.$$

Зная $\bar{p}$, можно уточнить верхнюю оценку

$$\max \bar{c}_\ell < \Gamma\sqrt{\bar{p}} \quad \Rightarrow \quad X < 1 + 2\ell bD^{-1}\Gamma\sqrt{\bar{p}}.$$

Рассмотрим уравнение $X^2 - a_\ell = 2\sqrt{2X + a_0}$ на отрезке $[\min X, \max X]$. Определим количество существующих решений на данном интервале. Если будут выполняться условия:

$$\min X^2 - a_\ell < 2\sqrt{2\min X + a_0},$$
$$\max X^2 - a_\ell > 2\sqrt{2\max X + a_0},$$

тогда на указанном промежутке существует единственное решение.

Итак, проверим первое условие. Для этого вычислим:

$$\begin{aligned}&\min X^2 - a_\ell\\&= (1 + 2\ell\Gamma\sqrt{p_\ell} bD^{-1})^2 - 4(\ell\Gamma b)^2 p_\ell D^{-2} + 1\\&= 2\min X,\end{aligned}$$

$$\begin{aligned}&2\sqrt{2\min X + a_0}\\&= 2\sqrt{1 + 4\ell\Gamma\sqrt{p_\ell} bD^{-1} + 4(\ell\Gamma b)^2 p_0 D^{-2}}\\&> 2\sqrt{1 + 4\ell\Gamma\sqrt{p_\ell} bD^{-1} + 4(\ell\Gamma b)^2 p_\ell D^{-2}}\\&= 2\min X.\end{aligned}$$

Таким образом, мы получили, что $2\sqrt{2\min X + a_0} > 2\min X = \min X^2 - a_\ell$. Рассмотрим второе условие:

$$\begin{aligned}&\max X^2 - a_\ell\\&= (1 + 2\ell\Gamma b\sqrt{\bar{p}} D^{-1})^2 - 4(\ell\Gamma b)^2 p_\ell D^{-2}\\&> 2(1 + 2\ell\Gamma b\sqrt{p_\ell} D^{-1}),\end{aligned}$$

$$\begin{aligned}&2\sqrt{2\max X + a_0}\\&= 2\sqrt{2 + 4\ell\Gamma b\sqrt{\bar{p}} D^{-1} + 4(\ell\Gamma b)^2 p_0 D^{-2} - 1}\\&< 2\sqrt{2 + 4\ell\Gamma b\sqrt{p_\ell} D^{-1} + 4(\ell\Gamma b)^2 p_\ell D^{-2} - 1}\\&= 2(1 + 2\ell\Gamma b\sqrt{p_\ell} D^{-1}).\end{aligned}$$

Видим, что $2\sqrt{2\max X + a_0} < \max X^2 - a_\ell$. Для доказательства данного неравенства мы ориентировались на то, что $p_0 > p_\ell$ (см. рис. 2). Посчитаем производные левой и правой части уравнения (8):

$$f_1' = [X^2 - a_\ell]' = 2X > 2,$$

$$f_2' = [2\sqrt{2X + a_0}]' = 2\frac{2}{2\sqrt{2X + a_0}} < 2.$$

Получаем, что $f_2' < f_1'$, $f_1(\min X) < f_2(\min X)$ и $f_1(\max X) > f_2(\max X)$. Это доказывает нам, что на промежутке $[\min X, \max X]$ существует единственное пересечение кривых.

Пример графической локализации искомого решения X приведен на рис. 6. Данные для вычислений брались из статьи [5]: $T = 673{,}15$ К; $V_1 = 1750$ см^3; $V_2 = 3000$ см^3; $D = 5{,}5 \times 10^{-5}$ см^2/сек; $b = 2{,}1 \times 10^{-22}$ см^4/сек; $s = 5{,}7 \times 10^{-5}$; $\mu = 2{,}474 \times 10^{22}/\sqrt{T}$; $\Gamma = \sqrt{\mu s/b}$; $\ell = 0{,}014$ см; $p_0(t_0) = 32$ Торр; $p_\ell(t_0) = 2$ Торр; $\bar{p} = \frac{p_0 V_1}{V_1 + V_2}$; $\bar{c}_0 = \Gamma\sqrt{p_0}$; $\bar{c}_\ell = \Gamma\sqrt{p_\ell}$.

При использовании пакетов прикладных программ (в данной работе использовался Scilab 6) целесообразнее рассматривать алгебраическое уравнение четвертой степени (8). При этом возникают еще три дополнительных корня. Ограничение $X \in [\min X, \max X]$ позволит автоматически выделить единственный корень уравнения, подходящий по физическому смыслу задачи.

После определения значения $X = X_\ell(t_0)$ находим $X_0(t_0)$ из соотношений (7).

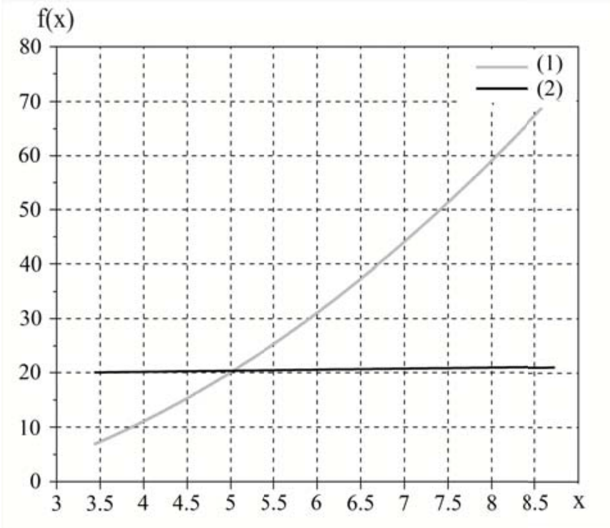


Рис. 6. Графическое решение уравнения: (1) $\max X^2 - a_\ell$, (2) $2\sqrt{2X + a_0}$

Fig. 6. Graphical solution of the equation: (1) $\max X^2 - a_\ell$, (2) $2\sqrt{2X + a_0}$

Алгоритм численного решения краевой задачи быстрой водородопроницаемости

Теперь представим схему численного моделирования быстрой водородопроницаемости на конкретной задаче.

1. Заказчик (ИМЕТ УрО РАН) представляет файл в формате .opj (для работы используется ПО OriginViewer) с экспериментальными данными (рис. 1).

Графики соответствуют давлениям: убывающая кривая соответствует входному объему V_1, возрастающая – выходному V_2. Например, для $Ta_{77}Nb_{23}$ $p_0 = 30$ торр, $T = 673{,}15$ К (см. рис. 2). Исходные данные содержатся в табличном виде и зашумлены. Изначально они экспортируются в Excel, а затем в пакет Scilab 6.

Задача параметрической идентификации модели состоит в оценке значений параметров D, b, s (коэффициенты диффузии, десорбции, абсорбции).

2. С учетом переходных процессов выделяем момент времени $t_0 > 0$, когда распределение концентрации растворимого водорода можно считать практически квазистационар-

ным. Вычислительные эксперименты с полной моделью показывают, что квазистационар при заданных условиях достигается в пределах минуты (на фоне измерений в течение часов).

Задаем текущие приближения D, b, s. Определяем $p(t_0)$. Сглаживаем экспериментальные данные, используя алгоритм, описанный в [1]. Находим начальные данные $X_0(t_0)$, $X_\ell(t_0)$ для модели быстрой водородопроницаемости (13)–(14). Система ОДУ интегрируется в пакете Scilab 6 на отрезке времени $[t_0, t_*]$ (t_* – заданное время, в частности, время окончания измерений). Полученные интегральные кривые представлены на рис. 7.

3. В силу полученных соотношений

$$X_0(t) = 1 + 2\ell c_0(t) b D^{-1},$$
$$X_\ell(t) = 1 + 2\ell c_\ell(t) b D^{-1}$$

находим концентрации $c_0(t)$, $c_\ell(t)$, $t \geqslant t_0$. Из граничных условий (4) генерируем модельные зависимости $p_{0,\ell}(t)$ и сравниваем их с экспериментальными данными (см. рис. 8).

4. В зависимости от полученного результата варьируем значения исходных параметров. Вычислительные эксперименты показывают характерные зависимости изменений графиков при изменении параметров модели (см. рис. 8–11).

5. При неудовлетворительной аппроксимации возвращаемся к п. 2.

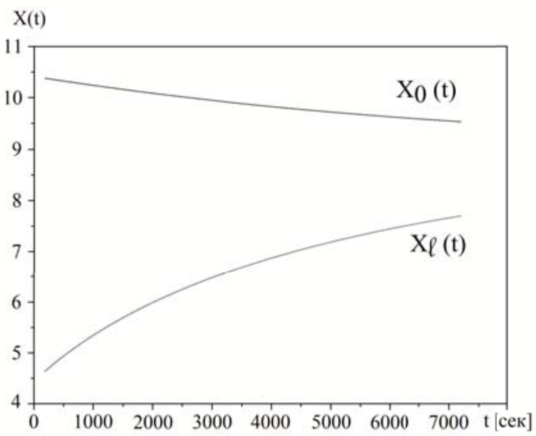


Рис. 7. Интегральные кривые системы ОДУ
Fig. 7. Integral curves of the ODE system

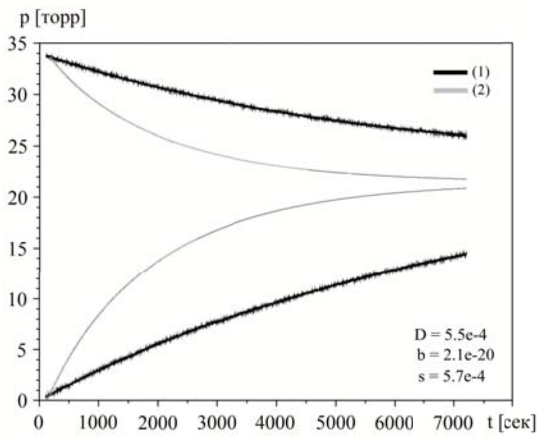


Рис. 8. Результат численного моделирования.
Здесь и на рис. 9–12: (1) – экспериментальные данные, (2) – модель
Fig. 8. The result of numerical simulation.
Here and in Fig. 9–12: (1) – experimental data, (2) – model

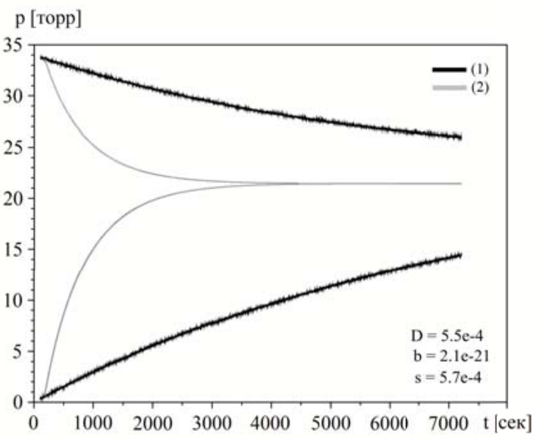


Рис. 9. Изменение коэффициента десорбции
Fig. 9. Change in the desorption coefficient

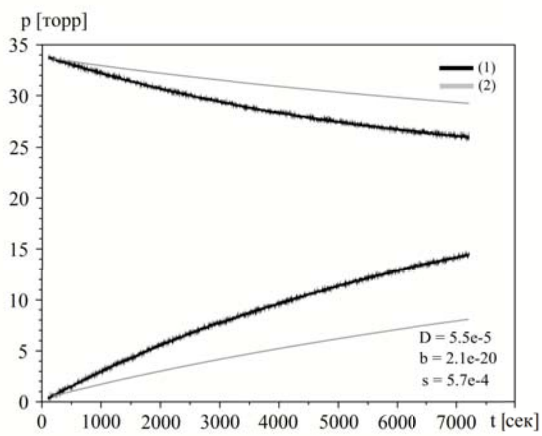


Рис. 10. Изменение коэффициента диффузии
Fig. 10. Change in the diffusion coefficient

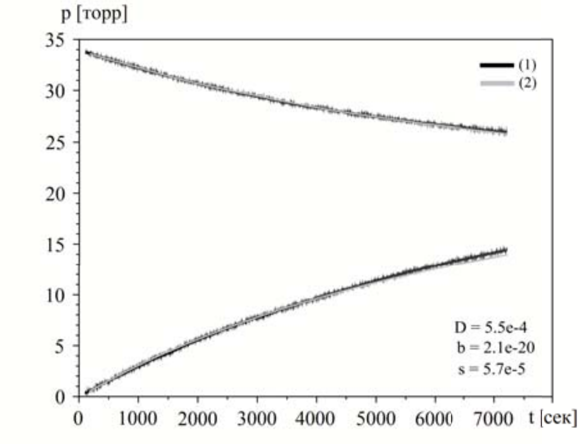


Рис. 11. Изменение коэффициента абсорбции
Fig. 11. Change in the absorption coefficient

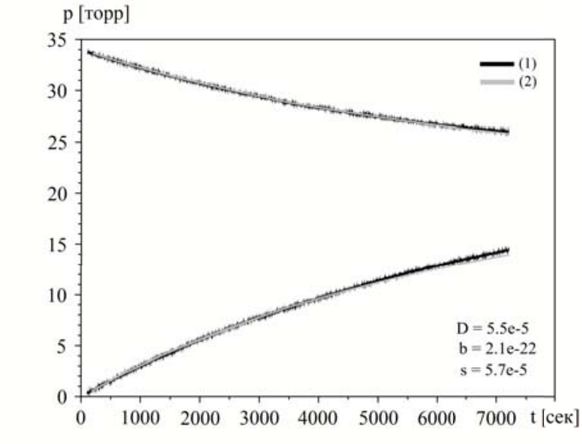


Рис. 12. Результаты численного моделирования
Fig. 12. Numerical simulation results

Заключение

Изучение взаимодействия изотопов водорода с различными конструкционными материалами широко распространено и применяется во многих реальных задачах. Мгновенно определить значение параметров и коэффициентов модели в ходе эксперимента практически невозможно. При этом их значения необходимы для дальнейших исследований. Математическое моделирование позволяет решить данную проблему.

В рамках текущей работы было автоматизировано нахождение единственного корня уравнения четвертой степени (8), который удовлетворяет физическим условиям поставленной задачи, что, в свою очередь, привело к упрощению нахождения начальных условий для системы ОДУ (13)–(14). Решение указанной системы позволяет построить графики давлений во входной и выходной емкости камеры. Результат представлен на рис. 12.

Аппроксимация кривых давлений производилась варьированием коэффициентов диффузии, абсорбции и десорбции из физических соображений. Один из вариантов дальнейшего развития – автоматизация разработанной схемы решения обратной задачи параметрической идентификации модели на основе экспериментальных данных.

Литература

1. *Волков Е. А.* Численные методы: учеб. пособие для вузов. М.: Наука, 1987. 248 с.

2. *Гордиенко Ю. Н., Кульсартов Т. В., Заурбекова Ж. А., Понкратов Ю. В., Гныря В. С., Никитенков Н. Н.* Применение метода водородопроницаемости в реакторных экспериментах по исследованию взаимодействия изотопов водорода с конструкционными материалами // Известия Томского политехнического университета. Инжиниринг георесурсов. 2014. Т. 324, № 2. С. 149–162.

3. *Заика Ю. В.* Интегральные операторы прогнозирования и идентификации моделей водородопроницаемости. Петрозаводск: КарНЦ РАН, 2013. 505 с.

4. *Заика Ю. В., Родченкова Н. И.* Краевая задача водородопроницаемости мембран газоразделения // Труды Карельского научного центра РАН. 2015. № 10. С. 55–68. doi: 10.17076/mat133

5. *Заика Ю. В., Родченкова Н. И.* Краевая задача со свободной границей: гидрирование циркониевого сплава // Труды Карельского научного центра РАН. 2016. № 8. С. 24–33. doi: 10.17076/mat392

6. *Колачев Б. А., Ильин А. А., Лавренко В. А., Левинский Ю. В.* Гидридные системы: Справ. изд. М.: Металлургия, 1992. 352 с.

7. *Писарев А. А., Цветков И. В., Маренков Е. Д., Ярко С. С.* Проницаемость водорода через металлы: учебн. пособие. М.: МИФИ, 2008. 144 с.

8. *Чертов А. Г.* Единицы физических величин: учебное пособие для вузов. М.: Высшая школа, 1977. 287 с.

9. *Kojakhmetov S., Sidorov N., Piven V., Sipatov I., Gabis I., Arinov B.* Alloys based on 5 group metals for hydrogen purification membranes // J. Alloys Compd. 2015. Vol. 645, suppl. 1. P. S36–S40. doi: 10.1016/j.jallcom.2015.01.242

References

1. *Volkov E. A.* Numerical methods: a university textbook. Moscow: Nauka; 1987. 248 p. (In Russ.)

2. *Gordienko Yu. N., Kul'sartov T. V., Zaurbekova Zh. A., Ponkratov Yu. V., Gnyrya V. S., Nikitenkov N. N.* Application of hydrogen permeation method in reactor experiments on investigation of hydrogen isotopeinteraction with structural

materials. *Izvestiya Tomskogo politekhnicheskogo universiteta = Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering.* 2014;2:149–162. (In Russ.)

3. *Zaika Y. V.* Integral operators for forecasting and identification of hydrogen permeability models. Petrozavodsk: KarRC RAS; 2013. 505 p. (In Russ.)

4. *Zaika Y. V., Rodchenkova N. I.* Boundary-value problem of hydrogen permeability of gas separation membranes. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2015;10:55–68. (In Russ.). doi: 10.17076/mat133

5. *Zaika Y. V., Rodchenkova N. I.* Boundary-value problem with free boundary: zirconium alloy hydrogenation. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2016;8:24–33. (In Russ.). doi: 10.17076/mat392

6. *Kolachev B. A., Il'in A. A., Lavrenko V. A., Levinskii Yu. V.* Hydride systems: a reference edition. Moscow: Metallurgiya; 1992. 352 p. (In Russ.)

7. *Pisarev A. A., Tsvetkov I. V., Marenkov E. D., Yarko S. S.* The permeability of hydrogen through metals. Moscow: MIFI; 2008. 144 p. (In Russ.)

8. *Chertov A. G.* Units of physical quantities: a university textbook. Moscow: Vysshaya shkola; 1977. 287 p. (In Russ.)

9. *Kojakhmetov S., Sidorov N., Piven V., Sipatov I., Gabis I., Arinov B.* Alloys based on 5 group metals for hydrogen purification membranes. *J. Alloys Compd.* 2015;645(1):S36–S40. doi: 10.1016/j.jallcom.2015.01.242

Поступила в редакцию / received: 05.05.2025; принята к публикации / accepted: 02.06.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Грудова Ксения Васильевна
младший научный сотрудник

e-mail: grudova@krc.karelia.ru

CONTRIBUTOR:

Grudova, Kseniya
Junior Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 17–23
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 17–23
DOI: 10.17076/mat2090

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.837

СТРАТЕГИИ ОПТИМАЛЬНОЙ ОСТАНОВКИ В ИГРЕ НА РАЗОРЕНИЕ С ЗАТРАТАМИ НА КАЖДОМ ШАГЕ

А. А. Ивашко

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Рассмотрена многошаговая игра двух лиц с конечным горизонтом, связанная с задачей о разорении. На каждом из n шагов два игрока, имеющие разный начальный капитал, разыгрывают единицу стоимости. Предполагается, что игроки несимметричны и имеют неодинаковые шансы на победу на каждом шаге. Игрок выигрывает в данной игре, если капитал его противника закончился, то есть произошло его разорение. Выигрыши игроков определяются в конце игры. При этом в выигрыше игрока учитываются затраты размером c, которые игрок понес на каждом шаге игры. Если противник разорился на шаге τ, то игрок получает выигрыш $1 - c\tau$. Если в течение промежутка времени n игра не закончилась, то выигрыш равен $-cn$. Рассмотрены различные варианты задачи: с неограниченным капиталом у одного из игроков и неограниченным капиталом у обоих игроков. Стратегией игрока является момент остановки игры для максимизации своего ожидаемого выигрыша. Для вычисления вероятности разорения и ожидаемых выигрышей игроков используются свойства несимметричного случайного блуждания, описывающего процесс данной игры. Найдены оптимальные стратегии остановки и ожидаемые выигрыши игроков с помощью метода динамического программирования. Проведено сравнение выигрышей в задаче без возможности остановиться до конечного момента времени n и с использованием стратегии оптимальной остановки. В задаче с неограниченным капиталом у обоих игроков оптимальные стратегии остановки были найдены с помощью процедуры наилучших ответов. Проведено численное моделирование полученных результатов для различных значений параметров задачи.

Ключевые слова: задача о разорении; случайное блуждание; оптимальная стратегия; метод наилучших ответов; игра двух лиц

Для цитирования: Ивашко А. А. Стратегии оптимальной остановки в игре на разорение с затратами на каждом шаге // Труды Карельского научного центра РАН. 2025. № 4. С. 17–23. doi: 10.17076/mat2090

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

A. A. Ivashko. OPTIMAL STOPPING STRATEGIES IN GAMBLER'S RUIN GAME WITH COSTS AT EACH STEP

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

We consider a two-person multi-stage finite-horizon game related to the ruin problem. At each of the n stages, two players with different initial capitals play one unit of capital. It is assumed that the players are asymmetric and have unequal chances of winning at each stage. A player wins if their opponent runs out of capital, i.e. is ruined. The players' payoffs are determined at the end of the game taking into account the costs c that the player incurred at each stage of the game. If the opponent goes broke at stage τ, then the player receives a payoff of $1 - c\tau$. If the game has not ended within the n interval, the players' payoffs are $-cn$. Various scenarios are examined: with one player having unlimited capital while the other has limited capital, and with both players having unlimited capital. The player's strategy is to stop the game so as to maximize their expected payoff. The ruin probability and the players' optimal stopping strategies are calculated using the properties of an asymmetric random walk describing the process of the game. Optimal stopping strategies and expected payoffs of the players were found using the dynamic programming method. Payoffs were compared in the problem without the option of stopping before the final time n and using the optimal stopping strategy. In the problem with both playershaving unlimited capital, optimal stopping strategies were found using the best response procedure. Numerical simulations of the obtained results are presented for different values of the problem parameters.

Keywords: gambler's ruin game; random walk; optimal stopping; best response algorithm; two-person game

For citation: Ivashko A. A. Optimal stopping strategies in gambler's ruin game with costs at each step. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 17–23. doi: 10.17076/mat2090

Funding. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

Введение

Рассматривается многошаговая игра двух лиц в дискретном времени и с конечным горизонтом n, связанная с задачей о разорении [3, 9]. Пусть у каждого из двух игроков (I и II) имеется некоторое количество капитала. Игроки на каждом шаге игры разыгрывают единицу капитала. Предполагается, что у игрока I вероятность выиграть равна p, а у игрока II – $q = 1 - p$. Соответственно, на каждом шаге капитал игрока либо увеличивается на единицу, либо уменьшается. Игра длится до тех пор, пока не закончится капитал у одного из игроков либо пока время игры не истечет. Выигрыши игроков определяются в конце игры следующим образом. Если капитал какого-то игрока закончился в момент времени τ, то есть произошло его разорение, другой игрок получает выигрыш, равный $1 - c\tau$, где c – затраты игрока на каждом шаге игры, $0 \leqslant c < 1$. Если в течение промежутка времени n игра не закончилась, то выигрыш равен $-cn$. В данной задаче необходимо найти оптимальные стратегии остановки для максимизации ожидаемых выигрышей игроков.

Задача о разорении игрока широко известна в литературе (см. обзор [1, 10]). Исследованы различные постановки, для которых найдены вероятности разорения игрока и ожидаемое время до окончания игры. Задачи с двумя игроками рассмотрены в работах [2, 4]. Постановки, в которых несколько игроков играют до разорения одного из них, исследованы в [5, 8].

Варианты задачи о разорении, в которых вероятность победы игрока на каждом следующем шаге зависит от соотношения побед и поражений на предыдущих шагах, были рассмотрены в работах [6, 7].

В данной работе исследованы две постановки задачи: а) у одного из игроков начальный капитал конечный, а у другого бесконечный; б) у обоих игроков капитал бесконечный. Были найдены границы оптимальной остановки и выигрыши игроков для обоих случаев. Проведено численное моделирование оптимальных стратегий и выигрышей игроков для различных значений параметров задачи.

Задача о разорении с неограниченным капиталом у одного из игроков

Рассматривается задача о разорении следующего вида. В начале игры задан интервал времени n. Имеются два игрока I и II, которые на каждом шаге игры разыгрывают единицу капитала. У одного из игроков (II) имеется начальный капитал, равный $a, a \geqslant 0$. У другого игрока (I) капитал не ограничен. На каждом шаге игроки играют друг с другом, при этом игрок I может выиграть с вероятностью p и проиграть с вероятностью $q = 1 - p$. Если капитал игрока II исчерпан в момент времени τ, то есть произошло его разорение, игра прекращается и игрок I получает выигрыш, равный $1 - c\tau$, где c – затраты игрока на каждом шаге игры. Если в течение промежутка времени n игра не закончилась, то игрок I получает $-nc$.

Для описания задачи о разорении и нахождения ее решения удобно использовать случайные блуждания. Случайное блуждание по целочисленной строке начинается с 0 и на каждом шаге i, $(i = 1, 2, ..., n)$ перемещается на $+1$ вправо или на -1 влево. При этом игрок может выиграть с вероятностью p и проиграть с вероятностью $q = 1 - p$.

Обозначим $X_1, ..., X_n$ последовательность случайных величин, принимающих значения $+1$ или -1, соответствующих рассмотренному выше случайному блужданию. Тогда $S_i = \sum_{j=1}^{i} X_j$, $i = 1, 2, ..., n$, $S_0 = 0$ – разность между числом $+1$ и -1 в течение i шагов. Для рассматриваемой задачи S_i представляет собой капитал, полученный игроком I за первые i шагов. Последовательность значений $X_1, ..., X_n$ формирует некоторую траекторию на плоскости. Эта траектория описывает изменение капитала игрока I за i шагов и заканчивается в точке с координатами (i, S_i).

Рассмотрим сначала задачу о разорении, в которой оба игрока играют до того момента, когда либо капитал игрока II закончится, либо закончится время игры.

Предположим, что a и n одной четности, так что траектория может закончиться в точке (n, a).

Игра продолжается до тех пор, пока не закончится ресурс у игрока II либо пока время игры не истечет. На рис. 1 изображено возможное изменение капитала игрока I на каждом из n шагов. Значение капитала игрока II на шаге i соответствует расстоянию от точки (i, S_i) до уровня a.

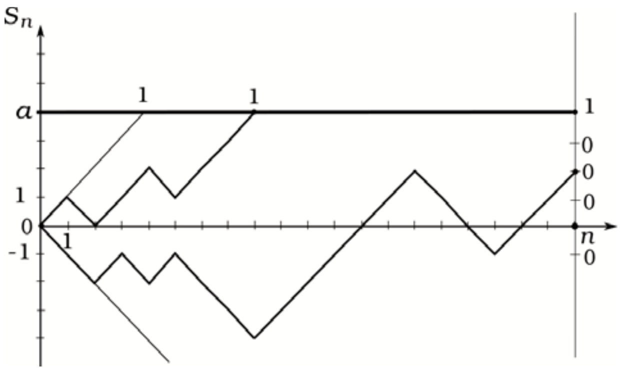


Рис. 1. Игра на разорение с неограниченным капиталом у одного из игроков
Fig. 1. Gambler's ruin problem where one of the players has infinite capital

Обозначим $V_n(a)$ выигрыш игрока I в данной задаче и найдем его значение.

Вероятность попасть из точки $(0, 0)$ в точку (n, j), где $n + j$ четное и $-n \leqslant j \leqslant n$, по любой траектории, для несимметричного случайного блуждания не зависит от пути и равна

$$P(n, j) = p^k(1-p)^{n-k} = (1-p)^n\left(\frac{p}{1-p}\right)^k, \quad (1)$$

где $k = \dfrac{n+j}{2}$, $n - k = \dfrac{n-j}{2}$.

Число всех траекторий из точки $(0, 0)$ в точку (n, j) равно $N_{n,j} = \binom{n}{k}$, $k = \dfrac{n+j}{2}$.

Если до момента времени n блуждание не достигнет значения a, то на последнем шаге игрок получит 0 в качестве выигрыша, затратив при этом $-cn$. Чтобы это произошло, траектория должна, не задев a, закончиться в точке $(n, a-2)$, $(n, a-4)$,..., $(n, -n)$. Согласно принципу отражения, число траекторий, заканчивающихся в точке $(n, a - 2s)$ и не пересекающих a, где $s = 1, ..., \dfrac{n-a}{2}$, равно

$$N_{n,a-2s} - N_{n,a+2s} = \binom{n}{\frac{n+a}{2} - s} - \binom{n}{\frac{n+a}{2} + s}.$$

Число траекторий, заканчивающихся в точке $(n, a - 2s)$, где $s = \dfrac{n-a}{2} + 1, ..., \dfrac{n+a}{2}$, равно

$$N_{n,a-2s} = \binom{n}{\frac{n+a}{2} - s}.$$

Таким образом, в случае недостижения случайным блужданием уровня a вклад $U_n(a)$

в выигрыш игрока I, с учетом (1), равен

$$U_n(a)=-cn\Bigg[\sum_{s=1}^{\frac{n-a}{2}}\left(\binom{n}{\frac{n+a}{2}-s}-\binom{n}{\frac{n+a}{2}+s}\right)$$

$$\times P(n,a-2s)+\sum_{s=\frac{n-a}{2}+1}^{\frac{n+a}{2}}\binom{n}{\frac{n+a}{2}-s}P(n,a-2s)\Bigg]$$

$$=-cn\Bigg[\sum_{s=0}^{\frac{n+a}{2}-1}\binom{n}{s}p^s(1-p)^{n-s}$$

$$+\left(\frac{p}{1-p}\right)^a\sum_{s=0}^{\frac{n-a}{2}-1}\binom{n}{s}p^s(1-p)^{n-s}\Bigg].$$

Теперь вычислим вклад $W_n(a)$ в выигрыш игрока I в случае достижения уровня a.

Первый раз траектория достигает уровня a в точке (a,a). Игрок I получает выигрыш $1-ac$. Вероятность такого выигрыша равна $P(a,a)=p^a$.

Далее блуждание может достичь уровня a в точке $(a+2,a)$. Таких траекторий будет $N_{a+1,a-1}-N_{a+1,a+1}$. Игрок I получит выигрыш $1-(a+2)c$. Вероятность такого выигрыша равна $(N_{a-1,a-1}-N_{a-1,a+1})P(a+2,a)=(N_{a-1,a-1}-N_{a-1,a+1})p^{a+1}(1-p)$.

Рассуждая подобным образом, находим выигрыш игрока I при выходе траектории блуждания на уровень a:

$$W_n(a)=\sum_{i=0}^{\frac{n-a}{2}}(1-(a+2i)c)\Big(N_{a+2i-1,a-1}$$

$$-N_{a+2i-1,a+1}\Big)P(a+2i,a)$$

$$=\sum_{i=0}^{\frac{n-a}{2}}\binom{a+2i}{i}\frac{a(1-(a+2i)c)}{a+2i}p^{a+i}(1-p)^i.$$

Объединяя выигрыши, окончательно получаем следующее утверждение.

Утверждение. *Выигрыш игрока I в игре, связанной с разорением игрока, имеет вид*

$$V_n(a)=\sum_{i=0}^{\frac{n-a}{2}}\binom{a+2i}{i}\frac{a(1-(a+2i)c)}{a+2i}p^{a+i}(1-p)^i$$

$$-c\cdot n\Bigg[\sum_{s=0}^{\frac{n+a}{2}-1}\binom{n}{s}p^s(1-p)^{n-s}$$

$$+\left(\frac{p}{1-p}\right)^a\sum_{s=0}^{\frac{n-a}{2}-1}\binom{n}{s}p^s(1-p)^{n-s}\Bigg].$$

Замечание. Вероятность разорения игрока II вычисляется по формуле

$$P_{ruin}(a)=1-\Bigg[\sum_{s=0}^{\frac{n+a}{2}-1}\binom{n}{s}p^s(1-p)^{n-s}$$

$$+\left(\frac{p}{1-p}\right)^a\sum_{s=0}^{\frac{n-a}{2}-1}\binom{n}{s}p^s(1-p)^{n-s}\Bigg].$$

Для больших значений n получим приближение для вероятности разорения:

$$P_{ruin}(a)\approx\begin{cases}\left(\dfrac{p}{1-p}\right)^a, & p<\dfrac{1}{2},\\[2mm] 1,\ p\geqslant\dfrac{1}{2}.\end{cases}$$

Найдем выражение для выигрыша $V_n(a)$ для больших значений n.

Положим, $V(x,y)$ – выигрыш игрока I начиная с шага x. Здесь $x=k+m$ и $y=k-m$, где k – количество побед, а m – количество поражений.

Функция выигрыша игрока I примет вид

$$\begin{gathered}V(x,y)=pV(x+1,y+1)\\+(1-p)V(x+1,y-1)-c,\end{gathered}\tag{2}$$

с начальными условиями

$$\begin{gathered}V(x,a)=1,\ 0\leqslant x\leqslant n,\\V(n,y)=0,\ -n\leqslant y<a.\end{gathered}$$

При достаточно больших n значение выигрыша $V(x,y)$ близко к $V(y)$ – решению уравнения

$$V(y)=pV(y+1)+(1-p)V(y-1)-c\tag{3}$$

при начальных условиях

$$V(a)=1;\ V(-n)=0.$$

Здесь $y=k-m$, где k – количество побед, а m – количество поражений.

Решая уравнение (3), получим, что при больших значениях n и $p>1/2$

$$V(y)=1-\frac{c(a-y)}{2p-1}.$$

Тогда выигрыш в начале игры равен

$$V(0)=1-\frac{ac}{2p-1}.$$

В случае, если $p\leqslant 1/2$, $V(y)\to-\infty$.

Оптимальная остановка в игре, связанной с разорением

В данном разделе рассматривается вариант игры на разорение, в котором игрок I, имеющий бесконечный капитал, может остановить игру в любой момент времени $\tau \leqslant n$ для максимизации своего ожидаемого выигрыша. Если в момент времени $\tau \leqslant n$ игрок II разорится, то игрок I получит выигрыш $1-c\tau$. В этом случае игрок I на каждом шаге продолжает игру, только если его ожидаемый выигрыш неотрицателен. Таким образом, оптимальная стратегия игрока I представляет собой множество состояний, в которых нужно останавливать или продолжать игру. Граница оптимальной остановки определяется кривой, ниже которой следует прекратить игру.

Найдем выигрыш игрока I, используя метод динамического программирования.

Заметим, что в выражении (2) при некоторых значениях (x, y) выигрыш игрока I $V(x, y) \leqslant 0$. В этих состояниях игроку I невыгодно продолжать игру и следует остановиться. Таким образом, приходим к задаче оптимальной остановки с уравнением оптимальности для выигрыша $V^*(x, y)$ вида

$$V^*(x,y) = \max\Big\{0, pV^*(x+1, y+1) + (1-p)V^*(x+1, y-1) - c\Big\}, \tag{4}$$

с граничными условиями

$$V^*(x, a) = 1, \text{ если } 0 \leqslant x \leqslant n,$$

$$V^*(n, y) = 0, \text{ если } -n \leqslant y < a.$$

Оптимальная граница остановки $s(i)$, $i = 1, ..., n$ находится последовательно, начиная с конца, из уравнения (4).

На рис. 2 изображена граница оптимальной остановки игрока I для $n = 50$, $c = 0{,}01$ и начального капитала $a = 2$ у игрока II.

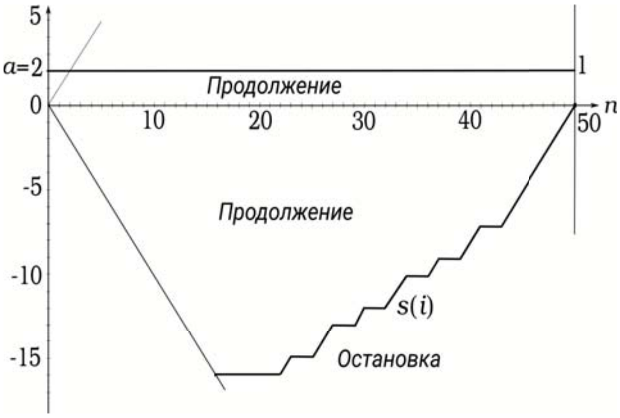


Рис. 2. Граница оптимальной остановки игрока I для $n = 50$, $c = 0,01$ и капиталом игрока II $a = 2$
Fig. 2. Optimal stopping boundary for player I for $n = 50$, $c = 0.01$ and the player II's capital $a = 2$

В таблице приведены оптимальные выигрыши $V^*(0, 0)$ игрока I и $V_n(a)$ (в круглых скобках) для различных значений n и a при $c = 0$ и $c = 0,01$. Для случая $c = 0$ значения оптимальных выигрышей $V^*(0, 0)$ и $V_n(a)$ одинаковы.

Из таблицы видно, что выигрыши растут с ростом n и уменьшаются с ростом начального капитала a у игрока II. Также отметим, что выигрыши уменьшаются с увеличением затрат c и $V^*(0, 0) \geqslant V_n(a)$. То есть выигрыш игрока I при использовании оптимальной стратегии остановки больше, чем в задаче о разорении.

Оптимальные выигрыши $V^*(0, 0)$ игрока I и $V_n(a)$ (в скобках) для различных значений n, a и $p = 2/3$
Optimal payoffs $V^*(0, 0)$ of player I and $V_n(a)$ (in parentheses) for different values of n, a and $p = 2/3$

	$c = 0$					$c = 0{,}01$				
a	2	4	6	8	10	2	4	6	8	10
$n = 10$	0,866	0,614	0,322	0,108	0,017	0,821 (0,819)	0,544 (0,535)	0,250 (0,228)	0,054 (0,009)	0 (-0,083)
$n = 50$	0,997	0,991	0,977	0,950	0,906	0,938 (0,938)	0,873 (0,872)	0,801 (0,780)	0,720 (0,717)	0,626 (0,620)
$n = 100$	0,9999	0,9998	0,9994	0,999	0,997	0,940 (0,940)	0,880 (0,880)	0,820 (0,820)	0,759 (0,759)	0,698 (0,697)

Игра на разорение с неограниченным капиталом у обоих игроков

Рассмотрим игру, в которой у обоих игроков капитал не ограничен. Пусть каждый игрок может остановить игру в любой момент времени $\tau \leqslant n$. Если какой-то из участников останавливается и выходит из игры, игра прекращается и оставшийся в игре игрок получает выигрыш, равный $1 - c\tau$, где c – затраты игрока на каждом шаге игры. Если в течение промежутка времени n игра не закончилась, игрок получает выигрыш $-nc$. Цель каждого игрока максимизировать свой ожидаемый выигрыш в такой игре.

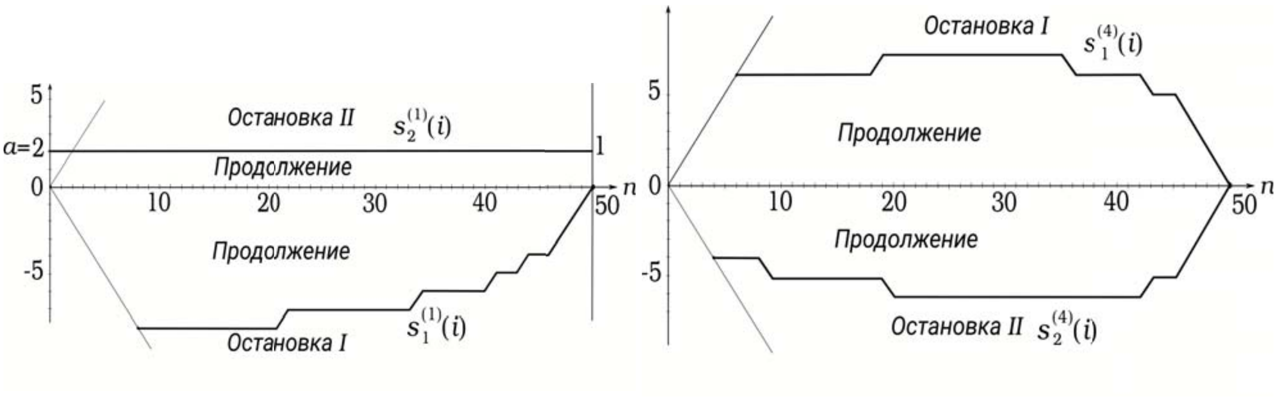
В такой постановке данная игра является игрой с оптимальной остановкой для двух игроков. Стратегией игрока l, $l = 1, 2$ является определение границы $s_l(i)$, $i = 0, 1, ..., n$ области его остановки. Если случайное блуждание опускается ниже этой границы, то игрок останавливает игру.

Найдем оптимальные стратегии игроков в данной игре, используя последовательность наилучших ответов игроков на стратегии друг друга. На каждой итерации k алгоритма наилучших ответов игрок l, зная границы остановки $s_{-l}^{(k)}(i)$ противника, может найти свою наилучшую стратегию $s_l^{(k)}(i) = BR(s_{-l}^{(k)}(i))$ (наилучший ответ), $i = 1, ..., n$. Процесс останавливается, когда стратегии игроков не меняются.

На рис. 3 представлен результат применения алгоритма наилучших ответов игроков для $n = 50$, $c = 0{,}01$ и $p = 0{,}51$. Предположим, что игроки используют стратегии следующего вида. На первом этапе игрок II останавливается, если траектория достигла уровня $a = 2$. Зная это, игрок I находит свою наилучшую стратегию. Эта игра эквивалентна игре, рассмотренной в предыдущем разделе. На рис. 3, (а) изображена граница $s_1^{(1)}(i)$ оптимальной остановки игрока I.

Затем на втором этапе игрок II находит свой наилучший ответ $s_2^{(2)}(i)$ на стратегию игрока I. Повторяя данную процедуру несколько раз, получим оптимальные стратегии обоих игроков (рис. 3, (б)).

В результате при использовании оптимальных стратегий игрок I получит выигрыш 0,338, а игрок II – 0,153.



(а) (б)

Рис. 3. Стратегии игроков в игре на разорение для $n = 50$, $c = 0{,}01$ и $p = 0{,}51$
Fig. 3. Players' strategies in the gambler's ruin problem for $n = 50$, $c = 0.01$ and $p = 0.51$

Заключение

В работе исследована многошаговая игра двух лиц, связанная с игрой на разорение в течение заданного промежутка времени n. На каждом шаге один из игроков может выиграть с вероятностью p и проиграть с вероятностью $q = 1 - p$. Рассмотрено два варианта игры: с ограниченным капиталом у одного из игроков и неограниченным капиталом у обоих игроков. Предполагается, что выигрыши игроков определяются в конце игры и учитывают затраты c на каждом шаге игры. Найдены выигрыши игроков в данной задаче для различных значений продолжительности игры и различных значений c. Для задачи с ограниченным капиталом у одного из игроков проведено численное моделирование оптимальной стратегии остановки и выигрышей игрока, имеющего неограниченный капитал. В задаче с неограниченным капиталом у обоих игроков оптимальные их стратегии найдены с использованием метода наилучших ответов.

Литература

1. *Asmussen S., Albrecher H.* Ruin Probabilities. Singapore: World Scientific, 2010. 621 p. doi: 10.1142/9789814282536

2. *Blass A., Braun G.* Random orders and gambler's ruin // Electron. J. Comb. 2005. Vol. 12. Art. R23. doi: 10.37236/1920

3. *Feller W.* An introduction to probability theory and its applications. Vol. 1. New York, London, Sydney: John Wiley & Sons, Inc., 1968. 526 p.

4. *Harper J. D., Ross K. A.* Stopping strategies and gambler's ruin // Mathematics Magazine. 2005. Vol. 78, iss. 4. P. 255–268. doi: 10.1080/0025570X.2005.11953340

5. *Hussain A., Hanif M., Naseer M.* Three-player gambler's ruin problem: Some extensions // Sci. Inquiry Rev. 2021. Vol. 5, no. 3. P. 1–11. doi: 10.32350/sir/53.01

6. *Mazalov V. V., Ivashko A. A.* Harmonic numbers in gambler's ruin problem // Mathematical Optimization Theory and Operations Research (MOTOR 2023). Lecture Notes in Computer Science. 2023. Vol. 13930. P. 278–287. doi: 10.1007/978-3-031-35305-5_19

7. *Mazalov V., Ivashko A.* Optimal stopping strategies in gambler's ruin game // Mathematical Optimization Theory and Operations Research: Recent Trends (MOTOR 2024). Communications in Computer and Information Science. 2024. Vol. 2239. P. 237–249. doi: 10.1007/978-3-031-73365-9_16

8. *Rocha A. L., Stern F.* The gambler's ruin problem with n players and asymmetric play // Stat. Probab. Lett. 1999. Vol. 44, iss. 1. P. 87–95. doi: 10.1016/s0167-7152(98)00295-8

9. *Shiryaev A.* Probability-1. NY: Springer New York, 2016. 486 p.

10. *Song S., Song J.* A note on the history of the gambler's ruin problem // Communications for Statistical Applications and Methods. The Korean Statistical Society. 2013. Vol. 20, iss. 2. P. 157–168. doi: 10.5351/csam.2013.20.2.157

References

1. *Asmussen S., Albrecher H.* Ruin probabilities. Singapore: World Scientific; 2010. 621 p. doi: 10.1142/9789814282536

2. *Blass A., Braun G.* Random orders and gambler's ruin. *The Electronic Journal of Combinatorics.* 2005;12:R 23. doi: 10.37236/1920

3. *Feller W.* An introduction to probability theory and its applications. Vol. 1. New York, London, Sydney: John Wiley & Sons, Inc.; 1968. 526 p.

4. *Harper J. D., Ross K. A.* Stopping strategies and gambler's ruin. *Mathematics Magazine.* 2005; 78(4):255–268. doi: 10.1080/0025570X.2005.11953340

5. *Hussain A., Hanif M., Naseer M.* Three-player gambler's ruin problem: Some extensions. *Sci. Inquiry Rev.* 2021;5(3):1–11. doi: 10.32350/sir/53.01

6. *Mazalov V. V., Ivashko A. A.* Harmonic numbers in gambler's ruin problem. *Mathematical Optimization Theory and Operations Research. MOTOR 2023. Lecture Notes in Computer Science.* 2023;13930:278–287. doi: 10.1007/978-3-031-35305-5_19

7. *Mazalov V., Ivashko A.* Optimal stopping strategies in gambler's ruin game *Mathematical Optimization Theory and Operations Research: Recent Trends. MOTOR 2024. Communications in Computer and Information Science.* 2024;2239:237–249. doi: 10.1007/978-3-031-73365-9_16

8. *Rocha A. L., Stern F.* The gambler's ruin problem with n players and asymmetric play. *Statistics and Probability Letters.* 1999;44(1):87–95. doi: 10.1016/s0167-7152(98)00295-8

9. *Shiryaev A.* Probability-1. NY: Springer New York; 2016. 486 p.

10. *Song S., Song J.* A Note on the history of the gambler's ruin problem. *Communications for Statistical Applications and Methods. The Korean Statistical Society.* 2013;20(2):157–168. doi: 10.5351/csam.2013.20.2.157

Поступила в редакцию / received: 16.04.2025; принята к публикации / accepted: 21.05.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Ивашко Анна Антоновна

канд. физ.-мат. наук, старший научный сотрудник

e-mail: aivashko@krc.karelia.ru

CONTRIBUTOR:

Ivashko, Anna

Cand. Sci. (Phys.-Math.), Senior Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 24–33
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 24–33
DOI: 10.17076/mat2099

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 81.32

ВЫБОР ОПТИМАЛЬНОГО РАДИУСА КЛАСТЕРИЗАЦИИ НА КАРТАХ КАТАСТРОФ

А. А. Крижановский

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Викиданные содержат информацию о вулканах и землетрясениях, позволяющую провести их пространственный анализ. Чтобы найти взаимосвязь между ними, нужно их сгруппировать в кластеры. Рассмотрено несколько подходов выбора оптимального радиуса кластеризации. Варьирование радиуса кластеризации показало, что (1) 90 км – оптимальное расстояние для выявления парных взаимодействий (максимум смешанных кластеров, содержащих и вулканы, и землетрясения); (2) 127 км – порог устойчивости кластерной структуры (превышение ведет к артефактам слияния). С помощью SPARQL-запросов и программ на языке Python были построены карты, отображающие возможные взаимосвязи между извержениями и сейсмическими событиями. Анализ показывает, что сейчас в Викиданных мало информации о датах извержения вулканов, хотя число объектов «извержение вулкана» со свойством «дата» и выросло за три года в несколько раз. Показаны возможности и текущие ограничения Викиданных для геонаучных исследований. Работа подтверждает, что краудсорсинговые базы данных могут дополнять традиционные научные ресурсы при исследовании глобальных геофизических закономерностей.

Ключевые слова: землетрясение; извержение вулкана; кластеризация; Викиданные

Для цитирования: Крижановский А. А. Выбор оптимального радиуса кластеризации на картах катастроф // Труды Карельского научного центра РАН. 2025. № 4. С. 24–33. doi: 10.17076/mat2099

A. A. Krizhanovsky. SELECTING THE OPTIMAL CLUSTERING RADIUS ON DISASTER MAPS

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

Wikidata contains data on volcanoes and earthquakes, enabling spatial analysis. To find the relationship between them, they need to be grouped into clusters. Several approaches to choosing the optimal clustering radius are considered. Varying the clustering radius revealed that (1) 90 km is the optimal distance for detecting pairwise interactions (maximizing mixed clusters containing both volcanoes and earthquakes), and (2) 127 km is the stability threshold of the cluster structure (exceeding it leads to merging artifacts). Using SPARQL queries and Python scripts, maps were generated to visualize potential connections between eruptions and seismic events. The analysis shows that Wikidata currently lacks sufficient information on volcanic eruption dates, although the number of “volcanic eruption” objects with date properties has increased severalfold over three years. The study highlights both the potential and current limitations of Wikidata for geoscientific research. The study confirms that crowdsourced databases can complement traditional scientific resources in researching global geophysical patterns.

Keywords: earthquake; volcanic eruption; clustering; Wikidata; SPARQL

For citation: Krizhanovsky A. A. Selecting the optimal clustering radius on disaster maps. *Trudy Karel′skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2025. No. 4. P. 24–33. doi: 10.17076/mat2099

Введение

Вулканы и землетрясения с древних времен привлекали внимание людей и фиксировались учеными. Одной из крупнейших современных баз данных, описывающих конкретные извержения вулканов и землетрясения, является Викиданные. В работе поставлена двоякая задача: найти пространственную взаимосвязь вулканической и сейсмической активности на основе информации Викиданных и оценить степень полноты самой этой базы.

Мы не претендуем на новые открытия в области вулканологии и сейсмологии. Но хотим взять ту часть базы Викиданных, которая описывает вулканы и землетрясения, и оценить полноту информации и возможность проведения экспериментальных исследований на основе этой огромной и доступной базы данных. То есть мы хотим проверить, насколько Викиданные готовы, чтобы на их основе проводить исследования. Вулканы и землетрясения – это два типа объектов из тысяч других в Викиданных. Положительный ответ позволит с большей уверенностью исследовать и другие стороны Вселенной на основе Викиданных.

В следующем разделе представим обзор литературы по вопросу взаимосвязи вулканов и землетрясений. Далее опишем доступные базы данных вулканов, в том числе Викиданные. Затем извлечем эти сведения из Викиданных с помощью SPARQL-запросов [1]. Остановимся на вопросе выбора оптимального порога при кластеризации пространственных данных. Затем мы опишем алгоритм программы поиска и визуализации на карте ближайших друг к другу вулканов и землетрясений и обсудим недостатки и потенциал Викиданных.

Обзор исследований, изучающих взаимосвязь землетрясений и вулканов

В работе [14] проверили гипотезу, что вулканы и землетрясения связаны во времени и пространстве. Гипотеза подтвердилась статистически, были найдены последовательности «извержение – землетрясения» [14]. Найдена связь между крупными извержениями и землетрясениями во времени и пространстве на примере исторических записей активности вулкана Мауна-Лоа, второго по объему и площади подошвы вулкана на Земле.

Часто отмечают, что крупные, не слишком отдаленные землетрясения влияют на извержения вулканов. В работе [2] проверялась гипотеза: следуют ли за землетрясениями извержения вулканов. Для индонезийского вулканического региона в целом получен отрицательный ответ. Однако если еще дополнительно учитывать время извержения, расстояние и силу землетрясения, то ответ будет иным, а именно: влияние землетрясения было статистически значимым для 7 из 35 вулканов, но только после учета внутреннего состояния вулкана [2]. Тектонические землетрясения (то есть вызванные смещением горных плит) могут приводить в действие (вызывать) большинство типов вулканов, но только при выполнении ряда вулканических и сейсмических условий [11].

Есть трудности в учете и анализе извержений вулканов:

- *Извержение имеет длительность*, а не точку во времени. «Извержение может длиться очень долгое время, например, вулкан Стромболи действует уже более 2500 лет» [13, с. 435].

- *Рост частоты исторических наблюдений.* Число активных вулканов в год стремительно растет последние пять веков. Но это объясняется ростом населения и лучшей коммуникацией, растет (читательский) интерес к вулканам. Это объясняет снижение роста числа вулканов во время двух мировых войн [13, с. 436].

- *Цикличность и эпизодичность.* Ряд вулканов, наблюдаемых сотни лет (например, Везувий, Колима, Майон), показывают повторяемость, цикличность извержений. С другой стороны, вулканологи при ограниченных ресурсах должны максимум внимания уделять именно опасным вулканам, но их выбор только на основе записей о предыдущих извержениях будет ошибкой, поскольку за последние 200 лет каждый год происходит в среднем одно или два извержения вулканов, отсутствующих в каких-либо записях [13, с. 446]. Другие вулканы извергаются после сотен лет тишины.

Внешние явления (землетрясения, оползни и приливы) могут потревожить магматический резервуар и вызвать извержение вулкана [3]. В среднем в год извергается 50 вулканов.

Извержения вулканов чаще происходят после землетрясений [6]. «Исторически сложилось так, что некоторые вулканы извергались после крупных землетрясений... Статистический анализ мировых вулканических и сейсмических записей показывает значительное увеличение вероятности извержения после землетрясения в течение периода от нескольких дней до нескольких лет» [3].

Базы данных вулканов

Проблемы баз данных вулканов в том, что нет единого стандарта, мало доступных баз данных [4, с. 739].

База вулканов всего мира WOVOdat (`https://wovodat.org/`) служит ресурсом для улучшения прогнозирования извержений. Сходство между Викиданными и WOVOdat в том, что «база данных связана ссылками с различными открытыми базами и каталогами, обеспечивающими многообразие дополнительных данных» [4, с. 739].

Наше исследование требует наличия обширной базы данных по вулканам и землетрясениям. Однако попытки создать глобальные вулканические базы данных начались лишь недавно [9]. Автор статьи [9] предлагает объединить две открытые базы данных о вулканах – GVP и LaMEVE – в единую. Это база данных извержений голоцена проекта Global Volcanism Project (GVP) Смитсоновского института (`http://volcano.si.edu`) и база данных мощных взрывных вулканических извержений LaMEVE [5], созданная в рамках проекта VOGRIPA (`http://www.bgs.ac.uk/vogripa`). Таким образом, в объединенной базе GVP+LaMEVE содержится информация о 9517 извержениях с указанием показателя вулканической эксплозивности (VEI) [9]. Сравните: 10 415 извержений вулканов в хронологическом порядке приводятся в книге [12].

Викиданные развиваются как связующий центр информации [15, с. 868], описанные выше две базы вулканов частично включены в Викиданные.

Видим на рис. 1, что объекты Викиданных, описывающие вулканы, могут содержать ссылки на те же вулканы в базах GVP и VOGRIPA. Ссылки на эти базы единообразно оформляются с помощью внешних идентификаторов Викиданных, а именно: Global Volcanism Program ID (P1886), VOGRIPA ID (P4708). Таким образом, Викиданные в некоторой степени объединяют эти базы данных вулканов и извержений.

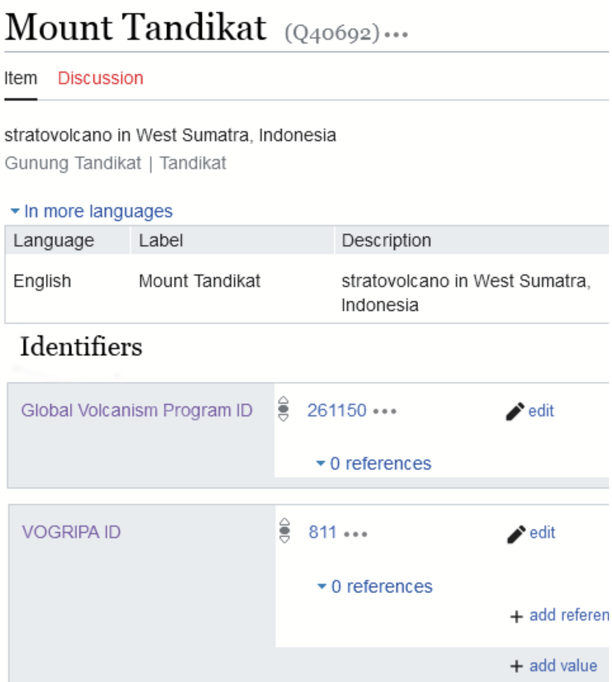


Рис. 1. Фрагмент карточки вулкана Mount Tandikat (Q40692) в Викиданных с идентификаторами *Global Volcanism Program ID* и *VOGRIPA ID*

Fig. 1. A fragment of the Mount Tandikat volcano card (Q40692) in Wikidata with the identifiers *Global Volcanism Program ID* and *VOGRIPA ID*

SPARQL-запросы по извлечению данных о вулканах и землетрясениях: подготовка данных

Викиданные пока не так полны, как нам хотелось бы, а именно: почти нет данных о времени и дате извержений вулканов, нет данных о внутреннем состоянии вулканов. Посмотрим, какую ценную информацию о вулканах и землетрясениях мы можем получить из Викиданных.

Обычно приводят листинги со SPARQL-запросами. Для экономии места вместо семи листингов представим данные в таблице. Чтобы увидеть листинг любого из этих запросов, щелкните по ссылке для перехода в Wikidata Query Service [1]. Откроется окно в браузере с кодом запроса. Далее щелкните по кнопке *Play*, чтобы выполнить запрос. После выполнения запроса на той же странице ниже будет результат: список, график или карта.

Наше исследование длится несколько лет, поэтому у нас есть данные за три года (2023–2025). Таблица показывает, что Викиданные постоянно меняются. Выросло число и вулканов, и землетрясений. Важным для исследования является рост числа вулканов и землетрясений, имеющих координаты (запросы 6 и 7 в таблице).

Было бы интересно узнать, в какой пропорции добавляются объекты, появившиеся за эти три года (например, произошли извержения), и в какой доле добавляются «старые» вулканы и землетрясения, проявившие себя до 2023 года.

SPARQL-запросы и результаты по вулканам и землетрясениям в Викиданных
SPARQL queries and the results for volcanoes and earthquakes in Wikidata

№ No.	Задача Task	Количество / Number			Ссылка Link
		2023	2024	2025	
1	Список всех землетрясений List of all earthquakes	2394	2430 ↑1.5 %	2451 ↑0.9 %	w.wiki/A362
2	Список всех вулканов List of all volcanoes	1891	–	2052	w.wiki/Dyad
3	Количество землетрясений по странам (график). Число стран с землетрясениями Number of earthquakes by country (chart). Number of countries with earthquakes	157	168	174	w.wiki/8BRe
4	Количество вулканов по странам (график). Число стран с вулканами Number of volcanoes by country (chart). Number of countries with volcanoes	85	86	86	w.wiki/DxFR
5	Число вулканов России, включая активные и потухшие; генерация карты Number of volcanoes in Russia, including active and extinct ones; map generation	–	322	326	w.wiki/DxEM
6	Список землетрясений с координатами без повторов List of earthquakes with coordinates without duplicates	1958	2055	2172	w.wiki/Dydp
7	Список вулканов на Земле с координатами без повторов List of volcanoes on Earth with coordinates without duplicates	1689	1825	2002	w.wiki/Dyeh

Необходимо пояснение к запросу 6, что значит «без повторов». Дело в том, что в Викиданных одному и тому же землетрясению могут быть приписаны несколько координат (например, координаты эпицентра по разным источникам). Чтобы получить по одной координате на землетрясение, мы используем агрегирующую функцию `SAMPLE`, которая возвращает одно (произвольное) значение из группы возможных (см. код запроса w.wiki/Dydp).

Аналогично «без повторов» получены вулканы с координатами в запросе 7. Также в код добавлено ограничение, что вулканы относятся к Земле, а не к другим астрономическим объектам (см. код запроса w.wiki/Dyeh).

Таким образом, с помощью запросов 6 и 7 из таблицы мы получили из Викиданных список вулканов и землетрясений с географическими координатами. CSV-файлы с этими данными доступны на сайте проекта (`https://github.com/componavt/volcano_wikidata/`). Эти данные будем использовать в следующем разделе.

Географическая кластеризация на основе расстояния между объектами: выбор оптимального радиуса кластеризации

Задача состоит в том, чтобы сгруппировать вулканы и землетрясения, находящиеся ближе некоторой заданной величины `dist_max`. Вулканы и землетрясения считаем объектами одного типа. Максимальное расстояние (радиус кластеризации) – такое, что если вулканы и землетрясения находятся на меньшем расстоянии, то они объединяются в один кластер.

Алгоритм *агломеративной иерархической кластеризации* с перебором радиуса кластеризации (переменная `dist_max`) включает шаги:

1. *Построение KD-дерева* по всем координатам вулканов и землетрясений для быстрого поиска ближайших соседей.
2. *Создание графа.* Для каждого значения `dist_max` из заданного диапазона строится невзвешенный граф. Вулканы и землетрясения — это вершины графа. Ребра добавляются между всеми парами точек, находящимися ближе `dist_max`. Расстояние вычисляется по формуле Хаверсина, учитывающей кривизну Земли.
3. *Выделение компонент связности* построенного графа как групп (кластеров). Вычисление для всех компонент: размера группы и диаметра (расстояние по формуле Хаверсина между максимально удаленными друг от друга вершинами).
4. *Анализ метрик групп:* их количество, среднее и максимальное число объектов в группе, диаметр в километрах (рис. 2).

Алгоритм кластеризации получился быстрый[1] благодаря поиску ближайших землетрясений и вулканов по KD-дереву [7].

Итак, какое нужно выбрать расстояние, чтобы вулканы и землетрясения были связаны? В работе [8] показано, что временная корреляция (от минут до часов) между землетрясениями и извержениями грязевого вулкана наиболее выражена при расстоянии между ними в пределах порядка 100 км при силе землетрясения в 6 и более баллов по шкале Меркалли.

Изменение параметров кластеров в зависимости от максимального допустимого расстояния между вулканами и землетрясениями представлено на рис. 2. На рисунке есть три оси Y: слева – для числа групп и размера наибольшей группы, справа – для среднего размера групп (со шкалой от 3 до 7,5 элемента в группе) и диаметра наибольшей группы (со шкалой от 1 до 4 тыс. км).

Для максимального расстояния выбран крупный шаг в 10 км для всего диапазона 80–160 км, но для участка 120–130 км использован шаг 1 км, что позволяет выделить на графике локальный изгиб зависимости.

Количество групп на рис. 2 уменьшается достаточно равномерно с увеличением порога кластеризации от 80 до 160 км. На рисунке это кажется не столь равномерным, поскольку мы выбрали неравномерную шкалу по оси X, чтобы растянуть самый интересный участок от 120 до 130 км, где происходят быстрые изменения размера и диаметра самого большого кластера.

На рис. 2 видно, что в диапазоне 127–128 км происходят большие изменения:

- группы вулканов и землетрясений начинают сливаться в один гигантский кластер, размер наибольшей группы увеличивается с 313 до 452 объектов;
- диаметр самой крупной группы скачком возрастает с 2,4 до 4 тыс. км.

При увеличении максимального расстояния от 127 до 128 км появляется суперкластер. Это позволяет определить порог кластеризации (оптимальный радиус) в 127 км – до момента появления очень больших кластеров.

Выясним далее, какие именно кластеры вулканов и землетрясений, существовавшие раздельно при радиусе 127 км, сливаются в общий кластер при 128 км.

Объединение кластеров вулканов и землетрясений в суперкластер

На рис. 2 видно, что самые большие кластеры при радиусе кластеризации (`dist_max`) равном 127 и 128 км содержат соответственно 313 и 452 элемента. Самая большая группа из 313 элементов не участвует в объединении.

Проанализированы только те группы, которые объединяются при переходе от 127 к 128 км. Найдены две группы, которые на этом шаге формируют суперкластер из 452 элементов:

- группа 1 из 228 объектов (144 вулкана, 84 землетрясения);
- группа 2 из 223 объектов (209 вулканов, 14 землетрясений).

Для визуализации этих групп вулканов и землетрясений на карте (рис. 3) реализованы следующие методы:

1. Алгоритм построения минимального остовного дерева (англ. Minimum Spanning Tree, MST) для выбора и рисования ребер, связывающих элементы кластера.

[1] Исходный код программы кластеризации на основе расстояния между объектами доступен онлайн: `https://github.com/componavt/volcano_wikidata/blob/main/src/volcano_earthquake_clustering.ipynb`.

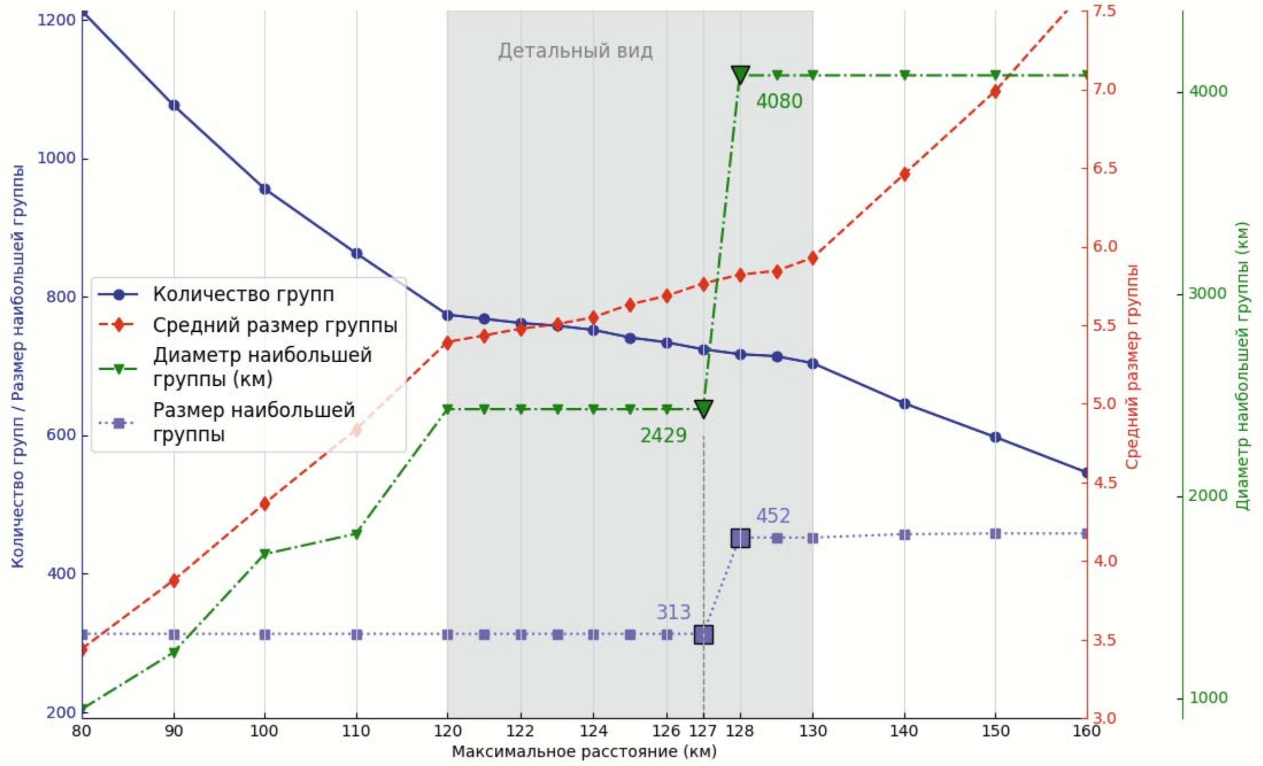


Рис. 2. Зависимость параметров групп от максимального расстояния в кластере между вулканами и землетрясениями
Fig. 2. Cluster metrics as a function of maximum pairing distance (volcano–earthquake)

2. Алгоритм жадного покрытия группы кругами с максимальным охватом для минимизации общего числа кругов (эти многочисленные круги радиуса `dist_max`, касаясь друг друга, наглядно показывают, что вершины в них относятся к одному кластеру).
3. Метод вычисления центра кластера на основе самой удаленной пары точек. Метод гарантирует, что даже самые удаленные точки окажутся внутри построенного круга.

Интерактивная карта (рис. 3) создается тем же кодом на языке Python[1].

Рисунок 3 полностью построен по Викиданным. То есть известные островные дуги Японская и Курило-Камчатская были выявлены с помощью алгоритма кластеризации на основе расстояния между объектами. Этот результат подтверждает:

- работоспособность предложенного алгоритма,
- достаточную полноту Викиданных для решения таких задач.

Описанный выше алгоритм кластеризации является аналогом алгоритма DBSCAN, который характерен тем, что позволяет выявлять кластеры сложной формы (рис. 3).

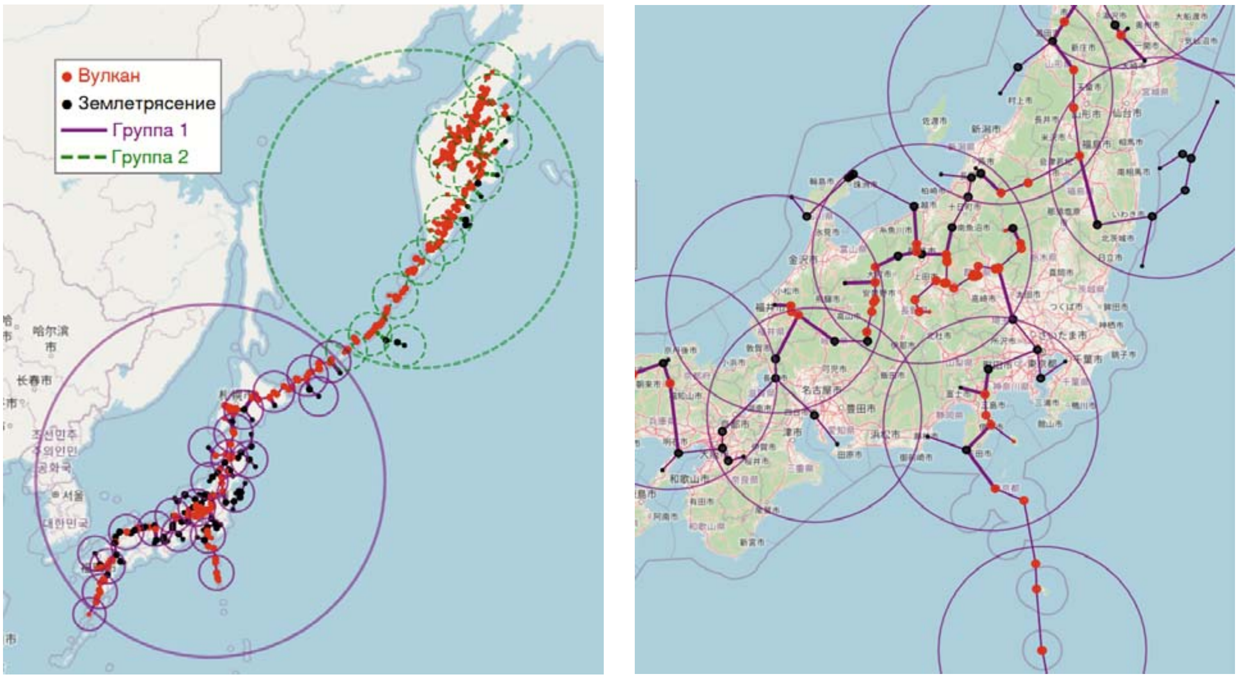
Нерешенный вопрос – как оценить качество найденных кластеров? Постараемся приблизиться к ответу в следующем разделе о числе смешанных кластеров, содержащих и вулканы, и землетрясения.

Смешанные группы в географической кластеризации на основе расстояния между объектами

Основной целью исследования был поиск взаимосвязи между вулканической и сейсмической активностью, поэтому особый интерес представляют смешанные группы, содержащие оба типа объектов.

На рис. 4 представлена зависимость числа трех возможных групп, получаемых при кластеризации, от максимального допустимого расстояния между вулканами и землетрясениями.

[1] Код для генерации интерактивной карты: `https://github.com/componavt/volcano_wikidata/blob/main/src/volcano_earthquake_clustering.ipynb`.



(а) (б)

Рис. 3. Два крупных кластера перед объединением – Японская островная дуга (группа 1) и Курило-Камчатская вулканическая дуга (группа 2):
(а) – диаметры групп составляют 2429 км (сплошная линия) и 1680 км (пунктирная линия);
(б) – фрагмент остовного дерева группы 1 (Япония). Толстые линии – ребра MST, соединяющие вулканы и землетрясения. Центральные вершины обозначены крупными точками
Fig. 3. Two large clusters before merging – Japanese island arc (Group 1) and Kuril-Kamchatka volcanic arc (Group 2):
(а) – cluster diameters are 2429 km (solid line) and 1680 km (dashed line);
(б) – a fragment of the spanning tree for Group 1 (Japan). MST edges connecting volcanoes and earthquakes are shown with thicker lines. Central vertices are slightly enlarged

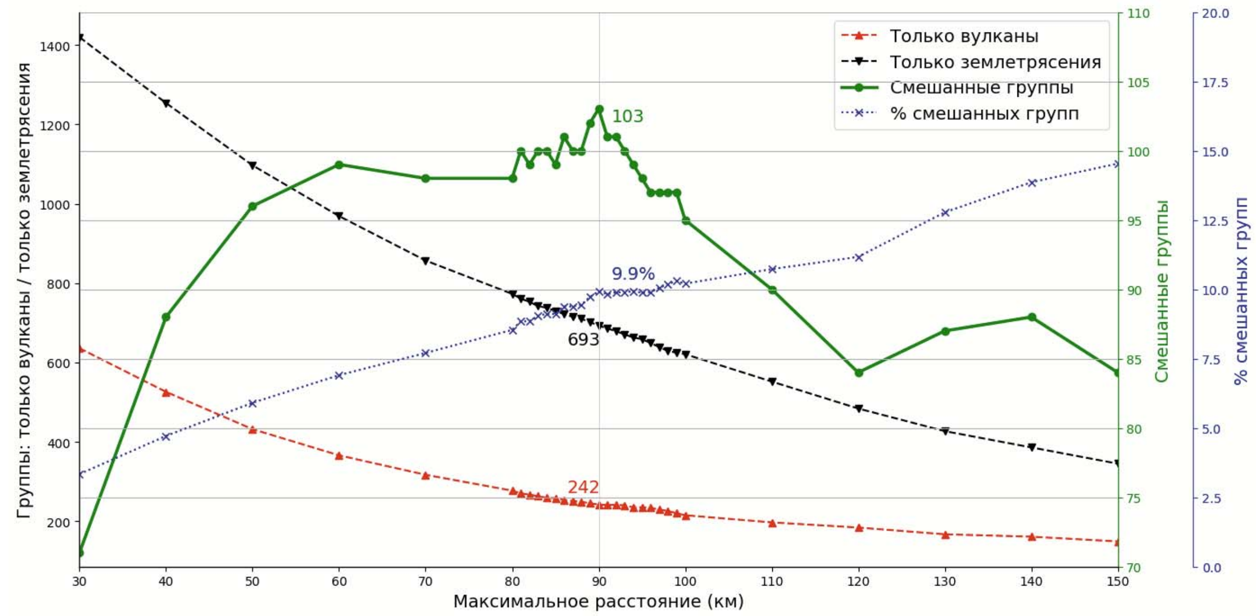


Рис. 4. Зависимость числа вулканических, сейсмических и смешанных групп от порога кластеризации
Fig. 4. Relationship between cluster counts (volcanic/seismic/mixed) and the distance threshold parameter

Максимум смешанных кластеров (103 группы, 9,9 %) наблюдается при `dist_max` = 90 км, что может отражать типичный радиус взаимодействия (*spatial correlation distance*) (∼100 км), зафиксированный для грязевых вулканов [8], где землетрясения статистически значимо провоцируют извержения.

При различных значениях максимального расстояния преобладают группы, содержащие только землетрясения. Например, при `dist_max` = 90 км кластеризация дает 693 сейсмические группы (66,8 %). Это указывает на более локализованное распределение землетрясений.

Заключение

Разработаны SPARQL-скрипты для получения информации о вулканах и землетрясениях из Викиданных. Разработана программа кластеризации и визуализации на карте ближайших вулканов и землетрясений. Разработанные алгоритмы и собранные данные доступны онлайн (`https://github.com/componavt/volcano_wikidata`).

Предложенный алгоритм географической кластеризации на основе расстояния между объектами аналогичен DBSCAN, но адаптирован для пространственных данных:

- использует единое пороговое расстояние (`dist_max`) вместо плотностных параметров DBSCAN (радиус окрестности и минимальное число точек для ядра);
- как и DBSCAN, он способен выявлять кластеры произвольной формы. Выявлены островные дуги Японская и Курило-Камчатская (рис. 3);
- не требует равномерного распределения объектов, критичного для классического DBSCAN.

Проведены анализ и сравнение радиусов кластеризации для разработанного алгоритма кластеризации. Получены два ключевых значения:

- 90 км – оптимальное расстояние, при котором доля смешанных кластеров (вулканы + землетрясения) максимальна (103 группы). Этот результат согласуется с известными данными о радиусе взаимодействия (∼100 км) для грязевых вулканов [8], но впервые расширен на обычные вулканы.
- 127–128 км – диапазон, где происходит качественное изменение структуры, а именно: слияние Японской и Курило-Камчатской дуг в единый суперкластер с резким увеличением диаметра (2,4 → 4 тыс. км), что свидетельствует о потере географической интерпретируемости.

Отмеченный выше максимум при пороге в 90 км – это 103 смешанные группы, которые составляют всего 10 % от числа всех групп. В будущей работе стоит модифицировать алгоритм кластеризации, чтобы в нем была больше доля групп смешанного типа.

У нашего алгоритма кластеризации есть недостаток – это сильная чувствительность к выбору `dist_max`. В диапазоне 120–130 км (рис. 2) даже 1 км изменения порога приводит к качественному скачку: слиянию изолированных кластеров в суперкластеры.

В работе мы исследовали ту часть Викиданных, которая описывает вулканы и землетрясения. Рассмотрены открытые базы данных вулканов GVP (*Global Volcanism Program*) и LaMEVE (*Large Magnitude Explosive Volcanic Eruptions*), поскольку Викиданные частично основываются на этих базах и объединяют их с помощью внешних идентификаторов.

Разработанные SPARQL-скрипты автоматизируют извлечение и предварительную очистку данных, а именно, обеспечивают:

- фильтрацию дубликатов координат (использование агрегирующей функции `SAMPLE` сократило число повторяющихся точек на 5,2 %);
- стандартизацию формата данных для последующей кластеризации;
- возможность регулярного обновления выборки при изменении Викиданных.

Дубликаты координат связаны с такой особенностью Викиданных, как многозначность. На примере объекта Antsiferov Island (Q614531) показано наличие двух альтернативных значений координат (расхождение ∼ 390 м). Подробнее о наличии множества значений, связанных с техническими ошибками или различными идеологиями и точками зрения, см. в работе [10, с. 2–5].

Викиданные служат не просто справочником, а полноценной платформой для исследований. Они позволяют получать пространственные данные для объектов заданного типа за несколько запросов, обеспечивают прозрачность источников через систему внешних идентификаторов.

Викиданные позволяют анализировать пространственные связи вулканов и землетрясений (реализовано в Python-программе), но временной анализ ограничен неполнотой датировок: только 3,3 % вулканов имеют данные об извержениях. Именно объект «извержение», а не «вулкан» может иметь в Викиданных дату. За 2023–2025 годы число датированных извержений в Викиданных выросло с 7 до 68, что свидетельствует об активном пополнении этих данных.

Работа демонстрирует, что Викиданные, несмотря на ограничения, могут служить достоверным источником для исследования пространственных взаимосвязей между вулканами и землетрясениями. Выявленные кластеры соответствуют известным геологическим структурам (Японская и Курило-Камчатская дуги), что подтверждает адекватность данных. Эти результаты демонстрируют, что краудсорсинговые базы данных могут дополнять традиционные научные ресурсы при исследовании глобальных геофизических закономерностей.

Литература

1. *Крижановский А. А., Балакирева М. С., Меньшикова Е. А., Паренченков Е. О., Потес А. С., Трубина Е. Д., Обрегон А.* Программирование Викиданных. Петрозаводск: ПетрГУ, 2024. 186 с. URL: `https://commons.wikimedia.org/?curid=146556492` (дата обращения: 04.05.2025).

2. *Bebbington M. S., Marzocchi W.* Stochastic models for earthquake triggering of volcanic eruptions // J. Geophys. Res.: Solid Earth. 2011. Vol. 116. Art. B05204. doi: 10.1029/2010JB008114

3. *Caricchi L., Townsend M., Rivalta E., Namiki A.* The build-up and triggers of volcanic eruptions // Nat. Rev. Earth Environ. 2021. No. 2. P. 458–476. doi: 10.1038/s43017-021-00174-8

4. *Costa F., Widiwijayanti C., Zar Win Nang T., Fajiculay E., Espinosa-Ortega T., Newhall C.* WOVOdat – the global volcano unrest database aimed at improving eruption forecasts // Disast. Prevent. Manag. 2019. Vol. 28, no. 6. P. 738–751. doi: 10.1108/DPM-09-2019-0301

5. *Crosweller H. S., Arora B., Brown S. K., Cottrell E., Deligne N. I., Ortiz Guerrero N., Hobbs L., Kiyosugi K., Loughlin S. C., Lowndes J., Nayembil M., Siebert L., Sparks R. S. J., Takarada S., Venzke E.* Global database on large magnitude explosive volcanic eruptions (LaMEVE) // J. Appl. Volcanol. 2012. Vol. 1. Art. 4. doi: 10.1186/2191-5040-1-4

6. *Eggert S., Walter T. R.* Volcanic activity before and after large tectonic earthquakes: Observations and statistical significance // Tectonophysics. 2009. No. 471. P. 14–26. doi: 10.1016/j.tecto.2008.10.003

7. *Maneewongvatana S., Mount D. M.* Analysis of approximate nearest neighbor searching with clustered point sets. 1999. arXiv preprint cs/9901013. URL: `https://arxiv.org/pdf/cs/9901013` (дата обращения: 04.05.2025).

8. *Mellors R., Kilb D., Aliyev A., Gasanov A., Yetirmishli G.* Correlations between earthquakes and large mud volcano eruptions // J. Geophys. Res. 2007. Vol. 112. Art. B04304. doi: 10.1029/2006JB004489

9. *Papale P.* Global time-size distribution of volcanic eruptions on Earth // Scientific Reports. 2018. Vol. 8. Art. 6838. doi: 10.1038/s41598-018-25286-y

10. *Santos V., Schwabe D., Lifschitz S.* Can you trust Wikidata? // Semantic Web. 2024. No. 1. P. 1–22. doi: 10.3233/SW-243577

11. *Seropian G., Kennedy B. M., Walter T. R., Ichihara M., Jolly A. D.* A review framework of how earthquakes trigger volcanic eruptions // Nat. Commun. 2021. Vol. 12. Art. 1004. doi: 10.1038/s41467-021-21166-8

12. *Siebert L., Simkin T., Kimberly P.* Volcanoes of the World. 3rd ed. Berkeley: University of California Press, 2010. 568 p.

13. *Simkin T.* Terrestrial volcanism in space and time // Annu. Rev. Earth Planet. Sci. 1993. Vol. 21, no. 1. P. 427–452. doi: 10.1146/annurev.ea.21.050193.002235

14. *Walter T. R., Amelung F.* Volcano-earthquake interaction at Mauna Loa volcano, Hawaii // J. Geophys. Res.: Solid Earth. 2006. Vol. 111. Art. B05204. doi: 10.1029/2005JB003861

15. *Zhao F.* A Systematic Review of Wikidata in Digital Humanities Projects // Dig. Scholarsh. Humanit. 2023. Vol. 38, iss. 2. P. 852–874. doi: 10.1093/llc/fqac083

References

1. *Krizhanovsky A. A., Balakireva M. C., Menshikova E. A., Parenchenkov E. O., Potes A. S., Trubina E. D., Obregon A.* Programming Wikidata. Petrozavodsk: PetrSU; 2024. 186 p. (In Russ.). URL: `https://commons.wikimedia.org/?curid=146556492` (accessed: 04.05.2025).

2. *Bebbington M. S., Marzocchi W.* Stochastic models for earthquake triggering of volcanic eruptions. *J. Geophys. Res.: Solid Earth.* 2011;116:B05204. doi: 10.1029/2010JB008114

3. *Caricchi L., Townsend M., Rivalta E., Namiki A.* The build-up and triggers of volcanic eruptions. *Nat. Rev. Earth Environ.* 2021;2:458–476. doi: 10.1038/s43017-021-00174-8

4. *Costa F., Widiwijayanti C., Zar Win Nang T., Fajiculay E., Espinosa-Ortega T., Newhall C.*

WOVOdat – the global volcano unrest database aimed at improving eruption forecasts. *Disast. Prevent. Manag.* 2019;28(6):738–751. doi: 10.1108/DPM-09-2019-0301

5. *Crosweller H. S., Arora B., Brown S. K., Cottrell E., Deligne N. I., Ortiz Guerrero N., Hobbs L., Kiyosugi K., Loughlin S. C., Lowndes J., Nayembil M., Siebert L., Sparks R. S. J., Takarada S., Venzke E.* Global database on large magnitude explosive volcanic eruptions (LaMEVE). *J. Appl. Volcanol.* 2012;1:4. doi: 10.1186/2191-5040-1-4

6. *Eggert S., Walter T. R.* Volcanic activity before and after large tectonic earthquakes: Observations and statistical significance. *Tectonophysics.* 2009;471:14–26. doi: 10.1016/j.tecto.2008.10.003

7. *Maneewongvatana S., Mount D. M.* Analysis of approximate nearest neighbor searching with clustered point sets. arXiv preprint cs/9901013. 1999. URL: `https://arxiv.org/pdf/cs/9901013` (accessed: 04.05.2025).

8. *Mellors R., Kilb D., Aliyev A., Gasanov A., Yetirmishli G.* Correlations between earthquakes and large mud volcano eruptions. *Geophys. Res.* 2007;112:B04304. doi: 10.1029/2006JB004489

9. *Papale P.* Global time-size distribution of volcanic eruptions on Earth. *Scientific Reports.* 2018;8:6838. doi: 10.1038/s41598-018-25286-y

10. *Santos V., Schwabe D., Lifschitz S.* Can you trust Wikidata? *Semantic Web.* 2024;1:1–22. doi: 10.3233/SW-243577

11. *Seropian G., Kennedy B. M., Walter T. R., Ichihara M., Jolly A. D.* A review framework of how earthquakes trigger volcanic eruptions. *Nat. Commun.* 2021;12:1004. doi: 10.1038/s41467-021-21166-8

12. *Siebert L., Simkin T., Kimberly P.* Volcanoes of the World. 3rd ed. Berkeley: University of California Press; 2010. 568 p.

13. *Simkin T.* Terrestrial volcanism in space and time. *Annu. Rev. Earth Planet. Sci.* 1993;21(1):427–452. doi: 10.1146/annurev.ea.21.050193.002235

14. *Walter T. R., Amelung F.* Volcano-earthquake interaction at Mauna Loa volcano, Hawaii. *J. Geophys. Res.: Solid Earth.* 2006;111:B05204. doi: 10.1029/2005JB003861

15. *Zhao F.* A Systematic Review of Wikidata in Digital Humanities Projects. *Dig. Scholarsh. Humanit.* 2023;38(2):852–874. doi: 10.1093/llc/fqac083

Поступила в редакцию / received: 05.05.2025; принята к публикации / accepted: 23.05.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Крижановский Андрей Анатольевич
канд. техн. наук, ведущий научный сотрудник

e-mail: andew.krizhanovsky@gmail.com

CONTRIBUTOR:

Krizhanovsky, Andrew
Cand. Sci. (Tech.), Leading Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 34–43
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 34–43
DOI: 10.17076/mat2100

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.248

АСИМПТОТИЧЕСКИЙ И ДОПРЕДЕЛЬНЫЙ РЕЖИМЫ МОДЕЛИ ГЕТЕРОГЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ

О. В. Лукашенко[1], С. Н. Астафьев[1], В. А. Иголкин[2], А. С. Румянцев[1*]

[1] *Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910), *ar0@krc.karelia.ru*
[2] *Институт математики и информационных технологий, Петрозаводский государственный университет (пр. Ленина, 33, Петрозаводск, Республика Карелия, Россия, 185910)*

В работе предложен подход к исследованию характеристик гетерогенных вычислительных сетей большого масштаба с помощью гауссовских случайных процессов, обладающих свойствами долгой памяти. Представлены перспективные направления развития указанного подхода.

Ключевые слова: процессы с долгой памятью; системы распределенных вычислений; асимптотический анализ; распределения с тяжелым хвостом

Для цитирования: Лукашенко О. В., Астафьев С. Н., Иголкин В. А., Румянцев А. С. Асимптотический и допредельный режимы модели гетерогенной вычислительной сети // Труды Карельского научного центра РАН. 2025. № 4. С. 34–43. doi: 10.17076/mat2100

O. V. Lukashenko[1], S. N. Astafiev[1], V. A. Igolkin[2], A. S. Rumyantsev[1*]. **ASYMPTOTIC AND PRE-LIMIT REGIMES OF A MODEL OF A HETEROGENEOUS COMPUTING NETWORK**

[1] *Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia), *ar0@krc.karelia.ru*
[2] *Institute of Mathematics and Information Technologies, Petrozavodsk State University (33 Lenin St., 185910 Petrozavodsk, Karelia, Russia)*

Gaussian long-range-dependence processes are suggested as a model for the performance analysis of large-scale heterogeneous computing networks. Possible directions for developing this promising approach further are suggested.

Keywords: long-range dependence; distributed computing systems; asymptotic analysis; heavy-tailed distributions

For citation: Lukashenko O. V., Astafiev S. N., Igolkin V. A., Rumyantsev A. S. Asymptotic and pre-limit regimes of a model of a heterogeneous computing network. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 34–43. doi: 10.17076/mat2100

Введение

Развитие беспроводных технологий сопровождается развитием вычислительных и коммуникационных возможностей цифровых систем. Высокий вычислительный потенциал современных цифровых устройств реализуется с развитием систем (сетей) распределенных и повсеместных вычислений, которые позволяют объединить разнородные вычислительные ресурсы для решения различных задач, от вспомогательных (таких, как управление умным домом [17]) до вычислительноемких [4]. При этом значительные вычислительные возможности таких систем сопровождаются низкой надежностью, высокой степенью разнородности используемых устройств (с точки зрения архитектуры, вычислительной мощности, скорости работы, надежности, используемого питания), трудностью координации вычислений из-за отсутствия централизованного управления и большого числа устройств в сети.

Анализ эффективности [3] и прогнозирование нагрузки [14] вычислительных сетей являются предметом интенсивных исследований. В анализе таких систем широко применяются стохастические модели, однако при исследовании систем большого масштаба целесообразно использовать асимптотические результаты, позволяющие перейти от индивидуальных особенностей отдельных вычислительных узлов к свойствам системы в целом. Следует отметить, что некоторые особенности предельных процессов, такие, например, как *долгая память* (выражающаяся в расходимости ряда автоковариаций процесса), существенно затрудняют прогноз искомых характеристик (например, оценки времени завершения проекта в распределенной вычислительной сети [14]). В данной работе основное внимание уделено предельным результатам для моделей вычислительных сетей с большим числом узлов, а именно, развивается подход, предложенный ранее в [14], где рассматривалась гауссовская аппроксимация процесса выполнения работы распределенной вычислительной сетью. В настоящей работе с помощью статистического моделирования проводится исследование корреляционных свойств допредельного процесса при относительно большом числе неоднородных вычислительных узлов.

Обзор литературы

В данном разделе приведен краткий обзор источников по математическим проблемам в области моделирования распределенных вычислительных сетей.

Среди распределенных вычислительных сетей, в которых ресурсы предоставляются добровольцами (систем добровольных вычислений), можно выделить подклассы вычислительных сетей из персональных компьютеров (desktop grid), облачных систем из персональных компьютеров (ОСПК, desktop clouds), систем мобильных добровольных вычислений (СМДВ, mobile volunteer computing) и систем добровольных параллельных вычислений (parallel volunteer computing) [12]. К особенностям всех указанных типов можно отнести высокую степень гетерогенности вычислительных ресурсов и низкую надежность узлов сети, что требует исследования алгоритмов оптимизации распределения ресурсов и назначения заданий. При этом в каждом подклассе систем цели оптимизации могут различаться: в частности, для СМДВ помимо равномерности распределения нагрузки важны минимизация энергопотребления узлов [9] (что, впрочем, актуально в целом для систем распределенных вычислений [6]), учет ограничений на занимаемую память, минимизация сетевого взаимодействия [10].

Традиционной задачей в области систем распределенных вычислений является разделение ресурсов. При этом важно находить баланс между уровнем избыточности (при котором одна и та же задача может решаться различными узлами для обеспечения надежности) и доступной степенью параллелизма [16]. При решении задач разделения ресурсов в ОСПК учитывается возможность выхода узлов из строя [1]. Для нахождения оптимальных способов разделения ресурсов применяются эвристические методы [5], сравнительный анализ имитационных моделей [8]. К передовым направлениям исследований можно отнести также управление ресурсами с использованием искусственного интеллекта [7].

В качестве математических моделей распределенных систем часто рассматриваются системы типа разветвления–соединения (fork–join), анализ которых классическими методами затруднен в связи с тем, что исследуемый (марковский) процесс является бесконечномерным [22]. В этой связи исследование таких систем ведется в различных предельных режимах: анализ насыщенной системы [11], исследование асимптотик хвоста распределения [18], нахождение стохастических границ для исследуемых характеристик [2, 13, 19] или их аппроксимация [15], исследование предельных процессов [14]. Наконец, отметим также применение машинного обучения для исследования производительности таких систем [23].

Кумулятивная работа в гетерогенной сети

В данном разделе рассматривается модель распределенной вычислительной сети, динамика которой описывается так называемым процессом кумулятивной работы $\{A(t),\, t \geqslant 0\}$, то есть суммарным объемом работы, выполненной вычислительной сетью за промежуток времени $[0, t]$.

Предположим, что вычислительная система состоит из M неоднородных вычислительных узлов, динамика которых описывается в терминах так называемых *on-off процессов*. Все вычислительные узлы относятся к одному из n типов ($n \leqslant M$), характеризующихся различными скоростями выполнения работы. Обозначим далее через M_i число узлов типа i, т. е. $\sum_{i=1}^{n} M_i = M$, а через R_i – объем работы, выполненный в единицу времени узлом типа i. Определим далее on-off процесс как случайный процесс $\left\{I^{(i)}(t),\ t \geqslant 0\right\}$, описывающий активность вычислительного узла типа i следующим образом:

$$I^{(i)}(t) = \begin{cases} R_i,\ t \in \text{ on-периоду}, \\ 0,\ t \in \text{ off-периоду}. \end{cases}$$

Дополнительно предполагается, что периоды активности (on) и периоды простоя (off) образуют последовательности независимых случайных величин, а также независимы между собой. Периоды активности и простоя чередуются попеременно. Индекс $i = 1, \ldots, n$ соответствует типу узла, при этом для выбора конкретного узла данного типа будем использовать нижний индекс $k = 1, \ldots, M_i$.

Отметим, что модели на основе *on-off процессов* рассматривались в приложениях теории массового обслуживания. Основное отличие состоит в том, что наряду с периодами длительности фаз работы и простоя вводятся дополнительные параметры R_i, характеризующие скорость выполнения работы.

Обозначим $A_i(t)$ суммарную работу, выполненную вычислительными узлами типа i:

$$A_i(t) = \int\limits_0^t \left(\sum_{k=1}^{M_i} I_k^{(i)}(u) \right) du.$$

Тогда *кумулятивная работа (агрегированная)*, выполненная всеми узлами за время $[0, t]$, равна

$$A(t) = \sum_{i=1}^{n} A_i(t) = \int\limits_0^t \left(\sum_{i=1}^{n} \sum_{k=1}^{M_i} I_k^{(i)}(u) \right) du. \quad (1)$$

Распределение процесса $A(t)$ фактически будет определяться функциями распределений $F_{on}^{(i)}$ и $F_{off}^{(i)}$ длительностей on и off периодов вычислительных узлов. Для каждого типа $i = 1, \ldots, n$ обозначим

$$\mu_{on}^{(i)},\, \sigma_{on}^{(i)},\, \mu_{off}^{(i)},\, \sigma_{off}^{(i)}$$

математическое ожидание и стандартное отклонение длительностей on и off периодов соответственно.

Далее в статье полагается, что каждое из распределений on- (off-) периодов удовлетворяет одному из двух условий (условия для распределений on и off периодов могут отличаться):

1. Хвост распределения, $\overline{F}_j^{(i)}$, $j \in \{on, off\}$, является *правильно меняющейся функцией* на бесконечности, т. е. при $x \to \infty$:

   $$\overline{F}_j^{(i)} := 1 - F_j^{(i)}(x) \sim x^{-\alpha_j^{(i)}} L_j^{(i)}(x), \quad (2)$$

   где $\alpha_j^{(i)} \in (1, 2)$, $f \sim g$ означает, что $f/g \to 1$, а функция $L_j^{(i)}$ медленно меняется на бесконечности, т. е. для произвольного $t > 0$ выполнено

   $$\lim_{x\to\infty} \frac{L_j^{(i)}(tx)}{L_j^{(i)}(x)} = 1.$$

   Отметим, что в этом случае дисперсии соответствующих on- (off-) периодов будут бесконечными.

2. Дисперсия периодов конечна, $\sigma_j^{(i)} < \infty$. В этом случае для унификации обозначений положим формально $\alpha_j^{(i)} = 2$.

Дальнейший интерес представляет изучение свойств распределения случайного процесса $A(t)$, в частности, распределение следующей случайной величины

$$\tau_D = \min\{t \geqslant 0:\ A(t) \geqslant D\},$$

которая интерпретируется как время, необходимое для решения вычислительной задачи трудоемкости D.

Найти требуемые распределения в явном виде не представляется возможным. Следовательно, приходится полагаться либо на имитационное моделирование, либо на асимптотические методы теории случайных процессов.

Предельное распределение кумулятивной работы

Как уже отмечалось выше, *on-off процессы* активно использовались в области моделирования коммуникационных систем в качестве моделей порождения сетевого трафика. Для агрегированных случайных процессов вида (1) были доказаны асимптотики [21], а именно функциональные предельные теоремы, утверждающие, что соответствующим образом нормированный кумулятивный процесс сходится (при некотором предельном режиме) к гауссовскому процессу с соответствующей корреляционной структурой.

В рамках рассматриваемой модели системы распределенных вычислений данные асимптотические результаты доказаны для случая, когда $R_i = 1$ для всех $i = 1, \ldots, n$. Интуитивно ясно, что более общий случай, когда скорости R_i принимают произвольные значения, не должен радикальным образом сказываться на доказательстве предельных теорем. В действительности данные параметры скажутся только на нормировке процесса.

Теперь приведем основной предельный результат для процессов кумулятивной работы $A_i(t)$, $i = 1, \ldots, n$, доказанный в работе [21]:

$$\lim_{T\to\infty} \lim_{M_i\to\infty} \left\{ \frac{\left(A_i(tT) - R_i M_i p_{on}^{(i)} tT\right)}{T^{H_i}\sqrt{L(T)M_i}} \right\}_{t\geqslant 0} =_d \{c_i B_{H_i}(t)\}_{t\geqslant 0}, \quad (3)$$

где $=_d$ означает равенство по распределению, константы $c_i > 0$ определяются параметрами on/off процесса; $p_{on}^{(i)}$ имеет вид

$$p_{on}^{(i)} = \frac{\mu_{on}^{(i)}}{\mu_{on}^{(i)} + \mu_{off}^{(i)}};$$

$\{B_H(t),\ t \geqslant 0\}$ – дробное броуновское движение (ДБД) с параметром $H \in (0, 1)$ (так называемый параметр Хёрста), центрированный гауссовский процесс с ковариационной функцией

$$\mathrm{Cov}\,(B_H(t), B_H(s)) = \frac{1}{2}\left(t^{2H} + s^{2H} - \mid t - s \mid^{2H}\right);$$

параметр Хёрста H_i, $i = 1, \ldots, n$, предельного ДБД выражается через параметры on/off процессов по следующей формуле:

$$H_i = \frac{3 - \min\left\{\alpha_{on}^{(i)},\, \alpha_{off}^{(i)}\right\}}{2} \in \left[\frac{1}{2}, 1\right). \quad (4)$$

Доказательство результата (3) приводится в работе [21] для случая $R_i = 1$, однако для произвольных положительных скоростей оно повторяется почти дословно.

Неформально данный предельный результат означает, что при достаточно больших значениях M_i и T_i распределение кумулятивной работы A_i, выполненной узлами типа i, сближается с распределением процесса

$$\{R_i M_i p_{on}^{(i)} tT + T^{H_i}\sqrt{L_i(T)M_i}\, c_i B_{H_i}(t)\}_{t\geqslant 0}. \quad (5)$$

В гетерогенном случае справедливо следующее обобщение результата (5) (также адаптированное к исследуемой задаче).

Теорема [21, Теорема 2]. *Пусть выполнены два условия.*

1. *Число узлов согласованно растет, т. е. при $M \to \infty$ выполнено*
$$\lim_{M\to\infty} M_i/M =: w_i > 0,\ i = 1, \ldots, n.$$
2. *Параметр шкалирования по времени растет, т. е. $T \to \infty$.*

Тогда агрегированный процесс $A(tT)$, при соответствующем центрировании и нормировании, сходится в смысле сходимости конечномерных распределений к суперпозиции независимых ДБД, т. е. для больших значений M_i, $i = 1, \ldots, n$, и большого T,

$$A(tT) \approx \mathcal{C}tT + \sum_{i=1}^{n} T^{H_i}\sqrt{L_i(T)M_i}\, c_i B_{H_i}(t), \quad (6)$$

где L_i – медленно меняющиеся на бесконечности функции (выраженные в терминах исходных параметров), B_{H_i} – независимые ДБД с параметрами Хёрста H_i, а параметр $\mathcal{C}$ имеет вид

$$\mathcal{C} = \sum_{i=1}^{n} R_i M_i p_{on}^{(i)}.$$

Асимптотический результат (6) дает теоретическую мотивировку рассматривать сумму независимых ДБД с линейным сносом в качестве модели процесса вычислений в гетерогенной вычислительной сети. При этом заметим, что если у хотя бы одного из n классов вычислительных узлов распределение длительности периода активности или периода простоя имеет вид (2), то в силу (4) соответствующее данному классу значение параметра Хёрста будет строго больше $1/2$, а значит, процесс будет обладать свойством долговременной зависимости [20].

Для простоты изложения далее положим $R_i = 1$, $i = 1, \dots, n$. Аппроксимация (6) может использоваться при достаточно большом числе узлов и на большом масштабе времени, однако ее практическое применение требует дополнительных исследований. В этой связи особую важность имеет исследование локальных (по времени) свойств модели вычислительной сети, и в частности, доступной вычислительной мощности. Нестрого говоря, в работе [21] показано, что ковариационные свойства процесса (6) являются следствием степенного характера убывания автоковариации процесса $\{G(t)\}_{t\geqslant 0}$, описывающего (центрированную и нормированную) *доступную вычислительную мощность* в момент t при большом числе источников,

$$\{G(t)\}_{t\geqslant 0} =_d \lim_{M\to\infty}\{G_M(t)\}_{t\geqslant 0}, \qquad (7)$$

где соответствующий допредельный процесс имеет вид

$$G_M(t) = \frac{1}{\sqrt{M}}\sum_{i=1}^{n}\sum_{k=1}^{M_i}\left(I_k^{(i)}(t) - \mathbb{E}I_k^{(i)}(t)\right) = \sum_{i=1}^{n}\sqrt{\frac{M_i}{M}}G_{M_i}^{(i)}(t), \qquad (8)$$

при этом процессы $\{G_{M_i}^{(i)}(t)\}_{t\geqslant 0}$ описывают (центрированную и нормированную) доступную вычислительную мощность узлов класса i,

$$G_{M_i}^{(i)}(t) = \frac{1}{\sqrt{M_i}}\sum_{k=1}^{M_i}\left(I_k^{(i)}(t) - \mathbb{E}I_k^{(i)}(t)\right).$$

Из (7) и (8) следует, что

$$G(t) = \sum_{i=1}^{n}\sqrt{w_i}G^{(i)}(t),$$

где $\{G^{(i)}(t)\}_{t\geqslant 0}$ есть предельные процессы для $i = 1, \dots, n$,

$$\{G^{(i)}(t)\}_{t\geqslant 0} =_d \lim_{M_i\to\infty}\{G_{M_i}^{(i)}\}_{t\geqslant 0}.$$

Обозначим γ_M автоковариационную функцию процесса (7),

$$\gamma_M(u) = \mathrm{Cov}(G_M(0), G_M(u)). \qquad (9)$$

С учетом результатов, представленных в работе [21], можно показать, что при больших u и больших M справедлива асимптотика

$$\gamma_M(u) \approx \gamma(u) = \mathrm{Cov}(G(0), G(u)) \approx \sum_{i=1}^{n} d_i w_i u^{2H_i - 2}, \qquad (10)$$

где d_i есть некоторые константы, а параметры H_i имеют вид (4). Это означает, что при (хотя бы одном) $H_i > 1/2$ имеет место расходимость ряда автоковариаций процесса. Таким образом, при достаточно большом числе узлов M автоковариационная функция γ_M должна проявлять степенной характер убывания.

Вычислительные эксперименты

Для иллюстрации указанных теоретических результатов проведены численные эксперименты. В качестве распределения для on- и off-периодов источника типа $i = 1, \dots, n$ здесь и далее используется распределение Парето первого типа, хвост которого имеет вид

$$\overline{F}_j^{(i)}(x) = x^{-\alpha_j^{(i)}}, x > 1, j \in \{on, off\}.$$

Отметим, что такой хвост распределения относится к классу правильно меняющихся функций, при этом соответствующая медленно меняющаяся функция тождественно равна 1. С учетом (10) целесообразно ожидать, что при больших M имеет место асимптотически линейное убывание автоковариации в двойных логарифмических координатах,

$$\log\gamma_M(u) \approx \beta_M \log u + c_M,$$

где c_M – некоторая константа, $\beta_M \approx 2H - 2$,

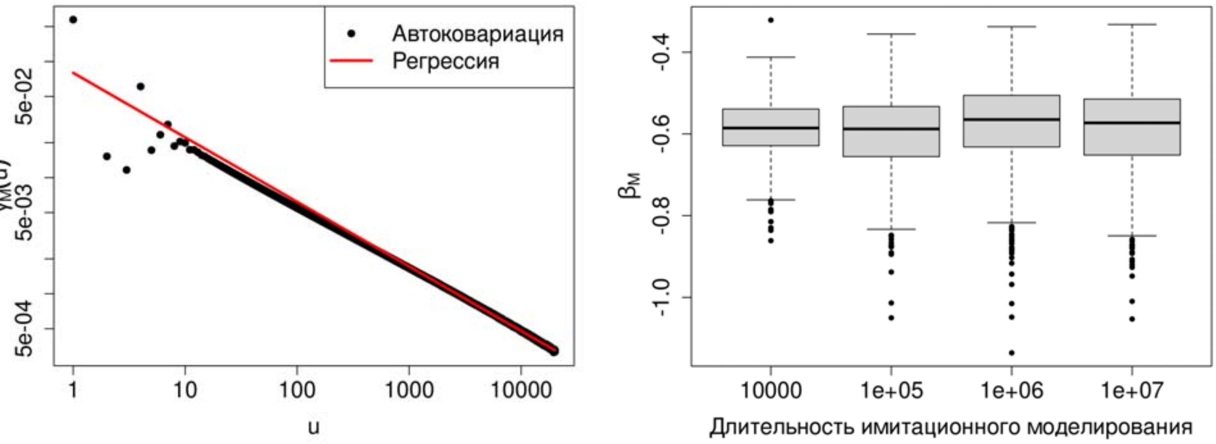
$$H = \max_{i=1,\dots,n} H_i.$$

С учетом (4) и (10)

$$\beta_M \approx 1 - \min_{i=1,\dots,n}\min\left\{\alpha_{on}^{(i)}, \alpha_{off}^{(i)}\right\}. \qquad (11)$$

Величину β_M можно оценить с помощью построения наклона линии регрессии для выборочной автоковариации, однако в неоднородном случае такой наклон необходимо строить для достаточно больших значений u и, соответственно, при достаточной длине траекторий.

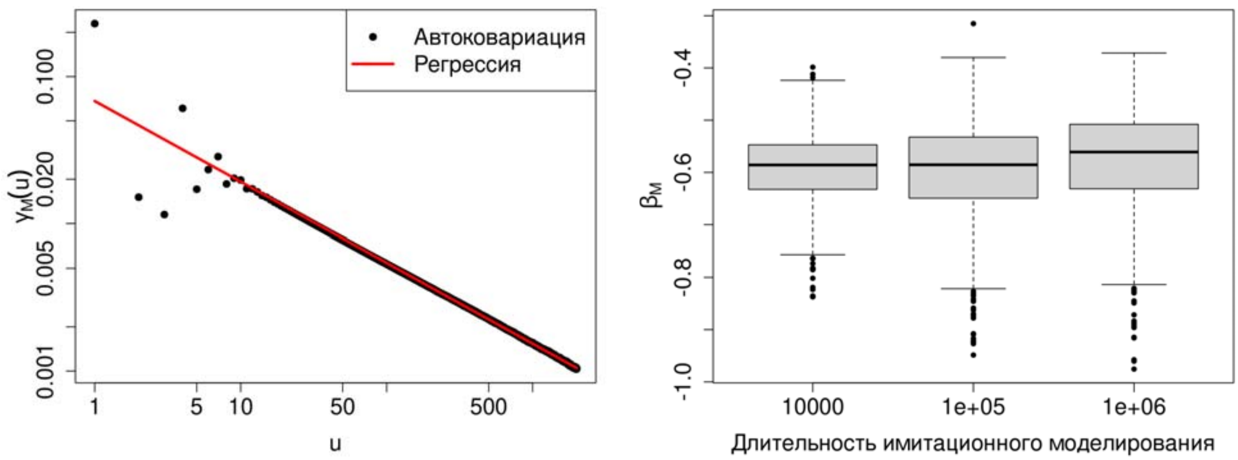
В первом эксперименте рассмотрен однородный случай, т. е. предполагается, что в системе один класс узлов (для упрощения обозначений опустим соответствующие верхние индексы и сохраним обозначение γ_M для *выборочной* автоковариации). Выборочная автоковариация $\gamma_M(u)$ для $M = 10^3$ и $M = 10^4$ узлов при $u = 1, \dots, 20000$ (с шагом, равным 1) вычислена по 988 и 1104 траекториям процесса длиной 10^7 и 10^6 единиц времени соответственно. Использовались следующие значения параметров узлов: $\alpha_{on} = 1{,}5$ (что соответствует условию (2)), $\alpha_{off} = 2{,}5$ (т. е. дисперсия off-периода конечна). Тогда из (4) и (11) следует, что $H = 0{,}75$, $\beta_M = -0{,}5$. График автоковариационной функции (9) построен при заранее выбранном M в логарифмических координатах по обеим осям. Из (11) следует, что при больших u, M график должен соответствовать прямой линии с наклоном $-0{,}5$. В качестве оценки β_M брался наклон линии регрессии по значениям $\gamma_M(u)$ для $u = 10^3, \ldots, 2 \cdot 10^4$. Как видно на рис. 1 (а) и 2 (а), график соответствует ожидаемому.



(а) (б)

Рис. 1. Результаты экспериментов с узлами одного типа, $M = 10^3$:
(а) – среднее (по числу траекторий) значение автоковариации. Наклон линии регрессии $\beta_M \approx -0{,}56$;
(б) – оценка распределения β_M в зависимости от длины траектории $\tau_j = 10^j$, для $j = 4, \ldots, 7$, построенная по интервалу параметра $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^{-3}]$

Fig. 1. Experimental results for single-type nodes, $M = 10^3$:
(a) – average (trajectorywise) autocovariance value. The slope of regression line $\beta_M \approx -0.56$;
(б) – distribution estimate for β_M vs. trajectory length $\tau_j = 10^j$, for $j = 4, \ldots, 7$, for the parameter value $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^{-3}]$

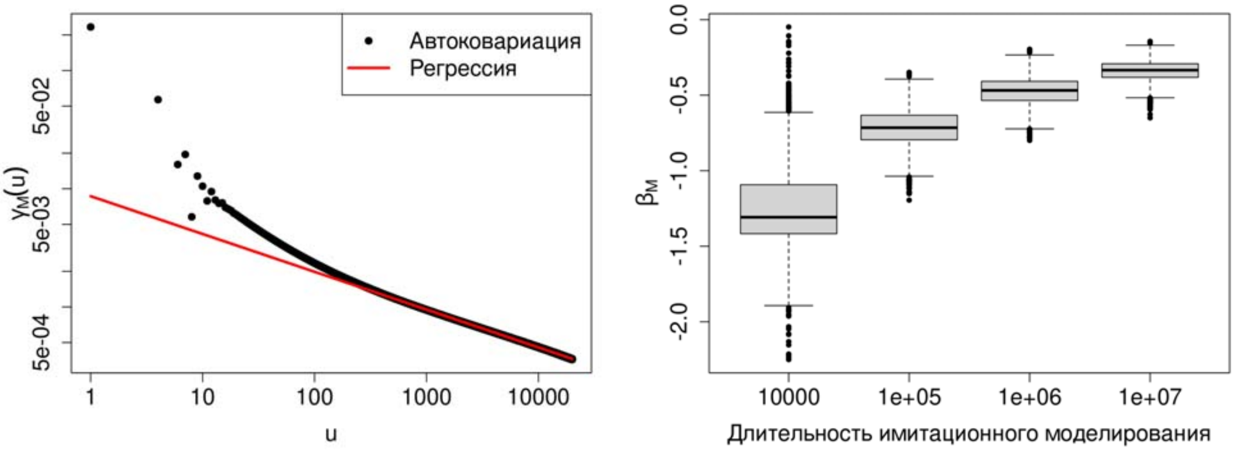


(а) (б)

Рис. 2. Результаты экспериментов с узлами одного типа, $M = 10^4$:
(а) – среднее (по числу траекторий) значение автоковариации. Наклон линии регрессии $\beta_M \approx -0{,}55$;
(б) – оценка распределения β_M в зависимости от длины траектории $\tau_j = 10^j$, для $j = 4$, 5, 6, построенная по интервалу параметра $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^{-3}]$

Fig. 2. Experimental results for single-type nodes, $M = 10^4$:
(a) – average (trajectorywise) autocovariance value. The slope of regression line $\beta_M \approx -0.55$;
(б) – distribution estimate for β_M vs. trajectory length $\tau_j = 10^j$, for $j = 4$, 5, 6, for the parameter value $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^{-3}]$

Поведение системы при нескольких типах узлов отличается от случая с одним типом узла. Во втором эксперименте рассмотрена неоднородная система с двумя типами узлов. С помощью имитационного моделирования построены 4886 независимых траекторий процесса $G_M(t)$, каждая длиной 10^7 единиц времени, при следующих значениях параметров: $M = 10^3$, $\alpha_{on}^{(1)} = 1{,}1$, $\alpha_{on}^{(2)} = 1{,}9$, $\alpha_{off}^{(1)} = 2{,}1$, $\alpha_{off}^{(2)} = 2{,}56$, $w_1 = 0{,}1$, $w_2 = 0{,}9$. Параметры системы подобраны таким образом, чтобы средние значения длительности on и off периодов для агрегированного процесса соответствовали первому эксперименту. Затем для исследования чувствительности оценки β_M к максимальному лагу автоковариации выполнено усечение траекторий до общей длины $\tau_j = 10^j$, а оценка построена для значений $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^3]$, $j = 4, \dots, 7$. Результаты представлены на рис. 3.



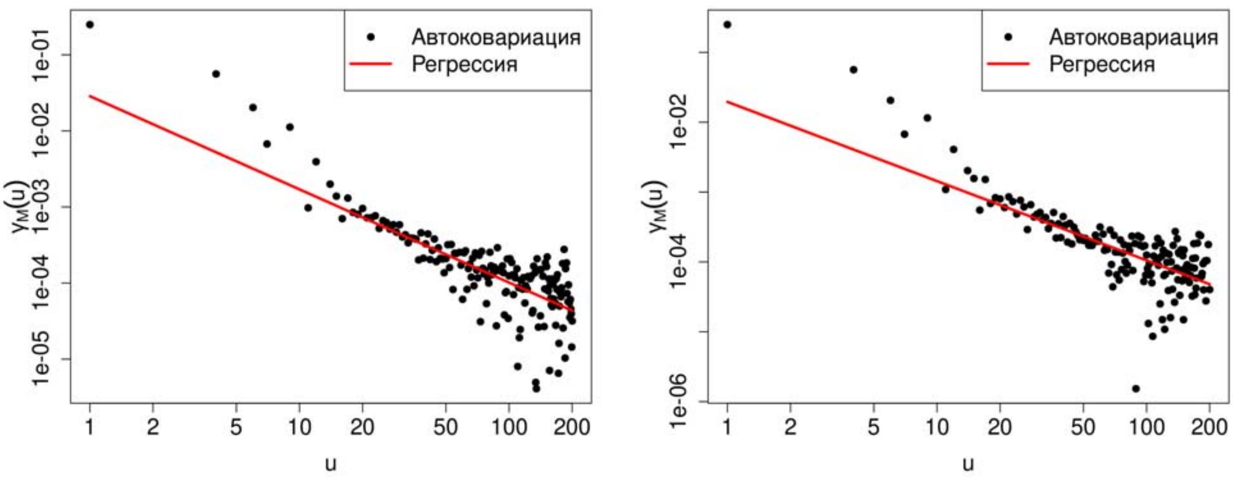
(а) (б)

Рис. 3. Результаты численных экспериментов для сети с узлами двух типов, $H_i > 0{,}5$:
(а) – среднее (по числу траекторий) значение автоковариации. Наклон линии регрессии $\beta_M \approx -0{,}32$;
(б) – оценка распределения β_M в зависимости от длины траектории $\tau_j = 10^j$, для $j = 4, \dots, 7$, построенная по интервалу параметра $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^{-3}]$;

Fig. 3. Numerical results for a network with two types of nodes, $H_i > 0.5$:
(а) – average (trajectorywise) autocovariance value. The slope of regression line $\beta_M \approx -0.32$;
(б) – distribution estimate for β_M vs. trajectory length $\tau_j = 10^j$, for $j = 4, \dots, 7$, for the parameter value $u \in [\tau_j \cdot 10^{-4}, \tau_j \cdot 2 \cdot 10^{-3}]$

На рис. 3 (а) хорошо видно отклонение усредненной (по траекториям) оценки автоковариации для небольших u от прямой линии (линии регрессии), однако, как и ожидается, для больших u график оценки в двойных логарифмических координатах приближенно соответствует прямой линии. В отличие от первого эксперимента из рис. 3 (б) видно, что абсолютное значение оценки и дисперсия оценки β_M уменьшаются с увеличением времени имитационного моделирования τ_j и, соответственно, увеличением диапазона значений u. При этом на рис. 3 (б) межквартильный размах для оценки при $j = 6, 7$ примерно соответствует размаху на рис. 1 (б) и 2 (б). Из (11) ожидается, что $\beta_M \approx -0{,}1$, что примерно соответствует оценке, представленной на рис. 3 (а). Отметим, однако, что теоретическое значение не достигается, но точность с увеличением максимального значения лага увеличивается. Можно предположить, что для получения более точной оценки требуется построение автоковариации для более высоких значений лага (что потребует увеличения длительности имитационного моделирования).

На рис. 4 представлены результаты третьего эксперимента с параметрами $\alpha_{on}^{(1)} = 2{,}1$, $\alpha_{on}^{(2)} = 2{,}9$, $\alpha_{off}^{(1)} = 2{,}3$, $\alpha_{off}^{(2)} = 2{,}56$, $w_1 = 0{,}1$, $w_2 = 0{,}9$ при общей длине траекторий 10^5. При таких параметрах $H_i = 0{,}5$, $i = 1, 2$, т. е. отсутствует долговременная зависимость и процесс $G_M(t)$ аппроксимируется винеровским процессом. Теоретический наклон прямой в этом случае составляет -1, что соответствует экспериментальным результатам.



$M = 10^3$, $\beta_M = -1{,}22$ $\quad$ $M = 10^4$, $\beta_M = -1{,}13$

Рис. 4. Результаты экспериментов с узлами двух разных типов, $H_i = 0{,}5$. Для каждого графика использовалось 100 независимых траекторий

Fig. 4. Numerical results for a network with two types of nodes, $H_i = 0.5$. 100 independent trajectories were used for each graph

Заключение

Выполнен анализ допредельного процесса кумулятивной работы в гетерогенной вычислительной сети большого масштаба с помощью статистического моделирования. Отметим, что на практике асимптотическое равенство (6) позволяет исследовать свойства таких сетей без детализации характеристик отдельных узлов.

С учетом известных аналитических результатов, а также задач, поставленных в данной статье, целесообразно предложить следующие направления дальнейших исследований:

- исследование процесса кумулятивной работы в гетерогенной системе в допредельном режиме (при относительно небольшом числе узлов);
- анализ скорости сходимости к дробному броуновскому движению процесса выполненной кумулятивной работы;
- анализ чувствительности ключевых характеристик вычислительных сетей (время выполнения фиксированного объема работы) к макро- (параметр Хёрста) и микро- (параметры ф. р. интервалов on-off процессов, скорости серверов) параметрам исследуемых систем;
- исследование статистических моделей соответствующих процессов, учитывающих эффект долгой памяти (например, в классе моделей FARIMA).

Отметим, что некоторые из указанных исследований могут быть выполнены с использованием имитационного моделирования.

Литература

1. *Alwabel A., Walters R. J., Wills G. B.* A resource allocation model for desktop clouds // Web Services: Concepts, Methodologies, Tools, and Applications, IGI Global. 2019. P. 258–279. doi: 10.4018/978-1-5225-7501-6.ch016

2. *Baccelli F., Makowski A. M., Shwartz A.* The Fork-Join queue and related systems with synchronization constraints: stochastic ordering and computable bounds // Adv. Appl. Probab. 1989. Vol. 21, no. 3. P. 629–660. doi: 10.2307/1427640

3. *Chakravarthy S. R., Rumyantsev A. S.* Efficient redundancy techniques in cloud and desktop grid systems using MAP/G/c-type queues // Open Eng. 2018. Vol. 1, no. 8. P. 17–31. doi: 10.1515/eng-2018-0004

4. *Chernov I. A., Nikitina N. N., Ivashko E. E.* Task scheduling in desktop grids: Open problems // Open Eng. 2017. Vol. 7, no. 1. P. 343–351. doi: 10.1515/eng-2017-0038

5. *Han B., Zhang R.* Stochastic matrix modelling and scheduling algorithm of distributed intelligent computing system // Math. Probl. Eng. 2022. Iss. 1. Art. 3730738. doi: 10.1155/2022/3730738

6. *Hu B., Cao Z., Zhou M.* Energy-minimized scheduling of real-time parallel workflows on heterogeneous distributed computing systems // IEEE Transactions on Services Computing. 2022. Vol. 15, no. 5. P. 2766–2779. doi: 10.1109/TSC.2021.3054754

7. *Ilager S., Muralidhar R., Buyya R.* Artificial Intelligence (AI)-centric management of resources in modern distributed computing systems // 2020 IEEE Cloud Summit. Harrisburg, USA, 2020. P. 1–10. doi: 10.1109/IEEECloudSummit48914.2020.00007

8. *Kurochkin I., Kondrashov N.* Comparison of various algorithms for scheduling tasks in a desktop grid system using a ComBos simulator // Communications in Computer and Information Science. 2020. Vol. 1304. P. 29–40. doi: 10.1007/978-3-030-66895-2_3

9. *Lin Z., Yang J., Wu C., Chen P.* Energy-efficient task offloading for distributed edge computing in vehicular networks // IEEE Transactions on Vehicular Technology. 2024. Vol. 73, iss. 9. P. 14056–14061. doi: 10.1109/TVT.2024.3395893

10. *Ma P., Garg S., Barika M.* Research allocation in mobile volunteer computing system: Taxonomy, challenges and future work // Future Gener. Comput. Syst. 2024. Vol. 154. P. 251–265. doi: 10.1016/j.future.2024.01.015

11. *Marin A., Rossi S., Sottana M.* Dynamic resource allocation in Fork-Join queues // ACM Transactions on Modeling and Performance Evaluation of Computing Systems. 2020. Vol. 5, no. 1. Art. 3. doi: 10.1145/3372376

12. *Mengistu T. M., Che D.* Survey and taxonomy of volunteer computing // ACM Computing Survey. 2020. Vol. 52, no. 3. P. 1–35. doi: 10.1145/3320073

13. *Mohanty M., Gautam G., Aggarwal V., Parag P.* Analysis of Fork-Join scheduling on heterogeneous parallel servers // IEEE/ACM Transactions on Networking. 2024. Vol. 32, iss. 6. P. 4798–4809. doi: 10.1109/TNET.2024.3432183

14. *Morozov E. V., Lukashenko O. V., Rumyantsev A. S., Ivashko E. E.* A Gaussian approximation of runtime estimation in a desktop grid project // 2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). 2017. P. 107–111. doi: 10.1109/ICUMT.2017.8255158

15. *Osman R., Harrison P. G.* Approximating closed Fork-Join queueing networks using product-form stochastic Petri-nets // J. Syst. Softw. 2015. Vol. 110. P. 264–278. doi: 10.1016/j.jss.2015.08.036

16. *Peng P., Soljanin E., Whiting P.* Diversity vs. parallelism in distributed computing with redundancy // 2020 IEEE International Symposium on Information Theory (ISIT). Los Angeles, USA, 2020. P. 257–262. doi: 10.1109/ISIT44484.2020.9174030

17. *Pokhrel S. R., Vu H. L., Cricenti A. L.* Adaptive admission control for IoT applications in home WiFi networks // IEEE Transactions on Mobile Computing. 2020. Vol. 19, no. 12. P. 2731–2742. doi: 10.1109/TMC.2019.2935719

18. *Raaijmakers Y., Borst S., Boxma O.* Fork-Join and redundancy systems with heavy-tailed job sizes // Queueing Syst. 2023. Vol. 103. P. 131–159. doi: 10.1007/s11134-022-09856-6

19. *Rizk A., Poloczek F., Ciucu F.* Stochastic bounds in Fork-Join queueing systems under full and partial mapping // Queueing Syst. 2016. Vol. 83. P. 261–291. doi: 10.1007/s11134-016-9486-x

20. *Samorodnitsky G.* Long range dependence // Foundations and trends in stochastic systems. 2006. Vol. 1, no. 3. P. 163–257. doi: 10.1561/0900000004

21. *Taqqu M. S., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modeling // Comput. Commun. Rev. 1997. Vol. 27. P. 5–23. doi: 10.1145/263876.263879

22. *Thomasian A.* Analysis of Fork/Join and related queueing systems // ACM Computing Surveys. 2014. Vol. 47, no. 2. P. 1–71. doi: 10.1145/2628913

23. *Vishnevsky V. M., Klimenok V. I., Sokolov A. M., Larionov A. A.* Investigation of the Fork-Join system with Markovian arrival process arrivals and phase-type service time distribution using machine learning methods // Mathematics. 2024. Vol. 12. P. 659. doi: 10.3390/math12050659

References

1. *Alwabel A., Walters R. J., Wills G. B.* A resource allocation model for desktop clouds. *Web Services: Concepts, Methodologies, Tools, and Applications, IGI Global.* 2019. P. 258–279. doi: 10.4018/978-1-5225-7501-6.ch016

2. *Baccelli F., Makowski A. M., Shwartz A.* The Fork-Join queue and related systems with synchronization constraints: stochastic ordering and computable bounds. *Adv. Appl. Probab.* 1989;21(3):629–660. doi: 10.2307/1427640

3. *Chakravarthy S. R., Rumyantsev A. S.* Efficient redundancy techniques in cloud and desktop grid systems using MAP/G/c-type queues. *Open Eng.* 2018;1(8):17–31. doi: 10.1515/eng-2018-0004

4. *Chernov I. A., Nikitina N. N., Ivashko E. E.* Task scheduling in desktop grids: Open problems. *Open Eng.* 2017;7(1):343–351. doi: 10.1515/eng-2017-0038

5. *Han B., Zhang R.* Stochastic matrix modelling and scheduling algorithm of distributed intelligent computing system. *Math. Probl. Eng.* 2022;1:3730738. doi: 10.1155/2022/3730738

6. *Hu B., Cao Z., Zhou M.* Energy-minimized scheduling of real-time parallel workflows on heterogeneous distributed computing systems. *IEEE Transactions on Services Computing.* 2022; 15(5):2766–2779. doi: 10.1109/TSC.2021.3054754

7. *Ilager S., Muralidhar R., Buyya R.* Artificial Intelligence (AI)-centric management of resources in modern distributed computing systems. *2020 IEEE Cloud Summit.* Harrisburg, USA; 2020. P. 1–10. doi: 10.1109/IEEECloudSummit48914.2020.00007

8. *Kurochkin I., Kondrashov N.* Comparison of various algorithms for scheduling tasks in a desktop grid system using a ComBos simulator. *Communications in Computer and Information Science.* 2020;1304:29–40. doi: 10.1007/978-3-030-66895-2_3

9. *Lin Z., Yang J., Wu C., Chen P.* Energy-efficient task offloading for distributed edge computing in vehicular networks. *IEEE Transactions on Vehicular Technology.* 2024;73(9):14056–14061. doi: 10.1109/TVT.2024.3395893

10. *Ma P., Garg S., Barika M.* Research allocation in mobile volunteer computing system: Taxonomy, challenges and future work. *Future Gener. Comput. Syst.* 2024;154:251–265. doi: 10.1016/j.future.2024.01.015

11. *Marin A., Rossi S., Sottana M.* Dynamic resource allocation in Fork-Join queues. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems.* 2020;5(1):3. doi: 10.1145/3372376

12. *Mengistu T. M., Che D.* Survey and taxonomy of volunteer computing. *ACM Computing Survey.* 2020;52(3):1–35. doi: 10.1145/3320073

13. *Mohanty M., Gautam G., Aggarwal V., Parag P.* Analysis of Fork-Join scheduling on heterogeneous parallel servers. *IEEE/ACM Transactions on Networking.* 2024;32(6):4798–4809. doi: 10.1109/TNET.2024.3432183

14. *Morozov E. V., Lukashenko O. V., Rumyantsev A. S., Ivashko E. E.* A Gaussian approximation of runtime estimation in a desktop grid project. *2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT).* 2017. P. 107–111. doi: 10.1109/ICUMT.2017.8255158

15. *Osman R., Harrison P. G.* Approximating closed Fork-Join queueing networks using product-form stochastic Petri-nets. *J. Syst. Softw.* 2015;110:264–278. doi: 10.1016/j.jss.2015.08.036

16. *Peng P., Soljanin E., Whiting P.* Diversity vs. parallelism in distributed computing with redundancy. *2020 IEEE International Symposium on Information Theory (ISIT).* Los Angeles, USA; 2020. P. 257–262. doi: 10.1109/ISIT44484.2020.9174030

17. *Pokhrel S. R., Vu H. L., Cricenti A. L.* Adaptive admission control for IoT applications in home WiFi networks. *IEEE Transactions on Mobile Computing.* 2020;19(12):2731–2742. doi: 10.1109/TMC.2019.2935719

18. *Raaijmakers Y., Borst S., Boxma O.* Fork-Join and redundancy systems with heavy-tailed job sizes. *Queueing Syst.* 2023;103:131–159. doi: 10.1007/s11134-022-09856-6

19. *Rizk A., Poloczek F., Ciucu F.* Stochastic bounds in Fork-Join queueing systems under full and partial mapping. *Queueing Syst.* 2016;83:261–291. doi: 10.1007/s11134-016-9486-x

20. *Samorodnitsky G.* Long range dependence. *Foundations and trends in stochastic systems.* 2006;1(3):163–257. doi: 10.1561/0900000004

21. *Taqqu M. S., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modeling. *Comput. Commun. Rev.* 1997;27:5–23. doi: 10.1145/263876.263879

22. *Thomasian A.* Analysis of Fork/Join and related queueing systems. *ACM Computing Surveys.* 2014;47(2):1–71. doi: 10.1145/2628913

23. *Vishnevsky V. M., Klimenok V. I., Sokolov A. M., Larionov A. A.* Investigation of the Fork-Join system with Markovian arrival process arrivals and phase-type service time distribution using machine learning methods. *Mathematics.* 2024;12:659. doi: 10.3390/math12050659

Поступила в редакцию / received: 05.05.2025; принята к публикации / accepted: 10.06.2025.
Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРАХ:

Лукашенко Олег Викторович
канд. физ.-мат. наук, старший научный сотрудник
e-mail: lukashenko@krc.karelia.ru

Астафьев Сергей Николаевич
аспирант, младший научный сотрудник
e-mail: seryymail@mail.ru

Иголкин Владислав Александрович
преподаватель
e-mail: easyufbln@gmail.com

Румянцев Александр Сергеевич
д-р физ.-мат. наук, ведущий научный сотрудник
e-mail: ar0@krc.karelia.ru

CONTRIBUTORS:

Lukashenko, Oleg
Cand. Sci. (Phys.-Math.), Senior Researcher

Astafiev, Sergey
Doctoral Student, Junior Researcher

Igolkin, Vladislav
Lecturer

Rumyantsev, Alexander
Dr. Sci. (Phis.-Math.), Leading Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 44–53
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 44–53
DOI: 10.17076/mat2024

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.212.2+519.175.4

О ЛОКАЛЬНОМ КЛАСТЕРНОМ КОЭФФИЦИЕНТЕ КОНФИГУРАЦИОННОГО ГРАФА

Ю. Л. Павлов

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Рассматриваются конфигурационные графы с N вершинами, степени которых независимы и одинаково распределены. Распределение случайной величины ξ, равной степени любой вершины графа, при $k \to \infty$ удовлетворяет условию

$$\mathbf{P}\{\xi = k\} \sim \frac{L}{k^\tau \ln^g k},$$

где L, $g > 0$, $\tau \in (2, 3)$. Изучаются локальные кластерные коэффициенты $c(s)$ таких графов, отражающие вероятность того, что две разные вершины, смежные с одной и той же вершиной степени s, тоже соединены ребром. Доказана предельная теорема для $c(s)$ при $N \to \infty$ и $s = o(N^{(\tau-2)/(\tau-1)})$.

Ключевые слова: конфигурационный граф; локальный кластерный коэффициент; предельные теоремы

Для цитирования: Павлов Ю. Л. О локальном кластерном коэффициенте конфигурационного графа // Труды Карельского научного центра РАН. 2025. № 4. С. 44–53. doi: 10.17076/mat2024

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

Yu. L. Pavlov. ON THE LOCAL CLUSTERING COEFFICIENT OF A CONFIGURATION GRAPH

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

We consider configuration graphs with N vertices whose degrees are independent and identically distributed. The distribution of the random variable ξ, which is defined as the degree of any vertex, is assumed to satisfy the condition

$$\mathbf{P}\{\xi = k\} \sim \frac{L}{k^\tau \ln^g k},$$

as $k \to \infty$, where L, $g > 0$, $\tau \in (2, 3)$. We study the local clustering coefficient $c(s)$ which can be interpreted as the probability that two different vertices adjacent to a vertex of degree s are also connected by an edge. The limit theorem is proved for the $c(s)$ as $N \to \infty$ and $s = o(N^{(\tau-2)/(\tau-1)})$.

K e y w o r d s: configuration graph; local clustering coefficient; limit theorems

F o r c i t a t i o n: Pavlov Yu. L. On the local clustering coefficient of a configuration graph. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2025. No. 4. P. 44–53. doi: 10.17076/mat2024

F u n d i n g. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

Введение

Случайные графы успешно используются в качестве моделей разнообразных сложных сетей коммуникаций (Интернет, социальные, электрические, транспортные, биологические, телефонные сети и т. п.). Описанию таких моделей посвящена обширная литература, наиболее полные обзоры публикаций можно найти, например, в книгах [8, 9, 13]. Многие сложные сети обладают схожими свойствами, что, естественно, находит отражение в соответствующих моделях и позволяет строить и исследовать графы, достаточно адекватно описывающие свойства и динамику различных сетей. При моделировании конкретных сетевых структур такой подход сводится к использованию типовых случайных графов и уточнению их параметров.

Наиболее известным общим свойством современных сложных сетей, выявленным в ходе многочисленных наблюдений, является то, что степени узлов сети можно считать независимыми одинаково распределенными случайными величинами, имеющими степенное распределение. С этим свойством связана и другая особенность, состоящая в высоком уровне кластеризации сети, когда узлы нередко образуют группы, концентрирующиеся вокруг узлов с высокой степенью и тесными связями друг с другом внутри таких групп, в то время как плотность связей между группами существенно ниже.

Одним из часто используемых для моделирования сетей видов случайных графов является конфигурационный граф [7]. Различают два вида конструкции такого графа – с заданным множеством степеней вершин и со случайными степенями. Будем считать, что конфигурационный граф содержит N вершин. В первом варианте предполагается, что степени вершин известны и образуют множество $\{d_1, d_2, \ldots, d_N\}$, где d_i – степень i-й вершины, $i = 1, \ldots, N$. Понятно, что сумма степеней $L_N = d_1 + \ldots + d_N$ должна быть четной. Степень каждой вершины равна числу инцидентных этой вершине различимых полуребер, т. е. ребер, для которых смежные вершины еще не определены. Граф строится путем попарного случайного соединения полуребер друг с другом для образования ребер. Обычно такое соединение производится равновероятно, что и предполагается в нашей статье. Построение графа со случайными степенями отличается от предыдущего варианта только тем, что степени вершин $1, \ldots, N$ являются реализациями независимых целочисленных неотрицательных случайных величин $\xi_1, \ldots, \xi_N$ соответственно и образуют случайный вектор $\mathbf{d} = (\xi_1, \ldots, \xi_N)$. Именно этот случай рассматривается в статье. Для того чтобы обеспечить четность суммы степеней вершин в случае нечетного значения суммы $\zeta_N = \xi_1 + \ldots + \xi_N$, в граф вводится вспомогательная вершина единичной степени или к произвольно выбранной (например, равновероятно) вершине графа добавляется одно полуребро. Известно, что появление такого дополнительного элемента не влияет на асимптотические свойства графа при $N \to \infty$, поэтому далее мы считаем, что сумма ζ_N четна. Понятно, что такая конструкция конфигурационного графа допускает появление петель и кратных ребер.

В последние годы значительное внимание уделяется исследованию конфигурационных графов со случайными независимыми одинаково распределенными степенями вершин. Обозначим ξ случайную величину, закон распределения которой совпадает с распределением $\xi_1, \ldots, \xi_N$ и имеет вид

$$p_k = \mathbf{P}\{\xi = k\} = \frac{h(k)}{k^\tau}, \qquad (1)$$

где $k = 1, 2, \ldots$, $\tau > 1$, а $h(x)$ – медленно меняющаяся на бесконечности функция. Известно (см. [6, 9, 10]), что для моделирования сложных сетей коммуникаций подходящими являются графы с распределением (1) степеней вершин, где $\tau \in (2, 3)$.

В дальнейшем мы будем использовать важное свойство медленно меняющихся функций, состоящее в том, что при достаточно больших

x и любом $\delta > 0$

$$x^{-\delta} < h(x) < x^{\delta}. \qquad (2)$$

Из (1) и (2) следует, что распределение случайной величины ξ имеет конечное математическое ожидание и бесконечную дисперсию. Обозначим

$$m = \mathbf{E}\xi. \qquad (3)$$

Количественными характеристиками степени кластеризации структуры графа являются кластерные коэффициенты. Их значения находятся на основе подсчета числа треугольников графа. Три различные вершины образуют треугольник, если все они соединены друг с другом ребрами. Такое определение означает, как нетрудно видеть, что наличие кратных ребер и петель не влияет на число треугольников. Существует несколько типов кластерных коэффициентов, наиболее известные из них – глобальный и локальный. Глобальный кластерный коэффициент равен отношению утроенного числа треугольников графа к числу пар смежных ребер (см., например, [9]). Таким образом, его можно интерпретировать как вероятность того, что две вершины графа, противоположные общей вершине двух смежных ребер, тоже соединены ребром. В статье [11] рассматривался локальный кластерный коэффициент вершин графа, имеющих степень s, $s \geqslant 2$. Для каждой вершины степени s можно найти число различных треугольников, содержащих эту вершину. Перебирая одну за другой все такие вершины, найдем сумму Δ_s, равную числу треугольников, содержащих хотя бы одну вершину степени s. В [11] локальный кластерный коэффициент $c(s)$ определяется равенством

$$c(s) = \frac{2\Delta_s}{N_s s(s-1)}, \qquad (4)$$

где N_s – число вершин степени s. Понятно, что если треугольник содержит две вершины степени s, то в сумме Δ_s он учтен дважды, и трижды, если все вершины треугольника имеют степень s. Иногда рассматривают локальный кластерный коэффициент отдельной вершины степени s, равный отношению числа различных треугольников, содержащих эту вершину, к потенциально возможному их числу, т. е. к $s(s-1)/2$ (см., например, [4]). Тогда выражение (4) можно понимать как средний локальный кластерный коэффициент вершин степени s. Заметим еще, что конфигурационные графы, в которых кратные ребра объединены в одно ребро, а петли удалены, получили название стертых (erased) (см., например, [12]).

В статье [11] изучалось асимптотическое поведение локальных кластерных коэффициентов. Там же отмечается, что при подсчете числа треугольников в соответствии с перечисленными выше правилами кластерный коэффициент (4) совпадает с локальным кластерным коэффициентом стертого конфигурационного графа. Более того, оказалось, что значения кластерных коэффициентов, найденные без объединения кратных ребер (т. е. треугольники с кратными ребрами считаются неоднократно), существенно выше и не соответствуют наблюдаемой кластерной структуре реальных сложных сетей коммуникаций.

В [11] предполагалось, что распределение ξ при $k \to \infty$ обладает свойством

$$p_k = \mathbf{P}\{\xi = k\} = Ck^{-\tau}(1 + o(1)), \qquad (5)$$

где C – положительная постоянная и $\tau \in (2, 3)$. Нетрудно видеть, что распределение со свойством (5) является частным случаем (1), если при $k \to \infty$

$$h(k) \to C. \qquad (6)$$

В [11] рассматривалось предельное поведение $c(s)$ в трех зонах возможных значений s:

1. $s \geqslant 2, s = o(N^{(\tau-1)/(\tau-2)})$;
2. $N^{(\tau-2)/(\tau-1)} = O(s), s = o(\sqrt{N})$;
3. $N = O(s^2)$.

Полученные в [11] асимптотические выражения для $c(s)$ в этих трех зонах отличаются друг от друга, но, естественно, содержат константу C.

Остался открытым вопрос о поведении $c(s)$ при $h(k) \to 0$. Об актуальности такой задачи свидетельствуют, например, работы [2, 3], в которых

$$h(k) \sim \frac{L}{\ln^g k}, \qquad (7)$$

где $L, g > 0$. В [2] показано, что распределения со свойством (7) случайной величины ξ возникают, в частности, в моделях со случайным параметром τ в (1).

В настоящей статье изучается предельное поведение локального кластерного коэффициента $c(s)$ в конфигурационном графе G с N вершинами, в котором распределение степеней вершин определено в (1), где $\tau \in (2, 3)$, а медленно меняющаяся функция $h(x)$ обладает свойством (7). Доказательство полученного ниже результата основано на идеях работы [11] и следует той же схеме. Поэтому в нем нередко используются доказанные в [11] вспомогательные утверждения, соответствующие ссылки на которые приводятся. Подробнее приводятся рассуждения, в которых условие (7) рассматривается вместо (6). Для удобства в ряде случаев мы используем те же обозначения, что и в [11].

В статье доказана теорема о предельном поведении $c(s)$ в первой из перечисленных выше зоне возможных значений s. Мы надеемся, что, следуя предложенной схеме доказательства, будет нетрудно найти предельное поведение локального кластерного коэффициента и при стремлении к нулю $h(x)$ со скоростью, отличающейся от (7).

Из доказанной ниже теоремы следует, что асимптотическое поведение $c(s)$ при выполнении условия (7) сходно с результатом [11]. Оказалось, что при $N \to \infty$, $s = o(N^{(\tau-2)/(\tau-1)})$ коэффициент $c(s)$, деленный на $N^{2-\tau} \ln^{1-2g} N$, пропорционален константе.

Статья состоит из семи разделов. Во втором разделе вводятся необходимые обозначения. В третьем разделе формулируется теорема о предельном поведении $c(s)$, эта теорема является основным результатом работы. В разделах 4–6 получены вспомогательные результаты. В разделе 4 доказаны лемма 1 о предельном поведении максимальной степени вершины графа и лемма 2 об оценке условного математического ожидания $c(s)$ относительно случайного вектора $\mathbf{d}$. Полученная в лемме 2 оценка дополняется в разделе 5, где показано, что не рассмотренная до этого второстепенная часть условного математического ожидания $c(s)$ пренебрежимо мала по сравнению с основной частью (лемма 3). В лемме 4 раздела 6 рассматривается условная дисперсия $c(s)$. Наконец, в последнем разделе статьи с помощью полученных вспомогательных результатов доказывается теорема.

Обозначения

Для последовательностей $f(N)$ и $g(N)$, $N = 1, 2, \ldots$, будем, как обычно, писать $f(N) = o(g(N))$, если $\lim\limits_{N\to\infty} f(N)/g(N) = 0$, а для последовательностей случайных величин $\eta_1, \eta_2, \ldots$ равенство $\eta_N = o_p(g(N))$ означает, что $\mathbf{P}\{|\eta_N/g(N)| > \varepsilon\} \to 0$ при $N \to \infty$ и любом $\varepsilon > 0$. Подобным же образом будем писать $f(N) = O(g(N))$ или $\eta_N = O_p(g(N))$, если существует константа $C > 0$ такая, что $|f(N)/g(N)| \leqslant C$ для всех N или $\mathbf{P}\{|\eta_N/g(N)| \leqslant C\} \to 1$ при $N \to \infty$ соответственно. По традиции символ $\xrightarrow{p}$ означает сходимость по вероятности.

В равенстве (4), как уже говорилось, при подсчете числа треугольников поочередно рассматриваются вершины графа и, если очередная вершина имеет степень s, в сумме Δ_s учитываются все различные треугольники, содержащие эту вершину. Для удобства дальнейшего изложения назовем такую вершину базовой для этих треугольников. В любом отдельно взятом треугольнике будем обозначать u и v другие его вершины (не базовые) и, соответственно, ξ_u и ξ_v – их степени. Из приведенного выше комментария к формуле (4) ясно, что вершины u и v тоже могут иметь степень s, при этом они для данного треугольника базовыми не являются. Подобные обозначения мы будем использовать далее и в случае необходимости одновременного рассмотрения более чем одного треугольника. Например, если имеются два треугольника с общей базовой вершиной, то отличные от нее другие вершины этих треугольников будем обозначать u и v для одного треугольника и w и z для другого. Заметим, что если из вершин u, v, w, z только три разных, то эти треугольники имеют общее ребро. Общее ребро возникает также и при разных базовых вершинах, если только две из вершин u, v, w, z разные. С целью избежать неоднозначности при использовании таких обозначений приводимые в статье выкладки сопровождаются соответствующими комментариями.

Пусть $0 < \varepsilon < 1$ и в треугольнике вершины u и v, смежные с базовой, таковы, что

$$\varepsilon m N \leqslant \xi_u \xi_v \leqslant mN/\varepsilon, \tag{8}$$

где, напомним, m означает математическое ожидание случайной величины ξ (см. (3)). Из (1), (3) и условия $\tau \in (2, 3)$ следует, что $m > 0$ и конечно. Пусть степень s базовой вершины удовлетворяет условию

$$s = o(N^{(\tau-2)/(\tau-1)}). \tag{9}$$

Следуя [11], обозначим $W_N^s(\varepsilon)$ множество, состоящее из пар вершин u и v, входящих в рассматриваемые треугольники с базовыми вершинами степени s, для которых выполнены условия (8) и (9). Каждому треугольнику и, следовательно, каждому элементу множества $W_N^s(\varepsilon)$ можно присвоить номер базовой вершины. Тогда в $W_N^s(\varepsilon)$ могут встретиться совпадающие пары, но имеющие разные номера, если такая пара образует общее ребро для разных треугольников.

Пусть $c(s, W_N^s(\varepsilon))$ означает вклад в $c(s)$ треугольников, содержащих вершины $u, v \in W_N^s(\varepsilon)$, и $c(s, \overline{W}_N^s(\varepsilon))$ – вклад других треугольников. Понятно, что

$$c(s) = c(s, W_N^s(\varepsilon)) + c(s, \overline{W}_N^s(\varepsilon)). \tag{10}$$

Эта сумма лежит в основе доказательств полученных результатов. Для того чтобы с ее помощью оценить поведение $c(s)$, сначала находится асимптотика первого слагаемого при выполнении условия (8). Затем доказывается, что второе слагаемое в (10) пренебрежимо мало по сравнению с первым.

Символы $C_1, C_2, \ldots$ далее означают некоторые положительные постоянные.

Основной результат

Ниже приводится формулировка основного результата статьи.

Теорема. *Пусть* $\tau \in (2,3)$, $s \geqslant 2$, $N \to \infty$, $s = o(N^{(\tau-2)/(\tau-1)})$. *Тогда*

$$\mathbf{P}\left\{4^g \leqslant \frac{(\tau-1)(s-1)N^{\tau-2}m^\tau c(s)}{(\tau-3)\Gamma(2-\tau)r^2 s \ln^{1-2g} N}\right.$$

$$\left.\leqslant \left(\frac{(\tau-1)^2}{\tau-2}\right)^g\right\} \to 1,$$

где $\Gamma(x)$ *– гамма-функция.*

Вклад $W_N^s(\varepsilon)$ в $c(s)$

Обозначим ξ_{max} максимальную степень вершины графа G и найдем предельное распределение этой случайной величины.

Лемма 1. *Пусть* $N \to \infty$. *Тогда*

$$\mathbf{P}\left\{\xi_{max} < t\left(\frac{N}{\ln^g N}\right)^{1/(\tau-1)}\right\}$$

$$\to \exp\left\{-\frac{r(\tau-1)^{g-1}}{t^{\tau-1}}\right\}.$$

Доказательство. Очевидно, что

$$\mathbf{P}\{\xi_{max} < x\} = \mathbf{P}\{\xi_1 < x, \ldots, \xi_N < x\}$$

$$= (1 - \mathbf{P}\{\xi \geqslant x\})^N. \tag{11}$$

Из (1) и (7) следует, что при $x \to \infty$

$$\mathbf{P}\{\xi \geqslant x\} = \sum_{k \geqslant x} \frac{L + o(1)}{k^\tau \ln^g k}$$

$$= L\int_x^\infty \frac{dy}{y^\tau \ln^g y}(1 + o(1)).$$

Полагая $y = zx$, отсюда выводим равенство

$$\mathbf{P}\{\xi \geqslant x\} = \frac{L + o(1)}{x^{\tau-1}}\int_1^\infty \frac{dz}{z^\tau \ln^g(xz)}.$$

Функция $\ln^{-g}(xz)$ является медленно меняющейся, поэтому из теоремы 2.6 в работе [5] находим, что

$$\mathbf{P}\{\xi \geqslant x\} = \frac{L + o(1)}{x^{\tau-1}\ln^g x}\int_1^\infty \frac{dz}{z^\tau}$$

$$= \frac{L + o(1)}{(\tau-1)x^{\tau-1}\ln^g x}.$$

Если

$$x = t\left(\frac{N}{\ln^g N}\right)^{1/(\tau-1)},$$

то для любого $t > 0$

$$N\mathbf{P}\left\{\xi \geqslant t\left(\frac{N}{\ln^g N}\right)^{1/(\tau-1)}\right\}$$

$$= \frac{L(\tau-1)^{g-1} + o(1)}{t^{\tau-1}}.$$

Отсюда и из (11) получаем утверждение леммы 1. □

Замечание. Из леммы 1 следует, что выбором достаточно большого C вероятность

$$\mathbf{P}\left\{\xi_{max} < C\left(\frac{N}{\ln^g N}\right)^{1/(\tau-1)}\right\}$$

можно сделать сколь угодно близкой к единице.

Обозначим $\mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d})$ условное математическое ожидание $c(s, W_N^s(\varepsilon))$ относительно множества степеней $\mathbf{d}$. Изучим предельное поведение этой величины.

Лемма 2. *Пусть* $N \to \infty$, $s \geqslant 2$, $s = o(N^{(\tau-2)/(\tau-1)})$. *Тогда*

$$\mathbf{P}\left\{4^g \leqslant \frac{(\tau-1)(s-1)m^\tau \mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d})}{(3-\tau)sL^2N^{2-\tau}Int(\varepsilon)\ln^{1-2g} N}\right.$$

$$\left.\leqslant \left(\frac{(\tau-1)^2}{\tau-2}\right)^g\right\} \to 1,$$

где

$$Int(\varepsilon) = \int_\varepsilon^{1/\varepsilon} \frac{1 - e^{-x}}{x^{\tau-1}}dx.$$

Доказательство. Будем следовать доказательству леммы 6 из [11]. Пусть u, v, w – вершины графа, ξ_u, ξ_v, ξ_w – их степени, а ζ_N, как и во введении, означает сумму степеней всех вершин графа. Введем функцию

$$g_N(\xi_u, \xi_v, \xi_w) = (1 - e^{-\xi_u\xi_v/\zeta_N})$$

$$\times(1 - e^{-\xi_u\xi_w/\zeta_N})(1 - e^{-\xi_v\xi_w/\zeta_N}). \tag{12}$$

Для оценки ζ_N можно использовать теорему 3 из [1]. Обозначим $B_N = (Nh(B_N))^{1/(\tau-1)}$. Из этой теоремы следует, что вероятность

$$\mathbf{P}\left\{-A \leqslant \frac{\zeta_N - mN}{B_N} \leqslant A\right\}$$

можно сделать сколь угодно близкой к единице выбором достаточно большого A. Отсюда и из (7) вытекает соотношение

$$|\zeta_N - mN| = o_p(N^{1/(\tau-1)}). \qquad (13)$$

Согласно лемме 3 из [11], при выполнении равенства (13)

$$\mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d}) \xrightarrow{p} \frac{\sum_{(u,v)\in W_N^s(\varepsilon)} g_N(s,\xi_u,\xi_v)}{s(s-1)}. \qquad (14)$$

Из (9) и замечания находим, что $s\xi_i = o(N)$ равномерно по $i = 1,\ldots,N$. Полагая $\xi_w = s$ в (12) и применяя формулу Тейлора, из (14) находим

$$\mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d})$$

$$\xrightarrow{p} \frac{s}{s-1} \sum_{(u,v)\in W_N^s(\varepsilon)} \frac{\xi_u\xi_v(1-e^{-\xi_u\xi_v/\zeta_N})}{\zeta_N^2}. \qquad (15)$$

Пусть (u,v) – произвольная пара вершин из $W_N^s(\varepsilon)$. Для них выполнены неравенства (8), и, используя замечание к лемме 1, получаем, что при достаточно малом C_1

$$\xi_u > \frac{\varepsilon mN}{\xi_v} > C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)}$$

и в то же время при достаточно большом C_2

$$\xi_u < C_2(N/\ln^g N)^{1/(\tau-1)}.$$

Поэтому для всех вершин u, входящих в пары вершин из $W_N^s(\varepsilon)$, будем далее считать, что

$$C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)} < \xi_u$$

$$< C_2(N/\ln^g N)^{1/(\tau-1)}. \qquad (16)$$

Пусть $0 < a < b < \infty$ и

$$amN \leqslant \xi_u\xi_v \leqslant bmN. \qquad (17)$$

Обозначим $\mathbb{E}[a,b]$ событие, состоящее в том, что условия (16) и (17) выполнены одновременно. Введем случайную меру

$$M^{(N)}([a,b]) = \frac{(mN)^{\tau-1}}{N^2\ln^{1-2g}N} \sum_{\substack{u,v=1\\u\neq v}}^{N} I\{\mathbb{E}[a,b]\}, \qquad (18)$$

здесь и далее $I\{A\}$ означает индикатор события A. Заметим, что суммирование в (18) проводится по всем вершинам графа. Поэтому

$$\mathbf{E}M^{(N)}([a,b]) = \frac{(mN)^{\tau-1}(N-1)}{N\ln^{1-2g}N}\mathbf{P}\{\mathbb{E}[a,b]\}.$$

Отсюда и из (7), (16), (17) находим:

$$\mathbf{E}M^{(N)}([a,b]) = L^2(mN)^{\tau-1}\ln^{2g-1}N$$

$$\int_{A_N}^{B_N}\int_{amN/x}^{bmN/x} \frac{(xy)^{-\tau}dydx}{((\ln x)(\ln y))^g}(1+o(1)),$$

где $A_N = C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)}, B_N = C_2(N/\ln^g N)^{1/(\tau-1)}$. Полагая $z = xy$, отсюда получаем

$$\mathbf{E}M^{(N)}([a,b]) = L^2(mN)^{\tau-1}\ln^{2g-1}N$$

$$\times\int_{A_N}^{B_N}\left(\int_{amN}^{bmN}\frac{z^{-\tau}dz}{((\ln x)(\ln(z/x)))^g}\right)\frac{dx}{x}(1+o(1)).$$

Теперь положим $z = tmN$. Тогда

$$\mathbf{E}M^{(N)}([a,b]) = L^2(mN)^{\tau-1}\ln^{2g-1}N$$

$$\times\int_{A_N}^{B_N}\left(\int_a^b\frac{t^{-\tau}(mN)^{1-\tau}dt}{\ln^g(tmN/x)}\right)\frac{dx}{x\ln^g x}(1+o(1)). \qquad (19)$$

Легко видеть, что в области интегрирования по x выполнено соотношение $mN/x \to \infty$. Поэтому, используя свойство (2.11) из [5], видим, что

$$\int_a^b t^{-\tau}\ln^{-g}(tmN/x)dt \sim \ln^{-g}(mN/x)\int_a^b t^{-\tau}dt.$$

Отсюда и из (19) следует равенство

$$\mathbf{E}M^{(N)}([a,b]) = L^2\ln^{2g-1}N$$

$$\times\int_{A_N}^{B_N}x^{-1}((\ln x)(\ln(mN/x))^{-g}$$

$$\times\int_a^b t^{-\tau}dtdx(1+o(1)). \qquad (20)$$

Нетрудно проверить, что в области интегрирования по x для любого ε, $0 < \varepsilon < 1$,

$$\left(\frac{2}{\ln(mN)}\right)^{2g} \leqslant ((\ln x)(\ln(mN/x)))^{-g}$$

$$\leqslant \frac{1+\varepsilon}{(\tau-2)^g}\left(\frac{\tau-1}{\ln N}\right)^{2g}. \qquad (21)$$

Собирая вместе (20), (21) и используя теорему о среднем, находим, что существует число γ такое, что

$$\gamma \in \left[4^g, \left(\frac{(\tau-1)^2}{\tau-2}\right)^g\right] \qquad (22)$$

и

$$\mathbf{E}M^{(N)}([a,b]) = \frac{L^2}{\ln N}\gamma$$
$$\times \int_{A_N}^{B_N} \left(\int_a^b \frac{dt}{t^\tau}\right)\frac{dx}{x}(1+o(1)) = L^2\gamma$$
$$\times \int_a^b \left(\frac{3-\tau}{\tau-1} - \frac{2g\ln\ln N}{(\tau-1)\ln N}\right.$$
$$\left. + \frac{\ln(C_2/C_1)}{\ln N}\right)\frac{dt}{t^\tau}(1+o(1)). \qquad (23)$$

Обозначим

$$\mu([a,b]) = L^2\gamma\frac{3-\tau}{\tau-1}\int_a^b \frac{dt}{t^\tau}. \qquad (24)$$

Тогда из (23) следует

$$\lim_{N\to\infty}\mathbf{E}M^{(N)}([a,b]) = \mu([a,b]). \qquad (25)$$

Далее рассмотрим дисперсию меры $M^{(N)}([a,b])$. Из (16)–(18) находим, что

$$\mathbf{D}M^{(N)}([a,b]) = m^{2(\tau-1)}N^{2\tau-6}\ln^{4g-2}N$$
$$\times \sum_{u,v=1}^{N}\sum_{w,z=1}^{N}\Big(\mathbf{P}\Big\{\xi_u\xi_v, \xi_w\xi_z \in [amN, bmN],$$
$$\xi_u, \xi_w \in \Big[C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)},$$
$$C_2(N/\ln^g)^{1/(\tau-1)}\Big]\Big\} - \mathbf{P}\Big\{\xi_u\xi_v \in [amN, bmN],$$
$$\xi_u \in \Big[C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)},$$
$$C_2(N/\ln^g N)^{1/(\tau-1)}\Big]\Big\}$$
$$\times\ \mathbf{P}\Big\{\xi_w\xi_z \in [amN, bmN],$$
$$\xi_w \in \Big[C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)},$$
$$C_2(N/\ln^g N)^{1/(\tau-1)}\Big]\Big\}\Big). \qquad (26)$$

Поскольку степени вершин независимы, из (26) следует, что если все вершины u, v, w, z различны, то вклад таких слагаемых в дисперсию равен нулю. Обозначим D_3 вклад в (26) слагаемых, где ровно две из четырех вершин совпадают. Тогда

$$D_3 \leqslant m^{2(\tau-1)}N^{2\tau-6}\ln^{4g-2}N$$
$$\times\sum_{u,v,w}\mathbf{P}\Big\{\xi_u\xi_v, \xi_u\xi_w \in [amN, bmN],$$
$$\xi_u \in \Big[C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)},$$
$$C_2(N/\ln^g N)^{1/(\tau-1)}\Big]\Big\}$$
$$\leqslant C_3N^{2\tau-3}\ln^{4g-2}N\mathbf{P}\Big\{\xi_1\xi_2, \xi_1\xi_3 \in [amN, bmN],$$
$$\xi_1 \in \Big[C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)},$$
$$C_2(N/\ln^g N)^{1/(\tau-1)}\Big]\Big\}$$
$$\leqslant C_4N^{2\tau-3}\ln^{4g-2}N\int_2^\infty x^{-\tau}$$
$$\times\left(\int_{amN/x}^{bmN/x} y^{-\tau}dy\right)^2 dx \leqslant C_5N^{-1}\ln^{4g-2}N. \qquad (27)$$

Пусть D_2 означает вклад в (26) слагаемых, в которых только две из четырех вершин разные. Тогда, как и при выводе (19), (23),

$$D_2 \leqslant C_6N^{2\tau-6}\ln^{4g-2}N\sum_{u,v}\mathbf{P}\{\mathbb{E}[a,b]\}$$
$$\leqslant C_7N^{2\tau-4}\ln^{4g-2}N\int_{C_1(N^{\tau-2}\ln^g N)^{1/(\tau-1)}}^{C_2(N/\ln^g N)^{1/(\tau-1)}}$$
$$\left(\int_{amN}^{bmN}\frac{dz}{z^\tau}\right)\frac{dx}{x} \leqslant C_8N^{\tau-3}\ln^{4g-1}N. \qquad (28)$$

В соответствии с определением меры (18), в сумме (26) нет слагаемых, в которых все четыре вершины совпадают. Поэтому из (26)–(28) находим, что

$$\mathbf{D}M^{(N)}([a,b]) = o_p(1). \qquad (29)$$

Применив неравенство Чебышева, отсюда и из (23)–(25) видим, что для любого сколь угодно малого $\varepsilon > 0$

$$\mathbf{P}\{|M^{(N)}([a,b]) - \mu([a,b])| > \varepsilon\} \to 0,$$

следовательно,

$$M^{(N)}([a,b]) \xrightarrow{p} \mu([a,b]). \qquad (30)$$

Как показано в [11, соотношение (5.26)], из (13) и (18) следует, что

$$\sum_{u,v\in W_N^s(\varepsilon)}\frac{\xi_u\xi_v(1-e^{-\xi_u\xi_v/\zeta_N})}{\zeta_N^2}$$
$$\xrightarrow{p}\frac{\ln N}{N^{\tau-2}m^\tau}\int_\varepsilon^{1/\varepsilon} x(1-e^{-x})dM^{(N)}(x), \qquad (31)$$

полагая $M^{(N)}(x) = M^{(N)}([x,x])$. Из леммы 5 [11], (24), (30) получаем

$$\int_\varepsilon^{1/\varepsilon} x(1-e^{-x})dM^{(N)}(x) \xrightarrow{p} \int_\varepsilon^{1/\varepsilon} x(1-e^{-x})d\mu(x)$$

$$= L^2\gamma\frac{3-\tau}{\tau-1}\int_{\varepsilon}^{1/\varepsilon} x^{1-\tau}(1-e^{-x})dx.$$

Отсюда и из (15), (22), (29)–(31) вытекает утверждение леммы 2. □

Вклад $\overline{W}_N^s(\varepsilon)$ в $c(s)$

В этом разделе рассматривается условное математическое ожидание $c(s,\overline{W}_N^s(\varepsilon)|\mathbf{d})$.

Лемма 3. *Пусть $N\to\infty$, $s\geqslant 2$ и выполнено условие (9). Тогда*

$$\frac{(s-1)\mathbf{E}_c(s,\overline{W}_N^s(\varepsilon)|\mathbf{d})}{sN^{2-\tau}\ln^{1-2g}N} = O_p(\varepsilon^{\min(3-\tau,\tau-2)}).$$

Доказательство. В [11, соотношение (6.4)] показано, что

$$\mathbf{E}c(s,\overline{W}_N^s(\varepsilon)|\mathbf{d}) = O_p\Big((N^2s^{-2}\mathbf{E}\Big(\min\Big(1,\frac{s\xi_u}{mN}\Big)$$

$$\times\min\Big(1,\frac{s\xi_v}{mN}\Big)\min\Big(1,\frac{\xi_u\xi_v}{mN}\Big)I\{u,v\in\overline{W}_N^s(\varepsilon)\}\Big). \tag{32}$$

Полагая $x=\xi_u$, $y=\xi_v$, заменяя суммирование интегрированием и учитывая (7), получаем, как в (6.5) из [11], что

$$\mathbf{E}\Big(\min\Big(1,\frac{s\xi_u}{mN}\Big)\min\Big(1,\frac{s\xi_v}{mN}\Big)\min\Big(1,\frac{\xi_u\xi_v}{mN}\Big)$$

$$\times I\{u,v\in\overline{W}_N^s(\varepsilon)\} = \iint_{x,y\in\overline{W}_N^s(\varepsilon)}\frac{(xy)^{-\tau}}{((\ln x)(\ln y))^g}$$

$$\times min\Big(1,\frac{sx}{mN}\Big)\min\Big(1,\frac{sy}{mN}\Big)\min\Big(1,\frac{xy}{mN}\Big)dxdy, \tag{33}$$

где область интегрирования, в соответствии с определением $\overline{W}_N^s(\varepsilon)$, замечанием и тем, что степень любой вершины треугольника не менее двух, выбрана так, что $2\leqslant x,y\leqslant N^{1/(\tau-1)}$ и нарушено условие $\varepsilon mN\leqslant xy\leqslant mN/\varepsilon$ (см. (8)).

Сначала рассмотрим случай $xy<\varepsilon mN$. Учитывая (7) и (9), находим, что

$$\iint_{x,y\in\overline{W}_N^s(\varepsilon)}\frac{(xy)^{-\tau}}{((\ln x)(\ln y))^g}$$

$$\times min\Big(1,\frac{sx}{mN}\Big)\min\Big(1,\frac{sy}{mN}\Big)\min\Big(1,\frac{xy}{mN}\Big)dxdy$$

$$\leqslant\frac{s^2}{(mN)^3}\int_2^{N^{1/(\tau-1)}}\int_2^{\varepsilon mN/x}\frac{(xy)^{2-\tau}dydx}{((\ln x)(\ln y))^g}. \tag{34}$$

Выберем α так, что $0<\alpha<(3-\tau)/2$. Правую часть (34) можно представить в виде суммы четырех интегралов

$$\frac{s^2}{(mN)^3}\int_2^{N^{1/(\tau-1)}}\int_2^{\varepsilon mN/x}\frac{(xy)^{2-\tau}dydx}{((\ln x)(\ln y))^g}.$$

$$=\frac{s^2}{(mN)^3}(I_1+I_2+I_3+I_4), \tag{35}$$

где области интегрирования I_1–I_4 соответственно равны

$$S_1=\{2\leqslant x<N^\alpha,\quad 2\leqslant y<N^\alpha\},$$
$$S_2=\{2\leqslant x<N^\alpha,\quad N^\alpha\leqslant y\leqslant\varepsilon mN/x\},$$
$$S_3=\{N^\alpha\leqslant x\leqslant N^{1/(\tau-1)},\quad 2\leqslant y<N^\alpha\},$$
$$S_4=\{N^\alpha\leqslant x\leqslant N^{1/(\tau-1)},\ N^\alpha\leqslant y\leqslant\varepsilon mN/x\}.$$

Нетрудно видеть, что

$$I_1\leqslant N^{2\alpha}. \tag{36}$$

В соответствии с замечанием $\xi_u,\xi_v<N^{1/(\tau-1)}$, и кроме того, должно выполняться неравенство $\varepsilon mN/x\leqslant N^{1/(\tau-1)}$. Поэтому для I_2 и I_3 получаем одинаковые оценки:

$$I_2,I_3=O\left(N^{(3-\tau)/(\tau-1)+\alpha}\right). \tag{37}$$

Наконец,

$$I_4\leqslant\frac{C_{22}}{\ln^{2g}N}\int_{N^\alpha}^{N^{1/(\tau-1)}}\frac{dx}{x}\int_{N^\alpha x}^{\varepsilon mN}z^{2-\tau}dz$$

$$=O\Big((\varepsilon N)^{3-\tau}\ln^{1-2g}N\Big). \tag{38}$$

Собирая вместе (32)–(38), приходим к выводу, что при $\xi_u\xi_v<\varepsilon mN$ и достаточно малом α

$$\frac{(s-1)\mathbf{E}(c(s,\overline{W}_N^s(\varepsilon))|\mathbf{d})}{sN^{2-\tau}\ln^{1-2g}N}=O_p(\varepsilon^{3-\tau}). \tag{39}$$

Пусть $\xi_u\xi_v>mN/\varepsilon$. Нетрудно видеть, учитывая (9), что в этом случае при достаточно малом ε и достаточно большом N величину (33) можно ограничить выражением

$$\frac{s^2}{(mN)^2}\int_{mN^{(\tau-2)/(\tau-1)}/\varepsilon}^{N^{1/(\tau-1)}}\int_{mN/(\varepsilon x)}^{N^{1/(\tau-1)}}\frac{(xy)^{1-\tau}dydx}{((\ln x)(\ln y))^g}$$

$$\leqslant C_{23}\frac{s^2}{N^2\ln^{2g}N}\int_{mN^{(\tau-2)/(\tau-1)}/\varepsilon}^{N^{1/(\tau-1)}}\frac{dx}{x}$$

$$\times\int_{mN/\varepsilon}^{N^{1/(\tau-1)}x}z^{1-\tau}dz=O\left(\frac{s^2\varepsilon^{\tau-2}\ln^{1-2g}N}{N^\tau}\right).$$

Отсюда и из (32) находим, что если $\xi_u\xi_v>mN/\varepsilon$, то

$$\frac{(s-1)\mathbf{E}c(s,\overline{W}_N^s(\varepsilon)|\mathbf{d})}{sN^{2-\tau}\ln^{1-2g}N}=O_p(\varepsilon^{\tau-2}).$$

Это равенство вместе с (39) завершает доказательство леммы 3. □

Дисперсия вклада $W_N^s(\varepsilon)$ в $c(s)$

В этом разделе мы получим соотношение

$$\frac{\mathbf{D}(c(s, W_N^s(\varepsilon))|\mathbf{d})}{(\mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d}))^2} \xrightarrow{p} 0. \tag{40}$$

Следующая лемма 4 является аналогом леммы 9 из [11] и доказывается так же, поэтому укажем только на небольшие отличия, связанные с использованием свойства (7) вместо (6).

Лемма 4. *Пусть $N \to \infty$, $s \geqslant 2$ и выполнено условие (9). Тогда верно соотношение (40).*

Доказательство. Обозначим $\Delta_{i,u,v}$ событие, состоящее в том, что вершины i, u, v образуют треугольник. В [11, равенство (5.59)] показано, что

$$\mathbf{D}(c(s, W_N^s(\varepsilon))|\mathbf{d}) = \frac{1}{s^2(s-1)^2 N_s^2} \sum_{\xi_i, \xi_j = s}$$

$$\times \sum_{(u,v),(w,z) \in W_N^s(\varepsilon)} (\mathbf{P}\{\Delta_{i,u,v}\Delta_{j,w,z}|\mathbf{d}\}$$

$$- \mathbf{P}\{\Delta_{i,u,v}|\mathbf{d}\}\mathbf{P}\{\Delta_{j,u,v}|\mathbf{d}\}). \tag{41}$$

В равенстве (41) нужно рассмотреть несколько случаев в зависимости от того, сколько разных вершин содержится в множестве (i, u, v, j, w, z). Обозначим $D_s^{(l)}$ вклад в (41) треугольников, образованных вершинами этого множества, если l из них разные. Понятно, что $3 \leqslant l \leqslant 6$.

В [11, соотношение (5.60)] доказано, что оценки $D_s^{(6)}$ и $D_s^{(5)}$ совпадают и

$$D_s^{(6)}, D_s^{(5)} = o_p\left((\mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d})^2\right). \tag{42}$$

Случай $D_s^{(4)}$ соответствует двум треугольникам с одним общим ребром. Пусть сначала $i = j$ и $u = z$. Повторяя почти дословно оценки (5.61)–(5.64) из [11] и используя (7) и лемму 1, находим, что в этом случае $D_s^{(4)}$ можно оценить следующим образом:

$$D_s^{(4)} = O_p\left(\varepsilon^{-(2\tau+1)} s^{\tau-1} N^{3-2\tau-(\tau-2)/(\tau-1)}\right).$$

Учитывая нормировку в лемме 2, отсюда получаем, что при любом сколь угодно малом, но фиксированном ε

$$\frac{D_s^{(4)}}{N^{4-2\tau} \ln^{2-4g} N} \to 0. \tag{43}$$

Пусть теперь $i \neq j$, $u = z$, $v = w$. Из [11, неравенство (5.65)] следует, что в этом случае

$$D_s^{(4)} = O_p\left(\varepsilon^{-2} N^{1-\tau} \ln N\right),$$

следовательно, соотношение (43) остается в силе. Наконец, оценка $D_s^{(3)}$ получена в (5.71) [11].

$$D_s^{(3)} = O_p\left(s^{\tau-4} N^{1-\tau} \ln N\right),$$

что, как нетрудно видеть, вместе с (42), (43) завершает доказательство леммы 4. □

Доказательство теоремы

Пусть выполнены условия теоремы. Понятно, что разность

$$\left| \int_\varepsilon^{1/\varepsilon} x^{1-\tau}\left(1 - e^{-x}\right) dx - \int_0^\infty x^{1-\tau}\left(1 - e^{-x}\right) dx \right|$$

можно сделать сколь угодно малой выбором достаточно малого ε. В [11, равенство (3.16)] показано, что

$$\int_0^\infty x^{1-\tau}\left(1 - e^{-x}\right) dx = -\Gamma(2-\tau).$$

Отсюда и из леммы 2 следует, что при достаточно больших N можно сделать сколь угодно близкой к единице вероятность события

$$4^g \leqslant \frac{(\tau-1)(s-1)m^\tau \mathbf{E}(c(s, W_N^s(\varepsilon))|\mathbf{d})}{(\tau-3)\Gamma(2-\tau) s L^2 N^{2-\tau} \ln^{1-2g} N}$$

$$\leqslant \left(\frac{(\tau-1)^2}{\tau-2}\right)^g,$$

и из (10) и леммы 3 видим, что это верно и для события

$$4^g \leqslant \frac{(\tau-1)(s-1)m^\tau \mathbf{E}(c(s)|\mathbf{d})}{(\tau-3)\Gamma(2-\tau) s L^2 N^{2-\tau} \ln^{1-2g} N}$$

$$\leqslant \left(\frac{(\tau-1)^2}{\tau-2}\right)^g.$$

Для завершения доказательства теоремы осталось вспомнить лемму 4, которая вместе с леммой 3 приводит к соотношению

$$\frac{c(s)}{\mathbf{E}(c(s)|\mathbf{d})} \xrightarrow{p} 1.$$

Литература

1. *Павлов Ю. Л.* Асимптотика числа ребер Интернет-графа // Труды Карельского научного центра РАН. 2021. № 6. С. 59–63. doi: 10.17076/mat1434

2. *Павлов Ю. Л.* Условные конфигурационные графы со случайным параметром распределения степеней // Математический сборник. 2018. Т. 209, вып. 2. С. 120–137. doi: 10.4213/sm8832

3. *Павлов Ю. Л., Чеплюкова И. А.* Предельные распределения числа вершин заданной степени конфигурационного графа с ограниченным числом ребер // Теория вероятностей и ее применения. 2021. Т. 66, вып. 3. С. 468–486. doi: 10.4213/tvp5332

4. *Прохоренкова Л. А., Крот А. В.* Локальный кластерный коэффициент в моделях предпочтительного присоединения // Доклады Академии наук. 2016. Т. 66, вып. 3. С. 19–22. doi: 10.7868/S0869565216310066

5. *Сенета Е.* Правильно меняющиеся функции. М.: Наука, 1985. 141 с.

6. *Albert R., Jeong H., Barabasi A.-L.* Internet: diameter of the world-wide web // Nature. 1999. Vol. 401. P. 130–131. doi: 10.1038/43601

7. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // Eur. J. Comb. 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8

8. *Durrett R.* Random graph dynamics. Cambridge: Cambridge University Press, 2007. 223 p. doi: 10.1017/CBO9780511546594

9. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press, 2017. 337 p. doi: 10.1017/9781316779422

10. *Hofstad R., Hoorn P., Litvak N., Stegehuis C.* Limit theorems for assortativity and clustering in null models for scale-free networks // Adv. Appl. Probab. 2020. Vol. 52, iss. 4. P. 1035–1084. doi: 10.1017/apr.2020.42

11. *Hofstad R., Leenwaarden J., Stegehuis C.* Triadic closure in configuration models with unbounded degree fluctuation // J. Stat. Phys. 2018. Vol. 173, iss. 4. P. 746–774. doi: 10.1007/s10955-018-1952-x

12. *Hoorn P., Litvak N.* Upper bounds for number of removed edges in the Erased Configuration Model // Algorithms and Models for the Web Graph. Lecture Notes in Computer Science. 2015. Vol. 9479. P. 54–65. doi: 10.1007/978-3-319-26784-5_5

13. *Newman M. E. J.* Networks: An introduction. New York: Oxford University Press, 2010. 772 p.

References

1. *Pavlov Yu. L.* Asymptotics of the number of edges of an internet graph. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2021;6:59–63. (In Russ.). doi: 10.17076/mat1423

2. *Pavlov Yu. L.* Conditional configuration graphs with discrete power-law distribution of vertex degrees. *Sbornik: Mathematics.* 2018;209(2):258–272. doi: 10.1070/SM8832

3. *Pavlov Yu. L., Cheplyukova I. A.* Limit distributions of the number of vertices of a given degree in a configuration graph with bounded number of edges. *Theory of Probab. Appl.* 2021;66(3):376–390. doi: 10.1137/S0040585X97T9900460

4. *Prokhorenkova L. A., Krot A. V.* Local clustering coefficients in preferential attachment models. *Doklady Mathematics.* 2016;94(3):623–626. doi: 10.1134/S1064562416060041

5. *Seneta E.* Regulary varying functions. Berlin: Springer; 1976. 252 p.

6. *Albert R., Jeong H., Barabasi A.-L.* Internet: diameter of the world-wide web. *Nature.* 1999;401:130–131. doi: 10.1038/43601

7. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Comb.* 1980;1(4):311–316. doi: 10.1016/S0195-6698(80)80030-8

8. *Durrett R.* Random graph dynamics. Cambridge: Cambridge University Press; 2007. 233 p. doi: 10.1017/9781316779422

9. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press; 2017. 337 p. doi: 10.1017/9781316779422

10. *Hofstad R., Hoorn P., Litvak N., Stegehuis C.* Limit theorems for assortativity and clustering in null models for scale-free networks. *Adv. Appl. Probab.* 2020;52(4):1035–1084. doi: 10.1017/apr.2020.42

11. *Hofstad R., Leenwaarden J., Stegehuis C.* Triadic closure in configuration models with unbounded degree fluctuation. *J. Stat. Phys.* 2018;173(4):746–774. doi: 10.1007/s10955-018-1952-x

12. *Hoorn P., Litvak N.* Upper bounds for number of removed edges in the Erased Configuration Model. *Algorithms and Models for the Web Graph. Lecture Notes in Computer Science.* 2015;9479:54–65. doi: 10.1007/978-3-319-26784-5_5

13. *Newman M. E. J.* Networks: An introduction. New York: Oxford University Press; 2010. 772 p.

Поступила в редакцию / received: 06.12.2024; принята к публикации / accepted: 04.04.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Павлов Юрий Леонидович

д-р физ.-мат. наук, профессор, главный научный сотрудник

e-mail: pavlov@krc.karelia.ru

CONTRIBUTOR:

Pavlov, Yury

Dr. Sci. (Phys.-Math.), Professor, Chief Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 54–59
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 54–59
DOI: 10.17076/mat2041

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.212.2+519.175.4

О СТРУКТУРЕ УСЛОВНЫХ КОНФИГУРАЦИОННЫХ ГРАФОВ С ОГРАНИЧЕННЫМ ЧИСЛОМ РЕБЕР

Ю. Л. Павлов

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Рассматриваются конфигурационные графы с N вершинами при условии, что сумма степеней вершин ограничена сверху числом n. Степени вершин независимы и одинаково распределены по неизвестному закону, зависящему от медленно меняющейся функции и имеющему конечные математическое ожидание и дисперсию. Такие модели могут использоваться для описания различных сетей коммуникаций и топологии Интернета. В статье при $N, n \to \infty$ найдены предельные распределения максимальной степени вершины и числа вершин заданной степени.

Ключевые слова: конфигурационный граф; степень вершины; предельные распределения

Для цитирования: Павлов Ю. Л. О структуре условных конфигурационных графов с ограниченным числом ребер // Труды Карельского научного центра РАН. 2025. № 4. С. 54–59. doi: 10.17076/mat2041

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

Yu. L. Pavlov. ON THE STRUCTURE OF CONDITIONAL CONFIGURATION GRAPHS WITH A BOUNDED NUMBER OF EDGES

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

We consider configuration graphs with N vertices under the condition that the sum of vertex degrees is bounded from above by n. The degrees of vertices are independent and identically distributed according to an unknown distribution law which depends on a slowly varying function and has finite expectation and variance. Such models can be used to describe various communication networks and Internet topology. The paper finds the limit distributions of the maximum vertex degree and the number of vertices with a given degree as $N, n \to \infty$.

Keywords: configuration graph; vertex degree; limit distributions

For citation: Pavlov Yu. L. On the structure of conditional configuration graphs with a bounded number of edges. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2025. No. 4. P. 54–59. doi: 10.17076/mat2041

Funding. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

Введение

Интерес к исследованию случайных графов постоянно возрастает в связи с их использованием в качестве моделей сложных сетей коммуникаций (см., например, [10]). Наблюдения за реальными сетями показали, что их топология может быть описана случайными графами с независимыми одинаково распределенными степенями вершин. Более того, было замечено, что число вершин степени k при достаточно больших k пропорционально $k^{-\tau}$, где $\tau > 1$. Это значит, что распределение случайной величины ξ, равной степени произвольной вершины графа, можно представить в виде

$$p_k = \mathbf{P}\{\xi = k\} = \frac{h(k)}{k^\tau}, \qquad k = 1, 2, \ldots, \quad (1)$$

где $h(k)$ – медленно меняющаяся на бесконечности функция. Одним из наиболее подходящих для моделирования сетей графов является так называемый конфигурационный граф [9]. Случайная величина ξ в таком графе равна числу полуребер вершины, т. е. числу инцидентных ей ребер, для которых смежные вершины еще не определены. Все полуребра занумерованы в произвольном порядке. Сумма степеней вершин любого графа должна быть четной, поэтому, если она нечетна, можно добавить вспомогательную вершину единичной степени. Построение графа завершается путем попарного равновероятного соединения полуребер друг с другом для образования ребер. В [12] отмечается, что появление вспомогательной вершины вместе с инцидентным ей ребром не влияет на асимптотическое поведение графа при стремлении числа вершин к бесконечности. Нетрудно видеть также, что такие графы могут содержать петли и кратные ребра.

В статье [5] рассматривались условные конфигурационные графы при условии, что число ребер известно. В этой работе распределение степеней вершин имело вид

$$p_k = \frac{1}{k^\tau} - \frac{1}{(k+1)^\tau}, \quad k = 1, 2, \ldots, \quad \tau > 0.$$

Такое же распределение степеней вершин рассматривалось [3] при более естественном условии, что число ребер графа неизвестно, но ограничено сверху. В статьях [2, 4] предполагалось, что при $k \to \infty$

$$p_k \sim \frac{d}{k^g (\ln k)^h},$$

где $d > 0$, $g > 1$, $h \geqslant 0$.

В этой статье мы рассматриваем конфигурационные графы, содержащие N вершин, в которых случайная величина ξ имеет распределение (1), где $\tau > 3$. Обозначим $\xi_1, \ldots, \xi_N$ степени вершин $1, \ldots, N$ соответственно. Эти случайные величины независимы и распределены так же, как и ξ. Такой случайный граф естественным образом индуцирует вероятностную меру на подмножестве реализаций, в которых $\xi_1 + \ldots + \xi_N \leqslant n$. Далее мы рассматриваем условные конфигурационные графы с указанным выше ограниченным числом ребер. Обозначим $\eta_1, \ldots, \eta_N$ случайные величины, равные числу степеней вершин $1, \ldots, N$ в таком условном случайном графе. Понятно, что распределения этих случайных величин отличаются от распределения ξ из-за введенного ограничения, и они зависимы.

Одно из хорошо известных элементарных свойств медленно меняющихся функций состоит в том, что при $x \to \infty$ и любом $\delta > 0$

$$x^{-\delta} < h(x) < x^\delta. \quad (2)$$

Из (1), (2) получаем, что случайная величина ξ имеет конечные математическое ожидание $m = \mathbf{E}\xi$ и дисперсию $\sigma^2 = \mathbf{D}\xi$:

$$m = \sum_{k=1}^{\infty} \frac{h(k)}{k^{\tau-1}}, \qquad \sigma^2 = \sum_{k=1}^{\infty} \frac{h(k)}{k^{\tau-2}} - m^2. \quad (3)$$

Обозначим $\eta_{(N)}$ и μ_r максимальную степень вершины и число вершин степени r соответственно. В [11] были доказаны предельные теоремы для этих случайных величин в случае, когда $h(x)$ в (1) является медленно меняющейся функцией с остаточным членом. В настоящей статье аналогичные результаты получены при условии, что если $x \to \infty$, то функция $h(x)$ не возрастает. Заметим, что в частном случае

$$h(x) = \frac{d}{\ln^\alpha x},$$

где $d > 0$, $\alpha \geqslant 0$, предельные распределения числа вершин заданной степени найдены в [4], а максимальной степени – в [6].

Статья организована следующим образом. Главные результаты (теоремы 1–3) формулируются в следующем разделе. В третьем разделе обсуждается связь рассматриваемой задачи с обобщенной схемой размещения частиц по ячейкам. Четвертый раздел содержит вспомогательные результаты, которые используются в последнем разделе для доказательства теорем 1–3.

Основные результаты

Доказаны следующие теоремы.

Теорема 1. *Пусть $N, n \to \infty$, так, что*

$$\frac{n - Nm}{\sqrt{N}} \geqslant C > -\infty, \tag{4}$$

а $r = r(N, n)$ выбраны так, что

$$\frac{Nh(r)}{(\tau-1)r^{\tau-1}} \to \gamma, \tag{5}$$

где γ – положительная константа. Тогда

$$\mathbf{P}\{\eta_{(N)} \leqslant r\} \to e^{-\gamma}.$$

Теорема 2. *Пусть $N, n \to \infty, Np_r(1-p_r) \to \infty$ и*

$$\frac{n - Nm}{\sqrt{N}} \to \infty. \tag{6}$$

Тогда

$$\mathbf{P}\{\mu_r = k\} = \frac{1 + o(1)}{\sqrt{2\pi Np_r(1-p_r)}} e^{-u_r^2/2}$$

равномерно по k таким, что $u_r = (k - Np_r)/\sqrt{Np_r(1-p_r)}$ лежит в любом фиксированном конечном интервале.

Теорема 3. *Пусть $N, n, r \to \infty$ и выполнено условие (5). Тогда для неотрицательных целых k*

$$\mathbf{P}\{\mu_r = k\} = \frac{(Np_r)^k}{k!} e^{-Np_r}(1 + o(1))$$

равномерно по $(k - Np_r)/\sqrt{Np_r}$ в любом фиксированном конечном интервале.

Связь с обобщенной схемой размещения

Техника доказательств теорем 1–3 основана на использовании обобщенной схемы размещения частиц по ячейкам [1]. Нетрудно видеть, что если $\eta_1 + \ldots + \eta_N = n$, то

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\}$$

$$= \mathbf{P}\{\xi_1 = k_1, \ldots, \xi_N = k_N | \xi_1 + \ldots + \xi_N = n\}.$$

Это равенство означает, что выполнены условия обобщенной схемы размещения. Легко видеть также, что для нашего условного конфигурационного графа

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\}$$

$$= \mathbf{P}\{\xi_1 = k_1, \ldots, \xi_N = k_N | \xi_1 + \ldots + \xi_N \leqslant n\}. \tag{7}$$

В статьях [7, 8] пара множеств случайных величин $(\eta_1, \ldots, \eta_N), (\xi_1, \ldots, \xi_N)$, удовлетворяющих равенству (7), названа аналогом обобщенной схемы размещения.

Введем вспомогательные случайные величины $\xi_1^{(r)}, \ldots, \xi_N^{(r)}$ такие, что

$$\mathbf{P}\{\xi_i^{(r)} = k\} = \mathbf{P}\{\xi_i = k | \xi_i \leqslant r\}, \tag{8}$$

где $i = 1, \ldots, N, k = 1, \ldots, r$. Пусть $\tilde{\xi}_1^{(r)}, \ldots, \tilde{\xi}_N^{(r)}$ – случайные величины с распределением

$$\mathbf{P}\{\tilde{\xi}_i^{(r)} = k\} = \mathbf{P}\{\xi_i = k | \xi_i \neq r\}, \tag{9}$$

$i = 1, \ldots, N, k = 1, 2, \ldots$ Пусть

$$\zeta_N = \xi_1 + \ldots + \xi_N, \quad \zeta_N^{(r)} = \xi_1^{(r)} + \ldots + \xi_N^{(r)},$$

$$\tilde{\zeta}_N^{(r)} = \tilde{\xi}_1^{(r)} + \ldots + \tilde{\xi}_N^{(r)} \tag{10}$$

и

$$P_r = \mathbf{P}\{\xi > r\}. \tag{11}$$

В [7, 8] показано, что из (7) вытекают следующие утверждения.

Лемма 1. *Справедливо равенство*

$$\mathbf{P}\{\eta_{(N)} \leqslant r\} = (1 - P_r)^N \frac{\mathbf{P}\{\zeta_N^{(r)} \leqslant n\}}{\mathbf{P}\{\zeta_N \leqslant n\}}.$$

Лемма 2. *Справедливо равенство*

$$\mathbf{P}\{\mu_r = k\} = \binom{N}{k} p_r^k (1 - p_r)^{N-k}$$

$$\times \frac{\mathbf{P}\{\tilde{\zeta}_{N-k}^{(r)} \leqslant n - kr\}}{\mathbf{P}\{\zeta_N \leqslant n\}}.$$

Вспомогательные результаты

Для того чтобы оценить поведение $(1 - P_r)^N$ в лемме 1, рассмотрим асимптотику NP_r.

Лемма 3. *В условиях теоремы 1* $NP_r \to \gamma$.

Доказательство. Из (1) и (11) следует

$$P_r = \sum_{i=1}^{\infty} \frac{h(r+i)}{(r+i)^{\tau}}. \qquad (12)$$

Тогда

$$P_r = h(r) \sum_{i=1}^{\infty} \frac{h(r+i)}{h(r)(r+i)^{\tau}}. \qquad (13)$$

Очевидно, что $r \to \infty$ (см. (5)). Легко видеть, что если $i < Ar$, где $A > 0$ – константа, то равномерно по i

$$\frac{h(r+i)}{h(r)} \to 1.$$

В таком случае

$$\sum_{i=1}^{\infty} \frac{h(r+i)}{h(r)(r+i)^{\tau}} = \sum_{i=1}^{[Ar]} \frac{1+o(1)}{(r+i)^{\tau}}$$

$$+ \sum_{i=[Ar]+1}^{\infty} \frac{h(r+i)}{h(r)(r+i)^{\tau}}.$$

Нетрудно проверить, что при достаточно больших A сумма

$$\sum_{i=1}^{[Ar]} \frac{1}{(r+i)^{\tau}}$$

сколь угодно мало отличается от $(\tau-1)r^{1-\tau}$. Кроме того, поскольку $h(x)$ не возрастает,

$$\sum_{i=[Ar]+1}^{\infty} \frac{h(r+i)}{h(r)(r+i)^{\tau}} \leqslant \sum_{i=[Ar]+1}^{\infty} \frac{1}{(r+i)^{\tau}},$$

а последняя величина тем меньше $r^{1-\tau}$, чем больше A. Суммируя сказанное и (12), (13), приходим к выводу, что

$$P_r \sim \frac{h(r)}{(\tau-1)r^{\tau-1}}, \qquad (14)$$

откуда и вытекает утверждение леммы 3. □

Пусть

$$m_r = \mathbf{E}\xi_1^{(r)}, \quad \tilde{m}_r = \mathbf{E}\tilde{\xi}_1^{(r)},$$

$$\sigma_r^2 = \mathbf{D}\xi_1^{(r)}, \quad \tilde{\sigma}_r^2 = \mathbf{D}\tilde{\xi}_1^{(r)}.$$

Из (1), (7), (8) находим

$$m_r = \frac{m - \sum_{k>r} kp_k}{1-P_r},$$

$$\sigma_r^2 = \frac{\sigma^2 + m^2 - \sum_{k>r} k^2 p_k}{1-P_r} - m_r^2,$$

$$\tilde{m}_r = \frac{m - rp_r}{1-p_r},$$

$$\tilde{\sigma}_r^2 = \frac{\sigma^2}{(1-p_r)^2}\left(1 - p_r - \frac{(m-r)^2}{\sigma^2} p_r\right). \qquad (15)$$

Используя условие $\tau > 3$ и (1)–(3), (15), видим, что первые два момента распределений (1), (8), (9) конечны. Тогда к суммам (10) можно применить центральную предельную теорему и установить справедливость следующих утверждений.

Лемма 4. *Пусть* $N \to \infty$. *Тогда распределения* $(\zeta_N - Nm)/(\sigma\sqrt{N})$ *слабо сходятся к стандартному нормальному закону.*

Лемма 5. *Пусть* $N \to \infty$. *Тогда распределения* $(\zeta_N^{(r)} - Nm_r)/(\sigma_r\sqrt{N})$ *слабо сходятся к стандартному нормальному закону.*

Лемма 6. *Пусть* $N \to \infty$. *Тогда распределения* $(\tilde{\zeta}_N - N\tilde{m}_r)/(\tilde{\sigma}_r\sqrt{N})$ *слабо сходятся к стандартному нормальному закону.*

Доказательства теорем 1–3

Очевидно, что в условиях теоремы 1 справедливо соотношение $r \to \infty$. По аналогии с (14) нетрудно видеть, что

$$\sum_{k>r} kp_k = O\left(\frac{h(r)}{r^{\tau-2}}\right). \qquad (16)$$

Из (5)

$$\frac{h(r)}{r^{\tau-2}} \sim \frac{\gamma(\tau-1)r}{N}. \qquad (17)$$

и

$$r = O(N^{1/(\tau-1)}).$$

Поскольку $\tau > 3$, имеем $r = o(\sqrt{N})$. Таким образом, из (16), (17)

$$\sum_{k>r} kp_k = o\left(\frac{1}{\sqrt{N}}\right). \qquad (18)$$

Следовательно, из (15) и леммы 3 находим, что

$$m_r = m(1 + o(N^{-1/2})). \qquad (19)$$

Из (1) и (2)

$$r^2 p_r \to 0. \qquad (20)$$

Положим

$$z_N(n) = \frac{n - Nm}{\sigma\sqrt{N}}, \quad z_N^{(r)}(n) = \frac{n - Nm_r}{\sigma_r\sqrt{N}},$$

$$\tilde{z}_N^{(r)}(n) = \frac{n - N\tilde{m}_r}{\tilde{\sigma}_r\sqrt{N}}. \qquad (21)$$

Соотношение (20) вместе с (4), (15), (18), (19), (21) показывает, что в условиях теоремы 1 величины $z_N(n)$ и $z_N^{(r)}(n)$ асимптотически эквивалентны. Следовательно, из лемм 4, 5 находим, что

$$\frac{\mathbf{P}\{\zeta_N^{(r)} \leqslant n\}}{\mathbf{P}\{\zeta_N \leqslant n\}} = 1 + o(1).$$

Теперь утверждение теоремы 1 легко следует из лемм 1, 3.

В условиях теоремы 2 $k = Np_r + u_r\sqrt{Np_r(1-p_r)}$ и из (23)

$$\tilde{z}_{N-k}^{(r)}(n - kr) = \frac{n - kr - (N-k)\tilde{m}_r}{\tilde{\sigma}_r\sqrt{N}}$$

$$= \frac{n - Nm}{\tilde{\sigma}_r\sqrt{N}} - \frac{u_r(r - \tilde{m}_r)\sqrt{p_r(1-p_r)}}{\tilde{\sigma}_r}. \qquad (22)$$

Из (1), (2), (21) и (22) выводим, что

$$\tilde{z}_{N-k}^{(r)}(n - kr) = z_N(n)\frac{\sigma}{\tilde{\sigma}_r} + O(1).$$

Следовательно, используя (6), видим, что $z_N(n)$ и $z_{N-k}^{(r)}(n-kr)$ стремятся к бесконечности одновременно и из лемм 4, 6

$$\frac{\mathbf{P}\{\tilde{\zeta}_{N-k}^{(r)} \leqslant n - kr\}}{\mathbf{P}\{\zeta_N \leqslant n\}} \to 1. \qquad (23)$$

Используя нормальное приближение биномиального распределения при $Np_r(1-p_r) \to \infty$:

$$\binom{N}{k} p_r^k (1-p_r)^{N-k} = \frac{1 + o(1)}{\sqrt{2\pi Np_r(1-p_r)}} e^{-u_r^2/2}, \qquad (24)$$

находим, что теорема 2 следует из (23), (24) и леммы 2.

Теорема 3 доказывается аналогично. Ясно, что $p_r \to 0$. Заметим, что (22) остается справедливым для рассматриваемых в теореме 3 значений k, если мы заменим u_r на $(k - Np_r)/\sqrt{Np_r}$. Учитывая (4), (15), (20), (21), получаем

$$\tilde{z}_{n-k}^{(r)}(n - kr) \sim z_N(n).$$

Отсюда и из лемм 4, 6 опять приходим к (23). Теорема 3 следует из (23) и леммы 2, если мы используем пуассоновское приближение биномиального распределения, справедливое при $p_r \to 0$:

$$\binom{N}{k} p_r^k (1-p_r)^{N-k} \sim \frac{(Np_r)^k}{k!} e^{-Np_r}.$$

Автор выражает глубокую благодарность рецензенту за важные замечания.

Литература

1. *Колчин В. Ф.* Случайные графы. М.: Физматгиз, 2000. 256 с.

2. *Павлов Ю. Л.* Условные конфигурационные графы со случайным параметром степенного распределения степеней // Математический сборник. 2018. Т. 209, вып. 2. С. 120–137. doi: 10.4213/sm8832

3. *Павлов Ю. Л., Хворостянская Е. В.* О предельных распределениях степеней вершин конфигурационных графов с ограниченным числом ребер // Математический сборник. 2016. Т. 207, вып. 3. С. 93–110. doi: 10.4213/sm8512

4. *Павлов Ю. Л., Чеплюкова И. А.* Предельные распределения числа вершин заданной степени конфигурационного графа с ограниченным числом ребер // Теория вероятностей и ее применения. 2021. Т. 66, вып. 3. С. 468–486. doi: 10.4213/tvp5332

5. *Павлов Ю. Л., Чеплюкова И. А.* Случайные графы Интернет-типа и обобщенная схема размещения // Дискретная математика. 2008. Т. 20, вып. 3. С. 3–18. doi: 10.4213/dm1008

6. *Чеплюкова И. А.* О максимальной степени вершины в условном конфигурационном графе // Труды Карельского научного центра РАН. 2020. № 7. С. 98–109. doi: 10.17076/mat1200

7. *Чупрунов А. Н., Фазекаш И.* Аналог обобщенной схемы размещения. Предельные теоремы для максимального объема ячейки // Дискретная математика. 2012. Т. 24, вып. 3. С. 122–129. doi: 10.4213/dm1203

8. *Чупрунов А. Н., Фазекаш И.* Аналог обобщенной схемы размещения. Предельные теоремы для числа ячеек заданного объема // Дискретная математика. 2012. Т. 24, вып. 1. С. 140–158. doi: 10.4213/dm1178

9. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // Eur. J. Comb. 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8

10. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press, 2017. 337 p. doi: 10.1017/9781316779422

11. *Pavlov Yu.* Asymptotics of the structure of conditional configuration graphs with bounded number of links // Stochastic modeling and applied research of tecnology. Proceedings of the Third International Workshop SMARTY'22. 2023. Vol. 3. P. 22–29. doi: 10.57753/SMARTY.2023.49.82.003

12. *Reittu H., Norros I.* On the power-law random graph model of massive data networks // Performance Evaluation. 2004. Vol. 55, iss. 1-2. P. 3–23. doi: 10.1016/S0166-5316(03)00097-X

References

1. *Kolchin V. F.* Random graphs. Cambridge: Cambridge University Press; 1999. 252 p. doi: 10.1017/CBO9780511721342

2. *Pavlov Yu. L.* Conditional configuration graphs with discrete power-law distribution of vertex degrees. *Sbornik: Mathematics.* 2018;209(2):258–275. doi: 10.1070/SM8832

3. *Pavlov Yu. L., Khvorostyanskaya E. V.* On the limit distributions of the degrees of vertices in configuration graphs with a bounded number of edges. *Sbornik: Mathematics.* 2016;207(3):400–417. doi: 10.1070/SM8512

4. *Pavlov Yu. L., Cheplyukova I. A.* Limit distributions of the number of vertices of a given degree in a configuration graph with bounded number of edges. *Theory Probab. Appl.* 2021;66(3):376–390. doi: 10.1137/S0040585X97T990460

5. *Pavlov Yu. L., Cheplyukova I. A.* Random graphs of Internet type and the generalised allocation scheme. *Discrete Mathematics and Applications.* 2008;18(5):447–463. doi: 10.1515/DMA.2008.033

6. *Cheplyukova I. A.* On the maximum vertex degree in a conditional configuration graph. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2020;7:98–110. (In Russ.). doi: 10.17076/mat1200

7. *Chuprunov A. N., Fasekas I.* An analogue of the generalised allocation scheme: limit theorems for the maximum dell load. *Discrete Mathematics and Applications.* 2012;22(3):307–314. doi: 10.1515/dma-2012-020

8. *Chuprunov A. N., Fasekas I.* An analogue of the generalised allocation scheme: limit theorems for the number of cells containing a given number of particles. *Discrete Mathematics and Applications.* 2012;22(1):101–422. doi: 10.1515/dma-2012-008

9. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Comb.* 1980;1(4):311–316. doi: 10.1016/S0195-6698(80)80030-8

10. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press; 2017. 337 p. doi: 10.1017/9781316779422

11. *Pavlov Yu.* Asymptotics of the structure of conditional configuration graphs with bounded number of links. *Stochastic modeling and applied research of tecnology. Proceedings of the Third International Workshop SMARTY'22.* 2023;3:22–29. doi: 10.57753/SMARTY.2023.49.82.003

12. *Reittu H., Norros I.* On the power-law random graph model of massive data networks. *Perform. Eval.* 2004;55(1-2):3–23. doi: 10.1016/S0166-5316(03)00097-X

Поступила в редакцию / received: 21.12.2024; принята к публикации / accepted: 10.04.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Павлов Юрий Леонидович

д-р физ.-мат. наук, профессор, главный научный сотрудник

e-mail: pavlov@krc.karelia.ru

CONTRIBUTOR:

Pavlov, Yury

Dr. Sci. (Phys.-Math.), Professor, Chief Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 60–64
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 60–64
DOI: 10.17076/mat2094

УДК 517.91

ДИНАМИКА ОПТИМАЛЬНОГО ПОВЕДЕНИЯ ДВУХВИДОВОГО БИОСООБЩЕСТВА С ДВУМЯ УЧАСТКАМИ ОБИТАНИЯ С УЧЕТОМ ВНУТРИВИДОВОЙ КОНКУРЕНЦИИ И МИГРАЦИИ

А. М. Сазонов

Институт прикладных математических исследований КарНЦ РАН,
ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)
Институт математики и информационных технологий,
Петрозаводский государственный университет (пр. Ленина, 33, Петрозаводск, Республика Карелия, Россия, 185910)

В статье рассматривается задача теории оптимального фуражирования, а именно, задача выбора популяцией наиболее пригодного участка обитания и нахождения условий ухода из него. Динамика взаимодействия популяций хищника и жертвы описывается модифицированной системой Лотки – Вольтерры с внутривидовой конкуренцией жертв. Предполагается, что жертвы имеют возможность мигрировать между участками обитания. В работе развивается подход Чарнова – Кривана, учитывая критику касательно невозможности для популяции обладания точной информацией о качестве всех участков. Для этого ставится задача нахождения оптимальных с точки зрения равновесия по Нэшу долей остающихся на участке жертв. Таким образом, оптимальная стратегия поведения определяется только для текущего участка обитания популяции. Исследуется полученная на основе равновесия по Нэшу гибридная система. Получены достаточные условия инвариантности одной из областей фазового пространства гибридной системы и устойчивости положения равновесия в данной области.

Ключевые слова: динамические системы; теория оптимального фуражирования; гибридные системы; равновесие по Нэшу

Для цитирования: Сазонов А. М. Динамика оптимального поведения двухвидового биосообщества с двумя участками обитания с учетом внутривидовой конкуренции и миграции // Труды Карельского научного центра РАН. 2025. № 4. С. 60–64. doi: 10.17076/mat2094

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

A. M. Sazonov. DYNAMICS OF THE OPTIMAL BEHAVIOUR OF A TWO-SPECIES BIOCOMMUNITY WITH TWO PATCHES TAKING INTO ACCOUNT INTRASPECIFIC COMPETITION AND MIGRATION

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)
Institute of Mathematics and Information Technologies, Petrozavodsk State University (33 Lenin St., 185910 Petrozavodsk, Karelia, Russia)

The paper examines a problem of the optimal foraging theory, namely, the problem of choosing the most suitable patch by a population and finding the conditions for leaving it. The dynamics of the interactionbetween the predators and the preys is described by a modified Lotka – Volterra system taking into accountthe intraspecific competition among the prey and the feasibility of migration for the predator and the prey. Prey are assumed to have the possibility to migrate between patches. The paper furthers the approach of Charnov – Krivan taking into account the criticism concerning the impossibility for the population to have full information on the quality of all patches. This is done by posing the problem of finding optimal in the sense of the Nash equilibrium shares of the preystaying in the patch. Thus, the optimal behavior strategy is determined only for the patch currently occupied by the population. The hybrid system based on the Nash equilibrium is studied. The sufficient conditions for the invariance of a domain of the hybrid system's phase space and stability of the equilibrium in this domain are determined.

Keywords: dynamical systems; optimal foraging theory; hybrid systems; Nash equilibrium

For citation: Sazonov A. M. Dynamics of the optimal behaviour of a two-species biocommunity with two patches taking into account intraspecific competition and migration. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2025. No. 4. P. 60–64. doi: 10.17076/mat2094

Funding. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

Введение

В статье рассматривается задача теории оптимального фуражирования, а именно, задача выбора наиболее пригодного участка обитания и условия ухода из него. Под участком понимается ограниченная территория, содержащая ресурсы питания (энергетические ресурсы). Основным результатом в решении задачи об уходе популяции из участка является классическая теорема Э. Чарнова [3] о маргинальных значениях, согласно которой уход популяции из участка происходит при снижении мгновенной скорости потребления до средней скорости потребления. Данная теория была развита В. Криваном, который предложил концепцию идеального свободного распределения [4], предполагающую, что популяция обладает точной информацией о качестве каждого участка и распределяется между участками, максимизируя скорость потребления энергии.

Однако критики данного подхода утверждают, что в реальных условиях популяция не имеет точной информации о качестве участков [6]. В работах [1, 5] предложено развитие концепции В. Кривана, учитывающее критику, представленную в [6]. Согласно предложенному подходу, в качестве принципа оптимальности при выборе участка используется равновесие по Нэшу. При этом, в отличие от подхода Кривана, оптимальная стратегия строится только по отношению к некоторому участку, а не определяется оптимальное распределение популяции по всем участкам. Решается задача поиска оптимальной доли популяции, остающейся на участке.

Настоящая работа развивает исследования [1, 5], в которых рассматривается миграция между двумя участками с учетом внутривидовой конкуренции жертв на каждом из участков. Ставится задача определения оптимальных в смысле равновесия по Нэшу долей популяции, остающихся на каждом из участков. При этом оставшаяся часть особей мигрирует на другой участок.

Задача оптимального поведения жертв

Постановка задачи

Рассмотрим следующие динамические системы, описывающие динамику взаимодействующих на участках популяций хищника и жертвы, учитывающую миграцию жертв между участками:

$$\begin{cases} \dot{x}_1 = x_1(a_1p_1 - b_1p_1^2x_1 - \mu_1(1-p_1)) \\ -c_1p_1x_1y_1 + \mu_2(1-p_2)x_2 \\ = f_1(x_1, x_2, y_1, p_1, p_2), \\ \dot{y}_1 = y_1(k_1c_1p_1x_1 - m_1) = g_1(x_1, y_1), \\ \dot{x}_2 = x_2(a_2p_2 - b_2p_2^2x_2 - \mu_2(1-p_2)) \\ -c_2p_2x_2y_2 + \mu_1(1-p_1)x_1 \\ = f_2(x_1, x_2, y_2, p_1, p_2), \\ \dot{y}_2 = y_2(k_2c_2p_2x_2 - m_2) = g_2(x_2, y_2), \end{cases} \tag{1}$$

где $x_i(t), y_i(t) > 0$ – количественные характеристики популяций жертв и хищников на участке i соответственно, $p_i \in [0,1]$ – доли жертв, остающихся на участке i, $a_i > 0$ – коэффициент прироста жертв в отсутствие хищников, $b_i > 0$ описывает внутривидовую конкуренцию жертв на участке i, $c_i > 0$ – коэффициент истребления хищником жертв на участке i, $\mu_i > 0$ – коэффициент миграции жертв из участка i за единицу времени, $m_i > 0$ – коэффициент естественной смертности хищников на участке i, $k_i \in (0,1)$ – доля полученной с потребляемой хищником биомассой энергии, которая расходуется им на воспроизводство на участке i, $i = 1, 2$. Естественно считать фазовым пространством системы (1) множество $\mathbb{R}^4_+ = \{(x_1, x_2, y_1, y_2) \in \mathbb{R}^4 : x_i > 0, y_i > 0, i = 1, 2\}$.

Поставим задачу нахождения долей p_1, p_2, максимизирующих мгновенные скорости роста популяций жертв $\dot{x}_1, \dot{x}_2$, то есть f_1, f_2. Соответствующие доли характеризуют оптимальное поведение популяций жертв на каждом из участков.

Таким образом, получаем игровую задачу с двумя участниками — популяциями жертв, где p_1, p_2 — стратегии популяций жертв на каждом из участков. Будем называть такую игру «миграция жертв». В качестве принципа оптимальности будем рассматривать равновесие по Нэшу $(p_1^*, p_2^*) \in [0,1]^2$, где $f_1(p_1^*, p_2^*) = \max\limits_{p_1} f_1(p_1, p_2^*)$, $f_2(p_1^*, p_2^*) = \max\limits_{p_2} f_2(p_1^*, p_2)$.

Равновесие по Нэшу

Теорема 1. *Равновесие по Нэшу (p_1^*, p_2^*) в игре «миграция жертв» имеет вид*

$$(p_1^*, p_2^*) = \begin{cases} (0,0), z \in D_1, \\ (\hat{p}_1, 0), z \in D_2, \\ (0, \hat{p}_2), z \in D_3, \\ (\hat{p}_1, \hat{p}_2), z \in D_4, \\ (1,0), z \in D_5, \\ (1, \hat{p}_2), z \in D_6, \\ (0,1), z \in D_7, \\ (\hat{p}_1, 1), z \in D_8, \\ (1,1), z \in D_9, \end{cases}$$

где $z = (x_1, x_2, y_1, y_2) \in \mathbb{R}^4_+$. Здесь

$$\hat{p}_i = \frac{a_i + \mu_i - c_iy_i}{2b_ix_i}, i = 1, 2,$$

$$D_1 = \{z \in \mathbb{R}^4_+ : y_1 \geqslant y_1^+, y_2 \geqslant y_2^+\},$$
$$D_2 = \{z \in \mathbb{R}^4_+ : y_1^- \leqslant y_1 < y_1^+, y_2 \geqslant y_2^+\},$$
$$D_3 = \{z \in \mathbb{R}^4_+ : y_1 \geqslant y_1^+, y_2^- \leqslant y_2 < y_2^+\},$$
$$D_4 = \{z \in \mathbb{R}^4_+ : y_1^- \leqslant y_1 < y_1^+, y_2^- \leqslant y_2 < y_2^+\},$$
$$D_5 = \{z \in \mathbb{R}^4_+ : y_1 < y_1^-, y_2 \geqslant y_2^+\},$$
$$D_6 = \{z \in \mathbb{R}^4_+ : y_1 < y_1^-, y_2^- \leqslant y_2 < y_2^+\},$$
$$D_7 = \{z \in \mathbb{R}^4_+ : y_1 \geqslant y_1^+, y_2 < y_2^-\},$$
$$D_8 = \{z \in \mathbb{R}^4_+ : y_1^- \leqslant y_1 < y_1^+, y_2 < y_2^-\},$$
$$D_9 = \{z \in \mathbb{R}^4_+ : y_1 < y_1^-, y_2 < y_2^-\},$$

где $y_i^+ = \frac{a_i+\mu_i}{c_i}$, $y_i^- = \frac{a_i+\mu_i-2b_ix_i}{c_i}$, $i = 1, 2$.

Доказательство. Рассмотрим уравнение

$$\frac{\partial f_1(p_1, p_2)}{\partial p_1} = a_1x_1 - 2b_1p_1x_1^2 + \mu_1x_1 - c_1x_1y_1 = 0.$$

Поскольку $x_1 > 0$, получим $p_1 = \frac{a_1+\mu_1-c_1y_1}{2b_1x_1} = \hat{p}_1$. Условие $p_2 \leqslant 1$ равносильно $y_1 \geqslant \frac{a_1+\mu_1-2b_1x_1}{c_1}$. Ясно, что если $y_1 \leqslant \frac{a_1+\mu_1-2b_1x_1}{c_1}$, то $p_1 = 1$. Если $z \in D_1 \cup D_2$, то есть $p_1 < 0$, очевидно, следует полагать $p_1 = 0$.

Подставив $p_1 = \hat{p}_1$ и $p_1 = 0$ в $f_2(x_1, x_2, y_2, p_1, p_2)$, в обоих случаях получим

$$\frac{\partial f_2(p_1, p_2)}{\partial p_2} = a_2x_2 - 2b_2p_2x_2^2 + \mu_2x_2 - c_2x_2y_2.$$

Отсюда получим $p_2 = \frac{a_2+\mu_2-c_2y_2}{2b_2x_2} = \hat{p}_2$. Условие $p_2 \leqslant 1$ равносильно $y_2 \geqslant \frac{a_2+\mu_2-2b_2x_2}{c_2}$. Ясно, что если $y_2 \leqslant \frac{a_2+\mu_2-2b_2x_2}{c_2}$, то $p_2 = 1$. Если $z \in D_1 \cup D_3$, то есть $p_2 < 0$, очевидно, следует полагать $p_2 = 0$. □

Замечание 1. Согласно теореме 1, на выбор стратегии жертв влияют численность популяций жертв (x_i), определяющая внутривидовую конкуренцию, и численность хищников (y_i), определяющая потребление хищниками жертв.

Оптимальная динамика и предельные множества

Система с переменной структурой

Рассмотрим систему (1) при полученных в теореме 1 равновесных по Нэшу стратегиях (p_1^*, p_2^*). Получим следующую гибридную систему, имеющую вид системы с переменной структурой, описывающую оптимальное миграционное поведение популяций:

$$\begin{cases} \dot{x}_1 = -\mu_1 x_1 + \mu_2 x_2, \\ \dot{y}_1 = -m_1 y_1, \\ \dot{x}_2 = -\mu_2 x_2 + \mu_1 x_1, \\ \dot{y}_2 = -m_2 y_2, \end{cases} \quad z \in D_1. \tag{2}$$

$$\begin{cases} \dot{x}_1 = \frac{(a_1+\mu_1-c_1y_1)^2}{4b_1} - \mu_1 x_1 + \mu_2 x_2, \\ \dot{y}_1 = k_1c_1y_1\left(\frac{a_1+\mu_1-c_1y_1}{2b_1} - \frac{m_1}{k_1c_1}\right), \\ \dot{x}_2 = -\mu_2 x_2 + \mu_1 x_1, \\ \dot{y}_2 = -m_2 y_2, \end{cases} \quad z \in D_2. \tag{3}$$

$$\begin{cases} \dot{x}_1 = -\mu_1 x_1 + \mu_2 x_2, \\ \dot{y}_1 = -m_1 y_1, \\ \dot{x}_2 = \frac{(a_2+\mu_2-c_2y_2)^2}{4b_2} - \mu_2 x_2 + \mu_1 x_1, \\ \dot{y}_2 = k_2c_2y_2\left(\frac{a_2+\mu_2-c_2y_2}{2b_2} - \frac{m_2}{k_2c_2}\right), \end{cases} \quad z \in D_3. \tag{4}$$

$$\begin{cases} \dot{x}_1 = \frac{(a_1+\mu_1-c_1y_1)^2}{4b_1} - \mu_1 x_1 + \mu_2 x_2, \\ \dot{y}_1 = k_1c_1y_1\left(\frac{a_1+\mu_1-c_1y_1}{2b_1} - \frac{m_1}{k_1c_1}\right), \\ \dot{x}_2 = \frac{(a_2+\mu_2-c_2y_2)^2}{4b_2} - \mu_2 x_2 + \mu_1 x_1, \\ \dot{y}_2 = k_2c_2y_2\left(\frac{a_2+\mu_2-c_2y_2}{2b_2} - \frac{m_2}{k_2c_2}\right), \end{cases} \quad z \in D_4.$$

$$\begin{cases} \dot{x}_1 = x_1(a_1 - b_1x_1 - c_1y_1) + \mu_2 x_2, \\ \dot{y}_1 = y_1(k_1c_1x_1 - m_1), \\ \dot{x}_2 = -\mu_2 x_2, \\ \dot{y}_2 = -m_2 y_2, \end{cases} \quad z \in D_5. \tag{5}$$

$$\begin{cases} \dot{x}_1 = x_1(a_1 - b_1x_1 - c_1y_1) + \mu_2 x_2, \\ \dot{y}_1 = y_1(k_1c_1x_1 - m_1), \\ \dot{x}_2 = \frac{(a_2+\mu_2-c_2y_2)^2}{4b_2} - \mu_2 x_2 + \mu_1 x_1, \\ \dot{y}_2 = k_2c_2y_2\left(\frac{a_2+\mu_2-c_2y_2}{2b_2} - \frac{m_2}{k_2c_2}\right), \end{cases} \quad z \in D_6.$$

$$\begin{cases} \dot{x}_1 = -\mu_1 x_1, \\ \dot{y}_1 = -m_1 y_1, \\ \dot{x}_2 = x_2(a_2 - b_2x_2 - c_2y_2) + \mu_1 x_1, \\ \dot{y}_2 = y_2(k_2c_2x_2 - m_2), \end{cases} \quad z \in D_7. \tag{6}$$

$$\begin{cases} \dot{x}_1 = \frac{(a_1+\mu_1-c_1y_1)^2}{4b_1} - \mu_1 x_1, \\ \dot{y}_1 = k_1c_1y_1\left(\frac{a_1+\mu_1-c_1y_1}{2b_1} - \frac{m_1}{k_1c_1}\right), \\ \dot{x}_2 = x_2(a_2 - b_2x_2 - c_2y_2) + \mu_1 x_1, \\ \dot{y}_2 = y_2(k_2c_2x_2 - m_2), \end{cases} \quad z \in D_8.$$

$$\begin{cases} \dot{x}_1 = x_1(a_1 - b_1x_1 - c_1y_1), \\ \dot{y}_1 = y_1(k_1c_1x_1 - m_1), \\ \dot{x}_2 = x_2(a_2 - b_2x_2 - c_2y_2), \\ \dot{y}_2 = y_2(k_2c_2x_2 - m_2), \end{cases} \quad z \in D_9. \tag{7}$$

Теорема 2. *Положения равновесия $P_i \notin D_i$, $i = 1, 2, 3, 5, 7$, положительные полутраектории соответствующих подсистем* (2)–(6) *покидают области D_i, $i = 1, 2, 3, 5, 7$.*

Множество D_9 инвариантно, положение равновесия $P_9 \in D_9$ устойчиво, если $b_i < k_ic_i < \frac{2b_i(\mu_i - m_i)}{a_i+\mu_i}$, $a_ik_ic_i > m_ib_i$, $2m_i < \mu_i - a_i$, $\max(a_i, m_i) < \mu_i$ $i = 1, 2$.

Доказательство. Очевидно, положение равновесия $P_i = (x_1^{(i)}, y_1^{(i)}, x_2^{(i)}, y_2^{(i)}) \notin D_i$, $i = 1, 2, 3, 5, 7$, поскольку для любого из указанных P_i существует $y_j^{(i)} = 0$, $j \in \{1, 2\}$, а минимально возможные значения переменных y_j в областях D_i: $\frac{a_j+\mu_j}{c_j} > 0$, $j = 1, 2$.

Отметим, что для каждой из систем (2)–(6) существует $p_j^* = 0$, $j \in \{1, 2\}$. Тогда для $p_j^* = 0$ при любом $R \geqslant \frac{a_j+\mu_j}{c_j}$ имеем $\dot{y}_j < 0$ при $y_j = R$. Следовательно, положительные полутраектории системы пересекают любую гиперплоскость $y_j = R$, где $R \geqslant \frac{a_j+\mu_j}{c_j}$ (включая граничную), в направлении убывания y_j. Отсюда получаем, что положительные полутраектории соответствующих подсистем (2)–(6) покидают области D_i, $i = 1, 2, 3, 5, 7$, через граничную гиперплоскость $y_j = \frac{a_j+\mu_j}{c_j}$, где $p_j^* = 0$, $j \in \{1, 2\}$.

Найдем теперь достаточные условия инвариантности области D_9. Ясно, что действующая в этой области система (7) представляет собой совокупность двух независимых двумерных систем (первые два и вторые два уравнения системы (7) соответственно). Следовательно, можно проводить качественный анализ каждой из подсистем отдельно. Рассмотрим систему, состоящую из первых двух уравнений системы (7). Скалярное произведение правых частей этой системы на нормаль $\tau_1 = (2b_1, c_1)$ к границе $y_1 = \frac{a_1+\mu_1-2b_1x_1}{c_1}$ соответствующей области имеет вид

$$2b_1\dot{x}_1 + c_1\dot{y}_1 = 2b_1(b_1 - k_1c_1)x_1^2 + (k_1c_1(a_1+\mu_1) + 2b_1(m_1 - \mu_1))x_1 - m_1(a_1+\mu_1). \tag{8}$$

Ясно, что выражение (8) отрицательно при всех $x_1 \geqslant 0$, если $b_1 < k_1c_1$, $k_1c_1 < \frac{2b_1(\mu_1-m_1)}{a_1+\mu_1}$, что требует $m_1 < \mu_1$. Условие $b_1 < \frac{2b_1(\mu_1-m_1)}{a_1+\mu_1}$ равносильно $2m_1 < \mu_1 - a_1$, что требует $a_1 < \mu_1$. Очевидно, для границы $y_2 = \frac{a_2+\mu_2-2b_2x_2}{c_2}$

области, соответствующей системе из двух последних уравнений (7), рассуждение полностью аналогично. Таким образом, поскольку $\mathbb{R}^4_+$ инвариантно для системы (7), получаем, что приведенные выше условия достаточны для инвариантности D_9.

Согласно [2] положение равновесия P_9 устойчиво для системы (7). Таким образом, из инвариантности D_9 следует утверждение об устойчивости P_9. $\square$

Замечание 2. Следует отметить, что приведенные в теореме 2 результаты с точки зрения предметной области означают, что полная миграция с любого из участков ($p_j^* = 0$, $j \in \{1, 2\}$) возможна только в течение конечного времени, поскольку положительные полутраектории соответствующих подсистем покидают свои области фазового пространства за конечное время. Полученные в теореме 2 достаточные условия инвариантности области D_9 и устойчивости положения равновесия P_9 с точки зрения предметной области являются условиями постоянного отсутствия миграции между участками для начальных значений численностей популяций хищников и жертв, принадлежащих D_9.

Заключение

В ходе рассмотрения поставленной задачи получены оптимальные с точки зрения равновесия по Нэшу доли остающихся на участке жертв. Исследована полученная на основе равновесия по Нэшу гибридная система. Согласно результатам анализа, полная миграция с любого из участков возможна только в течение конечного времени. Получены достаточные условия инвариантности области фазового пространства гибридной системы, соответствующей случаю отсутствия миграции, и устойчивости положения равновесия в данной области.

Литература

1. *Данилова И. В., Кириллов А. Н.* Динамика оптимального поведения двухвидового сообщества с учетом внутривидовой конкуренции и миграции // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2019. Т. 29, вып. 4. С. 518–531. doi: 10.18255/1818-1015-2018-3-268-275

2. *Свирежев Ю. М., Логофет Д. О.* Устойчивость биологических сообществ. М., 1978. 352 с.

3. *Charnov E. L.* Optimal foraging, the marginal value theorem // Theor. Popul. Biol. 1976. Vol. 9, no. 2. P. 129–136.

4. *Cressman R., Krivan V.* Two-patch population models with adaptive dispersal: the effects of varying dispersal speeds // J. Math. Biol. 2013. Vol. 67, no. 2. P. 329–358. doi: 10.1007/s00285-012-0548-3

5. *Ivanova A. S., Kirillov A. N.* Equilibrium and control in the biocommunity species composition preservation problem // Autom. Remote Control. 2017. Vol. 78, no. 8. P. 1500–1511. doi: 10.1134/S0005117917080100

6. *Matsumura S., Arlinghaus R., Dieckmann U.* Foraging on spatially distributed resources with suboptimal movement, imperfect information, and travelling costs: departures from the ideal free distribution // Oikos. 2010. Vol. 119, no. 9. P. 1469–1483. doi: 10.1111/j.1600-0706.2010.18196.x

References

1. *Danilova I. V., Kirillov A. N.* Optimal behavior dynamics of the two-species community with intraspecific competition and migration. *Vestnik Udmurtskogo Universiteta. Matematika. Mekhanika. Komp'yuternye Nauki = The Bulletin of Udmurt University. Mathematics. Mechanics. Computer Science.* 2019;29(4):518–531. (In Russ.)

2. *Svirezhev Yu. M., Logofet D. O.* Stability of biological communities. Moscow: Nauka; 1978. 352 p. (In Russ.)

3. *Charnov E. L.* Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* 1976;9(2):129–136.

4. *Cressman R., Krivan V.* Two-patch population models with adaptive dispersal: the effects of varying dispersal speeds. *J. Math. Biol.* 2013;67(2):329–358. doi: 10.1007/s00285-012-0548-3

5. *Ivanova A. S., Kirillov A. N.* Equilibrium and control in the biocommunity species composition preservation problem. *Autom. Remote Control.* 2017; 78(8):1500–1511. doi: 10.1134/S0005117917080100

6. *Matsumura S., Arlinghaus R., Dieckmann U.* Foraging on spatially distributed resources with suboptimal movement, imperfect information, and travelling costs: departures from the ideal free distribution. *Oikos.* 2010;119(9):1469–1483. doi: 10.1111/j.1600-0706.2010.18196.x

Поступила в редакцию / received: 24.04.2025; принята к публикации / accepted: 23.05.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Сазонов Александр Михайлович
канд. физ.-мат. наук, младший научный сотрудник
e-mail: sazon-tb@mail.ru

CONTRIBUTOR:

Sazonov, Alexander
Cand. Sci. (Phys.-Math.), Junior Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 65–73
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 65–73
DOI: 10.17076/mat2097

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.179.4

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ДЛЯ МАКСИМАЛЬНОГО ОБЪЕМА ДЕРЕВА В ЛЕСЕ ГАЛЬТОНА – ВАТСОНА С ОГРАНИЧЕННЫМ ЧИСЛОМ ВЕРШИН

Е. В. Хворостянская

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Рассматривается случайный лес, образованный траекториями однородного ветвящегося процесса Гальтона – Ватсона с N начальными частицами, в котором число прямых потомков каждой частицы имеет распределение $p_k = (k+1)^{-\tau} - (k+2)^{-\tau}$, $k = 0, 1, 2, \ldots$. Для леса Гальтона – Ватсона, общее число вершин которого не превосходит n, получены предельные распределения максимального объема дерева при $N, n \to \infty$ и различных значениях параметра τ, соответствующих критическому или докритическому ветвящемуся процессу.

Ключевые слова: лес Гальтона – Ватсона; максимальный объем дерева; предельное распределение

Для цитирования: Хворостянская Е. В. Предельные теоремы для максимального объема дерева в лесе Гальтона – Ватсона с ограниченным числом вершин // Труды Карельского научного центра РАН. 2025. № 4. С. 65–73. doi: 10.17076/mat2097

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

E. V. Khvorostyanskaya. LIMIT THEOREMS FOR THE MAXIMUM TREE SIZE IN THE GALTON – WATSON FOREST WITH A BOUNDED NUMBER OF VERTICES

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

We consider a random forest formed by trajectories of a homogeneous Galton – Watson branching process starting with N particles, where the number of immediate offspring of each particle has the distribution $p_k = (k+1)^{-\tau} - (k+2)^{-\tau}$, $k = 0, 1, 2, \ldots$. For the Galton – Watson forest containing at most n vertices, we find limit distributions of the maximum tree size as $N, n \to \infty$ for different values of the parameter τ corresponding to a critical or subcritical branching process.

Keywords: Galton – Watson forest; maximum tree size; limit distribution

F o r c i t a t i o n: Khvorostyanskaya E. V. Limit theorems for the maximum tree size in the Galton – Watson forest with a bounded number of vertices. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2025. No. 4. P. 65–73. doi: 10.17076/mat2097

F u n d i n g. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

Введение

В [3–7, 10] рассматривались леса Гальтона – Ватсона $\mathcal{F}_{N,n}$, сгенерированные однородным критическим ветвящимся процессом G_N, распределение числа прямых потомков каждой частицы которого является случайной величиной ξ с распределением вида

$$\mathbf{P}\{\xi = k\} = \frac{L(k+1)}{(k+1)^{\tau+1}},\ k = 0, 1, 2, \ldots, \quad (1)$$

где $\tau \in (1, 2)$, $L(x)$ – медленно меняющаяся на бесконечности функция, принимающая только положительные значения при $x \geqslant 1$. Предполагалось, что число некорневых вершин случайного леса известно и равно n. Нетрудно видеть, что для значений параметра $\tau \in (1, 2)$ распределение (1) имеет бесконечную дисперсию. Были доказаны предельные теоремы для максимального объема дерева и числа деревьев заданного объема при $N, n \to \infty$, при этом значение параметра τ распределения (1) в указанных работах определялось условием $\mathbf{E}\xi = 1$. Задача изучения характеристик случайных лесов, порожденных ветвящимися процессами с бесконечным вторым моментом распределения числа прямых потомков, возникла в связи с возможностью применения таких результатов при исследовании структуры и динамики конфигурационных графов, предназначенных для моделирования сложных сетей коммуникаций [17, 18]. Полученные ранее результаты о лесах Гальтона – Ватсона предполагают существование конечной дисперсии распределения случайной величины ξ [1, 8].

Пусть далее G_N – однородный ветвящийся процесс Гальтона – Ватсона с N начальными частицами, занумерованными числами $1, \ldots, N$. Распределение числа прямых потомков каждой частицы процесса G_N задается равенствами

$$p_k = \mathbf{P}\{\xi = k\} = \frac{1}{(k+1)^{\tau}} - \frac{1}{(k+2)^{\tau}}, \quad (2)$$

$k = 0, 1, 2, \ldots$. Характеристики случайного леса $\mathcal{F}_{N,n}$ с n некорневыми вершинами, порожденного критическим ветвящимся процессом G_N, изучались в [12, 14]. Нетрудно видеть, что распределение (2) является частным случаем распределения (1), при этом $L(x) = x\,(1 - (x/(x+1))^{\tau})$. Учитывая (2), несложно показать, что

$$m = \mathbf{E}\xi = \zeta(\tau, 2),$$

$$\sigma^2 = \mathbf{D}\xi = 2\zeta(\tau-1, 2) - 3\zeta(\tau, 2) - \zeta^2(\tau, 2),\ \ \tau > 2,$$

где $\zeta(s, v) = \sum\limits_{k=0}^{\infty}(k+v)^{-s}$ – обобщенная дзета-функция. Пусть τ_0 определяется равенством $\zeta(\tau_0, 2) = 1$, тогда $\tau_0 \approx 1,728$. Будем считать далее, что значения параметра τ распределения (2) находятся в интервале $\tau_0 \leqslant \tau \leqslant C < \infty$, то есть ветвящийся процесс G_N является критическим при $\tau = \tau_0$ или докритическим при $\tau > \tau_0$. Здесь и далее $C, C_1, C_2, \ldots$ – некоторые положительные постоянные, не всегда различные. Заметим, что для $\tau_0 \leqslant \tau \leqslant 2$ распределение (2) не имеет конечной дисперсии.

Выделим подмножество $F_{N,n}$ траекторий процесса G_N, каждая из которых содержит не более n вершин. Ветвящийся процесс G_N индуцирует на $F_{N,n}$ распределение вероятностей при условии, что общее число вершин не превосходит n. Построенный таким образом лес Гальтона – Ватсона, состоящий из N деревьев и содержащий не более n вершин, обозначим через $\mathcal{F}'_{N,n}$. В [13] получены предельные распределения числа деревьев заданного объема такого леса. В настоящей работе изучается предельное поведение случайной величины $\eta_{(N)}$, равной максимальному объему дерева в лесе $\mathcal{F}'_{N,n}$ при различном характере стремления N, n к бесконечности, при этом значения параметра τ могут быть как фиксированными, так и изменяющимися вместе с N: $\tau = \tau(N)$.

Заметим, что подобная задача рассматривалась в работах [9, 11], где были доказаны предельные теоремы для характеристик леса Гальтона – Ватсона с ограниченным числом вершин, порожденного ветвящимся процессом с пуассоновским распределением числа прямых потомков каждой частицы.

Основные результаты

Легко видеть, что процесс G_N состоит из N независимых подпроцессов $G^{(1)}, G^{(2)}, \ldots, G^{(N)}$, каждый из которых начинается с одной частицы. Пусть $\nu^{(1)}, \nu^{(2)}, \ldots, \nu^{(N)}$ – независимые одинаково распределенные случайные величины, равные числу частиц, существовавших за все время эволюции в подпроцессах $G^{(1)}, G^{(2)}, \ldots, G^{(N)}$. Введем обозначения:

$$q_k = \mathbf{P}\left\{\nu^{(1)} = k\right\},\ k = 1, 2, \ldots,$$

$$a = \mathbf{E}\nu^{(1)},\quad \tau \neq \tau_0,\quad b^2 = \mathbf{D}\nu^{(1)},\quad \tau > 2.$$

Известно [8, п. 2.3, (2)], что

$$a = \frac{1}{1-m},\quad b^2 = a^3\sigma^2,$$

при этом $a \to \infty$ при $\tau = \tau(N) \to \tau_0$.

Определим последовательность $B_N(a)$ следующим образом:

$$B_N(a) = \begin{cases} \sqrt{a^3 N \ln N} \text{ при } \\ \quad \tau = 2 \text{ или} (\tau-2)\ln N \to 0; \\ \sqrt{a^3 N \ln N \left(1-e^{-\alpha}\right)/\alpha} \text{ при } \\ \quad (\tau-2)\ln N \to 2\alpha,\ \alpha \neq 0; \\ \sqrt{2a^3 N/(\tau-2)} \text{ при } \\ \quad (\tau-2)\ln N \to +\infty; \\ \left(2a^3 N/|\tau-2|\right)^{1/\tau} \text{ при } \\ \quad (\tau-2)\ln N \to -\infty; \end{cases} \qquad (3)$$

и пусть

$$B_{a,N} = \left(a^{\tau+1}N\right)^{1/\tau}.$$

Обозначим через $g(x)$ плотность устойчивого распределения с параметром τ, $1 < \tau < 2$, и характеристической функцией

$$f(t) = \exp\left\{-\Gamma(1-\tau)|t|^\tau e^{-i\pi\tau t/(2|t|)}\right\}, \qquad (4)$$

и пусть $p(x)$ – плотность устойчивого распределения с параметром $1/\tau$ и характеристической функцией

$$h(t) = \exp\left\{-\left(\frac{|t|}{-\Gamma(1-\tau)}\right)^{1/\tau} e^{-i\pi t/(2\tau|t|)}\right\}. \qquad (5)$$

Положим

$$C(\tau) = \frac{1}{\tau\Gamma(1-1/\tau)\left(-\Gamma(1-\tau)\right)^{1/\tau}}.$$

$$\delta(u,v) = \begin{cases} 1, & u/v \geqslant 1, \\ 0, & u/v < 1. \end{cases}$$

Справедливы следующие утверждения.

Теорема 1. *Пусть $N, n \to \infty$, $n \geqslant N$, $\tau \geqslant 2$ фиксировано или $\tau = \tau(N) \to 2 \pm 0$. Тогда для любого фиксированного $z > 0$*

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{B_{a,N}} \leqslant z\right\} \to e^{-z^{-\tau}}.$$

Теорема 2. *Пусть $N \to \infty$, $n \geqslant N$, $N^{\tau-1}/a \to \infty$, $\tau_0 < \tau \leqslant C_1 < 2$. Тогда для любого фиксированного $z > 0$*

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{B_{a,N}} \leqslant z\right\} = 1 + \frac{\sum\limits_{k=1}^{\infty} \frac{(-\tau)^k}{k!} \int\limits_{-\infty}^{v_N} \widehat{I}_k(z,x)\,dx}{\int\limits_{-\infty}^{v_N} g(x)dx} + o(1),$$

где $v_N = (n - aN)/B_{a,N}$, при $k = 1, 2, \ldots$

$$\widehat{I}_k(u,v) = \int\limits_{x_1,\ldots,x_k \geqslant u} \frac{g\left(v - x_1 - \ldots - x_k\right)}{\left(x_1 \ldots x_k\right)^{\tau+1}} dx_1 \ldots dx_k. \qquad (6)$$

Теорема 3. *Пусть $N \to \infty$, $n \geqslant N$, $\tau = \tau_0$, $n/N^\tau \to \gamma$, где γ – некоторая положительная постоянная. Тогда для любого фиксированного $z > 0$*

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{N^\tau} \leqslant z\right\}$$

$$= 1 + \frac{\sum\limits_{k=1}^{[\gamma/z]} \frac{(-C(\tau))^k}{k!} \int\limits_{kz}^{\gamma} I_k(z,x)\,dx}{\int\limits_0^{\gamma} p(x)dx} \delta(\gamma, z) + o(1),$$

где при $k = 1, 2, \ldots$

$$I_k(u,v) = \int\limits_{X_k(u,v)} \frac{p\left(v - x_1 - \ldots - x_k\right)}{\left(x_1 \ldots x_k\right)^{1+1/\tau}} dx_1 \ldots dx_k, \qquad (7)$$

$$X_k(u,v) = \left\{x_i \geqslant u,\ i=1,\ldots,k,\ x_1 + \ldots + x_k \leqslant v\right\}.$$

Теорема 4. *Пусть $N \to \infty$, $n \geqslant N$, $\tau = \tau_0$, $n/N^\tau \to \infty$. Тогда для любого фиксированного $z > 0$*

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{N^\tau} \leqslant z\right\} = 1 + \sum_{k=1}^{\infty} \frac{(-C(\tau))^k}{k!} \int\limits_{kz}^{\infty} I_k(z,x)\,dx + o(1),$$

где $I_k(u,v)$ определено в (7).

Заметим, что в теореме 2 и далее запись $\tau_0 < \tau \leqslant C_1 < 2$ означает, что $\tau \to \tau_0 + 0$ или $\tau_0 < C_2 \leqslant \tau \leqslant C_1 < 2$.

Вспомогательные утверждения

Пусть ν_N – случайная величина, равная общему числу частиц, существовавших в процессе G_N до его вырождения:

$$\nu_N = \nu^{(1)} + \nu^{(2)} + \ldots + \nu^{(N)}.$$

Введем случайные величины $\nu_1(\mathcal{F}'), \ldots, \nu_N(\mathcal{F}')$, равные объемам деревьев леса $\mathcal{F}'_{N,n}$ с корнями $1, 2, \ldots, N$ соответственно. Очевидно, что эти случайные величины зависимы и для натуральных $k_1, \ldots, k_N$ таких, что $k_1 + \ldots + k_N \leqslant n$, справедливо равенство

$$\begin{aligned} &\mathbf{P}\left\{\nu_1(\mathcal{F}') = k_1, \ldots, \nu_N(\mathcal{F}') = k_N\right\} \\ &= \mathbf{P}\left\{\nu^{(1)} = k_1, \ldots, \nu^{(N)} = k_N \mid \nu_N \leqslant n\right\}. \end{aligned} \tag{8}$$

Таким образом, для двух наборов случайных величин $\nu_1(\mathcal{F}'), \ldots, \nu_N(\mathcal{F}')$ и $\nu^{(1)}, \ldots, \nu^{(N)}$ выполнены условия аналога обобщенной схемы размещения не более чем n частиц по N ячейкам [15, 16]. Обобщенная схема размещения была введена и подробно изучена В. Ф. Колчиным (см., например, [2]), в ней предполагалось, что число размещаемых частиц равно n.

Введем независимые случайные величины $\nu_1^{(\leqslant r)}, \ldots, \nu_N^{(\leqslant r)}$, распределение которых задается равенствами

$$\mathbf{P}\left\{\nu_i^{(\leqslant r)} = k\right\} = \mathbf{P}\left\{\nu^{(1)} = k \mid \nu^{(1)} \leqslant r\right\},$$

где $i = 1, \ldots, N,\ k = 1, 2, \ldots$. Согласно теореме 1 из [15], из (8) следует равенство

$$\mathbf{P}\left\{\eta_{(N)} = k\right\} = (1-P_r)^N \frac{\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\}}{\mathbf{P}\left\{\nu_N \leqslant n\right\}}, \tag{9}$$

где

$$P_r = \mathbf{P}\left\{\nu^{(1)} > r\right\}, \quad \zeta_N^{(\leqslant r)} = \nu_1^{(\leqslant r)} + \ldots + \nu_N^{(\leqslant r)}.$$

Таким образом, для получения предельного распределения случайной величины $\eta_{(N)}$ достаточно знать асимптотику вероятностей P_r, $\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\}$, $\mathbf{P}\left\{\nu_N \leqslant n\right\}$.

Следующие утверждения (леммы 1–4) получены в [13, п. 2.2].

Лемма 1. *Пусть $N \to \infty$, $n \geqslant N$, $\tau > 2$ фиксировано. Тогда*

$$\mathbf{P}\{\nu_N \leqslant n\} = \frac{1 + o(1)}{\sqrt{2\pi}} \int_{-\infty}^{z_N} e^{-x^2/2} dx, \quad z_N = \frac{n - aN}{b\sqrt{N}}.$$

Лемма 2. *Пусть $N \to \infty$, $n \geqslant N$, $\tau = 2$ или $\tau = \tau(N) \to 2 \pm 0$. Тогда*

$$\mathbf{P}\{\nu_N \leqslant n\} = \frac{1 + o(1)}{\sqrt{2\pi}} \int_{-\infty}^{y_N} e^{-x^2/2} dx, \quad y_N = \frac{n - aN}{B_N(a)}.$$

Лемма 3. *Пусть $N \to \infty$, $n \geqslant N$, $\tau_0 < \tau \leqslant C_3 < 2$, $N^{\tau-1}/a \to \infty$. Тогда*

$$\mathbf{P}\{\nu_N \leqslant n\} = \int_{-\infty}^{v_N} g(x) dx (1 + o(1)), \quad v_N = \frac{n - aN}{B_{a,N}}.$$

Лемма 4. *При $N \to \infty$, $n \geqslant N$, $\tau = \tau_0$ справедливо равенство*

$$\mathbf{P}\left\{\nu_N \leqslant n\right\} = \int_0^{n/N^\tau} p(x) dx (1 + o(1)).$$

Введем обозначения:

$$\begin{aligned} E\left(u, v\right) &= C(\tau) \int_v^\infty e^{iuy} y^{-1-1/\tau} dy, \\ \widehat{E}\left(u, v\right) &= \tau \int_v^\infty e^{iuy} y^{-1-\tau} dy. \end{aligned} \tag{10}$$

Согласно лемме 3 из [12] справедливо следующее утверждение.

Лемма 5. *Если $N \to \infty$, $\tau = \tau_0$, $r = zN^\tau + O(1)$, где z – некоторая положительная постоянная, то $NP_r \to E\left(0, z\right) = \tau C(\tau) z^{-1/\tau}$.*

Ниже в лемме 6 получена асимптотика P_r при $\tau > \tau_0$.

Лемма 6. *Если $N \to \infty$, $\tau > \tau_0$, $N^{\tau-1}/a \to \infty$, $r = zB_{a,N} + O(1)$, где z – некоторая положительная постоянная, то $NP_r \to \widehat{E}\left(0, z\right) = z^{-\tau}$.*

Доказательство. Нетрудно видеть, что при условии $N^{\tau-1}/a \to \infty$ для заданных значений r выполнено соотношение $r^{1-1/\tau}/a \to \infty$. В [13] показано, что если $k \to \infty$ и $\tau_0 < C_4 \leqslant \tau \leqslant C_5 < \infty$ или $\tau \to \tau_0$ так, что $k^{1-1/\tau}/a \to \infty$, то справедливо равенство

$$q_k = \tau \left(\frac{a}{k}\right)^{\tau+1} (1 + o(1)). \tag{11}$$

Используя (11) и заменяя суммирование интегрированием, несложно показать, что

$$P_r = \frac{\tau(1+o(1))}{NB_{a,N}} \sum_{k=r+1}^{\infty} \left(\frac{k}{B_{a,N}}\right)^{-\tau-1}$$

$$= \frac{\tau}{N}(1+o(1)) \int_z^{\infty} y^{-\tau-1} dy$$

$$= \frac{\widehat{E}\,(0,z)}{N}(1+o(1)) = \frac{z^{-\tau}}{N}(1+o(1)).$$

Отсюда следует утверждение леммы. □

Найдем асимптотику вероятностей $\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\}$ при $N \to \infty$ и различных значениях τ.

Лемма 7. *Пусть $N \to \infty$, $n \geqslant N$, $\tau > 2$ фиксировано, $r = zB_{a,N} + O(1)$, где z – некоторая положительная постоянная. Тогда*

$$\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\} = \frac{1+o(1)}{\sqrt{2\pi}} \int_{-\infty}^{z_N} e^{-x^2/2} dx, \ z_N = \frac{n-aN}{b\sqrt{N}}.$$

Доказательство. Пусть $\varphi(t)$ обозначает характеристическую функцию случайной величины $\nu^{(1)}$. В [13, формула (44)] показано, что при $u \to 0$

$$\varphi(u) = 1 + iau - \left(b^2 + a^2\right) u^2/2 + o\left(u^2\right).$$

С помощью этого равенства находим, что для характеристической функции $\psi_r^{(1)}(u)$ случайной величины $\nu_1^{(\leqslant r)}$ выполнено соотношение

$$\begin{aligned} \psi_r^{(1)}(u) = 1 + iau - \left(b^2 + a^2\right) u^2/2 \\ + o\left(u^2\right) + O\left(uP_r\right) + S_r(u), \end{aligned} \tag{12}$$

где

$$S_r(u) = \sum_{k=r+1}^{\infty} \left(1 - e^{iuk}\right) q_k. \tag{13}$$

Несложно показать, что $\left|1 - e^{ix}\right| \leqslant C_6|x|$ для любого x. Учитывая (11), получаем оценку

$$|S_r(u)| \leqslant C_7|u| \sum_{k=r+1}^{\infty} k^{-\tau} \leqslant C_8|u|N^{-1+1/\tau}. \tag{14}$$

Из (12), (14) и леммы 6 следует, что при $N \to \infty$ и любом фиксированном t

$$\psi_r^{(1)}\left(\frac{t}{b\sqrt{N}}\right) = 1 + \frac{iat}{b\sqrt{N}} - \frac{b^2+a^2}{2b^2N}t^2 + o\left(\frac{1}{N}\right).$$

Используя это равенство, для характеристической функции $\psi_{r,N}(t)$ случайной величины $\left(\zeta_N^{(\leqslant r)} - aN\right) / \left(b\sqrt{N}\right)$ выводим соотношение

$$\psi_{r,N}(t) = e^{-t^2/2}(1+o(1)).$$

Отсюда следует утверждение леммы. □

Лемма 8. *Пусть $N \to \infty$, $n \geqslant N$, $\tau = 2$ или $\tau = \tau(N) \to 2 \pm 0$, $r = zB_{a,N} + O(1)$, где z – некоторая положительная постоянная. Тогда*

$$\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\} = \frac{1+o(1)}{\sqrt{2\pi}} \int_{-\infty}^{y_N} e^{-x^2/2} dx, \ y_N = \frac{n-aN}{B_N(a)}.$$

Доказательство. Пусть $\tau = 2$ или $\tau \to 2$ так, что $(\tau - 2)\ln N \to 0$. В [13, формула (17)] при $u \to 0$ доказано равенство

$$\varphi(u) = 1 + iau + a^3u^2 \ln|u| + o\left(u^2 \ln|u|\right).$$

С помощью этого соотношения, аналогично (12), находим, что для характеристической функции $\psi_r^{(1)}(u)$ случайной величины $\nu_1^{(\leqslant r)}$ справедливо равенство

$$\begin{aligned} \psi_r^{(1)}(u) = 1 + iau + a^3u^2 \ln|u| \\ + o\left(u^2 \ln|u|\right) + O\left(uP_r\right) + S_r(u), \end{aligned} \tag{15}$$

где $S_r(u)$ определено в (13). Легко видеть, что выполнено (14).

Обозначим через $\widetilde{\psi}_{r,N}(t)$ характеристическую функцию случайной величины $\left(\zeta_N^{(\leqslant r)} - aN\right) / B_N(a)$:

$$\widetilde{\psi}_{r,N}(t) = \exp\left\{-\frac{iaNt}{B_N(a)}\right\}\left(\psi_r^{(1)}\left(\frac{t}{B_N(a)}\right)\right)^N. \tag{16}$$

Используя (3), (14)–(16), лемму 6 и формулу Тейлора, несложно показать, что при любом фиксированном t

$$\widetilde{\psi}_{r,N}(t) = \exp\left\{-\frac{iaNt}{B_N(a)} \right. \tag{17}$$

$$\left. + N \ln\left(1 + \frac{iat}{B_N(a)} - \frac{t^2}{2N} + o\left(\frac{1}{N}\right)\right)\right\} = e^{-t^2/2} + o(1).$$

Пусть $\tau \to 2$ так, что $(\tau - 2)\ln N \to 2\alpha$, $\alpha \neq 0$. В [13, формула (24)] показано, что при $u \to 0$

$$\varphi(u) = 1 + iau + \frac{1-e^{-\alpha}}{\alpha} a^3u^2 \ln|u| + o\left(u^2 \ln|u|\right).$$

С помощью этого равенства, (3), (14), (16), леммы 6 и формулы Тейлора несложно проверить, что при любом фиксированном t справедливо (17).

Пусть $\tau \to 2$ так, что $(\tau - 2)\ln N \to +\infty$. Согласно формуле (25) из [13] для характеристической функции $\varphi(u)$ при $u \to 0$ выполнено соотношение

$$\varphi(u) = 1 + iau - \frac{a^3u^2}{\tau-2}(1+o(1)).$$

Используя это равенство, (13), (14) и лемму 6, находим, что

$$\psi_r^{(1)}(u) = 1+iau-\frac{a^3u^2}{\tau-2}(1+o(1))+O\left(\frac{u}{N^{1-1/\tau}}\right).$$

Отсюда и из (16) при любом фиксированном t и $B_N(a) = \sqrt{2a^3N/(\tau-2)}$ выводим (17).

Пусть $\tau \to 2$ так, что $(\tau-2)\ln N \to -\infty$. В [13, формула (26)] доказано, что

$$\varphi(u) = 1 + iau + \frac{a^3|u|^\tau}{\tau-2}(1+o(1)).$$

С помощью этого соотношения, (3), (13), (14), (16) и леммы 6 получаем (17).

Из равенства (17) следует утверждение леммы. □

Лемма 9. *Пусть $N \to \infty$, $\tau_0 < \tau \leqslant C_9 < 2$, $N^{\tau-1}/a \to \infty$, $r = zB_{a,N} + O(1)$, где z – некоторая положительная постоянная. Тогда распределение случайной величины $\left(\zeta_N^{(\leqslant r)} - Na\right)/B_{a,N}$ слабо сходится к распределению с плотностью*

$$\widehat{\omega}(x) = e^{z^{-\tau}}\sum_{k=0}^{\infty}\frac{(-\tau)^k}{k!}\widehat{I}_k(z,x),$$

где $\widehat{I}_0(u,v) = g(v)$, $\widehat{I}_k(u,v)$ при $k = 1,2,\ldots$ определены в (6).

Доказательство. Аналогично [13, формула (14)] можно показать, что если $u \to 0$ так, что $a^\tau|u|^{\tau-1} \to 0$, то

$$\begin{aligned}\varphi(u) &= 1+iau\\ &- \Gamma(1-\tau)a^{\tau+1}|u|^\tau e^{-i\pi\tau u/(2|u|)} + o\big(a^{\tau+1}u^\tau\big).\end{aligned} \tag{18}$$

Обозначим через $\rho_{r,N}(t)$ характеристическую функцию случайной величины $\left(\zeta_N^{(\leqslant r)} - aN\right)/B_{a,N}$. Нетрудно видеть, что

$$\begin{aligned}\rho_{r,N}(t) &= \exp\left\{-\frac{iaNt}{B_{a,N}}\right\}\varphi^N\left(\frac{t}{Ba,N}\right)\\ &\times\left(\frac{1-\widehat{Q}_r(t)}{1-P_r}\right)^N,\end{aligned} \tag{19}$$

где

$$\widehat{Q}_r(t) = (1+o(1))\sum_{k=r+1}^{\infty}e^{itk/B_{a,N}}q_k.$$

При выполнении условий леммы $r^{1-1/\tau}/a\to\infty$ и справедливо соотношение (11). Учитывая (11) и переходя от суммирования к интегрированию, находим, что

$$\begin{aligned}\widehat{Q}_r(t) &= \frac{\tau}{N}(1+o(1))\int_z^\infty e^{ity}y^{-\tau-1}dy\\ &= \frac{\widehat{E}(t,z)}{N}(1+o(1)),\end{aligned}$$

где $\widehat{E}(t,z)$ определено в (10). Отсюда, из (18), (19) и леммы 6 получаем, что при любом фиксированном t

$$\begin{aligned}\rho_{r,N}(t) = \exp\Bigg\{&z^{-\tau} - \widehat{E}(t,z)\\ &-\Gamma(1-\tau)|t|^\tau e^{-i\pi\tau t/(2|t|)} + O\left(\frac{a^2N}{B_{a,N}^2}\right) + o(1)\Bigg\},\end{aligned}$$

и поскольку $a^2N/B_{a,N}^2 \to 0$, выполнено равенство

$$\rho_{r,N}(t) = \exp\left\{z^{-\tau} - \widehat{E}(t,z)\right\}f(t)(1+o(1)),$$

где $f(t)$ задано формулой (4).

Нетрудно видеть, что функция $\hat{l}(t) = \widehat{E}(u,v)/\widehat{E}(0,v)$ является характеристической функцией распределения с плотностью

$$\theta(x) = \frac{\tau x^{-\tau-1}}{\widehat{E}(0,v)}, \qquad x \geqslant v.$$

Раскладывая $e^{-\widehat{E}(t,z)}$ в ряд по степеням $\widehat{E}(t,z)$, получаем, что

$$\begin{aligned}\rho_{r,N}(t) &= e^{z^{-\tau}}f(t)(1+o(1))\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\widehat{E}^k(t,z)\\ &= e^{z^{-\tau}}(1+o(1))\sum_{k=0}^{\infty}\frac{(-1)^k\widehat{E}^k(0,z)}{k!}f(t)\hat{l}^k(t).\end{aligned}$$

Произведение характеристических функций $f(t)\hat{l}^k(t)$ при $k = 1,2,\ldots$ является характеристической функцией суммы независимых случайных величин, плотность $\widehat{p}_k(v)$ распределения такой суммы является сверткой плотностей:

$$\widehat{p}_k(v) = \int\limits_{x_1,\ldots,x_k\geqslant z} g\left(v - x_1 - \ldots - x_k\right)$$

$$\times\,\theta(x_1)\ldots\theta(x_k)dx_1\ldots dx_k = \frac{\tau^k \widehat{I}_k\left(z,v\right)}{\widehat{E}^k\left(0,z\right)}.$$

Заметим, что $\widehat{p}_0(v) = g(v)$. Тогда

$$\rho_{r,N}(t) = e^{z^{-\tau}}(1+o(1))$$
$$\times \sum_{k=0}^{\infty} \frac{(-1)^k \widehat{E}^k\left(0,z\right)}{k!} \int\limits_{-\infty}^{\infty} e^{itx}\widehat{p}_k(x)dx$$

$$= e^{z^{-\tau}}(1+o(1)) \int\limits_{-\infty}^{\infty} e^{itx} \left(\sum_{k=0}^{\infty} \frac{(-\tau)^k}{k!} \widehat{I}_k\left(z,x\right)\right) dx.$$

Отсюда следует утверждение леммы. $\square$

Лемма 10. *Пусть $N\to\infty$, $\tau=\tau_0$, $r/N^\tau\to z$, где z – некоторая положительная постоянная. Тогда распределение случайной величины $\zeta_N^{(\leqslant r)}/N^\tau$ слабо сходится к распределению с плотностью*

$$\omega(x) = e^{E(0,z)} \sum_{k=0}^{\infty} \frac{\left(-C(\tau)\right)^k}{k!} I_k\left(z,x\right),$$

где $I_0(u,v) = p(v)$, $I_k\left(u,v\right)$ при $k = 1,2,\ldots$ определены в (7).

Доказательство. Обозначим через $\varphi_{r,N}(t)$ характеристическую функцию случайной величины $\zeta_N^{(\leqslant r)}/N^\tau$. При любом фиксированном t справедливо равенство

$$\varphi_{r,N}(t) = \varphi^N\left(\frac{t}{N^\tau}\right)\left(\frac{1-Q_r(t)}{1-P_r}\right)^N,$$

где

$$Q_r(t) = (1+o(1)) \sum_{k=r+1}^{\infty} e^{itk/N^\tau} q_k.$$

В [14, формула (16)] показано, что при $k\to\infty$

$$q_k = \frac{C(\tau)}{k^{1+1/\tau}}(1+o(1)).$$

Следуя доказательству равенства (20) в [12], находим, что при любом фиксированном t

$$\varphi_{r,N}(t) = e^{E(0,z)-E(t,z)} \varphi^N\left(\frac{t}{N^\tau}\right)(1+o(1))$$
$$= e^{E(0,z)-E(t,z)} h(t)(1+o(1)),$$

где $h(t)$ определено в (5). Аналогично доказательству леммы 9 несложно показать, что

$$\varphi_{r,N}(t) = e^{E(0,z)} \sum_{k=0}^{\infty} \frac{(-1)^k E^k\left(0,z\right)}{k!} l_k(t)(1+o(1)),$$

где $l_0(t) = h(t)$, а при $k = 1,2,\ldots$ функция $l_k(t) = h(t)\left(E\left(t,z\right)/E\left(0,z\right)\right)^k$ является характеристической функцией суммы независимых случайных величин, при этом плотность $p_k(v)$ распределения такой суммы имеет вид $p_k(v) = \left(C(\tau)/E\left(0,z\right)\right)^k I_k\left(z,v\right)$. Тогда

$$\varphi_{r,N}(t) = e^{E(0,z)}(1+o(1))$$
$$\times \int\limits_{0}^{\infty} e^{itx} \left(\sum_{k=0}^{\infty} \frac{\left(-C(\tau)\right)^k}{k!} I_k\left(z,x\right)\right) dx,$$

откуда следует утверждение леммы. $\square$

Доказательство теорем 1–4

Легко видеть, что утверждение теоремы 1 при фиксированных $\tau > 2$ следует из равенства (9) и лемм 1, 6, 7, а при $\tau = 2$ или $\tau = \tau(N) \to 2\pm 0$ – из формулы (9) и лемм 2, 6, 8.

Пусть выполнены условия теоремы 2. С помощью леммы 9 при $r = zB_{a,N} + O(1)$ выводим соотношение

$$\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\} = e^{z^{-\tau}}(1+o(1))$$
$$\times \left(\int\limits_{-\infty}^{v_N} g(x)dx + \sum_{k=1}^{\infty} \frac{(-\tau)^k}{k!} \int\limits_{-\infty}^{v_N} \widehat{I}_k\left(z,x\right)dx\right).$$

Отсюда, из равенства (9) и лемм 3, 6 нетрудно получить утверждение теоремы 2.

При выполнении условий теоремы 3 из леммы 10 находим, что

$$\mathbf{P}\left\{\zeta_N^{(\leqslant r)} \leqslant n\right\} = e^{E(0,z)}(1+o(1)) \int\limits_{0}^{n/N^\tau} p(x)dx$$

$$+\, e^{E(0,z)}(1+o(1)) \sum_{k=1}^{\infty} \frac{\left(-C(\tau)\right)^k}{k!} \int\limits_{kz}^{n/N^\tau} I_k\left(z,x\right)dx$$

$$= e^{E(0,z)}(1+o(1)) \int\limits_{0}^{\gamma} p(x)dx + e^{E(0,z)}(1+o(1))$$

$$\times\,\delta\left(\gamma,z\right) \sum_{k=1}^{[\gamma/z]} \frac{\left(-C(\tau)\right)^k}{k!} \int\limits_{kz}^{\gamma} I_k\left(z,x\right)dx.$$

Из этого равенства, (9) и лемм 4, 5 следует утверждение теоремы 3.

Аналогично теореме 3 доказывается теорема 4, учитывая, что $n/N^{\tau} \to \infty$.

Литература

1. *Казимиров Н. И., Павлов Ю. Л.* Одно замечание о лесах Гальтона – Ватсона // Дискретная математика. 2000. Т. 12, вып. 1. С. 47–59. doi: 10.4213/dm320

2. *Колчин В. Ф.* Случайные отображения. М.: Наука, 1984, 208 с.

3. *Павлов Ю. Л.* Максимальное дерево случайного леса в конфигурационном графе // Математический сборник. 2021. Т. 212, вып. 9. С. 146–163. doi: 10.4213/sm9481

4. *Павлов Ю. Л.* Об объемах деревьев леса Гальтона – Ватсона в промежуточном случае // Дискретная математика. 2025. Т. 37, вып. 1. С. 39–51. doi: 10.4213/dm1860

5. *Павлов Ю. Л.* Об объемах деревьев леса Гальтона – Ватсона с бесконечной дисперсией в критическом случае // Дискретная математика. 2024. Т. 36, вып. 2. С. 33–49. doi: 10.4213/dm1813

6. *Павлов Ю. Л.* О максимальном дереве леса Гальтона – Ватсона с бесконечной дисперсией распределения числа потомков // Дискретная математика. 2023. Т. 35, вып. 2. С. 78–92. doi: 10.4213/dm1765

7. *Павлов Ю. Л.* О максимальном объеме дерева случайного леса // Дискретная математика. 2022. Т. 34, вып. 4. С. 69–83. doi: 10.4213/dm1729

8. *Павлов Ю. Л.* Случайные леса. Петрозаводск: Карел. науч. центр РАН, 1996. 259 с.

9. *Павлов Ю. Л., Хворостянская Е. В.* О максимальном объеме дерева в лесе Гальтона – Ватсона с ограниченным числом вершин // Дискретная математика. 2014. Т. 26, вып. 3. С. 90–100. doi: 10.4213/dm1293

10. *Павлов Ю. Л., Чеплюкова И. А.* Объемы деревьев случайного леса и конфигурационные графы // Труды МИАН им. В. А. Стеклова. 2022. Т. 316. С. 298–315. doi: 10.4213/tm4216

11. *Хворостянская Е. В.* Предельные распределения числа деревьев заданного объема в лесе Гальтона – Ватсона с ограниченным числом вершин // Ученые записки ПетрГУ. Серия: Естественные и технические науки. 2014. № 8(145). Т. 1. С. 115–118.

12. *Хворостянская Е. В.* Предельные теоремы для максимального объема дерева леса Гальтона – Ватсона в критическом случае // Дискретная математика. 2022. Т. 34, вып. 2. С. 120–136. doi: 10.4213/dm1703

13. *Хворостянская Е. В.* Предельные теоремы для числа деревьев заданного объема леса Гальтона – Ватсона с ограниченным числом вершин // Дискретная математика. 2025. Т. 37, вып. 2. С. 120–136. doi: 10.4213/dm1855

14. *Хворостянская Е. В.* О числе деревьев заданного объема в лесе Гальтона – Ватсона в критическом случае // Теория вероятностей и ее применения. 2023. Т. 68, вып. 1. С. 75–92. doi: 10.4213/tvp5573

15. *Чупрунов А. Н., Фазекаш И.* Аналог обобщенной схемы размещения. Предельные теоремы для максимального объема ячейки // Дискретная математика. 2012. Т. 24, вып. 3. С. 122–129. doi: 10.4213/dm1203

16. *Чупрунов А. Н., Фазекаш И.* Аналог обобщенной схемы размещения. Предельные теоремы для числа ячеек заданного объема // Дискретная математика. 2012. Т. 24, вып. 1. С. 140–158. doi: 10.4213/dm1178

17. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press, 2017. 337 p. doi: 10.1017/9781316779422

18. *Hofstad R.* Random Graphs and Complex Networks. Vol. 2. Cambridge: Cambridge University Press, 2024. 498 p. doi: 10.1017/9781316795552

References

1. *Kazimirov N. I., Pavlov Yu. L.* A remark on Galton – Watson forests. *Discrete Mathematics and Applications.* 2000;10(1):49–62. doi: 10.4213/dm320

2. *Kolchin V. F.* Random mappings. New York: Springer; 1986. 206 p.

3. *Pavlov Yu. L.* The maximum tree of a random forest in the configuration graph. *Sbornik: Mathematics.* 2021;212(9):1329–1346. doi: 10.1070/SM9481

4. *Pavlov Yu. L.* On the sizes of trees in a Galton – Watson forest in the intermediate case. *Diskretnaya matematika = Discrete Mathematics and Applications.* 2025;37(1):39–51. (In Russ.). doi: 10.4213/dm1860

5. *Pavlov Yu. L.* On the sizes of trees in a Galton – Watson forest with infinite variance in the critical case. *Diskretnaya matematika = Discrete Mathematics and Applications.* 2024;36(2):33–49. (In Russ.). doi: 10.4213/dm1813

6. *Pavlov Yu. L.* On the maximal Galton – Watson forest tree with infinite variance of the offspring. *Diskretnaya matematika = Discrete Mathematics and Applications.* 2023;35(2):78–92. (In Russ.). doi: 10.4213/dm1765

7. *Pavlov Yu. L.* On the maximal size of tree in a random forest. *Discrete Mathematics and Applications.* 2024;34(4):221–232. doi: 10.1515/dma-2024-0019

8. *Pavlov Yu. L.* Random Forests. Utrecht: VSP; 2000. 122 p. doi: 10.1515/9783110941975

9. *Pavlov Yu. L., Hvorostyanskaya E. V.* On the maximum size of a tree in the Galton – Watson forest with a bounded number of vertices. *Discrete Mathematics and Applications.* 2014;24(6):363–371. doi: 10.1515/dma-2014-0032

10. *Pavlov Yu. L., Cheplyukova I. A.* Sizes of trees in a random forest and configuration graphs. *Proceedings of the Steklov Institute of Mathematics.* 2022;316:280–297. doi: 10.1134/S0081543822010205

11. *Khvorostyanskaya E. V.* Limit distribution of given size trees' number in Galton – Watson forest with limited number of vertices. *Uchenye zapiski PetrGU. Serya: Estestvennye i tehnicheskie nauki = Proceedings of Petrozavodsk State University. Ser.: Natural & Engineering Sciences.* 2014;8(145.1):115–118. (In Russ.)

12. *Khvorostyanskaya E. V.* Limit theorems for the maximal tree size of a Galton – Watson forest in the critical case. *Discrete Mathematics and Applications.* 2023;33(4):205–217. doi: 10.1515/dma-2023-0019

13. *Khvorostyanskaya E. V.* Limit theorems for the number of trees of a given size in a Galton – Watson forest with a bounded number of vertices. *Diskretnaya matematika = Discrete Mathematics and Applications.* 2025;37(2):120–136. (In Russ.). doi: 10.4213/dm1855

14. *Khvorostyanskaya E. V.* On the number of trees of a given size in a Galton – Watson forest in the critical case. *Theory of Probability and its Applications.* 2023;68(1):62–76. doi: 10.1137/S0040585X97T991283

15. *Chuprunov A. N., Fasekas I.* An analogue of the generalised allocation scheme: limit theorems for the maximum dell load. *Discrete Mathematics and Applications.* 2012;22(3):307–314. doi: 10.1515/dma-2012-020

16. *Chuprunov A. N., Fasekas I.* An analogue of the generalised allocation scheme: limit theorems for the number of cells containing a given number of particles. *Discrete Mathematics and Applications.* 2012;22(1):101–422. doi: 10.1515/dma-2012-008

17. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press; 2017. 337 p. doi: 10.1017/9781316779422

18. *Hofstad R.* Random Graphs and Complex Networks. Vol. 2. Cambridge: Cambridge University Press; 2024. 498 p. doi: 10.1017/9781316795552

Поступила в редакцию / received: 29.04.2025; принята к публикации / accepted: 02.06.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Хворостянская Елена Владимировна

канд. физ.-мат. наук, старший научный сотрудник

e-mail: cher@krc.karelia.ru

CONTRIBUTOR:

Khvorostyanskaya, Elena

Cand. Sci. (Phys.-Math.), Senior Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 74–77
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 74–77
DOI: 10.17076/mat2042

УДК 519.21

МОМЕНТНЫЕ ХАРАКТЕРИЗАЦИИ БИНОМИАЛЬНОГО И ОТРИЦАТЕЛЬНОГО БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

А. Н. Чупрунов, К. Н. Яковлев*

*Факультет прикладной математики, физики и информационных технологий, Чувашский государственный университет (ул. Университетская, 38, Чебоксары, Чувашская Республика, Россия, 428034), *ktq0x@yandex.ru*

Пусть случайная величина $\xi(\beta)$, $0 < \beta < R$, имеет распределение степенного ряда, $m(\beta)$ – ее математическое ожидание, $\sigma^2(\beta)$ – ее дисперсия. Здесь R – радиус сходимости ряда, определяющего распределение случайной величины $\xi(\beta)$. Показано, что случайная величина $\xi(\beta)$ имеет биномиальное распределение с параметрами n, $\frac{\beta}{1+\beta}$ тогда и только тогда, когда $\sigma^2(\beta) = \frac{1}{1+\beta}m(\beta)$, $0 < \beta < R$, и $\lim\limits_{\beta\to 0+0} \frac{m(\beta)}{\beta} = n$, случайная величина $\xi(\beta)$ имеет отрицательное биномиальное распределение с параметрами r, β тогда и только тогда, когда $\sigma^2(\beta) = \frac{1}{1-\beta}m(\beta)$, $0 < \beta < 1$, и $\lim\limits_{\beta\to 0+0} \frac{m(\beta)}{\beta} = r$.

Ключевые слова: распределение степенного ряда; биномиальное распределение; отрицательное биномиальное распределение; математическое ожидание; дисперсия

Для цитирования: Чупрунов А. Н., Яковлев К. Н. Моментные характеризации биномиального и отрицательного биномиального распределения // Труды Карельского научного центра РАН. 2025. № 4. С. 74–77. doi: 10.17076/mat2042

Финансирование. Финансовое обеспечение исследования осуществлялось из средств федерального бюджета на выполнение государственного задания Чувашского государственного университета.

A. N. Chuprunov, K. N. Iakovlev*. MOMENT CHARACTERIZATIONS OF BINOMIAL AND NEGATIVE BINOMIAL DISTRIBUTION

*Faculty of Applied Mathematics, Physics and Information Technologies, Chuvash State University (38 Universitetskaya St., 428034 Cheboksary, Chuvash Republic, Russia), *ktq0x@yandex.ru*

Let $\xi(\beta)$, $0 < \beta < R$, be a random variable with a power series distribution. Here, R is the radius of convergence of the series that determines the distribution of the random variable $\xi(\beta)$. We will denote by $m(\beta)$ and $\sigma^2(\beta)$ the expectation and the variance of $\xi(\beta)$. It is proved that $\xi(\beta)$ has a binomial distribution with the parameters n and $\frac{\beta}{1+\beta}$ if and only if $\sigma^2(\beta) = \frac{1}{1+\beta}m(\beta)$, $0 < \beta < R$, while $\lim\limits_{\beta\to 0+0} \frac{m(\beta)}{\beta} = n$, $\xi(\beta)$ has a negative binomial distribution with the parameters r, β if and only if $\sigma^2(\beta) = \frac{1}{1-\beta}m(\beta)$, $0 < \beta < 1$, and $\lim\limits_{\beta\to 0+0} \frac{m(\beta)}{\beta} = r$.

Keywords: power series distribution; binomial distribution; negative binomial distribution; expectation; variance

For citation: Chuprunov A. N., Iakovlev K. N. Moment characterizations of binomial and negative binomial distribution. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 74–77. doi: 10.17076/mat2042

Funding. The studies were funded from the federal budget through state assignment to the Chuvash State University.

Введение

Напомним понятие случайной величины с распределением степенного ряда. Пусть ряд $B(\beta) = \sum_{k=0}^{\infty} \frac{b_k}{k!}\beta^k$ такой, что $b_k \geqslant 0$, $k = 0, 1, 2, \dots$, хотя бы одно из чисел b_k положительно и ряд имеет положительный радиус сходимости R. Случайная величина $\xi = \xi(\beta)$ имеет распределение степенного ряда, если ее распределение имеет вид

$$p_k = p_k(\beta) = \mathbf{P}\{\xi = k\} = \frac{b_k \beta^k}{k! B(\beta)}, \quad k = 0, 1, 2, \dots,$$

где число $0 < \beta < R$.

Замечание. Пусть случайная величина $\xi(\beta)$ имеет распределение степенного ряда, определенное функцией $B(\beta)$, случайная величина $\xi'(\beta)$ имеет распределение степенного ряда, определенное функцией $CB(\beta)$, где $C > 0$ – некоторая константа. Тогда распределения случайных величин $\xi(\beta)$ и $\xi'(\beta)$ совпадают.

В [2] изучалась сходимость сумм независимых случайных величин, имеющих распределение степенного ряда. Распределениями степенного ряда являются пуассоновское распределение, биномиальное распределение, отрицательное биномиальное распределение, логарифмическое распределение и ряд других распределений, которые используются в дискретной теории вероятностей. Многие схемы дискретной теории вероятностей определяются случайными величинами, имеющими распределение степенного ряда [3]. В [4] получены некоторые аналоги принципа инвариантности для сумм независимых случайных величин, имеющих распределение степенного ряда.

В этой статье дано описание биномиального и отрицательного биномиального распределений в классе распределений степенного ряда.

Основные результаты

Мы будем рассматривать случайные величины $\xi(\beta)$, имеющие распределения степенного ряда с одними и теми же числами $b_k \geqslant 0$, $k = 0, 1, 2 \dots$. Тогда их математические ожидания $m(\beta)$, дисперсии $\sigma^2(\beta)$, а также факториальные моменты

$$m_{(r)}(\beta) = \mathbf{E}\xi(\beta)(\xi(\beta) - 1)\dots(\xi(\beta) - r + 1)$$

являются непрерывными неотрицательными функциями, определенными на отрезке $(0, R)$. Заметим, что $m(\beta) = m_{(1)}(\beta)$.

Факториальные моменты случайной величины, имеющей распределение степенного ряда, имеют вид (см. [2])

$$m_{(r)}(\beta) = \frac{\beta^r B^{(r)}(\beta)}{B(\beta)}. \tag{1}$$

Из (1) следует, что

$$m(\beta) = \frac{\beta B'(\beta)}{B(\beta)},$$

$$\sigma^2(\beta) = \frac{\beta^2 B''(\beta)}{B(\beta)} + \frac{\beta B'(\beta)}{B(\beta)} - \left(\frac{\beta B'(\beta)}{B(\beta)}\right)^2.$$

Поэтому

$$m'(\beta) = \frac{\sigma^2(\beta)}{\beta}. \tag{2}$$

Пусть случайная величина ξ имеет биномиальное распределение с параметрами n, p. Тогда

$$\mathbf{P}\{\xi = k\} = p_k = C_n^k p^k q^{n-k}$$

$$= C_n^k \left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)^{n-k} = \frac{C_n^k \beta^k}{(1+\beta)^n},$$

$0 \leqslant k \leqslant n$, где β такое, что $p = \frac{\beta}{1+\beta}$, при этом $q = 1 - p = \frac{1}{1+\beta}$. Поэтому случайная величина ξ имеет распределение степенного ряда, определенное функцией $B(\beta) = (1+\beta)^n$. Ее математическое ожидание и дисперсия:

$$m(\beta) = \frac{n\beta}{1+\beta}, \quad \sigma^2(\beta) = \frac{n\beta}{(1+\beta)^2}, \quad 0 < \beta.$$

Следовательно, ее математическое ожидание и дисперсия связаны соотношением

$$\sigma^2(\beta) = \frac{1}{1+\beta} m(\beta), \tag{3}$$

и

$$\lim_{\beta \to 0+0} \frac{m(\beta)}{\beta} = n. \tag{4}$$

Теорема 1. *Если случайная величина ξ с распределением степенного ряда удовлетворяет условиям (3), (4), то она имеет биномиальное распределение с параметрами n, $p = \frac{\beta}{1+\beta}$.*

Пусть случайная величина ξ имеет отрицательное биномиальное распределение с параметрами r, β. Тогда ξ имеет распределение степенного ряда, определенное функцией $B(\beta) = \frac{1}{(1-\beta)^r}$, $0 < \beta < 1$. Ее математическое ожидание и дисперсия равны [1]:

$$m(\beta) = \frac{r\beta}{1-\beta}, \quad \sigma^2(\beta) = \frac{r\beta}{(1-\beta)^2}, \quad 0 < \beta < 1. \tag{5}$$

Из (5) следует, что ее дисперсия и математическое ожидание связаны соотношением

$$\sigma^2(\beta) = \frac{1}{1-\beta} m(\beta), \quad 0 < \beta < 1, \tag{6}$$

и

$$\lim_{\beta \to 0+0} \frac{m(\beta)}{\beta} = r. \tag{7}$$

Теорема 2. *Если случайная величина ξ с распределением степенного ряда удовлетворяет условиям (6), (7), то она имеет отрицательное биномиальное распределение с параметром r, β.*

Доказательство теоремы 1. Используя (3) и (2), составляем дифференциальное уравнение

$$m'(\beta) = \frac{1}{\beta(1+\beta)} m(\beta).$$

Поэтому

$$\ln'(m(\beta)) = \ln'\left(\frac{\beta}{1+\beta}\right)$$

и

$$m(\beta) = \frac{C\beta}{1+\beta}$$

для некоторого $C > 0$. Из (4) следует, что $C = n$. Но тогда

$$\frac{\beta B'(\beta)}{B(\beta)} = \frac{n\beta}{1+\beta},$$

$$\ln'(B(\beta)) = n \ln'(1+\beta)$$

и

$$B(\beta) = C_1(1+\beta)^n$$

для некоторого $C_1 > 0$. С учетом замечания это влечет теорему. □

Доказательство теоремы 2. Используя (6) и (2), составляем дифференциальное уравнение

$$m'(\beta) = \frac{1}{\beta(1-\beta)} m(\beta).$$

Поэтому

$$\ln'(m(\beta)) = \ln'\left(\frac{\beta}{1-\beta}\right)$$

и

$$m(\beta) = \frac{C\beta}{1-\beta}$$

для некоторого $C > 0$. Из (7) следует, что $C = r$. Но тогда

$$\frac{\beta B'(\beta)}{B(\beta)} = \frac{r\beta}{1-\beta},$$

$$\ln'(B(\beta)) = -r \ln'(1-\beta)$$

и

$$B(\beta) = \frac{C_1}{(1-\beta)^r}$$

для некоторого $C_1 > 0$. С учетом замечания это влечет теорему. □

Литература

1. *Ивченко Г. И., Медведев Ю. И.* Дискретные распределения. Вероятностно-статистический справочник. Одномерные распределения. М.: Ленанд, 2015. 256 с.

2. *Колчин А. В.* Предельные теоремы для обобщенной схемы размещения // Дискретная математика. 2003. Т. 18, вып. 4. С. 148–157.

3. *Колчин В. Ф.* Случайные графы. М.: Физматлит, 2000. 256 с.

4. *Чупрунов А. Н., Фазекаш И.* Принцип инвариантности для чисел частиц в ячейках обобщенной схемы размещения // Дискретная математика. 2023. Т. 35, вып. 3. С. 81–99. doi: 10.4213/dm1738

References

1. *Ivchenko G. I., Medvedev Yu. I.* Discrete distributions. Probability-statistical handbook. Univariate distributions. Moscow: Lenand; 2015. 256 p. (In Russ.)

2. *Kolchin A. V.* On limit theorems for the generalised allocation scheme. *Discrete Mathematics and Applications*. 2003;13(6):627–636. doi: 10.1515/156939203322733336

3. *Kolchin V. F.* Random Graphs. Cambridge: Cambridge University Press; 1999. 252 p. doi: 10.1017/CBO9780511721342

4. *Chuprunov A. N., Fazekas I.* Invariance principle for numbers of particles in cells of a general allocation scheme. *Diskretnaya matematika = Discrete Mathematics and Applications*. 2023;35(3):81–99. doi: 10.4213/dm1738 (In Russ.)

Поступила в редакцию / received: 23.12.2024; принята к публикации / accepted: 04.04.2025.
Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРАХ:

Чупрунов Алексей Николаевич
д-р физ.-мат. наук, профессор

e-mail: achuprunov@mail.ru

Яковлев Кирилл Николаевич
студент

e-mail: ktq0x@yandex.ru

CONTRIBUTORS:

Chuprunov, Alexey
Dr. Sci. (Phys.-Math.), Professor

Iakovlev, Kirill
Student

Труды Карельского научного центра РАН. 2025. № 4. С. 78–88
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 78–88
DOI: 10.17076/mat2050

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 517.9

О ЧИСЛЕ ОШИБОК В ФАЙЛЕ ПРИ КОДИРОВАНИИ С КОНТРОЛЬНОЙ СУММОЙ

А. Н. Чупрунов, К. Н. Яковлев*

*Факультет прикладной математики, физики и информационных технологий, Чувашский государственный университет (ул. Университетская, 38, Чебоксары, Чувашская Республика, Россия, 428034), *ktq0x@yandex.ru*

Пусть случайная величина $\mu_0(2n, K, N)$ есть число пустых ячеек среди первых K ячеек в схеме размещения $2n$ различимых частиц по N различным ячейкам с четным числом частиц в каждой ячейке. Показано, что если $\frac{n}{N} \to \infty$, $2Ke^{-\frac{2n}{N}} \to \lambda$, где $0 < \lambda < \infty$, то случайная величина $\mu_0(2n, K, N)$ сходится к пуассоновской случайной величине с параметром λ, если $\frac{n}{N} \to \infty$, $Ke^{-\frac{2n}{N}} \to \infty$, то центрированная и нормированная случайная величина $\mu_0(2n, K, N)$ сходится к гауссовской случайной величине с нулевым средним и единичной дисперсией. Даны приложения этих результатов к изучению ошибок первого и второго рода в критерии проверки гипотезы о вероятности пустой ячейки. Обсуждается использование этого критерия для проверки гипотезы о количестве ошибок в файле при кодировании с контрольной суммой.

Ключевые слова: схема размещения различимых частиц по различным ячейкам; обобщенная схема размещения; гауссовская случайная величина; пуассоновская случайная величина; предельная теорема

Для цитирования: Чупрунов А. Н., Яковлев К. Н. О числе ошибок в файле при кодировании с контрольной суммой // Труды Карельского научного центра РАН. 2025. № 4. С. 78–88. doi: 10.17076/mat2050

Финансирование. Финансовое обеспечение исследования осуществлялось из средств федерального бюджета на выполнение государственного задания Чувашского государственного университета.

A. N. Chuprunov, K. N. Iakovlev*. ON THE NUMBER OF ERRORS IN A FILE WITH CHECKSUM CODING

*Faculty of Applied Mathematics, Physics and Information Technologies, Chuvash State University (38 Universitetskaya St., 428034 Cheboksary, Chuvash Republic, Russia), *ktq0x@yandex.ru*

Let the random variable $\mu_0(2n, K, N)$ be the number of empty cells from the first K cells in a scheme of allocating $2n$ distinguishable particles to N different cells with an even number of particles in each cell. It is proved that if $\frac{n}{N} \to \infty$ and $2Ke^{-\frac{2n}{N}} \to \lambda$, where $0 < \lambda < \infty$, then $\mu_0(2n, K, N)$ converge to a Poisson random variable with the parameter λ, if $\frac{n}{N} \to \infty$ and $Ke^{-\frac{2n}{N}} \to \infty$, then the centered and normed random variables $\mu_0(2n, K, N)$ converge to a Gaussian random variable with the expectation 0 and the variance 1. We suggest applications for there results in the study of first- and second-order errors in the empty cell probability hypothesis test. We discuss using this test to verify the hypothesis about the number of errors in a file with checksum coding.

Keywords: allocation scheme of distinguishable particles in different cells; generalized allocation scheme; Gaussian random variable; Poisson random variable; limit theorem

For citation: Chuprunov A. N., Iakovlev K. N. On the number of errors in a file with checksum coding. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 78–88. doi: 10.17076/mat2050

Funding. The studies were funded from the federal budget through state assignment to the Chuvash State University.

Введение и основные результаты

В. Ф. Колчин в [1] ввел понятие обобщенной схемы размещения. Пусть $\xi_1, \xi_2, \ldots, \xi_N$ – независимые неотрицательные целочисленные одинаково распределенные случайные величины. Случайные величины $\eta_1, \ldots, \eta_N$ называются обобщенной схемой размещения n частиц по N ячейкам, если их совместное распределение имеет вид

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\}$$
$$= \mathbf{P}\left\{\xi_1 = k_1, \ldots, \xi_N = k_N \;\middle|\; \sum\nolimits_{i=1}^{N} \xi_i = n\right\},$$

где $k_1, k_2, \ldots, k_N$ – такие неотрицательные целые числа, что $k_1 + k_2 + \cdots + k_N = n$.

Различные схемы дискретной теории вероятностей, такие как случайные леса, случайные перестановки, случайные размещения, урновые схемы, являются частными случаями обобщенной схемы размещения [2]. Большое количество статей посвящено доказательству предельных теорем для случайных величин, связанных с обобщенной схемой размещения (см. монографии [2, 4, 5]).

Случайные величины $\eta_1, \ldots, \eta_N$ называются схемой размещения n различимых частиц по N различным ячейкам, если их совместное распределение имеет вид

$$\mathbf{P}\{\eta_1 = k_1, \ldots \eta_N = k_N\} = \frac{n!}{k_1! k_2! \cdots k_N!} \frac{1}{N^n},$$

где $k_1, k_2, \ldots k_N$ такие неотрицательные целые числа, что $k_1 + k_2 + \cdots + k_N = n$. В [3] собраны предельные теоремы для схемы размещения различимых частиц по различным ячейкам. Если $\xi_1, \xi_2, \ldots, \xi_N$ – независимые пуассоновские одинаково распределенные случайные величины, имеющие произвольный параметр β, то обобщенная схема размещения является схемой размещения различимых частиц по различным ячейкам [2].

Пусть r – неотрицательное целое число, K – натуральное число такое, что $0 < K \leqslant N$. Тогда случайная величина

$$\mu_r(n, K, N) = \sum_{i=1}^{K} I_{\{\eta_i = r\}}$$

является количеством ячеек среди первых K ячеек, которые содержат r частиц в обобщенной схеме размещения $\eta_1, \ldots, \eta_N$.

При $K = N$ случайную величину $\mu_r(n, K, N)$ обозначают через $\mu_r(n, N)$. Предельным теоремам для случайных величин $\mu_r(n, N)$ в различных схемах размещения посвящено большое количество исследований (см. цитируемые выше монографии). Пуассоновские и гауссовские предельные теоремы для случайных величин $\mu_r(n, K, N)$ в различных схемах размещения имеются в работах [6, 9, 11, 12].

Пусть $\eta_1, \ldots, \eta_N$ – схема размещения $2n$ различимых частиц по N различным ячейкам, $\eta'_1, \ldots, \eta'_N$ – схема $\eta_1, \ldots, \eta_N$ при условии: в каждой ячейке содержится четное количество частиц. То есть схема $\eta'_1, \ldots, \eta'_N$ имеет распределение

$$\mathbf{P}\{\eta'_1 = 2k_1, \ldots, \eta'_N = 2k_N\}$$
$$= \mathbf{P}\{\eta_1 = 2k_1, \ldots, \eta_N$$
$$= 2k_N \mid \eta_i \in A_2, 1 \leqslant i \leqslant N\},$$

где $k_1, k_2, \ldots, k_N$ такие неотрицательные целые числа, что $2k_1 + 2k_2 + \cdots + 2k_N = 2n$, $A_2 = \{2k : k = 0, 1, 2 \ldots\}$ – множество неотрицательных четных чисел.

Обозначим $ch(\beta) = \frac{e^{\beta} + e^{-\beta}}{2}$, $sh(\beta) = \frac{e^{\beta} - e^{-\beta}}{2}$, $th(\beta) = \frac{e^{\beta} - e^{-\beta}}{e^{\beta} + e^{-\beta}}$, $\beta \in \mathbf{R}$, гиперболический косинус, гиперболический синус, гиперболический тангенс соответственно.

Теорема 1. *Схема $\eta'_1, \dots \eta'_N$ является обобщенной схемой размещения $2n$ частиц по N ячейкам со случайными величинами $\xi'_1, \dots \xi'_N$, имеющими распределение*

$$\mathbf{P}\{\xi'_i = 2k\} = \frac{\beta^{2k}}{(2k)!ch(\beta)}, \quad k = 0, 1, 2 \dots,$$
$$1 \leqslant i \leqslant N,$$

при любом $\beta > 0$.

Теорема 1 доказана в [7]. В этой работе мы будем изучать случайную величину

$$\mu_0(2n, K, N) = \sum_{i=1}^{K} I_{\{\eta'_i = 0\}}$$

– количество пустых ячеек среди первых K ячеек в схеме $\eta'_1, \dots, \eta'_N$.

Мы будем рассматривать суммы случайных величин

$$S_{KN} = \sum_{i=1}^{K} \xi_i^{\{0\}} + \sum_{i=K+1}^{N} \xi_i^*, \quad S_N = \sum_{i=1}^{N} \xi_i^*,$$

где $\xi_1^{\{0\}}, \dots \xi_K^{\{0\}}, \xi_{K+1}^*, \dots \xi_N^*$ и $\xi_1^*, \dots \xi_N^*$ – последовательности независимых случайных величин, имеющих распределения

$$\mathbf{P}\{\xi_i^* = k\} = \frac{\beta^{2k}}{(2k)!ch(\beta)}, \quad k = 0, 1, 2 \dots,$$
$$1 \leqslant i \leqslant N,$$

$$\mathbf{P}\{\xi_i^{\{0\}} = k\} = \mathbf{P}\{\xi_i^* = k \mid \xi_i^* > 0\} = \frac{\mathbf{P}\{\xi_i^* = k\}}{1 - p_0},$$

$$k = 1, 2 \dots, \quad \mathbf{P}\{\xi_i^{\{0\}} = 0\} = 0, \quad 1 \leqslant i \leqslant K,$$

и

$$p_0 = p_0(\beta) = \mathbf{P}\{\xi_i^* = 0\} = (ch(\beta))^{-1}.$$

Заметим, что $\xi'_i(\beta) = 2\xi_i^*(\beta)$ и $\mathbf{E}\xi'_i(\beta) = 2\mathbf{E}\xi_i^*(\beta)$.

При изучении случайной величины $\mu_0(2n, K, N)$ мы будем использовать следующую лемму. Ее доказательство в идейном плане близко к доказательству леммы 1.2.1 из [2].

Лемма 1. *Справедливо равенство*

$$\mathbf{P}\{\mu_0(2n, K, N) = k\}$$

$$= C_K^k p_0^k (1 - p_0)^{K-k} \frac{\mathbf{P}\{S_{(K-k)(N-k)} = n\}}{\mathbf{P}\{S_N = n\}}, \quad (1)$$

$$k = 0, 1, 2, \dots K.$$

Доказательство. Обозначим множество $J_K = \{1, 2, \dots, K\}$, суммы случайных величин

$$S'_{KN} = \sum_{i=1}^{K} \xi_i^{\{00\}} + \sum_{i=K+1}^{N} \xi'_i, \quad S'_N = \sum_{i=1}^{N} \xi'_i,$$

где $\{\xi_1^{\{00\}}, \dots, \xi_K^{\{00\}}, \xi'_1, \dots, \xi'_N\}$ – семейство независимых случайных величин, случайные величины $\xi_i^{\{00\}}$, $1 \leqslant i \leqslant K$, имеют распределение

$$\mathbf{P}\{\xi_i^{\{00\}} = 2k\} = \frac{\beta^{2k}}{(2k)!ch(\beta)(1 - p_0)},$$

$$k = 1, 2 \dots, \quad \mathbf{P}\{\xi_i^{\{00\}} = 0\} = 0,$$

события

$$A_J = \left\{\xi'_i = 0, i \in J, \xi'_i > 0, i \in J_K \setminus J, S'_N = 2n\right\},$$

$$J \in \mathbf{J}_k,$$

где $\mathbf{J}_k$ – класс всех подмножеств множества J_K, состоящих из k элементов. Имеем

$$\mathbf{P}(\mu_0(2n, K, N) = k) = \frac{\sum_{J \subset J_K, |J| = k} \mathbf{P}\{A_J\}}{\mathbf{P}\{S'_N = 2n\}}$$

$$= \frac{\sum\limits_{J \in \mathbf{J}_k} \mathbf{P}\left\{\xi'_i = 0, i \in J, \xi'_i > 0, i \in J_K \setminus J\right\} \mathbf{P}\left\{S'_N = 2n \mid \xi'_i = 0, \ i \in J, \ \xi'_i > 0, \ i \in J_K \setminus J\right\}}{\mathbf{P}\{S'_N = 2n\}}$$

$$= C_K^k p_0^k (1 - p_0)^{K-k} \frac{\mathbf{P}\left\{\sum_{i=1}^{N} \xi'_i = 2n \mid \xi'_i = 0, \ i \in J, \ \xi'_i > 0, \ i \in J_K \setminus J\right\}}{\mathbf{P}\{S'_N = 2n\}}$$

$$= C_K^k p_0^k (1 - p_0)^{K-k} \frac{\mathbf{P}\{S'_{(K-k)(N-k)} = 2n\}}{\mathbf{P}\{S'_N = 2n\}} = C_K^k p_0^k (1 - p_0)^{K-k} \frac{\mathbf{P}\{S_{(K-k)(N-k)} = n\}}{\mathbf{P}\{S_N = n\}}.$$

□

Заметим, что

$$p_0 = 2e^{-\beta}\left(1+e^{-2\beta}\right)^{-1}.$$

Поэтому $p_0 = 2e^{-\beta}(1+o(1))$ при $\beta \to \infty$. Рассмотрим случайную величину $\xi^*(\beta)$, $\beta > 0$, с распределением

$$\mathbf{P}\{\xi^*(\beta) = k\} = \frac{\beta^{2k}}{(2k)!ch(\beta)}, \quad k = 0, 1, 2 \ldots.$$

Математическое ожидание, дисперсия, четвертый центрированный момент случайной величины $\xi^*(\beta)$ соответственно равны [7]:

$$\begin{aligned} m(\beta) &= \frac{\beta}{2}th(\beta) = \frac{\beta}{2}\left(1 - \frac{2e^{-2\beta}}{1+e^{-2\beta}}\right) \\ &= \frac{\beta}{2}\left(1 + O(e^{-2\beta})\right), \end{aligned} \tag{2}$$

$$\sigma^2(\beta) = \frac{\beta}{4}\left(1 + (4\beta - 2)\frac{e^{-2\beta}}{1+e^{-2\beta}} - \frac{4\beta e^{-4\beta}}{(1+e^{-2\beta})^2}\right)$$

$$= \frac{\beta}{4}\left(1 + O(\beta e^{-2\beta})\right), \tag{3}$$

$$\mathbf{E}(\xi^*(\beta) - \mathbf{E}\xi^*(\beta))^4 = \frac{3}{16}\beta(1+o(1)) \quad \text{при} \quad \beta \to \infty. \tag{4}$$

Следовательно, математическое ожидание и дисперсия случайной величины $S_N = \sum_{i=1}^N \xi_i^*$ соответственно равны:

$$m_N = m_N(\beta) = \frac{1}{2}N\beta th(\beta),$$

$$\sigma_N^2 = \sigma_N^2(\beta) = \frac{1}{4}N\beta\left(1 + O(\beta e^{-2\beta})\right).$$

Так как функция $m(\beta)$ непрерывна и возрастает, $m(\beta) \to 0$ при $\beta \to 0$, $m(\beta) \to \infty$ при $\beta \to \infty$, то уравнение

$$m(\beta) = \frac{n}{N} \tag{5}$$

имеет решение $\beta = \beta\left(\frac{n}{N}\right)$. Это решение единственно и положительно. Схема размещения различимых частиц по различным ячейкам с четным числом частиц в каждой ячейке не зависит от β, и мы будем считать, что β удовлетворяет (5).

Основным результатом работы является следующая теорема.

Теорема 2. *Пусть $n, K, N \to \infty$ и $\frac{n}{N} \to \infty$.*

1) Предположим, что $2Ke^{-\frac{2n}{N}} \to \lambda$, где $0 < \lambda < \infty$. Тогда

$$\mathbf{P}\{\mu_0(2n, K, N) = k\} = e^{-\lambda}\frac{\lambda^k}{k!}(1+o(1)), \quad k = 0, 1, 2 \ldots, \tag{6}$$

2) Предположим, что $Ke^{-\frac{2n}{N}} \to \infty$. Тогда

$$\mathbf{P}\{\mu_0(2n, K, N) = k\} = \frac{1}{\sqrt{2\pi Kp_0(1-p_0)}}e^{-\frac{(k-Kp_0)^2}{2Kp_0(1-p_0)}}(1+o(1)) \tag{7}$$

равномерно при $\left|\frac{k-Kp_0}{\sqrt{Kp_0(1-p_0)}}\right| < C$, где $C > 0$ произвольно.

Доказательство теоремы 2 основано на оценке числителя и знаменателя дроби, присутствующей в (1). То есть на новых локальных предельных теоремах (лемме 2 и лемме 6).

Будем обозначать через Φ функцию распределения гауссовской случайной величины, имеющей нулевое среднее и единичную дисперсию. Теорема 2 влечет

Следствие. *Пусть $n, K, N \to \infty$ так, что $\frac{n}{N} \to \infty$ и $Ke^{-\frac{2n}{N}} \to \infty$. Тогда*

$$\mathbf{P}\left\{\frac{\mu_0(2n, K, N) - Kp_0}{\sqrt{Kp_0\,(1-p_0)}} < t\right\} \to \Phi(t), \quad t \in \mathbf{R}.$$

Приложение к математической статистике

Величина $\frac{\mu_0(2n,K,N)}{K}$ является несмещенной оценкой вероятности p_0, и в условиях следствия она асимптотически нормальна. Используем это утверждение для изучения вероятностей ошибок первого и второго рода в тесте о вероятности пустой ячейки.

Мы наблюдаем количество частиц в первых K ячейках в схеме размещения $2n$ различимых частиц по N различным ячейкам с четным числом частиц в каждой ячейке, где число n неизвестно. Предположим, что среди первых K ячеек оказалось k пустых ячеек. Для проверки гипотезы $H_0 : p_0 = ch^{-1}(\beta_0)$ при альтернативной гипотезе $H_1 : p_0 = ch^{-1}(\beta_1)$ мы предлагаем следующий критерий. Пусть $0 < \alpha < 1$ и число u_α такое, что $\Phi(u_\alpha) = \alpha$. Обозначим

$$p_{00} = ch^{-1}(\beta_0), \quad p_{01} = ch^{-1}(\beta_1),$$

$$C_0 = Kp_{00} + u_\alpha\sqrt{Kp_{00}(1-p_{00})},$$

$$C_1 = Kp_{01} + u_\alpha\sqrt{Kp_{01}(1-p_{01})}.$$

Мы предполагаем, что $\beta_0 = \beta_0(n_0, N)$ – решение уравнения $m(\beta_0) = \frac{n_0}{N}$, $\beta_1 = \beta_1(n_1, N)$ –

решение уравнения $m(\beta_1) = \frac{n_1}{N}$ и $n_0 > n_1$. Тогда $\beta_0 > \beta_1$ и $C_0 < C_1$ для больших K. Гипотеза H_0 принимается, если $k < C_0$, гипотеза H_1 принимается, если $k > C_1$. Тогда вероятность ошибки первого рода

$$\alpha(n_0, K, N) = \mathbf{P}\left\{\frac{\mu_0(2n_0, K, N) - Kp_{00}}{\sqrt{Kp_{00}(1-p_{00})}} > u_\alpha\right\},$$

и по следствию при $n_0, K, N \to \infty$ таких, что $Ke^{-\frac{2n_0}{N}} \to \infty$, $\frac{n_0}{N} \to \infty$, имеем

$$\alpha(n_0, K, N) \to 1 - \alpha.$$

Вероятность ошибки второго рода

$$\beta(n_1, K, N) = \mathbf{P}\left\{\frac{\mu_0(2n_1, K, N) - Kp_{00}}{\sqrt{Kp_{00}(1-p_{00})}} < u_\alpha\right\}$$

$$= \mathbf{P}\left\{\frac{\mu_0(2n_1, K, N) - Kp_{01}}{\sqrt{Kp_{01}(1-p_{01})}} < \frac{\sqrt{K}(p_{00} - p_{01})}{\sqrt{p_{01}(1-p_{01})}} + u_\alpha\sqrt{\frac{p_{00}(1-p_{00})}{p_{01}(1-p_{01})}}\right\}$$

$$= \mathbf{P}\left\{\frac{\mu_0(2n_1, K, N) - Kp_{01}}{\sqrt{Kp_{01}(1-p_{01})}} < \frac{\sqrt{Kp_{01}}\left(\frac{p_{00}}{p_{01}} - 1\right)}{\sqrt{p_{01}(1-p_{01})}} + u_\alpha\sqrt{\frac{p_{00}(1-p_{00})}{p_{01}(1-p_{01})}}\right\}.$$

Так как

$$\frac{p_{00}}{p_{01}} = \exp\left(2\frac{n_1 - n_0}{N}\right)(1 + o(1))$$

$$\text{при} \quad \frac{n_1}{N} \to \infty,$$

то по следствию

$$\beta(n_1, K, N) \to 0$$

при $n_1, n_0, K, N \to \infty$ так, что $Ke^{-\frac{2n_1}{N}} \to \infty$, $\frac{n_1}{N} \to \infty$ и

$$\frac{n_0 - n_1}{N} > \delta$$

для некоторого $\delta > 0$. Заметим, что этот тест можно рассматривать как тест на проверку гипотезы $H_0 : n = n_0$ при альтернативной гипотезе $H_1 : n = n_1$.

Приложение к кодированию с контрольной суммой

Рассмотрим файл, состоящий из N блоков с информацией, закодированной с использованием контрольных сумм. Блоком является последовательность $i_1, i_2, \dots i_m$, где $i_j = 1$ или $i_j = 0$, $1 \leqslant j \leqslant m$, и $i_m = \left(\sum_{j=1}^{m-1} i_j\right) mod\, 2$, а m – фиксированное число. То есть каждый блок содержит четное количество единиц. Число i_m называется контрольной суммой. Предположим, что при передаче файла по каналу связи произошли ошибки: некоторые единицы заменились нулями и некоторые нули заменились единицами. Кодирование с контрольными суммами позволяет идентифицировать блоки с нечетным количеством ошибок: если в блоке $i_1, i_2, \dots i_m$ выполняется равенство $1 = \left(\sum_{j=1}^{m} i_j\right) mod\, 2$, то блок $i_1, i_2, \dots i_m$ имеет нечетное количество ошибок. Пусть случайные величины $\eta'_1, \dots \eta'_N$ – количества ошибок в блоках файла, полученного по каналу связи. Мы предполагаем, что выполнены следующие условия:

1) общее количество ошибок в файле равно $2n$ $(\eta'_1 + \dots + \eta'_N = 2n)$;

2) ошибки распределяются по блокам равновероятно и независимо;

3) метод контрольных сумм не обнаружил блоки с ошибками, т. е. все блоки содержат четное количество ошибок и случайные величины $\eta'_1, \dots \eta'_N$ принимают только четные значения.

Тогда случайные величины $\eta'_1, \dots \eta'_N$ являются схемой размещения $2n$ различимых частиц по N различным ячейкам при условии: в каждой ячейке содержится четное количество частиц, и мы можем применить рассмотренный выше критерий для проверки гипотезы о количестве ошибок в файле.

Вспомогательные леммы и доказательства

Лемма 5 из работы [10] влечет следующую лемму.

Лемма 2. *Пусть $C > 0$. Предположим, что $N \to \infty$ и $\beta \to \infty$. Тогда*

$$\sigma_N \mathbf{P}\{S_N = n\} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(n - m_N)^2}{2\sigma_N^2}\right)(1 + o(1))$$

равномерно по n таким, что $\frac{|n - m_N|}{\sigma_N} < C$.

Рассмотрим случайную величину $\xi^{\{0\}}(\beta)$, $\beta > 0$, с распределением

$$\mathbf{P}\{\xi^{\{0\}}(\beta) = k\} = \mathbf{P}\{\xi^*(\beta) = k \mid \xi^*(\beta) > 0\} = \frac{\beta^{2k}}{(1 - p_0)(2k)!ch(\beta)}, \quad k = 1, 2 \dots . \qquad (8)$$

Обозначим через $m_0 = m_0(\beta)$ и $\sigma_0^2 = \sigma_0^2(\beta)$ ее математическое ожидание и дисперсию.

Лемма 3. *Справедливы равенства*

$$m_0(\beta) = \frac{\beta}{2(1-p_0)} th(\beta) = \frac{\beta}{2}(1 + O(e^{-\beta})), \quad (9)$$

$$\sigma_0^2(\beta) = \frac{\beta}{4}\left(1 + O(\beta e^{-\beta})\right) \quad (10)$$

и

$$\mathbf{E}(\xi^{\{0\}}(\beta) - \mathbf{E}\xi^{\{0\}}(\beta))^4 = \frac{3}{16}\beta(1 + o(1)) \quad (11)$$

при $\beta \to \infty$.

Доказательство. Равенство (8) влечет

$$\mathbf{E}(\xi^{\{0\}}(\beta))^k = \frac{\mathbf{E}(\xi^*(\beta))^k}{1-p_0} \quad \text{и} \quad (12)$$
$$p_0 = 2e^{-\beta}(1 + O(e^{-2\beta})) \quad \text{при} \quad \beta \to \infty.$$

Следовательно, (2) влечет (9). Из (12), (2) и (3) следует

$$\sigma_0^2(\beta) = \frac{\mathbf{E}(\xi^*(\beta))^2}{1-p_0} - \left(\frac{\mathbf{E}\xi^*(\beta)}{1-p_0}\right)^2$$
$$= \sigma^2(\beta)(1 + O(e^{-\beta})) \quad \text{при} \quad \beta \to \infty.$$

Поэтому справедливо (10). Также, используя (12), получаем

$$\mathbf{E}(\xi^{\{0\}}(\beta) - \mathbf{E}\xi^{\{0\}}(\beta))^4$$
$$= \sum_{k=0}^{4} C_4^k \mathbf{E}(\xi^{\{0\}}(\beta))^k \left(-\frac{\mathbf{E}(\xi^*(\beta))}{1-p_0}\right)^{4-k}$$
$$= \left(\sum_{k=0}^{4} C_4^k \mathbf{E}(\xi^*(\beta))^k \left(-\mathbf{E}(\xi^*(\beta))\right)^{4-k}\right)$$
$$\times(1+O(e^{-\beta})) = \mathbf{E}(\xi^*(\beta) - \mathbf{E}\xi^*(\beta))^4(1+O(e^{-\beta}))$$

при $\beta \to \infty$. Итак, формула (11) вытекает из формулы (4). □

Следующая лемма получена в [8] (см. Theorem A).

Лемма 4. *Пусть* ξ_i' *– независимые случайные величины с дисперсиями* σ_i^2, $1 \leqslant i \leqslant N$, *и* $\sigma^2 = \sum_{i=1}^{N} \sigma_i^2$. *Тогда*

$$\sup_{t\in\mathbf{R}} \left| \mathbf{P}\left\{ \frac{1}{\sigma} \sum_{i=1}^{N} (\xi_i' - \mathbf{E}\xi_i') < t \right\} - \Phi(t) \right|$$
$$< 2c \left(\frac{\sum_{i=1}^{N} \mathbf{E}(\xi_i' - \mathbf{E}\xi_i')^4}{\sigma^4} \right)^{\frac{1}{2}},$$

где c – константа в неравенстве Берри–Эссеена.

Заметим, что

$$m_{KN} = m_{KN}(\beta) = K m_0(\beta) + (N-K) m(\beta)$$

и

$$\sigma_{KN}^2 = \sigma_{KN}^2(\beta) = K\sigma_0^2(\beta) + (N-K)\sigma^2(\beta)$$

– математическое ожидание и дисперсия суммы S_{KN}.

Используя лемму 3 и лемму 4, получаем следующую лемму.

Лемма 5. *Справедливо неравенство*

$$\sup_{t\in\mathbf{R}} \left| \mathbf{P}\left\{ \frac{S_{KN} - m_{KN}}{\sigma_{KN}} < t \right\} - \Phi(t) \right|$$
$$< 2c \left(\frac{3(1+o(1))}{N} \right)^{\frac{1}{2}} \quad (13)$$

и

$$\sigma_{KN}^2 = \frac{N\beta}{4}(1 + o(1)) \quad \text{при} \quad \beta \to \infty. \quad (14)$$

Доказательство. Равенство (14) вытекает из (3) и (10). Используя (14), (4) и (11) для оценки правой части неравенства из леммы 4, получаем (13). □

Обозначим:

$$\varphi(t) = \frac{ch(\beta e^{i\frac{t}{2}})}{ch(\beta)} = \frac{e^{\beta e^{i\frac{t}{2}}} + e^{-\beta e^{i\frac{t}{2}}}}{e^{\beta} + e^{-\beta}}$$

– характеристическая функция случайной величины $\xi^*(\beta)$,

$$\varphi_0(t) = \frac{e^{\beta e^{i\frac{t}{2}}} + e^{-\beta e^{i\frac{t}{2}}} - 2}{e^{\beta} + e^{-\beta} - 2} = \left(\frac{e^{\frac{\beta}{2} e^{i\frac{t}{2}}} - e^{-\frac{\beta}{2} e^{i\frac{t}{2}}}}{e^{\beta/2} - e^{-\beta/2}} \right)^2$$

– характеристическая функция случайной величины $\xi^{\{0\}}(\beta)$, $\varphi^c(t) = \varphi(t)e^{-itm(\beta)}$, $\varphi_0^c(t) = \varphi_0(t)e^{-itm_0(\beta)}$, $1 \leqslant i \leqslant N$, $\varphi_{KN}^c(t)$ – характеристическая функция случайной величины $S_{KN} - m_{KN}$ и $\varphi_{KN}(t)$ – характеристическая функция случайной величины $\frac{S_{KN} - m_{KN}}{\sigma_{KN}}$.

Лемма 6. *Пусть* $N, \beta \to \infty$. *Тогда*

$$\sigma_{KN}\mathbf{P}\{S_{KN} = n\}$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(n - m_{KN})^2}{2\sigma_{KN}^2}\right)(1 + o(1)) \quad (15)$$

равномерно по n таким, что $\frac{|n - m_{KN}|}{\sigma_{KN}} < C_1$, *где* $C_1 > 0$ *– произвольное фиксированное число.*

Доказательство. Обозначим

$$z = \frac{n - m_{KN}}{\sigma_{KN}}.$$

Так же, как в теореме Гнеденко, мы представляем вероятность $\mathbf{P}\{S_{KN} = n\}$ в виде интеграла

$$\mathbf{P}\{S_{KN} = n\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-it\sigma_{KN}z}\varphi^c_{KN}(t)dt.$$

После замены переменных $x = t\sigma_{KN}$ мы получаем

$$\sigma_{KN}\mathbf{P}\{S_{KN} = n\} = \frac{1}{2\pi}\int_{-\pi\sigma_{KN}}^{\pi\sigma_{KN}} e^{-ixz}\left(\varphi^c_0\left(\frac{x}{\sigma_{KN}}\right)\right)^K \times \left(\varphi^c\left(\frac{x}{\sigma_{KN}}\right)\right)^{N-K} dx.$$

Пусть $0 < \varepsilon < 1$, $B > 0$. Так как

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixz}e^{-\frac{x^2}{2}}dx,$$

мы представляем разность в виде суммы четырех интегралов:

$$R_N = 2\pi\left(\sigma_{KN}\mathbf{P}\{S_{KN} = n\} - \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\right) = I_1 + I_2 + I_3 + I_4, \qquad (16)$$

где

$$I_1 = \int_{|x|<B} e^{-ixz}\varphi_{KN}(x)dx - \int_{|x|<B} e^{-ixz}e^{-\frac{x^2}{2}}dx,$$

$$I_2 = -\int_{|x|>B} e^{-ixz}e^{-\frac{x^2}{2}}dx,$$

$$I_3 = \int_{B<|x|\leqslant\varepsilon\sigma_{KN}} e^{-ixz}\left(\varphi^c_0\left(\frac{x}{\sigma_{KN}}\right)\right)^K \times \left(\varphi^c\left(\frac{x}{\sigma_{KN}}\right)\right)^{N-K} dx$$

и

$$I_4 = \int_{\varepsilon\sigma_{KN}<|x|\leqslant\pi\sigma_{KN}} e^{-ixz}\left(\varphi^c_0\left(\frac{x}{\sigma_{KN}}\right)\right)^K \times \left(\varphi^c\left(\frac{x}{\sigma_{KN}}\right)\right)^{N-K} dx.$$

Так как

$$I_1 = \int_{|x|<B} e^{-ixz}\left(\varphi_{KN}(x) - e^{-\frac{x^2}{2}}\right)dx,$$

лемма 5 влечет

$$I_1 \to 0 \qquad (17)$$

для любого фиксированного $B > 0$. Так как

$$|I_2| < \int_{|x|>B} e^{-\frac{x^2}{2}}dx,$$

то

$$I_2 \to 0, \quad \text{при} \quad B \to \infty. \qquad (18)$$

Заметим, что

$$|I_3| \leqslant \int_{B<|t|\leqslant\varepsilon\sigma_{KN}} \left| e^{-itz}\left(\varphi^c_0\left(\frac{t}{\sigma_{KN}}\right)\right)^K \times \left(\varphi^c\left(\frac{t}{\sigma_{KN}}\right)\right)^{N-K}\right| dt$$

$$= \int_{B<|t|\leqslant\varepsilon\sigma_{KN}} \left| e^{\beta\left(e^{\frac{it}{2\sigma_{KN}}}-1\right)}\right|^N \times \left|\frac{1 - e^{-\beta e^{i\frac{t}{2\sigma_{KN}}}}}{1 - e^{-\beta}}\right|^{2K}\left|\frac{1 + e^{-2\beta e^{i\frac{t}{2\sigma_{KN}}}}}{1 + e^{-2\beta}}\right|^{N-K} dt. \qquad (19)$$

Так как

$$\cos(t) - 1 \leqslant -\frac{t^2}{2} + \frac{t^4}{24} \leqslant -\frac{11t^2}{24},$$

$$|t| \leqslant 1, \ \ e^t - 1 \leqslant te^t, \ \ t \geqslant 0,$$

то

$$\left| e^{\beta\left(e^{\frac{it}{2\sigma_{KN}}}-1\right)}\right|^N \leqslant e^{-N\beta\frac{11t^2}{24\cdot 4\sigma^2_{KN}}} = e^{-\frac{11t^2(1+o(1))}{24}}, \quad B < |t| \leqslant \varepsilon\sigma_{KN} \qquad (20)$$

при $\beta \to \infty$. Поскольку

$$1 - \cos(t) \leqslant \frac{t^2}{2}, \quad t \in \mathbf{R}, \quad 1 + t < e^t, \ \ t > 0,$$

$$\exp\left(-2\beta\cos\left(\frac{\varepsilon}{2}\right)\right) = o(1),$$

$$\beta\exp\left(-2\beta\cos\left(\frac{\varepsilon}{2}\right)\right) = o(1),$$

имеем

$$
\begin{aligned}
&\left|\frac{1+e^{-2\beta e^{i\frac{t}{2\sigma_{KN}}}}}{1+e^{-2\beta}}\right| \leqslant \left|\frac{1+e^{-2\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}}{1+e^{-2\beta}}\right| = \left|1+\frac{e^{-2\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}-e^{-2\beta}}{1+e^{-2\beta}}\right| \\
&= \left|1+e^{-2\beta}\frac{e^{2\beta\left(1-\cos\left(\frac{t}{2\sigma_{KN}}\right)\right)}-1}{1+e^{-2\beta}}\right| \\
&\leqslant \left|1+e^{-2\beta}\frac{2\beta\left(1-\cos\left(\frac{t}{2\sigma_{KN}}\right)\right)\exp\left(2\beta\left(1-\cos\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)}{1+e^{-2\beta}}\right| \\
&\leqslant \left|1+e^{-2\beta}\frac{2\beta\left(\frac{t^2}{8\sigma_{KN}^2}\right)\exp\left(2\beta\left(1-\cos\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)}{1+e^{-2\beta}}\right| \\
&\leqslant \left|1+\frac{2\beta\left(\frac{t^2}{8\sigma_{KN}^2}\right)\exp\left(2\beta\left(-\cos\left(\frac{\varepsilon}{2}\right)\right)\right)}{1+e^{-2\beta}}\right| \\
&\leqslant \exp\left(2\beta\left(\frac{t^2}{8\frac{N\beta}{4}(1+o(1))}\right)\exp\left(-2\beta\cos\left(\frac{\varepsilon}{2}\right)\right)\right) = \exp\left(\frac{t^2}{N}o(1)\right)
\end{aligned} \tag{21}
$$

и

$$
\begin{aligned}
&\left|\frac{1-e^{-\beta e^{i\frac{t}{2\sigma_N}}}}{1-e^{-\beta}}\right|^2 \\
&= \frac{\left|1-e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\cos\left(-\beta\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)-ie^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\sin\left(-\beta\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)\right|^2}{(1-e^{-\beta})^2} \\
&= \frac{\left(1-e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\cos\left(\beta\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)^2+\left(e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\sin\left(\beta\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)^2}{(1-e^{-\beta})^2} \\
&= \frac{1-2e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\cos\left(\beta\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)+e^{-2\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}}{(1-e^{-\beta})^2} \\
&= \frac{\left(1-e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\right)^2+2e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\left(1-\cos\left(\beta\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)}{(1-e^{-\beta})^2} \\
&\leqslant 1+\frac{2e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\left(2\sin^2\left(\beta\frac{1}{2}\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)}{(1-e^{-\beta})^2} \\
&\leqslant \exp\left(\frac{2e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\left(2\sin^2\left(\beta\frac{1}{2}\sin\left(\frac{t}{2\sigma_{KN}}\right)\right)\right)}{(1-e^{-\beta})^2}\right) \\
&\leqslant \exp\left(\frac{2e^{-\beta\cos\left(\frac{t}{2\sigma_{KN}}\right)}\left(2\left(\beta\frac{1}{2}\left(\frac{t}{2\sigma_{KN}}\right)\right)^2\right)}{(1-e^{-\beta})^2}\right) \leqslant \exp\left(\frac{1}{4}\frac{e^{-\beta\cos\left(\frac{\varepsilon}{2}\right)}\frac{\beta^2t^2(1+o(1))}{N\beta}}{(1-e^{-\beta})^2}\right) = \exp\left(\frac{t^2}{N}o(1)\right).
\end{aligned} \tag{22}
$$

Используя (20), (21) и (22) для оценки правой части неравенства (19), получаем

$$|I_3| \leqslant 2\int_B^\infty \exp\left(-\frac{11x^2(1+o(1))}{24}\right)dx. \quad (23)$$

Неравенство (23) влечет

$$|I_3| \to 0 \text{ при } B \to \infty. \quad (24)$$

Имеем

$$|\varphi(x)| = \left|\frac{e^{\beta e^{i\frac{x}{2}}} + e^{-\beta e^{i\frac{x}{2}}}}{e^\beta + e^{-\beta}}\right|$$

$$\leqslant \frac{e^{\beta\cos(\frac{x}{2})} + e^{-\beta\cos(\frac{x}{2})}}{e^\beta + e^{-\beta}} \leqslant \frac{e^{\beta\cos(\varepsilon/2)} + e^{-\beta\cos(\varepsilon/2)}}{e^\beta + e^{-\beta}}$$

$$= \frac{ch(\beta\cos(\varepsilon/2))}{ch(\beta)}$$

и

$$|\varphi_0(x)| = \left(\frac{\left|e^{\beta e^{i\frac{x}{2}}/2} - e^{-\beta e^{i\frac{x}{2}}/2}\right|}{e^{\beta/2} - e^{-\beta/2}}\right)^2$$

$$\leqslant \left(\frac{e^{\beta\cos(\frac{x}{2})/2} + e^{-\beta\cos(\frac{x}{2})/2}}{e^{\beta/2} - e^{-\beta/2}}\right)^2$$

$$\leqslant \left(\frac{e^{\beta\cos(\varepsilon/2)/2} + e^{-\beta\cos(\varepsilon/2)/2}}{e^{\beta/2} - e^{-\beta/2}}\right)^2$$

$$= \left(\frac{ch(\beta\cos(\varepsilon/2)/2)}{sh(\beta/2)}\right)^2,$$

$\varepsilon < |x| \leqslant \pi$. Следовательно,

$$|I_4| \leqslant 2\pi\sigma_{KN}\left(\frac{ch(\beta\cos(\varepsilon/2))}{ch(\beta)}\right)^{N-K}$$

$$\times\left(\frac{ch(\beta\cos(\varepsilon/2)/2)}{sh(\beta/2)}\right)^{2K}$$

$$\leqslant 2\pi\sigma_{KN}\left(e^{-(1-\cos(\varepsilon/2))\beta}\frac{1+e^{-2\beta\cos(\varepsilon/2)}}{1+e^{-2\beta}}\right)^{N-K}$$

$$\times\left(e^{-(1-\cos(\varepsilon/2))\beta/2}\frac{1+e^{-\beta\cos(\varepsilon/2)}}{1-e^{-\beta}}\right)^{2K}.$$

Так как

$$\frac{1+e^{-2\beta\cos(\varepsilon/2)}}{1+e^{-2\beta}} \to 1 \quad \text{и} \quad \frac{1+e^{-\beta\cos(\varepsilon/2)}}{1-e^{-\beta}} \to 1$$

при $\beta \to \infty$, существует $\beta_0 > 0$ такое, что

$$\frac{1+e^{-2\beta\cos(\varepsilon/2)}}{1+e^{-2\beta}} < e^{\frac{1}{2}(1-\cos(\varepsilon/2))\beta}$$

и

$$\left(\frac{1+e^{-\beta\cos(\varepsilon/2)}}{1-e^{-\beta}}\right)^2 < e^{\frac{1}{2}(1-\cos(\varepsilon/2))\beta}$$

при $\beta > \beta_0$. Тогда

$$|I_4| \leqslant 2\pi\sigma_{KN}e^{-\frac{1}{2}(1-\cos(\varepsilon/2))\beta N}$$

$$\leqslant 2\pi\sqrt{\beta N}e^{-\frac{1}{2}(1-\cos(\varepsilon/2))\beta N}(1+o(1))$$

при $\beta > \beta_0$. Поэтому

$$I_4 \to 0. \quad (25)$$

Используя (17), (18), (24) и (25) для оценки правой части равенства (16), получаем (15). □

Заметим, что при $K = N$ лемма 6 влечет лемму 2.

Доказательство теоремы 2. Условия (2), (5) влекут $\frac{\beta}{2} > \frac{n}{N}$, т. е. $\beta > \frac{2n}{N}$. Поэтому $\beta \to \infty$ и

$$\sigma_N^2 = \frac{1}{2}n(1+o(1)),$$

$$\sigma^2_{(K-k)(N-k)} = \frac{1}{2}\frac{N-k}{N}n(1+o(1)),$$

$$\beta = \frac{2n}{N}(1+o(1)),$$

$$p_0 = (ch(\beta))^{-1} = 2e^{-\frac{2n}{N}}(1+o(1)).$$

Следовательно,

$$\frac{(n-m_N)^2}{\sigma_N^2} = 0, \quad (26)$$

$$\frac{\sigma_N}{\sigma_{(N-k)(K-k)}} = \sqrt{\frac{N\beta(1+o(1))}{(N-k)\beta}} = \sqrt{\frac{N(1+o(1))}{N-k}} \quad (27)$$

и

$$\frac{(n-m_{(K-k)(N-k)})^2}{\sigma^2_{(K-k)(N-k)}} = 2\frac{N-k}{N} \times \frac{\left(n - \frac{(K-k)}{1-p_0}\frac{n}{N} - (N-K)\frac{n}{N}\right)^2}{n(1+o(1))}$$

$$= 2\frac{N-k}{N}\frac{\left(K\frac{n}{N} - \frac{(K-k)n}{N(1-p_0)}\right)^2}{n(1+o(1))} \quad (28)$$

$$= 2\frac{N-k}{N}\frac{n}{N^2}\left(K - \frac{K-k}{1-p_0}\right)^2(1+o(1))$$

$$= 2\frac{N-k}{N}\frac{n}{N^2}\left(\frac{k-Kp_0}{1-p_0}\right)^2(1+o(1)).$$

Пусть выполнены условия п. 1, число k фиксировано. Тогда

$$\ln(\lambda + o(1)) = \ln 2 + \ln K - 2\frac{n}{N}$$

и

$$2\frac{n}{N^2} = -\frac{\ln(\lambda + o(1))}{N} + \frac{\ln 2}{N} + \frac{\ln K}{N}.$$

Поэтому

$$\frac{n}{N^2} \to 0$$

и из (28) вытекает

$$\frac{(n - m_{(K-k)(N-k)})^2}{\sigma^2_{(K-k)(N-k)}} = o(1). \tag{29}$$

Кроме того,

$$Kp_0 = 2Ke^{-\frac{2n}{N}} + o(1) = \lambda + o(1).$$

Используя лемму 2, лемму 6 и оценку (26), (27), (29), заключаем

$$\frac{\mathbf{P}\{S_{(K-k)(N-k)} = n\}}{\mathbf{P}\{S_N = n\}} = \frac{\sigma_N}{\sigma_{(K-k)(N-k)}}$$

$$\times \frac{\exp\left(-\frac{(n-m_{(K-k)(N-k)})^2}{2\sigma^2_{(K-k)(N-k)}}\right)(1+o(1))}{\exp\left(-\frac{(n-m_N)^2}{2\sigma^2_N}\right)(1+o(1))} \tag{30}$$

$$= 1 + o(1).$$

Используя (30) для оценки дроби в (1) и пуассоновскую предельную теорему, получаем (6). П. 1 доказан.

Пусть выполнены условия п. 2. Так как $p_0 \to 0$, имеем

$$\frac{\sigma^2_N}{\sigma^2_{(K-k)(N-k)}(\beta)} = \frac{N}{N-k}(1+o(1))$$

$$= \left(1 - \frac{k}{N}\right)^{-1}(1+o(1))$$

$$= \left(1 - \frac{k - Kp_0}{N} + \frac{Kp_0}{N}\right)^{-1}(1+o(1)) \tag{31}$$

$$= \left(1 - \frac{k - Kp_0}{\sqrt{Kp_0(1-p_0)}}\frac{\sqrt{Kp_0(1-p_0)}}{N} + \frac{Kp_0}{N}\right)^{-1}$$

$$\times(1+o(1)) = 1 + o(1).$$

Так как $\beta > 2\frac{n}{N}$, то $p_0 < 2e^{-2\frac{n}{N}}$. Используя условия п. 2 для оценки сомножителей в (28), заключаем

$$\frac{(n - m_{(K-k)(N-k)})^2}{\sigma^2_{(K-k)(N-k)}}$$

$$= 2\frac{N-k}{N}\frac{nKp_0}{N^2(1-p_0)}\frac{(k - Kp_0)^2}{Kp_0(1-p_0)}(1+o(1))$$

$$< \frac{2C^2}{1-p_0}\frac{n}{N}e^{-2\frac{n}{N}}(1+o(1)).$$

Следовательно,

$$\frac{(n - m_{(K-k)(N-k)})^2}{\sigma^2_{(K-k)(N-k)}} = o(1). \tag{32}$$

Используя лемму 2, лемму 6 и оценки (26), (31), (32), получаем

$$\frac{\mathbf{P}\{S_{(K-k)(N-k)} = n\}}{\mathbf{P}\{S_N = n\}} = \frac{\sigma_N}{\sigma_{(K-k)(N-k)}}$$

$$\times \frac{\exp\left(-\frac{(n-m_{(K-k)(N-k)})^2}{2\sigma^2_{(K-k)(N-k)}}\right)(1+o(1))}{\exp\left(-\frac{(n-m_N)^2}{2\sigma^2_N}\right)(1+o(1))} = 1 + o(1). \tag{33}$$

Оценка (33) дроби в (1) и локальная теорема Муавра–Лапласа влекут (7). □

Авторы выражают глубокую благодарность рецензенту за полезные замечания.

Литература

1. *Колчин В. Ф.* Один класс предельных теорем для условных распределений // Литовский математический сборник. 1968. Т. 8, вып. 1. С. 53–63. doi: 10.15388/LMJ.1968.20181

2. *Колчин В. Ф.* Случайные графы. М.: Физматлит, 2000. 256 с.

3. *Колчин В. Ф., Севастьянов Б. А., Чистяков В. П.* Случайные размещения. М.: Физматлит, 1976. 224 с.

4. *Павлов Ю. Л.* Случайные леса. Петрозаводск: Карел. науч. центр РАН, 1996. 259 с.

5. *Тимашев А. Н.* Асимптотические разложения в вероятностной комбинаторике. М.: ТВП; Ред. «ОПиПМ», 2011. 256 с.

6. *Хакимуллин Е. Р., Энатская Н. Ю.* Предельные теоремы для числа пустых ячеек // Дискретная математика. 1997. Т. 9, вып. 2. С. 120–130. doi: 10.4213/dm468

7. *Abdushukurov F. A., Chuprunov A. N.* On the number of empty cells in a non-homogenious allocation scheme // Lobachevskii Journal of Mathematics. 2021. Vol. 42, iss. 2. P. 269–279. doi: 10.1134/S1995080221020025

8. *Chickrin D. E., Chuprunov A. N., Kokunin P. A.* Gaussian limit theorems for the number of given value cells in the non-homogeneous generalized allocation scheme // Journal of Mathematical Sciences. 2020. Vol. 246, iss. 4. P. 476–487. doi: 10.1007/s10958-020-04753-w

9. *Chickrin D. E., Chuprunov A. N., Kokunin P. A.* Limit theorems for a number of particles from a fixed set of cells // Lobachevskii Journal of Mathematics. 2019. Vol. 40, iss. 5. P. 624–629. doi: 10.1134/S1995080219050044

10. *Chuprunov A. N., Fazekas I.* On numbers of particles in cells in an allocation scheme having even number of particles in each cell // MDPI, Mathematics. 2022. Vol. 10, iss. 7. P. 1–22. doi: 10.3390/math10071099

11. *Chuprunov A. N., Fazekas I.* On the number of empty cells in the allocation scheme of indistinguishable particles // Annales Universitatis Mariae Curie-Sklodowska. 2020. Vol. 74, iss. 1. P. 15–29. doi: 10.17951/a.2020.74.1.15-29

12. *Chuprunov A. N., Fazekas I.* Poisson limit theorems for the generalized allocation scheme // Ann. Univ. Sci. Budapest, Sect. Comp. 2019. Vol. 49. P. 77–96.

References

1. *Kolchin V. F.* A certain class of limit theorems for conditional distributions. *Litovskii matematicheskii sbornik = Lithuanian Mathematical Journal.* 1968;8(1):53–63. (In Russ.). doi: 10.15388/LMJ.1968.20181

2. *Kolchin V. F.* Random Graphs. Cambridge: Cambridge University Press; 1999. 252 p. doi: 10.1017/CBO9780511721342

3. *Kolchin V. F., Sevastianov B. A., Chistiakov V. P.* Random allocations. Washington: V. H. Winston; 1978. 262 p.

4. *Pavlov Yu. L.* Random Forests. Utrecht: VSP; 2000. 122 p. doi: 10.1515/9783110941975

5. *Timashev A. N.* Asymptotic expansions in probabilistic combinatorics. Moscow: TVP; OPiPM; 2011. 256 p. (In Russ.)

6. *Khakimullin E. R., Enatskaya N. Yu.* Limit theorems for the number of empty cells. *Discrete Math. Appl.* 1997;7(2):209–219. doi: 10.4213/dm468

7. *Abdushukurov F. A., Chuprunov A. N.* On the number of empty cells in a non-homogenious allocation scheme. *Lobachevskii Journal of Mathematics.* 2021;42(2):269–279. doi: 10.1134/S1995080221020025

8. *Chickrin D. E., Chuprunov A. N., Kokunin P. A.* Gaussian limit theorems for the number of given value cells in the non-homogeneous generalized allocation scheme. *J. Math. Sci.* 2020;246(4):476–487. doi: 10.1007/s10958-020-04753-w

9. *Chickrin D. E., Chuprunov A. N., Kokunin P. A.* Limit theorems for a number of particles from a fixed set of cells. *Lobachevskii Journal of Mathematics.* 2019;40(5):624–629. doi: 10.1134/S1995080219050044

10. *Chuprunov A. N., Fazekas I.* On numbers of particles in cells in an allocation scheme having even number of particles in each cell. *MDPI, Mathematics.* 2022;10(7):1–22. doi: 10.3390/math10071099

11. *Chuprunov A. N., Fazekas I.* On the number of empty cells in the allocation scheme of indistinguishable particles. *Annales Universitatis Mariae Curie-Sklodowska.* 2020;74(1):15–29. doi: 10.17951/a.2020.74.1.15-29

12. *Chuprunov A. N., Fazekas I.* Poisson limit theorems for the generalized allocation scheme. *Ann. Univ. Sci. Budapest, Sect. Comp.* 2019;49:77–96.

Поступила в редакцию / received: 28.12.2024; принята к публикации / accepted: 12.05.2025.
Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРАХ:

Чупрунов Алексей Николаевич
д-р физ.-мат. наук, профессор

e-mail: achuprunov@mail.ru

Яковлев Кирилл Николаевич
студент

e-mail: ktq0x@yandex.ru

CONTRIBUTORS:

Chuprunov, Alexey
Dr. Sci. (Phys.-Math.), Professor

Iakovlev, Kirill
Student

Труды Карельского научного центра РАН. 2025. № 4. С. 89–96
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 89–96
DOI: 10.17076/mat2047

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.115:519.2

КОМБИНАТОРНЫЙ АНАЛИЗ ИСХОДОВ СХЕМЫ В ОБРАТНОЙ ЗАДАЧЕ РАЗМЕЩЕНИЯ ЧАСТИЦ ПО ЯЧЕЙКАМ

Н. Ю. Энатская

Московский институт электроники и математики,
Национальный исследовательский университет «Высшая школа экономики»
(ул. Таллинская, 34, Москва, Россия, 123458)

Рассматривается схема C размещения $r = r(k)$ неразличимых частиц по n различимым ячейкам до момента наступления события A_k, когда впервые оказывается k $(k \leqslant r)$ непустых ячеек. Доасимптотический анализ схемы проводится авторским перечислительным методом (ПМ) по следующим направлениям: бесповторное перечисление и определение числа ее исходов, решение задачи нумерации, состоящей в установлении взаимно-однозначного соответствия между номерами и видами исходов схемы, определение вероятностного распределения на множестве ее исходов, и предлагается процедура их моделирования. Для всех других парных качеств по их различимостям составляющих схему элементов (ячеек и частиц) предлагается методика пересчета начальных результатов рассматриваемой здесь схемы C по перечислению их исходов, дающих возможность проведения для них алгоритмических исследований остальных направлений по ПМ.

К л ю ч е в ы е с л о в а: обратная задача размещения; перечислительный метод; задача нумерации; моделирование

Д л я ц и т и р о в а н и я: Энатская Н. Ю. Комбинаторный анализ исходов схемы в обратной задаче размещения частиц по ячейкам // Труды Карельского научного центра РАН. 2025. № 4. С. 89–96. doi: 10.17076/mat2047

N. Yu. Enatskaya. COMBINATORIAL ANALYSIS OF THE OUTCOMES OF THE SCHEME IN THE INVERSE PROBLEM OF ALLOCATING PARTICLES TO CELLS

National Research University Higher School of Economics, Moscow Institute of Electronics and Mathematics (34 Tallinskaya St., 123458 Moscow, Russia)

We consider a scheme C of allocating $r = r(k)$ indistinguishable particles to n distinguishable cells until the occurrence of the event A_k, when for the first time there are k $(k \leqslant r)$ non-empty cells. The pre-asymptotic analysis of the scheme is carried out by an original enumerative method (EM) along the following lines: non-repetitive enumeration and determination of the number of its outcomes, solutions to the numbering problem offinding one-to-one correspondence between the numbers and types of the scheme's outcomes, determining the probability distribution on the set of its outcomes, and a procedure for their modeling is proposed. For all other paired qualities according to distinguishability of the elements (cells and particles) that make up the scheme, a method is proposed for recalculating the initial results of the C scheme.

Keywords: inverse allocation problem; enumeration method; numbering problem; modeling

For citation: Enatskaya N. Yu. Combinatorial analysis of the outcomes of the scheme in the inverse problem of allocating particles to cells. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 89–96. doi: 10.17076/mat2047

Введение

Обратная задача о размещении частиц по ячейкам состоит в определении времени ожидания, т. е. числа r размещенных частиц, до момента выполнения определенного условия, связанного с заполнением ячеек.

В литературе широко представлены асимптотические исследования различных схем размещения частиц по ячейкам при неограниченном росте числа частиц в прямой и обратной постановках (см., например, [1–4]).

В представляемой здесь схеме размещения r частиц по ячейкам (без ограничения уровней их заполнения) анализируются все ее возможные исходы, приводящие к событию A_k, когда *впервые* оказывается заданное число k непустых ячеек.

Эта схема размещения r частиц по n ячейкам рассматривается во всех вариантах их парных качеств по различимости с обозначениями:

схема A – частицы и ячейки различимы;

схема B – частицы различимы, ячейки неразличимы;

схема C – частицы неразличимы, ячейки различимы;

схема D – частицы неразличимы, ячейки неразличимы.

Анализ схем проводим начиная с условий схемы C с приведением методики дальнейшего пересчета результатов для остальных схем.

Постановка задачи – изучить свойства исходов схемы C по всем направлениям перечислительного метода (ПМ) с пересчетом результатов для исходов других указанных выше схем: бесповторного их перечисления, определения их числа, решения задачи нумерации, определения их вероятностного распределения и построения процедуры их моделирования.

Основные обозначения, определения и понятия

ПМ – перечислительный метод исследования моделей по приведенным в аннотации направлениям;

МГ – представление случайного процесса перечисления исходов схемы в виде вероятностного графа;

итерация – шаг перехода в графе перечисления исходов схемы к следующему этапу (шагу) их перечисления;

траектория исхода – последовательность исходов итераций, ведущая к нему в графе от начала их перечисления;

бесповторное перечисление исходов схемы – единственность траекторий для каждого ее исхода;

пучок в графе процесса – совокупность переходов из каждого исхода каждой итерации;

размер пучка на каждой итерации – число переходов из каждого исхода итерации, т. е. число исходов из него на следующей итерации;

пучковая структура графа – перечисление размеров пучков всех итераций;

ЗН – задача нумерации – установление взаимно-однозначного соответствия между всеми номерами и видами исходов схемы;

ПЗН – прямая задача нумерации – нахождение вида исхода схемы по его номеру;

ОЗН – обратная задача нумерации – нахождение номера исхода схемы по его виду;

УМ – универсальный метод моделирования исходов схемы с их вероятностным распределением;

уровень заполнения ячейки – число частиц в ней.

1. Вспомогательные результаты

Приведем кратко результаты анализа основных используемых здесь схем в принятых для них обозначениях в указанных публикациях.

1.1. Перечислительный метод

В основе доасимптотического анализа рассматриваемых схем лежит ПМ (см. [5]). ПМ расширяет направления исследования комбинаторных схем и предлагает приемы их реализации. Его суть состоит в организации получения *качественной* информации об исходах схемы и переводе ее в *количественную* – результатов ее анализа в доасимптотической области изменения параметров. Эта качественная информация представляет собой исходы итерационного случайного процесса реализации комбинаторной схемы путем последовательного поединичного добавления элементов схемы до заданного значения или этапов перечисления составляющих схему более простых ранее изученных схем.

Инструментами перевода качественной информации о видах всех исходов схемы являются метод графов (МГ) (см. [5, п. 1.1]), состоящий в графическом представлении процедуры итерационного процесса перечисления исходов схемы, задача нумерации (ЗН) (см. [5, п. 1.1]), устанавливающая взаимно-однозначное соответствие между их номерами и видами исходов (соответственно ПЗН – прямая и ОЗН – обратная задача нумерации), и универсальное моделирование (УМ) (см. [5, п. 1.1 и Приложение А]) исходов, дающее его единый АЛГОРИТМ 1, состоящее в разыгрывании номеров исходов, виды которых определяются по решению задачи нумерации, учитывающей специфику схемы.

Одним из важных приемов ПМ для получения новых результатов доасимптотического анализа комбинаторных схем является операция по перечислению (ОПП), введение которой дает возможность вычисления характеристик схем, зависящих от видов предсостояний исходов при их формировании итерационным случайным процессом как функций от них по перечислению исходов итерации получения этих предсостояний.

Целью применения ПМ является изучение схемы по указанным в аннотации направлениям.

1.2. Схема последовательных действий (ПД) (с числом действий k)

Схема ПД с результатом анализа в [5, п. 1.8] (с обозначением в этом пункте через k числа действий) возникает, когда каждому следующему действию подвергаются исходы предыдущего действия и числа исходов на каждом следующем шаге (действии) одинаковы, т. е. зависят только от характера действия. Если i-е действие ($i = \overline{1,k}$) совершается d_i числом способов, то число исходов схемы $N = \prod_{i=1}^{k} d_i$.

Вид исхода после совершения i действий будет формироваться из принятых видов исходов последовательных действий, которые соответственно обозначаются через R_{ij_i}, где i – номер действия, j_i – номер исхода в результате его совершения, а конкретный вид R_{ij_i} определяется характером действия. Исход совершения r действий ($r \leqslant k$) обозначен в виде $R^{(r)} = \{R_{1j_1}, R_{2,j_2}, \ldots, R_{rj_r}\}$. Тогда окончательный исход схемы получен при $r = k$.

Теорема 1. Прямая ЗН. *Пусть в схеме с параметрами $d_1, \ldots, d_k$ дан номер $N^{(k)}$ ее исхода. Тогда вид исхода $R^{(k)} = \{R_{1j_1}, \ldots, R_{kj_k}\}$, определяемый номерами $(j_1, \ldots j_k)$ исходов его компонент в пучках графа перечисления исходов схемы, при $i = \overline{1,k}$ находится по формуле*

$$j_i = t_i + I(t_i)d_i,$$

где $t_i = N^{(i)} mod\ d_i$; $I(Z) = 0$ при $Z \neq 0$ и $I(Z) = 1$ при $Z = 0$;

$$N^{(i-1)} = \left[\frac{N^{(i)} + d_i - 1}{d_i}\right],$$

где $[Z]$ – целая часть числа Z и $i = k, k-1, \ldots, 1$; $N^{(0)} = 1$.

Теорема 2. Обратная ЗН. *Пусть в схеме с параметрами $d_1, \ldots, d_k$ дан вид ее исхода $R^{(k)} = \{R_{1j_1}, \ldots, R_{kj_k}\}$, определяющий номера $(j_1, \ldots j_k)$ исходов его компонент в пучках графа перечисления исходов схемы, при $i = \overline{1,k}$. Тогда его номер вычисляется по формуле*

$$N^{(k)} = \sum_{l=1}^{k-1} (j_l - 1) \prod_{i=l+1}^{k} d_i + j_k.$$

Доказательства теорем 1 и 2 приведены в [5, п. 1.8].

1.3. Обобщенная схема последовательных действий (ОПД) (с числом действий k)

Эта схема является основной структурой, дающей общее описание процесса формирования и перечисления исходов комбинаторных схем. Исследование свойств этих исходов создает основу доасимптотического анализа комбинаторных схем по направлениям ПМ, где действиями представлены последовательные итерации случайного процесса их формирования. (Схема ПД является частным случаем схемы ОПД.)

В [5, п. 1.9] приводятся результаты комбинаторного анализа схемы ОПД (с обозначением в этом пункте через k числа действий).

Схема ОПД с результатом анализа в [5, п. 1.9] возникает, когда каждому следующему действию подвергаются исходы предыдущего действия и числа исходов на каждом следующем шаге (действии) неодинаковы, т. е. зависят от характера действия и вида предыдущего исхода. Результатом этого являются разные размеры пучков внутри итерации в графе перечисления исходов схемы при переходе от исходов предыдущего действия к последующему, т. е. числа исходов из каждого состояния на следующей итерации.

Анализ схемы ОПД приводит к конкретным результатам только по результатам подобных исследований комбинаторных схем действий.

В схеме проводится k последовательных действий, i-е из которых $(i = \overline{1,k})$ на i-м шаге совершается $N^{(i)}$ способами. Тогда число исходов этих k действий складывается из $N^{(k-1)}$ пучков размерами $\bar{d}^{(k)} = (d_1^{(k)}, d_2^{(k)}, \ldots, d_{N^{(k-1)}}^{(k)})$, т. е. общее число $N^{(k)}$ исходов схемы получается из рекуррентного соотношения при $i = k$ и $N^{(0)} = 1$

$$N^{(i)} = \sum_{l=1}^{N^{(i-1)}} d_l^{(i)}.$$

Вид исхода после совершения i действий будет формироваться из принятых видов исходов последовательных действий, которые будем соответственно обозначать через R_{ij_i}, где i – номер действия, а j_i – номер исхода в результате его совершения.

Задача нумерации решается для нашей схемы при решенной ЗН для каждого из k действий и известной пучковой структуре графа перечисления исходов нашей схемы, т. е. с известными числами исходов (размерами пучков) при каждом действии на каждой итерации. Под траекторией T исхода схемы понимается последовательность исходов, ведущая в графе их перечисления от начала к исследуемому на последней итерации.

Прямая и обратная задачи нумерации решены следующими теоремами.

Теорема 3. *Пусть совершается k действий и задан номер исхода $N_*^{(k)}$. Тогда его вид, определяемый номерами исходов траектории T в содержащих их пучках $\{j_i\}$ от первой до k-й итераций, вычисляется по рекуррентной формуле для j_i $(i = \overline{1,k})$*

$$j_i = N_*^{(i)} - \sum_{l=1}^{N_*^{(i-1)}-1} d_l^{(i)},$$

где все пучковые структуры действий $\bar{d}^{(i)}$ заданы и

$$N_*^{(k-1)} = \delta + \max\ t : \left(\sum_{l=1}^{t} d_l^{(k)} = G_k \leqslant N_*^{(k)} \right),$$

где $\delta = 0$ при $G_k = N_*^{(k)}$ и $\delta = 1$ при $G_k < N_*^{(k)}$. Заменяя k на i, доходим по рекурренте до первого шага.

По решенной ЗН для всех действий находим из $\{j_i\}$ виды их исходов, из которых получаем искомый вид исхода $R_*^{(k)}$.

Теорема 4. *Пусть совершается k действий и задан вид исхода $R_*^{(k)}$, где $\{j_1, \ldots, j_k\}$ известные (из решенной ЗН для действий) номера исходов ПД, реализующих последовательные состояния траектории, ведущей к изучаемому исходу. Тогда его номер $N_*^{(k)}$ среди итоговых исходов схемы определяется по рекуррентной формуле при $i = k$, $\quad i = \overline{1,k}$:*

$$N_*^{(i)} = \sum_{l=1}^{N_*^{(i-1)}-1} d_l^{(i)} + j_i,$$

начиная с $i = 1$ при $N_*^{(1)} = j_1$.

Доказательства теорем 3 и 4 приведены в [5, п. 1.9].

2. Анализ схемы C

2.1. Перечисление и число $N(C)$ исходов схемы C

В схеме размещения неразличимых частиц по различимым ячейкам все исходы имеют вид наборов составов уровней заполнения ячеек в порядке ячеек. Соответственно, в рассматриваемой схеме C при появлении *впервые* ровно k непустых ячеек после размещения r-й частицы исходы представляют определенную часть исходов общей схемы и имеют тот же вид наборов составов уровней заполнения ячеек в порядке ячеек $\bar{h} = (h_1, \ldots, h_n)$, но со следующей очевидной спецификой в составах уровней заполнения ячеек для события A_k:

а) наличия ровно k ненулевых уровней заполнения в n ячейках;

б) присутствия хотя бы одной ячейки с единичным уровнем заполнения, любую из которых (из-за неразличимости частиц) можно считать результатом размещения k-й частицы.

В связи с этой спецификой предлагается следующий АЛГОРИТМ бесповторного прямого перечисления исходов схемы C двумя этапами:

1) выбора всеми C_n^k способами k непустых ячеек из n ячеек в порядке нумерации исходов схемы сочетаний по МГ (см. [5, п. 1.1.1.1]);

2) при каждом результате этапа 1 нахождения всех исходов схемы как разности множеств исходов размещения r частиц по k ячейкам без пустых ячеек и с не менее чем двумя частицами в каждой ячейке при нулевых уровнях заполнения остальных ячеек.

Первый этап, очевидно, следует из смысла схемы C, а второй – из специфики ее исходов. Уменьшаемое в разности дает в k фиксированных ячейках все составы ненулевых уровней заполнения, а вычитаемое – исключает из них исходы, не содержащие единичных уровней.

Нумерация исходов схемы производится в порядке их формирования в схеме ПД – последовательной реализации двух представленных выше этапов их перечисления.

Пучковая структура графа перечисления исходов схемы C представляет собой последовательности размеров пучков этапов перечисления: одним пучком размера (C_n^k) первого этапа и (C_n^k) пучков второго этапа одинакового размера, не зависящего от конкретного исхода первого этапа (выбора k непустых ячеек из n), размерами каждый по $N(C)/C_n^k$, где $N(C)$ – число исходов схемы C.

Теорема 5. *Число $N(C)$ исходов схемы C определяется по формуле*

$$N(C) = C_n^k (C_{r-1}^{k-1} - C_{r-k-1}^{k-1}). \qquad (1)$$

Доказательство. Число N исходов схемы будем находить в соответствии с процедурой перечисления исходов схемы, где сомножители в (1) являются числами исходов двух описанных выше этапов этой процедуры. Первый сомножитель не требует пояснения, а во втором сомножителе (при каждом фиксированном наборе k непустых ячеек) объясним виды численностей в уменьшаемом и вычитаемом. В уменьшаемом – числе размещений без пустых ячеек r частиц по k ячейкам в схеме сочетаний с повторением принудительно помещаем в каждую из k ячеек по одной частице, а остальные $r-k$ частиц размещаем по k ячейкам по схеме сочетаний с повторением числом способов $C_{k+r-k-1}^{r-k} = C_{r-1}^{k-1}$. А в вычитаемом – числе размещений с не менее чем двумя частицами в каждой ячейке – принудительно помещаем в каждую из k ячеек по две частицы, а остальные $r-2k$ частиц размещаем по k по схеме сочетаний с повторением числом способов $C_{k+r-2k-1}^{r-2k} = C_{r-k-1}^{k-1}$. В связи с тем, что доказательство вида второго сомножителя проведено при любой фиксации k непустых ячеек, отсюда следует формула (1). $\square$

Замечание. Для вычисления числа $N = N(C)$ исходов схемы C при известных для k непустых ячеек составах их уровней заполнения – множества $\{D\}$ (как наиболее просто визуально перечисляемых исходов в схеме D с их маркировками по совпадению этих уровней $\bar{\mu} = (\mu_1, \ldots, \mu_k)$) (с использованием известного числа исходов схемы перестановок с повторением) получаем формулу для числа $N = N(C)$ исходов схемы C:

$$N = N(C) = C_n^k \sum_{(\{D\})} k!/\prod_{j=1}^{k} \mu_j!,$$

где суммирование проводится в ПМ по ОПП (по операции сложения по перечислению) маркировок уровней заполнения ячеек по совпадению уровней во всех исходах $\{D\}$ схемы D.

Приведем числовой пример перечисления и нахождения чисел исходов схемы.

Пример 1. Пусть $n = 5$, $k = 3$, $r = 6$. По АЛГОРИТМУ перечислить исходы схемы C по шагам:

1) этот шаг совпадает с первым этапом перечисления исходов и по [5, п. 1.3] дает следующие $C_5^3 = 10$ исходов схемы сочетаний в виде номеров $k = 3$ непустых ячеек:
$(1, 2, 3)$, $(1, 2, 4)$, $(1, 2, 5)$, $(1, 3, 4)$, $(1, 3, 5)$, $(1, 4, 5)$, $(2, 3, 4)$, $(2, 3, 5)$, $(2, 4, 5)$, $(3, 4, 5)$;

2) найдем уменьшаемое второго этапа по схеме сочетаний с повторением без пустых ячеек, числом способов $C_{r-1}^{k-1} = C_5^2 = 10$ как места $(k-1) = 2$ внутренней перегородки между $k = 3$ ячейками среди всех мест для них и $r = 6$ лежащими в ряд частицами:
$(2, 4), (2, 5), (2, 6), (2, 7), (3, 5), (3, 6), (3, 7), (4, 6), (4, 7), (5, 7)$;

3) найдем вычитаемое второго этапа по схеме сочетаний с повторением, в каждой ячейке $\geqslant 2$ частиц, числом способов $C_{r-2k+k-1}^{k-1} = C_2^2 = 1$ как места $(k-1) = 2$ внутренней перегородки между $k = 3$ ячейками среди всех двух мест для них, где в каждую из $k = 3$ ячейках уже лежат по 2 частицы: (3,6).

Результат второго этапа, т. е. все исходы схемы в $k = 3$ фиксированных непустых ячейках, дают наборы уровней заполнения ячеек в результате расстановок внутренних перегородок между непустыми ячейками по вариантам разности множеств шагов 2 и 3:
$(2, 4), (2, 5), (2, 6), (2, 7), (3, 5), (3, 7), (4, 6), (4, 7), (5, 7)$, которые соответственно приводят к наборам уровней заполнения $k = 3$ непустых ячеек в порядке ячеек по числам мест из $(k + r - 1) = 8$ между местами перегородок, а именно 9 исходов:
$(1, 1, 4), (1, 2, 3), (1, 3, 2), (1, 4, 1), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1), (4, 1, 1)$.

Окончательно все исходы схемы появления события A_k получаются в виде наборов найденных выше уровней заполнения $k = 3$ непустых ячеек (в порядке ячеек) на местах этих непустых $k = 3$ ячеек, определенных на первом этапе с расстановкой нулевых уровней на остальных $n - k = 5 - 3 = 2$ местах. Например, при уровнях заполнения (1,1,4) $k = 3$ непустых ячеек (в порядке ячеек) на местах (1,2,3) исход схемы имеет вид (1,1,4,0,0), а на местах (1,2,4) – (1,1,0,4,0).

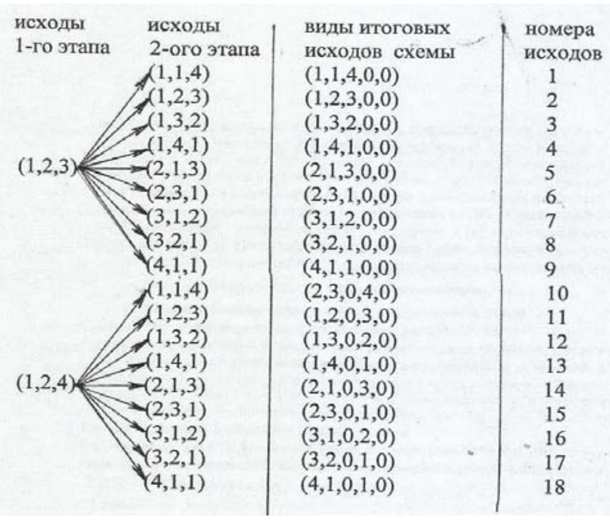
Пучковая структура графа перечисления исходов схемы в примере 1 есть:
$(10), (9, 9, 9, 9, 9, 9, 9, 9, 9, 9)$.

Найдем число исходов схемы по (1) и по процедуре перечисления исходов схемы.

По (1) $N = C_5^3(C_{6-1}^{3-1} - C_{6-3-1}^{3-1}) = C_5^3(C_5^3 - C_5^3) = 10 \cdot 9 = 90$.

При таком большом объеме исходов схемы C приведем фрагмент графа перечисления ее исходов для первых двух фиксаций мест непустых ячеек на рисунке.

По процедуре перечисления и по рисунку число исходов схемы N получается при каждой из 10 фиксаций мест непустых ячеек по 9 в каждом, откуда $N = 90$, что совпадает с полученным выше по (1) результатом.



Фрагмент графа перечисления исходов схемы в примере 1

A fragment of the graph of outcomes enumeration of the scheme in example 1

2.2. Решение задачи нумерации для исходов схемы C

Схема C представляет собой схему ПД (см. п. 1.2 или подробнее [5, п. 1.8]), и при изученных схеме сочетаний ([5, п. 1.3]) и известной из п. 2.1 пучковой структуре графа перечисления исходов схемы решается по теоремам 1 и 2 соответственно в прямой и обратной постановках. Проиллюстрируем решение ЗН на числовом примере.

Пример 2. Решаем ЗН в условиях примера 1 $n = 5$, $k = 3$, $r = 6$.

ПЗН. Дан номер исхода $N_* = 15$, найти вид исхода схемы R_*.
Решение. По теореме 1 из $N_* = N^{(2)} = 15$ находим

$$N^{(1)} = \left[\frac{N^{(2)} + 9 - 1}{9}\right] = 2,$$

$$j_1 = 2 \; mod \; 9 + 0 = 2, \;\; j_2 = 15 \; mod \; 9 + 0 = 6.$$

Для перечисления исходов первого этапа второй исход – это (1,2,4), а 6-й исход второго этапа – это (2,3,1), откуда получаем искомый вид исхода $R_* = (2, 3, 0, 1, 0)$, что совпадает с результатом по рисунку.

ОЗН. Дан вид исхода $R_* = (2, 3, 0.1, 0)$, найти номер исхода схемы $N_* = N^{(2)}$.
Решение. По теореме 2 из $R_* = (2, 3, 0.1, 0)$ находим $j_1 = 2$, $j_2 = 6$, отсюда получаем $N_* = N^{(2)} = (2 - 1)9 + 6 = 15$, что совпадает с результатом по рисунку.

2.3. Установление вероятностей для исходов схемы C

При равновероятных исходах используемой здесь схемы сочетаний вероятность $P(C)$ появления исхода схемы C среди всех C_{n+r-1}^r исходов схемы размещения r неразличимых частиц по n различимым ячейкам определяется отношением чисел их исходов:

$$P(C) = C_n^k(C_{r-1}^{k-1} - C_{r-k-1}^{k-1})/C_{n+r-1}^r.$$

Организация перечисления исходов при равновероятности исходов схемы сочетаний приводит к допустимым и равновероятным исходам (см. рисунок) с вероятностями каждого, равными обратному числу их исходов. Так, в примере 1 вероятности исходов схемы C получаются равными по $1/90$.

2.4. Моделирование исходов схемы C

Его предлагается проводить УМ по шагам:

1) разыгрывание случайного номера исхода схемы C с установленным в п. 2.3 вероятностным распределением;

2) получение в качестве смоделированного исхода схемы C по результату решения ПЗН его вида по разыгранному номеру шага 1.

3. Методика пересчета перечисления исходов схем A, B, D

Полученный результат перечисления исходов схемы C в п. 2.1 дает начальную информацию для остальных схем A, B, D вида составов уровней заполнения ячеек в порядке ячеек в условиях неразличимости частиц и различимости ячеек. Учитывая отличие качеств по различимости составляющих схему ячеек и частиц в исследуемых схемах от схем с ранее изученным перечислением исходов, предложим алгоритмические преобразования их перечислений для перечисления ее исходов. Далее при различимых частицах номера частиц называем частицами.

3.1. Перечисление исходов схемы A из схемы C

В схеме A частицы и ячейки различимы. Вид исхода схемы A – составы частиц в ячейках в порядке ячеек.

Отличие схемы A от схемы C – в различимости частиц. Поэтому для перечисления исходов схемы A в схеме в графе перечисления исходов схемы C добавляется итерация деления по k ячейкам r частиц на составы ячеек численностями уровней их заполнений в заданном в схеме C порядке (число таких делений известно как число исходов схемы перестановок с повторением). Для дальнейшего анализа исходов схемы A отметим, что она представляет собой схему ОПД (см. [5, п. 1.9], или по решению ЗН см. п. 1.3)

В схеме D частицы и ячейки неразличимы. Вид исхода схемы D – составы уровней заполнения ячеек в заранее определенном порядке, например, в порядке роста уровней заполнения.

3.2. Перечисление исходов схемы B из схемы A

В схеме B частицы различимы, ячейки неразличимы. Вид исхода схемы B – составы частиц в ячейках в заранее определенном порядке, например, в порядке роста минимальной частицы в ячейке.

Отличие схемы B от схемы A – в неразличимости ячеек. Поэтому для перечисления исходов схемы B во всех исходах схемы A среди непустых ячеек упорядочивают их составы, например, по возрастанию номеров содержащихся в ячейках частиц с отбраковкой повторов в исходах.

3.3. Перечисление исходов схемы D из схемы C

В схеме D частицы и ячейки неразличимы. Вид исхода схемы D – составы уровней заполнения ячеек в заранее определенном порядке, например, в порядке роста уровней заполнения среди непустых ячеек.

Отличие схемы D от схемы C – в неразличимости ячеек. Поэтому для перечисления исходов схемы D во всех исходах схемы C среди непустых ячеек упорядочивают их составы, например, по возрастанию уровней заполнения ячеек с отбраковкой повторов в исходах.

Более подробно о пересчете результатов ПМ в схемах размещения частиц по ячейкам см. в [5, п. 2.21].

Приведем примеры таких пересчетов перечислений исходов схем A, B, D из перечисления исходов начальной схемы C.

Пример 3. Пусть в схеме C, как в примере 1, $n = 5$, $k = 3$, $r = 6$.

Предлагается производить последовательные пересчеты перечисления исходов схем следующими парами с одним отличием качества составляющих их элементов – ячеек и частиц: 1) из схемы C для схемы A, 2) из схемы A для схемы B, 3) из схемы C для схемы D.

Построим соответствующие данным исходов схемы C фрагменты перечислений исходов схем A, B, D.

1) Пересчет исходов из схемы C для схемы A

Из исхода схемы C (1,4,1,0,0) по схеме перестановок с повторением можно получить $6!/1!4!1!0!0! = 30$ исходов. Рассмотрим, например, 5 из них:

$((1),(2,3,4,5),(6)), ((2),(3,4,5,6),(1)),$
$((3),(1,2,4,5),(6)), ((6),(2,3,4,5),(1)),$
$((6),(1,2,4,5),(3)).$

2) Пересчет исходов из схемы A для схемы B

Из исхода схемы A в 1) получим после отбраковки повторов 3 исхода схемы B:

((1),(2,3,4,5),(6)), ((1),(2),(3,4,5,6),),
((1,2,4,5),(3),(6)).

3) Пересчет исходов из схемы C для схемы D

Из 9 исходов схемы C:
(1,1,4,0,0),(1,2,3,0,0),(1,3,2,0,0),(1,4,1,0,0),
(2,1,3,0,0),(2,3,1,0,0),(3,1,2,0,0),(3,2,1,0,0),
(4,1,1,0,0)
получим после отбраковки повторов 2 исхода схемы D: (1,1,4), (1,2,3).

По очевидным исходам схемы D из замечания можно получить из них при маркировках по совпадениям уровней соответственно (2,1) и (1,1,1) число всех исходов схемы C в условиях примера:

$$N = N(k)$$
$$= C_5^3(3!/2!1! + 3!/1!1!1!) = 10(3+6) = 90,$$

что совпадает с ранее полученным результатом.

Выводы

1. Для изучения схемы C применен авторский перечислительный метод и получены новые результаты по всем его направлениям с добавлением новых приемов, отражающих специфику ее исходов.

2. Выведена явная аналитическая формула для числа исходов схемы C.

3. Получено вероятностное распределение исходов при фиксированном времени ожидания в схеме C.

4. Приведена методика пересчета начальных результатов анализа в виде перечисления исходов схемы C для схем A, B, D с другими парными качествами по различимости ячеек и частиц, дающая возможность проведения для них алгоритмических исследований остальных направлений по ПМ.

Литература

1. *Иванов В. А.* Предельные теоремы в схеме размещения со случайными уровнями // Математические заметки. 1982. Т. 31, вып. 4. С. 619–631.

2. *Колчин В. Ф., Севастьянов Б. А., Чистяков В. П.* Случайные размещения. М.: Наука, 1976. 223 с.

3. *Севастьянов Б. А.* Предельные теоремы в одной схеме размещения частиц по ячейкам // Теория вероятностей и ее применения. 1966. Т. 11, № 4. С. 696–700.

4. *Хакимуллин Е. Р., Энатская Н. Ю.* Предельные теоремы для числа пустых ячеек // Дискретная математика. 1997. Т. 9, № 2. С. 120–130.

5. *Энатская Н. Ю.* Доасимптотический анализ комбинаторных схем. М.: URSS, 2023. 536 с.

References

1. *Ivanov V. A.* Limit theorems in an allocation scheme with random levels. *Mathematical Notes.* 1982;31(4):619–631. (In Russ.)

2. *Kolchin V. F., Sevastyanov B. A., Chistyakov V. P.* Random allocations. Moscow: Nauka; 1976. 223 p. (In Russ.)

3. *Sevastyanov B. A.* Limit theorems in one scheme of particles placement in cells. *Probability Theory and its Applications.* 1966;11(4):696–700. (In Russ.)

4. *Khakimullin E. R., Enatskaya N. Yu.* Limit theorems for the number of empty cells. *Discrete Mathematics.* 1997;9(2):120–130. (In Russ.)

5. *Enatskaya N. Yu.* Pre-asymptotic analysis of combinatorial schemes. Moscow: URSS; 2023. 536 p. (In Russ.)

Поступила в редакцию / received: 22.12.2024; принята к публикации / accepted: 10.04.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Энатская Наталия Юрьевна

канд. физ.-мат. наук, доцент Департамента прикладной математики

e-mail: nat1943@mail.ru

CONTRIBUTOR:

Enatskaya, Natalia

Cand. Sci. (Phys.-Math.), Associate Professor

Труды Карельского научного центра РАН. 2025. № 4. С. 97–106
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 97–106
DOI: 10.17076/mat2045

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.115:519.2

ОБРАТНАЯ ЭКСТРЕМАЛЬНАЯ ЗАДАЧА В СХЕМЕ РАЗМЕЩЕНИЯ ЧАСТИЦ КОМПЛЕКТАМИ

Н. Ю. Энатская

Московский институт электроники и математики,
Национальный исследовательский университет «Высшая школа экономики»
(ул. Таллинская, 34, Москва, Россия, 123458)

Рассматриваются схемы размещения частиц по ячейкам неразличимыми (схема A) и различимыми (схема B) комплектами с достижением заданного минимального уровня заполнения ячеек в их исходах. Анализ схем проводится перечислительным методом на основе построения итерационного случайного процесса прямого бесповторного нумерованного перечисления их исходов в доасимптотической области изменения параметров по следующим направлениям: перечисления исходов и нахождения их числа, решения задачи нумерации в прямой и обратной постановках, состоящих в установлении взаимнооднозначного соответствия между номерами и видами исходов схемы, определения вероятностного распределения на множестве исходов по введенным вероятностям итерационных переходов процесса их перечисления и разработки процедуры их моделирования.

Ключевые слова: размещение частиц; минимальный уровень заполнения ячеек; задача нумерации; моделирование

Для цитирования: Энатская Н. Ю. Обратная экстремальная задача в схеме размещения частиц комплектами // Труды Карельского научного центра РАН. 2025. № 4. С. 97–106. doi: 10.17076/mat2045

N. Yu. Enatskaya. AN INVERSE EXTREME PROBLEM IN PARTICLE GROUP ALLOCATION SCHEME

National Research University Higher School of Economics, Moscow Institute of Electronics and Mathematics (34 Tallinskaya St., 123458 Moscow, Russia)

The article studies the schemes of particle allocation to cells in indistinguishable (scheme A) and distinguishable (scheme B) groups with the outcomes of achieveming a given minimum cell filling level. The schemes are analyzed by the enumerative method based on the construction of an iterative random process of direct non-repetitive numbered enumeration of their outcomes in the pre-asymptotic region of parameter change along the following lines: enumeration of outcomes and finding their number, solving the enumeration problem in direct and inverse formulations to establish one-to-one correspondence between the numbers and types of outcomes of the scheme, determining the probability distribution on the set of outcomes based on given probabilities of iterative transitions of the process of their enumeration, and developing a procedure for their modeling.

Keywords: particle allocation; minimum cell filling level; numbering problem; modeling

For citation: Enatskaya N. Yu. An inverse extreme problem in particle group allocation scheme. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 97–106. doi: 10.17076/mat2045

Введение

Обратная экстремальная задача о размещении частиц комплектами по ячейкам состоит в представлении всех исходов схемы (с достижением заданного минимального значения уровня заполнения (т. е. числа частиц) хотя бы в одной ячейке каждого исхода) и исследовании их свойств. Направления анализа исходов схемы определяются перечислительным методом (ПМ).

В научной литературе широко представлены результаты асимптотического анализа классической схемы размещения по ячейкам частиц комплектами (см., например, [1]).

Здесь авторским ПМ по расширенным направлениям: перечисления исходов и нахождения их числа, решения задачи нумерации, определения вероятностного распределения на множестве исходов по введенным вероятностям итерационных переходов процесса их перечисления и разработки процедуры их моделирования проводится доасимптотический анализ схем размещения по ячейкам частиц различимыми и неразличимыми комплектами.

Размещение по ячейкам частиц комплектами означает, что частицы каждого комплекта размещаются по разным ячейкам.

Введем параметры схем $A = A(n, r, m)$, $B = B(n, r, m)$: n – число ячеек, m – число комплектов частиц, r – размер комплекта, k – заданный минимальный уровень заполнения каждой ячейки в результате размещения m комплектов частиц по n ячейкам.

В схемах ячейки различимы, частицы комплекта неразличимы, но в схеме A комплекты неразличимы, а в схеме B – различимы, т. е., например, разных цветов, определяемых их номерами для составляющих их частиц.

Исходы схемы A представляют все составы уровней заполнения ячеек с данным минимальным значением в порядке ячеек.

Исходы схемы B представляют все составы номеров частиц в ячейках с данным минимальным значением их уровней заполнения в порядке ячеек.

Постановка задачи – провести доасимптотический анализ схем A и B по указанным в аннотации направлениям.

Основные обозначения, определения и понятия

ПМ – перечислительный метод исследования моделей по приведенным в аннотации направлениям;

МГ – представление случайного процесса перечисления исходов схемы в виде вероятностного графа;

итерация – шаг перехода в графе перечисления исходов схемы к следующему этапу (шагу) их перечисления;

траектория исхода – последовательность исходов итераций, ведущая к нему в графе от начала их перечисления;

бесповторное перечисление исходов схемы – единственность траекторий для каждого ее исхода;

пучок в графе процесса – совокупность переходов из каждого исхода каждой итерации;

размер пучка на каждой итерации – число переходов из каждого исхода итерации, т. е. число исходов из него на следующей итерации;

пучковая структура графа – перечисление размеров пучков всех итераций;

ЗН – задача нумерации – установление взаимно-однозначного соответствия между всеми номерами и видами исходов схемы;

ПЗН – прямая задача нумерации – нахождение вида исхода схемы по его номеру;

ОЗН – обратная задача нумерации – нахождение номера исхода схемы по его виду;

СУ – система уравнений для взаимного пересчета вероятностей исходов схемы и вероятностей итерационных переходов в графе процесса их перечисления;

УМ – универсальный (единый) метод моделирования исходов схемы с их вероятностным распределением.

1. Вспомогательные результаты

Здесь представим основные методы исследования, основные используемые далее конструкции схем и соотношения между параметрами изучаемых схем. Для остальных упоминаемых здесь ранее полученных результатов ограничимся ссылками на источники, а новые вспомогательные результаты докажем.

1.1. Перечислительный метод

В основе доасимптотического анализа рассматриваемых схем лежит ПМ (см. [2]), суть которого состоит в организации получения *качественной* информации об исходах схемы и переводе ее в *количественную* – результатов ее анализа в доасимптотической области значений параметров. Эта качественная информация представляет собой исходы итерационного случайного процесса реализации комбинаторной схемы путем последовательного поединичного добавления элементов схемы до заданного значения или этапов перечисления составляющих схему более простых ранее изученных схем. Инструментами перевода качественной информации о видах всех исходов схемы являются метод графов (МГ) (см. [2, п. 1.1]), состоящий в графическом представлении процедуры итерационного процесса перечисления исходов схемы, задача нумерации (ЗН) (см. [2, п. 1.1]), устанавливающая взаимно-однозначное соответствие между их номерами и видами исходов (соответственно ПЗН – прямая и ОЗН – обратная задача нумерации), СУ для взаимного пересчета вероятностей процесса перечисления исходов схемы с их итоговыми вероятностями и универсальное моделирование (УМ) (см. [2, п. 1.1 и Приложение А]) исходов, дающее его единый прием, состоящее в разыгрывании номеров исходов, виды которых определяются по решению задачи нумерации, учитывающему специфику схемы. Целью применения ПМ является изучение схемы по указанным в аннотации направлениям.

1.2. Схема последовательных действий

В [2, п. 1.8] приводятся результаты комбинаторного анализа схемы последовательных действий (ПД).

Схема ПД возникает, когда каждому следующему действию подвергаются исходы предыдущего действия и числа исходов на каждом следующем шаге (действии) одинаковы, т. е. зависят только от характера действия. Результатом этого являются одинаковые размеры пучков внутри итерации в графе перечисления исходов схемы при переходе от исходов предыдущего действия к последующему, т. е. числа исходов из каждого состояния на следующей итерации.

Анализ схемы ПД приводит к конкретным результатам только по результатам подобных исследований комбинаторных схем действий.

Пусть i-е действие ($i = \overline{1, K}$) совершается n_i числом способов.

Общее число N исходов схемы известно и задано формулой

$$N = \prod_{i=1}^{K} n_i.$$

Вид исхода после совершения i действий будет формироваться из принятых видов исходов последовательных действий, которые будем соответственно обозначать через R_{ij_i}, где i – номер действия, а j_i – номер исхода в результате его совершения, а конкретный вид R_{ij_i} определяется характером действия. Исход в результате совершения i действий ($i \leqslant K$) обозначен в виде $R^{(i)} = \{R_{1j_1}, R_{2,j_2} \ldots, R_{rj_i}\}$. Тогда окончательный исход схемы получен при $i = K$ в виде $R^{(K)} = \{R_{1j_1}, R_{2,j_2} \ldots, R_{Kj_K}\}$.

Для явного перечисления исходов схемы по МГ строится случайный процесс пошагового последовательного поединичного добавления действий к исходам всех предшествующих действий, изображаемого графом.

Нумерация исходов на каждом шаге проведена в порядке роста номеров упорядоченных в схеме действий и в порядке роста номеров исходов, заданных по каждому действию, который при конкретизации действий известен.

При решении ЗН запись вида исхода на каждом шаге представляет собой траекторию переходов процесса из состояния в состояние, т. к. первый индекс каждой компоненты указывает номер шага (действия), а второй – номер исхода в пучке этого шага. Считаем решенными ЗН для схем всех действий.

Прямая задача нумерации

Теорема 1. *Пусть в схеме с параметрами* $n_1, \ldots, n_K$ *дан номер* $N^{(K)}$ *ее исхода. Тогда вид исхода* $R^{(K)} = \{R_{1j_1}, \ldots, R_{Kj_K}\}$, *определяемый номерами* $(j_1, \ldots j_K)$ *исходов его компонент в пучках графа перечисления исходов схемы, при* $i = \overline{1, K}$ *находится по формуле*

$$j_i = t_i + I(t_i)n_i,$$

где $t_i = N^{(i)} mod\ n_i$; $I(Z) = 0$ *при* $Z \neq 0$ *и* $I(Z) = 1$ *при* $Z = 0$;

$$N^{(i-1)} = \left[\frac{N^{(i)} + n_i - 1}{n_i}\right],$$

где $[Z]$ *– целая часть числа* Z *и* $i = K, K-1, \ldots, 1$; $N^{(0)} = 1$.

Обратная задача нумерации

Теорема 2. *Пусть в схеме с параметрами $n_1, \ldots, n_K$ дан вид ее исхода $R^{(K)} = \{R_{1j_1}, \ldots, R_{Kj_K}\}$, определяющий номера $(j_1, \ldots j_K)$ исходов его компонент в пучках графа перечисления исходов схемы, при $i = \overline{1, K}$. Тогда его номер вычисляется по формуле*

$$N^{(K)} = \sum_{l=1}^{K-1} (j_l - 1) \prod_{i=l+1}^{K} n_i + j_K.$$

1.3. Обобщенная схема последовательных действий (ОПД)

Эта схема является основной структурой, дающей общее описание процесса формирования и перечисления исходов комбинаторных схем. Исследование свойств этих исходов создает основу доасимптотического анализа комбинаторных схем по направлениям ПМ, где действиями представлены последовательные итерации случайного процесса их формирования. (Схема ПД является частным случаем схемы ОПД.)

В [2, п. 1.9] приводятся результаты комбинаторного анализа схемы ОПД.

Схема ОПД с результатом анализа в [2, п. 1.9] возникает, когда каждому следующему действию подвергаются исходы предыдущего действия и числа исходов на каждом следующем шаге (действии) неодинаковы, т. е. зависят от характера действия и вида предыдущего исхода. Результатом этого являются разные размеры пучков внутри итерации в графе перечисления исходов схемы при переходе от исходов предыдущего действия к последующему, т. е. числа исходов из каждого состояния на следующей итерации.

Анализ схемы ОПД приводит к конкретным результатам только по результатам подобных исследований комбинаторных схем действий.

В схеме проводится K последовательных действий, i-е из которых $(i = \overline{1, K})$ на i-м шаге совершается $N^{(i)}$ способами. Тогда число исходов этих K действий складывается из $N^{(K-1)}$ пучков размерами $\bar{d}^{(K)} = (d_1^{(K)}, d_2^{(K)}, \ldots, d_{N^{(K-1)}}^{(K)})$, т. е. общее число $N = N^{(K)}$ исходов схемы получается из рекуррентного соотношения при $i = K$ и $N = N^{(0)} = 1$

$$N^{(i)} = \sum_{l=1}^{N^{(i-1)}} d_l^{(i)}.$$

Вид исхода после совершения i действий будет формироваться из принятых видов исходов последовательных действий, которые будем соответственно обозначать через R_{ij_i}, где i – номер действия, а j_i – номер исхода в результате его совершения.

Задача нумерации решается для нашей схемы при решенной ЗН для каждого из K действий и известной пучковой структуре графа перечисления исходов нашей схемы, т. е. с известными числами исходов (размерами пучков) при каждом действии на каждой итерации.

Прямая и обратная задачи нумерации решены следующими теоремами.

Теорема 3. *Пусть совершается K действий и задан номер исхода $N_*^{(K)}$. Тогда его вид, определяемый номерами исходов траектории T в содержащих их пучках $\{j_i\}$ от первой до K-й итераций, вычисляется по рекуррентной формуле для j_i $(i = \overline{1, K})$*

$$j_i = N_*^{(i)} - \sum_{l=1}^{N_*^{(i-1)}-1} d_l^{(i)},$$

где все пучковые структуры действий $\bar{d}^{(i)}$ заданы и

$$N_*^{(K-1)} = \delta + \max\ t : \left(\sum_{l=1}^{t} d_l^{(K)} = A_K \leqslant N_*^{(K)} \right),$$

где $\delta = 0$ при $A_k = N_*^{(K)}$ и $\delta = 1$ при $A_K < N_*^{(K)}$. Заменяя K на i, доходим по рекурренте до первого шага.

По решенной ЗН для всех действий находим из $\{j_i\}$ виды их исходов, из которых получаем искомый вид исхода $R_*^{(K)}$.

Теорема 4. *Пусть совершается K действий и задан вид исхода $R_*^{(K)} = \{j_1, \ldots, j_K\}$. Тогда его номер $N_*^{(K)}$ определяется по рекуррентной формуле при $i = K$, $\quad i = \overline{1, K}$*

$$N_*^{(i)} = \sum_{l=1}^{N_*^{(i-1)}-1} d_l^{(i)} + j_i,$$

начиная с $i = 1$ при $N_*^{(1)} = j_1$.

1.4. Перечисление и число N исходов в схеме размещения r неразличимых частиц по n различимым ячейкам не более чем по k в каждой ячейке

Эта процедура рассмотрена в [2, п. 2.13.2], и аналитический результат фактически получен, но не представлен в виде теоремы. Поэтому для его дальнейшего использования он будет приведен здесь как утверждение теоремы с доказательством.

Процесс бесповторного перечисления исходов схемы составляет последовательную расстановку $(n-1)$ внутренних границ ячеек между лежащими в ряд r неразличимыми частицами среди $M = (n+r-1)$ мест для этих перегородок $\bar{t} = (t_1, \ldots, t_{n-1})$ с ограничением для последовательных расстояний между ними $\leqslant (k+1)$ путем выбора диапазонов варьирования их мест. Тогда составы $\bar{\eta} = (\eta_1, \ldots, \eta_i, \ldots, \eta_n)$ уровней заполнения ячеек будут определяться по формуле $\eta_i = t_i - t_{i-1} - 1$, $i = \overline{1, n}$, $t_0 = 0$, $t_n = n$.

Теорема 5. *Допустимые значения* $\bar{t} = (t_1, \ldots, t_{n-1})$ *при* $i = \overline{1, n}$, $t_0 = 0$, $t_n = M + 1$ *определяются по рекуррентной формуле*

$$l_i = max(t_{i-1} + 1, r + i - k(n-i))$$

$$\leqslant t_i \leqslant min(t_{i-1} + k + 1, r + i) = L_i.$$

Число $N(r, n, k)$ *исходов схемы равно сумме размеров пучков предпоследней итерации в графе перечисления исходов:*

$$N(r, n, k) = \sum_j d_j,$$

где d_j *– размер* j*-го пучка предпоследней* $(n-1)$*-й итерации.*

Доказательство. Для обеспечения возможности реализации схемы введем дополнительные ограничения на все упорядоченные по возрастанию элементы (места границ из $M = (n + r - 1)$ между r неразличимыми частицами) $\bar{t} = (t_1, \ldots, t_{n-1})$, обеспечивающие ограничение уровня заполнения всех ячеек со следующим смыслом всех дополнительных условий для диапазонов их изменения при $i = \overline{1, n-1}$, $t_0 = 0$:

1) расстояние от предыдущего элемента не более $k + 1$, откуда
$$t_i \leqslant t_{i-1} + k + 1; \qquad (1)$$

2) правее значения t_i среди m мест поместятся остальные $n-1-i$ границ, откуда
$$t_i \leqslant M - (n - 1 - i) = r + i; \qquad (2)$$

3) значение следующего элемента больше предыдущего, откуда
$$t_i \geqslant t_{i-1} + 1; \qquad (3)$$

4) правее значения t_i среди m мест окажется мест не более, чем для $n-1-i$ остальных границ и $k(n-i)$ максимально допустимого в остальных ячейках схемы A числа частиц, откуда
$$\begin{aligned} t_i &\geqslant M - (n - 1 - i) - k(n - i) \\ &= r + i - k(n - i). \end{aligned} \qquad (4)$$

Теперь, объединяя соотношения (1) и (2) в верхнюю границу для t_i, а (3) и (4) – в нижнюю границу для t_i, получаем следующий диапазон изменения значений для t_i и, в частности, для t_1 и t_{n-1} рекурренты:

$$\begin{aligned} l_i &= max(t_{i-1} + 1, r + i - k(n - i)) \\ &\leqslant t_i \leqslant min(t_{i-1} + k + 1, r + i) = L_i. \end{aligned} \qquad (5)$$

Тогда

$$\begin{aligned} l_{n-1} &= max(t_{n-2} + 1, n + r - 1 - k) \leqslant t_{n-1} \\ &\leqslant min(t_{n-2} + k + 1, r + n - 1) = L_{n-1}. \end{aligned} \qquad (6)$$

Теперь перечисление всех исходов схемы через $\bar{t}$ будем производить последовательными действиями поединичного варьирования значений по возрастанию элементов $t_1, \ldots, t_{n-1}$ в порядке роста индексов t_i в указанных в (5) и (6) зависимых по шагам перечисления исходов диапазонах. Поэтому анализ схемы будет следовать из результатов анализа обобщенной схемы последовательных действий в п. 1.3 (или [2, п. 1.9]), для которого остается указать пучковую структуру, т. е. размеры пучков в графе перечисления исходов на i-м шаге, равные $(L_i - l_i + 1)$ при поединичном варьировании значений t_{i-1} в возрастающем порядке. Таким образом, задача анализа схемы сводится к указанию пошаговой пучковой структуры графа перечисления ее исходов. Тогда число N исходов схемы равно сумме размеров пучков последнего шага графа. Число шагов равно $(n - 1)$. □

Перевод видов исходов схемы в терминах $\bar{t}$ в более наглядный – перечисления уровней заполнения ячеек в их порядке $\bar{\eta} = (\eta_1, \ldots, \eta_n)$ производится по формуле $\eta_j = t_j - t_{j-1} - 1$, где $j = \overline{1, n}$, $t_0 = 0$, $t_n = m + 1$.

Поясним порядок описанного прямого перечисления исходов схемы на примере.

Пример 1. Пусть $n = 3$, $r = 5$, $k = 3$. Отсюда $S = 4$, $m = 7$. Перечислим все исходы схемы, вычисляя пределы варьирования их элементов в указанных в (5) и (6) диапазонах и размеры пучков по шагам перечисления исходов:

1-й шаг:

$l_1 = max(1, 0) \leqslant t_1 \leqslant min(4, 6)$, откуда $1 \leqslant t_1 \leqslant 4$, размер пучка равен 4;

2-й шаг:

$l_2 = max(t_1 + 1, 4) \leqslant t_2 \leqslant min(t_1 + 4, 6)$, откуда при

$t_1 = 1$ получаем $4 \leqslant t_2 \leqslant 5$, размер 1-го пучка равен 2;

$t_1 = 2$ получаем $4 \leqslant t_2 \leqslant 6$, размер 2-го пучка равен 3;

$t_1 = 3$ получаем $4 \leqslant t_2 \leqslant 7$, размер 3-го пучка равен 4;

$t_1 = 4$ получаем $5 \leqslant t_2 \leqslant 7$, размер 4-го пучка равен 3.

Отсюда число исходов схемы $N = 2 + 3 + 4 + 3 = 12$.

1.5. Метод сопровождающих схем

Определим понятие сопровождающей схемы (СС) для данной, как схемы, частью исходов которой являются исходы данной схемы.

Метод сопровождающих схем (МСС) состоит в приеме пересчета отдельных результатов анализа СС по ПМ для данной схемы. Применение МСС рационально, когда эти результаты для СС или известны, или получаются проще, чем непосредственно для данной схемы.

Для пересчета результатов введем обозначения: S, S^* – соответственно сопровождающая и данная схемы, числа исходов в них – N_S и N_{S^*}, а Q – признак, выделяющий исходы схемы S^* из исходов схемы S.

Будем считать известным перечисление исходов, результаты решения ЗН и вероятности итоговых исходов в схеме S.

Перечисление исходов данной схемы производится отбраковкой исходов СС по отсутствию признака, выделяющего исходы данной схемы из ее исходов.

Решение ЗН исходов данной схемы по МСС

ПЗН. Пусть дан номер $N^*_{S^*}$ исхода схемы S^*. Требуется найти его вид $R^*_{S^*}$.

Решение. В порядке перечисления исходов схемы S будем проверять для ее исходов наличие признака Q до его $N^*_{S^*}$-го выполнения. Тогда известный по ПЗН в схеме S его вид и будет искомым видом $R^*_{S^*}$.

ОЗН. Пусть дан вид $R^*_{S^*}$ исхода схемы S^*. Требуется найти его номер $N^*_{S^*}$.

Решение. В порядке перечисления видов исходов схемы S будем проверять для них выполнение признака Q и после этого сравнивать их виды с данным видом ее исхода $R^*_{S^*}$ схемы S^* до совпадения с ним. Тогда порядковый номер последнего такого совпадения даст искомый номер $N^*_{S^*}$.

Пересчет вероятностей итоговых исходов данной схемы по МСС

Пусть заданы вероятности исходов схемы S. Для нахождения вероятностей исходов схемы A изобразим граф бесповторного перечисления исходов схемы S с удалением всех траекторий графа, не ведущих к исходам схемы A. Тогда вероятности оставшихся траекторий, т. е. всех исходов схемы A (в графе их бесповторного перечисления), будем вычислять произведениями вероятностей составляющих их итерационных переходов. В свою очередь, эти вероятности итерационных переходов итерационно пересчитываются по пучкам их исходного графа схемы S для оставшихся после отбраковки переходов делением их вероятности в схеме S на сумму вероятностей оставшихся в пучке итерационных переходов.

1.6. Соотношения между параметрами изучаемых схем

Предполагается, что $r \leqslant n$, $rm \geqslant kn$, откуда (и из условия размещения частиц комплектами) следует, что $k \leqslant m$.

При заданных основных параметрах схем n, m, r для диапазонов возможных значений k и чисел i ячеек их достижений приведем леммы.

Лемма 1. $0 \leqslant k \leqslant [rm/n]$.

Доказательство очевидно по характеру схемы, т. к. $kn \leqslant rm$.

Лемма 2.

$$l = max(1, n(k+1) - rm) \\ \leqslant i \leqslant min(n, [m(n-r)/(m-k)]) = L,$$

где i – число ячеек в исходах схемы с заданным минимальным уровнем k заполнения ячеек.

Доказательство утверждения проведем отдельно для l и L из следующих соображений:

$(rm - kn)$ частиц должно хватить (хотя бы по одной) на размещение в $(n - i)$ ячейках, откуда $rm - kn \geqslant (n - i)$, или $i \geqslant n - rm + kn$, кроме этого, по условию обязательного достижения заданного значения k хотя бы в одной ячейке, $i \geqslant 1$; окончательно получаем: $l = max(1, n(k+1) - rm) \leqslant i$, что и утверждается;

$(rm - kn)$ частиц должны поместиться в $(n-i)$ ячейках не более чем по $m-k$ в каждой ячейке, откуда $rm - kn \leqslant (n-i)(m-k)$, или $i \leqslant m(n-r)/(m-k)$, кроме этого, $i \leqslant n$; окончательно получаем: $i \leqslant \; min(n, [m(n-r)/(m-k)]) = L$, где $[Z]$ – целая часть числа Z, что и утверждается.

2. Анализ схемы A

2.1. Бесповторное перечисление и число исходов схемы A

В [2, п. 2.22] доказано, что исходы схемы A без нижнего ограничения на уровни заполнения ячеек совпадают с исходами схемы $A^* = A^*(n, rm, m)$ с параметрами (n, rm, m) – размещения по n различимым ячейкам неразличимых rm частиц с не более чем по m частиц в каждой ячейке. Тогда исходы схемы A представляют собой часть исходов схемы $A^* = A^*(n, rm, m)$ со спецификой размещения частиц комплектами в схеме A (частицы каждого комплекта размещаются по разным ячейкам, а значит, число непустых ячеек каждого исхода $\geqslant r$) и требованием схемы A минимального уровня k заполнения ячеек с его достижением хотя бы в одной ячейке каждого ее исхода.

Исходы схемы A представляют собой все допустимые в ней составы уровней заполнения ячеек в порядке ячеек с данным минимальным значением k.

Процедура перечисления исходов схемы A^* известна (см. п. 1.4) и проводится по схеме ОПД п. 1.3 (или подробнее см. [2, п. 1.9]), где действиями являются последовательные варьирования диапазонов допустимых уровней заполнения ячеек, полученных из теоремы 5.

Предложим два способа бесповторного перечисления исходов схемы A.

Первый способ

Бесповторное перечисление N_A исходов схемы A будем производить этапами – последовательными действиями по АЛГОРИТМУ 1 по шагам:

1) варьируем числа $i \in (l, L)$ ячеек с уровнем заполнения, равным k по Лемме 2 п. 1.6, и перечисляем (по схеме сочетаний) i ячеек, помещая в них ровно по k частиц, а в остальные ячейки – по $k+1$ частиц (числом способов C_n^i);

2) перечисляем (по теореме 5) уровни заполнения ячеек $\bar{\eta}$ (при размещении $(rm-kn-n+i)$ частиц по $(n-i)$ ячейкам с $\leqslant (m-k-1)$ частицами в каждой ячейке) (числом способов $N(rm - nk - n + i, n - i, m - k - 1))$ и добавляем в них размещенные по ячейкам частицы в шаге 1).

Тогда число N_A вычисляется в соответствии с этой процедурой перечисления исходов по формуле:

$$N_A = \sum_{i=l}^{L} C_n^i N(rm-nk-n+i, n-i, m-k-1), \tag{7}$$

где число $N(r, n, k)$ определено теоремой 5.

Обоснование АЛГОРИТМА 1.

Шаг 1) перебирает наборы ячеек с достижением уровня заполнения k во всех допустимых количествах i, помещая в них по k частиц, а в остальные $n-i$ ячейках по $k+1$ частице;

шаг 2) перечисляет все исходы схемы A^* с минимальным уровнем заполнения ячеек равным k, т. е. – все исходы схемы A.

Второй способ

Для другого способа бесповторного перечисления N_A исходов схемы A предложим АЛГОРИТМ 2 по шагам:

1) считая, что в ячейках лежат по k частиц, для остальных $rm-nk$ частиц перечислим все исходы схемы $A^*(k, rm-nk, m-k)$ с известным числом N_{A^*} способов из теоремы 5, добавляя к каждому уровню в исходах по k – получаем множество исходов C_1;

2) считая, что в ячейках лежат по k частиц, для остальных $rm-n(k+1)$ частиц перечислим все исходы схемы $A^*(k, rm-n(k+1), m-k-1)$ с известным числом N_{A^*} способов из теоремы 5, добавляя к каждому уровню в исходах по $k+1$ – получаем множество исходов C_2;

3) находим искомое перечисление исходов схемы A из разности множеств $C_1 - C_2$.

Обоснование АЛГОРИТМА 2.

Шаг 1) перечисляет все исходы схемы A^* с минимальным уровнем заполнения ячеек $\geqslant k$;

шаг 2) перечисляет все исходы схемы A^* с минимальным уровнем заполнения ячеек $\geqslant k+1$;

шаг 3) перечисляет все исходы схемы A^* с минимальным уровнем заполнения ячеек равным k, т. е. – все исходы схемы A.

Тогда число N_A вычисляется в соответствии с этой процедурой перечисления исходов по формуле:

$$\begin{gathered} N_A = C_n^k(N(k, rm-nk, m-k) \\ -N(k, rm-n(k+1), m-k-1)), \end{gathered} \tag{8}$$

где число $N(r, n, k)$ определено теоремой 5.

Таким образом, порядок нумерации бесповторного перечисления исходов схемы A производится в порядке варьирования уровней заполнения ячеек по теореме 5.

Приведем числовой пример вычисления числа исходов схемы.

Пример 2. Пусть $n = 4$, $r = 2$, $m = 6$, $k = 2$. Опишем перечисление исходов и найдем их число двумя способами – по (7) и (8).

Первый способ

Найдем диапазон значений i чисел ячеек с уровнем заполнения k по Лемме 2:

$$max(1, 4(2+1) - 12) \\ \leqslant i \leqslant min(4, [6(4-2)/(6-2)]),$$

откуда получаем $1 \leqslant i \leqslant 3$.

При $i = 1$ имеем $C_4^1 = 4$ варианта места для единственной выделенной ячейки с двумя частицами, а в остальные три добавляем по одной частице; после этого остается $12 - 11 = 1$ частица, которую размещаем по трем невыделенным ячейкам по схеме сочетаний с повторением (без результата теоремы 3, т. к. уровни заполнения ячеек после этого не могут превышать допустимого, равного 6) $C_{3+1-1}^1 = C_3^1 = 3$ способами – получаем с $i = 1$ всего $4 \cdot 3 = 12$ исходов схемы.

При $i = 2$ имеем $C_4^2 = 6$ вариантов двух мест для двух выделенных ячеек с двумя частицами, а в остальные две добавляем по одной частице; после этого остается $12 - 10 = 2$ частицы, которые размещаем по двум невыделенным ячейкам по схеме сочетаний с повторением (без результата теоремы 3, т. к. уровни заполнения ячеек после этого не могут превышать допустимого, равного 6) $C_{2+2-1}^2 = C_3^2 = 3$ способами – получаем с $i = 2$ всего $6 \cdot 3 = 18$ исходов схемы.

При $i = 3$ имеем $C_4^3 = 4$ варианта трех мест для трех выделенных ячеек с двумя частицами, а в оставшуюся одну добавляем одну частицу; после этого остается $12 - 9 = 3$ частицы, которые размещаем в одну невыделенную ячейку по схеме сочетаний с повторением (без результата теоремы 3, т. к. уровни заполнения ячеек после этого не могут превышать допустимого, равного 6) $C_{1+1-1}^2 = C_1^1 = 1$ способом – получаем с $i = 3$ всего $4 \cdot 1 = 4$ исхода схемы.

Окончательно по (7) получаем $N_A = 12 + 18 + 4 = 34$.

Второй способ

После размещения во все четыре ячейки по две частицы остается $12-8 = 4$ частицы, которые будем размещать по четырем ячейкам по схеме сочетаний с повторением (без результата теоремы 3, т. к. уровни заполнения ячеек после этого не могут превышать допустимого, равного 6) $C_{4+4-1}^4 = C_7^4 = 35$ способами – получаем 35 исходов схемы с уровнями заполнения ячеек $\geqslant 2$.

После размещения во все четыре ячейки по три частицы остается $12 - 12 = 0$ частиц, на этом процесс формирования исхода с уровнями заполнения ячеек $\geqslant 3$ завершен одним способом.

Окончательно по (8) получаем $N_A = 35 - 1 = 34$, что совпадает с результатом по формуле (7).

По описанным процедурам обоих способов вычисления числа N_A здесь, в примере, легко получить перечисление всех 34 исходов схемы.

2.2. ЗН для исходов схемы A

При решенной ЗН для схемы A^*, являющейся сопровождающей схемой для схемы A, ЗН для исходов схемы A решается по МСС (см. п. 1.5).

2.3. Вероятностное распределение исходов схемы A

а) При заданном вероятностном распределении исходов СС A^* по описанной процедуре МСС находятся вероятности исходов схемы A;

б) при заданных итерационных переходных вероятностей процесса размещения частиц комплектами по ячейкам в схеме A вероятностное распределение ее итоговых исходов находится как вероятности их траекторий.

2.4. Моделирование исходов схемы A

Его предлагается проводить по УМ с полученным по п. 2.3 вероятностным распределением ее исходов путем разыгрывания их случайных номеров, по которым из результатов решения ПЗН для исходов схемы A определяются их виды.

Анализ схемы B

3.1. Бесповторное перечисление исходов схемы B

Отличие схемы B от схемы A состоит в различимости комплектов. Будем обозначать все неразличимые частицы каждого комплекта номерами своих комплектов (от 1 до m). Исходы схемы представляют собой наборы частиц в ячейках в порядке ячеек с допустимыми уровнями заполнений ячеек по условию схемы, т. е. $\geqslant k$. Все бесповторные составы этих уровней определены в схеме A. Для бесповторного перечисления исходов схемы B строится граф итерационным случайным процессом для схемы ПД, где действиями – итерациями будут последовательные поединичные добавления независимых равновероятных размещений следующих комплектов частиц по ячейкам в соответствии с условиями схемы – по схеме сочетаний до последнего m-го комплекта. Получение результатов решения ЗН требует определения пучковой структуры графа данного процесса перечисления ее исходов. Она для данного перечисления исходов схемы определяется пучковой структурой i-й итерации ($i = \overline{1, m}$) с одинаковыми размерами пучков по C_n^r по всем итерациям.

При этом бесповторность исходов каждой (а значит, и итоговой) итерации такого перечисления обеспечивается неизменяемыми далее, заложенными в процессе перечисления исходов отличиями:

по пучкам – в разных добавленных исходах схемы сочетаний последней размещенной на выполняемой итерации частицы,

а в пучках одной итерации – в разных наборах заполнений ячеек при размещении предшествующих комплектов.

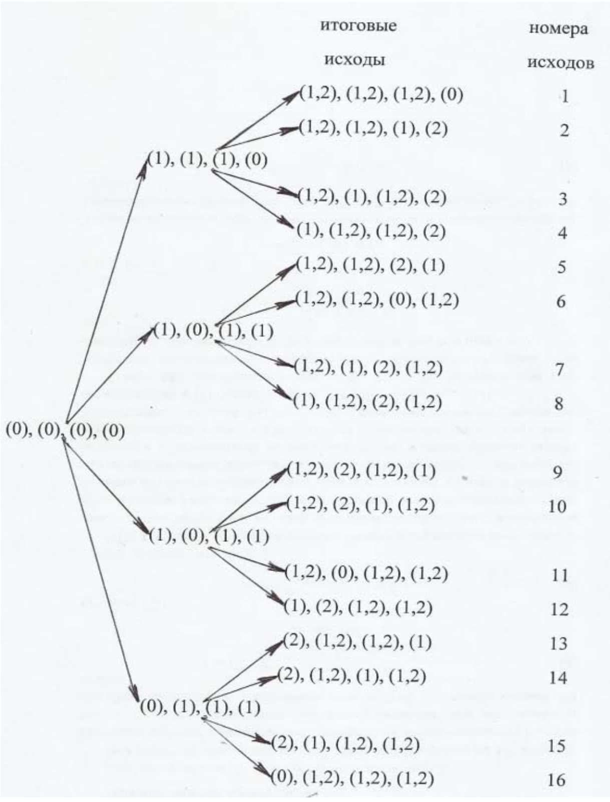
Исходы итераций будем записывать в виде последовательностей составов частиц в ячейках в порядке нумерации ячеек через запятую, заключая в круглые скобки составы ячеек и перечисляя их через запятую в ячейке в порядке размещения комплектов.

Для отражения различимости комплектов частиц при неразличимости частиц внутри каждого комплекта для описания результата исхода любой итерации перечислительного процесса исходов схемы будем далее обозначать все частицы i-го комплекта через i, где $i = \overline{1, m}$.

Таким образом, нумерация бесповторного перечисления исходов схемы B производится в порядке их формирования последовательными действиями по заданной процедуре.

Приведем числовой пример наглядного графического представления и записи бесповторного формирования с итогом перечисления всех разных исходов схемы.

Пример 3. Пусть $n = 4$, $r = 3$, $m = 2$. Тогда число размещений одного комплекта $C_4^3 = 4$, и они (в обозначениях наборов номеров непустых ячеек в порядке нумерации ячеек каждый и в порядке перечисления по методу графов [2]) представляют собой следующие 123, 124, 134, 234, что для i-го комплекта ($i = \overline{1, 2}$) в терминах п. 3.1 при перечислении заполнений $n = 4$ ячеек после его размещения соответствует исходам видов: $(\mathrm{i}, \mathrm{i}, \mathrm{i}, 0)$, $(\mathrm{i}, \mathrm{i}, 0, \mathrm{i})$, $(\mathrm{i}, 0, \mathrm{i}, \mathrm{i})$, $(0, \mathrm{i}, \mathrm{i}, \mathrm{i})$ в данном здесь порядке. Приведенный ниже граф перечисления последовательных исходов итераций схемы размещения частиц комплектами с заданными параметрами с учетом их суммирования по результатам размещения со всеми предшествующими комплектами представлен на следующем рисунке.



Граф перечисления исходов схемы
The graph of outcomes enumeration of the scheme

Здесь число N исходов схемы по рисунку совпадает с его теоретическим значением $N = (C_4^3)^2 = 16$.

3.2. Решение ЗН для исходов схемы B

Будем решать ЗН в схеме $B - S^*$ по МСС, где в качестве схемы S предлагается схема независимого последовательного m-итерационного размещения различимых комплектов частиц по ячейкам без нижнего ограничения на уровни заполнения ячеек. Тогда исходы нашей схемы S^* составляют часть исходов схемы S при выполнении ограничений на уровни заполнения ячеек $\geqslant k$.

Эта схема S соответствует схеме ПД с действиями – итерациями последовательного поединичного добавления результатов размещения по схеме сочетаний (см. [2, п. 1.3]) комплектов частиц по ячейкам, и в общем виде решена в п. 1.2.

Схема S – это частный случай схемы ПД с совпадающими размерами по C_n^r всех пучков на всех итерациях в графе процесса перечисления ее исходов.

Приведем числовой пример решения ЗН в схеме S в условиях примера 2 с наглядной проверкой по рисунку, считая решенной ЗН для действий – схем сочетания размещения каждого комплекта (см. п. 1.2).

Пример 4. Пусть $n = 4$, $r = 3$, $m = 2$. Тогда размеры всех пучков в графе перечисления исходов схемы равны $C_4^3 = 4$.

Прямая ЗН. Пусть дан номер исхода схемы $N^* = N^{(2)} = 14$. Найти его вид $R^* = R^{(2)}$.

Граф перечисления исходов схемы S в условиях примера приведен на рисунке.

Шаги решения:

1) по данному $N^{(2)} = 14$ находим по п. 1.3 номер искомого исхода схемы в пучке 2-й итерации $j_2 = 14 \ mod \ 4 = 2$, чему соответствует добавляемый результат размещения второго комплекта по ячейкам вида $R^{(2)} = ((2),(2),(0),(2))$;

2) по данному $N^{(2)} = 14$ находим по п. 1.3 номер предшествующего ему исхода на первой итерации $N^{(1)} = [(14 + 4 - 1)/4] = 4$, откуда его номер в пучке первой итерации $j_1 = 4 \ mod \ 4 + 4 = 4$, чему соответствует его вид $R^{(1)} = ((0),(1),(1),(1))$;

3) тогда, суммируя результаты независимых размещений двух комплектов, получаем итоговый искомый вид исхода схемы $R^* = ((2),(1,2),(1),(1,2))$, что совпадает с результатом по рисунку.

Обратная ЗН. Пусть дан вид исхода схемы $R^* = R^{(2)} = ((2),(1,2),(1),(1,2))$. Найти его номер $N^* = N^{(2)}$.

Шаги решения:

1) по данному $R^{(2)} = ((2),(1,2),(1),(1,2))$ находим $R^{(1)} = ((0),(1),(1),(1))$, а размещение второго комплекта по ячейкам имеет вид: $((2),(2),(0),(2))$, откуда $j_1 = 4,\ j_2 = 2$;

2) по п. 1.3 получаем $N^* = N^{(2)} = (4 - 1)4 + 2 = 14$, что совпадает с результатом по рисунку.

Таким образом, ЗН в СС для схемы B решена. Пересчет ее результатов в схеме B производится по МСС.

3.3. Распределение вероятностей на множестве исходов схемы B

а) При заданном вероятностном распределении исходов СС для схемы B (размещения частиц комплектами с тем же качеством ее элементов, что и в схеме B без нижнего ограничения на уровни заполнения ячеек) по описанной процедуре МСС находятся вероятности исходов схемы B;

б) при заданных итерационных переходных вероятностей процесса размещения частиц комплектами по ячейкам в схеме B вероятностное распределение ее итоговых исходов находится как вероятности их траекторий.

3.4. Моделирование исхода схемы B

Моделирование исхода схемы B предлагается проводить приемом УМ (см. [2]) путем разыгрывания его номера одним случайным числом с установленным вероятностным распределением и по нему и по результату решения ЗН находить вид смоделированного исхода схемы.

Литература

1. *Колчин В. Ф., Севастьянов Б. А., Чистяков В. П.* Случайные размещения. М.: Наука, 1976. 223 с.

2. *Энатская Н. Ю.* Доасимптотический анализ комбинаторных схем. М.: URSS, 2023. 536 с.

References

1. *Kolchin V. F., Sevastyanov B. A., Chistyakov V. P.* Random allocations. Moscow: Nauka; 1976. 223 p. (In Russ.)

2. *Enatskaya N. Yu.* Pre-asymptotic analysis of combinatorial schemes. Moscow: URSS; 2023. 536 p. (In Russ.)

Поступила в редакцию / received: 22.12.2024; принята к публикации / accepted: 10.04.2025.

Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Энатская Наталия Юрьевна

канд. физ.-мат. наук, доцент Департамента прикладной математики

e-mail: nat1943@mail.ru

CONTRIBUTOR:

Enatskaya, Natalia

Cand. Sci. (Phys.-Math.), Associate Professor

Труды Карельского научного центра РАН. 2025. № 4. С. 107–113
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 107–113
DOI: 10.17076/mat2073

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.175.4

ON THE CLUSTERING COEFFICIENT OF CONFIGURATION GRAPHS

I. A. Cheplyukova

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

Configuration graphs with N vertices are considered. The degrees of the vertices are independent identically distributed random variables following the power-law distribution with a positive parameter τ, where $\tau = \tau(N)$ varies as $0 < c_1 \leqslant \tau \leqslant c_2 < \infty$ and can take both fixed values and non-fixed values. Theorems describing the limit behaviour of the clustering coefficient for such graphs as $N \to \infty$ are formulated.

Keywords: configuration graph; clustering coefficient; limit theorems

For citation: Cheplyukova I. A. On the clustering coefficient of configuration graphs. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2025. No. 4. P. 107–113. doi: 10.17076/mat2073

Funding. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

И. А. Чеплюкова. О КОЭФФИЦИЕНТЕ КЛАСТЕРИЗАЦИИ КОНФИГУРАЦИОННЫХ ГРАФОВ

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Рассматриваются конфигурационные графы с N вершинами. Степени вершин графа являются независимыми одинаково распределенными случайными величинами, распределение которых является степенным распределением с положительным параметром τ, где $\tau = \tau(N)$ изменяется в диапазоне $0 < c_1 \leqslant \tau \leqslant c_2 < \infty$ и может принимать не только фиксированные значения. Получены теоремы, описывающие предельное поведение коэффициента кластеризации для таких графов с числом вершин N, стремящимся к бесконечности.

Ключевые слова: конфигурационный граф; кластерный коэффициент; предельные теоремы

Для цитирования: Cheplyukova I. A. On the clustering coefficient of configuration graphs // Труды Карельского научного центра РАН. 2025. № 4. С. 107–113. doi: 10.17076/mat2073

Ф и н а н с и р о в а н и е. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

Introduction

Random graphs have been widely used to model complex communication networks such as mobile networks, Internet, transport, social networks, etc. (e. g. [5, 7]). One of the most commonly used classes of random graphs is the configuration model. This model was first introduced in [3].

This paper deals with the configuration graph proposed in [11] with the vertex degrees being independent identically distributed random variables following the distribution:

$$\mathbf{P}\{\xi \geqslant k\} = \frac{h(k)}{k^{\tau}},\ \tau \in (1,2),\ k = 1, 2, \dots, \quad (1)$$

where the random variable ξ is equal to the degree of any vertex and $h(x)$ is a slowly varying function, i. e. $h(ax)/h(x) \to 1$ for every $a > 0$ and $x \to \infty$.

The construction of the configuration model can be described as follows. Random variables equal to vertex degrees are drawn independently from the distribution (1). The degree of each vertex in the configuration graph is equal to the number of its incident semiedges. All semiedges are numbered in an arbitrary order. Obviously, the sum of vertex degrees has to be even. Otherwise, an auxiliary vertex with degree one is added. The graph is constructed by joining all semiedges pairwise equiprobably to form edges. Because pairing is done without restrictions, multiple edges and loops can appear.

There are many works (e. g. [4, 11]) where results describing the limit behaviour of configuration graphs were obtained. Attention in the studies of configuration graphs has been given not only to the properties of the degree structure, but also to other numerical characteristics (see [8]). One of such graph characteristics is the clustering coefficient.

For the graph G the clustering coefficient C_G can be defined as follows (see [8])

$$C_G = \frac{3 \times \textit{number of graph triangles}}{\textit{number of connected triples of vertices}},$$

where a "connected triple" means a single vertex connected by edges to two others. In effect, C_G measures the fraction of triples that have their third edge filled in to complete a triangle. Here we define this notion for random graphs. We will use the terminology adopted in [7]. Consider the graph $G = G(V,E)$ with N vertices where V is the set of its vertices and E is the set of its edges. We say that the distinct vertices, (i,j,k) form an occupied triangle when the edges ij, jk and ki are all occupied. Note that (i,j,k) is the same triangle as (i,k,j) and as any other permutation. Following [7] (see equation (4.7.1)), we define the clustering coefficient of a random graph G to be

$$C_G = \frac{\mathbf{E}(\Delta_G)}{\mathbf{E}(W_G)},$$

where

$$\Delta_G = \sum_{i,j,k \in V} I\{ij, jk, ki \text{ occupied}\},$$

$$W_G = \sum_{i,j,k \in V} I\{ij, ik \text{ occupied,}\},$$

$I\{A\}$ is the indicator of an event A. Thus, Δ_G is six times the number of triangles in G, and W_G is two times the number of adjacent edges in G, and C_G is the ratio of the number of expected triangles to the expected number of adjacent edges. Note that the number of triangles formed by three vertices can be more than one if the edges between the vertices are multiples.

It is proved in [8] that for this configuration graph the clustering coefficient C_G is

$$C_G = \frac{\left(\sum\limits_{k=1}^{\infty} k q_k\right)^2}{Nm}, \quad (2)$$

where N is equal to the number of vertices of the graph,

$$p_k = \mathbf{P}\{\xi = k\},$$
$$q_k = (k+1)p_{k+1}/m, \qquad k = 0, 1, 2, \dots,$$
$$m = \sum_{k=1}^{\infty} k p_k.$$

It was noted in [11] that the function $h(k)$ does not affect the main asymptotic properties of the configuration graph as $N \to \infty$. So, the authors of [11] suggest to use the simplest case $h(k) = 1$. In this case

$$p_k = \mathbf{P}\{\xi = k\} = k^{-\tau} - (k+1)^{-\tau}, \quad k = 1, 2, \dots. \quad (3)$$

It follows from (3) that as $k \to \infty$

$$p_k \sim \frac{\tau}{k^{\tau+1}}. \quad (4)$$

In [9] Pavlov considered a configuration graph in which the distribution of vertex degrees must meet only the following condition as $k \to \infty$:

$$\mathbf{P}\{\xi = k\} \sim \frac{d}{k^g \ln^h k}, \tag{5}$$

where $d > 0$, $g > 1$, $h \geqslant 0$. Obviously, by virtue of (4), the distribution (3) satisfies the condition (5) as $d = \tau$, $g = \tau + 1$, $h = 0$. In [9], the limit theorems for the clustering coefficient C_G of such graphs with fixed parameters g and h were formulated. However, it was noted that the vertex degree distribution can change as the network size grows (e. g. [2]). In [10], conditional configuration graphs are considered where the degrees of the vertices are independent identically distributed random variables following the power-law distribution and the parameter of this distribution is a random variable uniformly distributed on the interval $[a, b]$, $0 < a < b < \infty$. The limit distributions of the number of vertices with a given degree were obtained in [10].

Here we consider a configuration graph in which vertex degrees have the distribution (3), where the distribution parameter τ can vary in the interval $0 < c_1 \leqslant \tau \leqslant c_2 < \infty$ and is not necessarily fixed. The aim is to study the limit behavior of the clustering coefficient C_G for such configuration graphs (Theorem 1 and Theorem 2).

Theorem 1 describes the limit behavior of the clustering coefficient C_G for the case of the parameter $\tau > 2$ as $N \to \infty$. In this case, the variance $\mathbf{D}\xi$ of a random variable equal to the vertex degree is finite. Theorem 2 describes the limit behavior of C_G as $0 < c_1 \leqslant \tau \leqslant 2$. It is not hard to see that in this case the variance $\mathbf{D}\xi$ is infinite. Here, we will use the notion "asymptotically almost sure" (a.a.s.), which means the following. Let A_N be an event of a random graph with N vertices having a certain property. We say that A happens a.a.s., if

$$\lim_{N\to\infty} \mathbf{P}\{A_N\} = 1.$$

Let the random variables $\xi_1, \xi_2, \ldots, \xi_N$ be equal to the degrees of vertices with numbers $1, 2, \ldots, N$ and

$$\xi_{(N)} = \max\{\xi_1, \ldots, \xi_N\}.$$

Observe that the maximum vertex degree of our graphs is proportional to $N^{1/\tau}$ a.a.s. (see Lemma below). For the case of $0 < c_1 \leqslant \tau \leqslant 2$ we restrict ourselves to considering random graphs that satisfy this condition. Namely, in Theorem 2 we will consider conditional configuration graphs under the condition

$$\xi_{(N)} \leqslant uN^{1/\tau}, \qquad 0 < u < \infty.$$

Theorem 2 shows that the asymptotically clustering coefficient of a random graph depends on the maximum vertex degree.

The article has the following structure. Section 2 formulates the main results (Theorems 1 and 2). In section 3 these theorems are proved.

Main results

Consider a configuration graph in which the vertex degrees $\xi_1, \xi_2, \ldots, \xi_N$ are independent identically distributed random variables with the distribution (3). If the parameter $\tau > 2$, then the following assertion is true.

Theorem 1. *Let $N \to \infty$. Then the relations*

$$C_G = \begin{cases} 4\left(\zeta(\tau-1) - \zeta(\tau)\right)^2 / (N\zeta^3(\tau)), \\ \qquad\qquad \text{if} \quad \tau = \tau(N) \geqslant c_3 > 2; \\ 4(\tau-2)^{-2}/(N\zeta^3(\tau))(1+o(1)), \\ \qquad\qquad \text{if} \quad \tau = \tau(N) \searrow 2, \end{cases}$$

hold, where $\zeta(x)$ is the value of the Riemann zeta function at the point x.

Corollary 1. *It follows from Theorem 1 that*

1. *if $\tau \geqslant c_3 > 2$ or $\tau = \tau(N) \searrow 2$, $(\tau-2)^2 N \to \infty$, then $C_G \to 0$;*
2. *if $\tau = \tau(N) \searrow 2$, $(\tau-2)^2 N \to const \neq 0$, then C_G tends to some positive constant;*
3. *if $\tau = \tau(N) \searrow 2$, $(\tau-2)^2 N \to 0$, then $C_G \to \infty$.*

Let $0 < c_1 < \tau \leqslant 2$. Consider conditional configuration graphs under the condition that

$$\xi_{(N)} \leqslant uN^{1/\tau}, 0 < u < \infty.$$

For such conditional graphs the assertions below are true.

Theorem 2. *Let $N \to \infty$. Then*

1. *if $\tau = 2$, then*

$$C_G = \frac{e^{-u^{-2}}}{\zeta^3(2)} \frac{\ln^2 N}{N} \qquad a.a.s.;$$

2. *if $\tau \nearrow 2$, then*

$$C_G = \frac{4\exp\{-u^{-2}\}(N^{1-\tau/2} - 2/\tau)^2}{\zeta^3(2)(2-\tau)^2 N}$$

a.a.s.;

3. *if* $1 < c_4 \leqslant \tau \leqslant c_5 < 2$, *then*

$$C_G = \frac{\exp\{-u^{-\tau}\}}{\zeta^3(\tau)} \left(\frac{u^{2-\tau}\tau}{2-\tau}\right)^2 N^{(4-3\tau)/\tau}$$

a.a.s.;

4. *if* $\tau \searrow 1$, *then*

$$C_G = \exp\{-u^{-1}\}(\tau-1)^3 u^2 N^{(4-3\tau)/\tau}$$

a.a.s.;

5. *if* $\tau = 1$, *then*

$$C_G = \exp\{-u^{-1}\}\frac{u^2 N}{\ln^3 N} \qquad a.a.s.;$$

6. *if* $0 < c_1 \leqslant \tau < 1$, *then*

$$C_G = \frac{e^{-u^{-\tau}}(1-\tau)^3 u^{1+\tau} N^{1/\tau}}{\tau(2-\tau)^2}$$

a.a.s.

Corollary 2. *It follows from Theorem 2 that the following assertions are true a.a.s.*

1. *if* $(\tau - 4/3)\ln N \to +\infty$, *then* $C_G = 0$;
2. *if* $|(\tau - 4/3)\ln N| \leqslant c_6 < \infty$, *then* C_G *is equal to the constant and this constant depends on* u;
3. *if* $(\tau - 4/3)\ln N \to -\infty$, *then* $C_G = \infty$.

Proofs of the main results

First, we will prove Theorem 1. Let $\tau \geqslant c_3 > 2$. In this case, $m = \zeta(\tau)$ and it follows from (2) that

$$C_G = \frac{\left(\sum\limits_{k=1}^{\infty} k q_k\right)^2}{\zeta(\tau)N} = \frac{1}{\zeta^3(\tau)N}$$
$$\times \left(\sum_{k=1}^{\infty}(k^2+k)\left(\frac{1}{(k+1)^\tau} - \frac{1}{(k+2)^\tau}\right)\right)^2$$
$$= \frac{1}{\zeta^3(\tau)N}\left(\sum_{k=2}^{\infty}(2k-2)\frac{1}{k^\tau}\right)^2. \qquad (6)$$

This yields the assertion of Theorem 1 for the case $\tau \geqslant c_3 > 2$.

Let $\tau = \tau(N) \searrow 2$. We will use the known expansion of the zeta-function about point 1 (see [1]):

$$\zeta(1+y) = y^{-1} + c + O(y), \quad y > 0,\ y \to 0, \ (7)$$

where c is the Euler–Mascheroni constant. It follows from here and (6) that

$$C_G = \frac{4}{\zeta^3(\tau)N}(\zeta(\tau-1) - \zeta(\tau))^2$$
$$= \frac{4\zeta^2(\tau-1)}{\zeta^3(\tau)N}(1+o(1))$$
$$= \frac{4\left((\tau-2)^{-1} + c + O(\tau-2)\right)^2}{\zeta^3(\tau)N}(1+o(1))$$
$$= \frac{4(\tau-2)^{-2}}{\zeta^3(\tau)N}(1+o(1)).$$

Therefore, Theorem 1 is fully proven.

Let us now prove Theorem 2. Here we consider the conditional configuration graphs under the condition that

$$\xi_{(N)} \leqslant uN^{1/\tau}, 0 < u < \infty.$$

Clearly, for a detailed study of the clustering coefficient C_G it is desirable to know the limit distribution of the maximum degree of graph vertices. The next Lemma follows from (3) and the known classical result [6].

Lemma. *Let* $N \to \infty$. *Then*

$$\mathbf{P}\{\xi_{(N)} \leqslant xN^{1/\tau}\} = e^{-x^{-\tau}} + o(1), \quad x > 0$$

holds.

According to Lemma for the maximal degree $\xi_{(N)}$, the inequation $\xi_{(N)} > N^{1/(\tau+\varepsilon)}$, $\varepsilon > 0$ holds a.a.s. Moreover, it follows from Lemma that

$$\mathbf{P}\left\{N^{1/(\tau+\varepsilon)} \leqslant \xi_{(N)} \leqslant \frac{1}{\delta}N^{1/\tau}\right\}$$

can be made arbitrarily close to 1 by choosing sufficiently small positive ε and δ. We can then say that Lemma predicates that the maximum vertex degree $\xi_{(N)}$ is proportional to $N^{1/\tau}$ a.a.s. It thus seems reasonable to consider the set of graphs in which $\xi_{(N)} \leqslant uN^{1/\tau}$, where

$$N^{-\frac{\varepsilon}{\tau(\tau+\varepsilon)}} < u < 1/\delta.$$

So, by virtue of Lemma, the condition for the maximal degree of our graphs in Theorem 2 is natural.

It follows from (2) that the clustering coefficient C_G for such conditional configuration graphs can be obtained from the following relation:

$$C_G = \frac{\left(\sum\limits_{k=1}^{\infty} k q'_k\right)^2}{Nm'}, \qquad (8)$$

where

$$p'_k = \mathbf{P}\{\xi = k | \xi_{(N)} \leqslant uN^{1/\tau}\},$$

$$q_k' = (k+1)p_{k+1}'/m', \quad k = 0,1,2,\ldots,$$

$$m' = \sum_{k=1}^{\infty} kp_k'.$$

Then, combining (3) and (8) we get

$$C_G = \frac{1}{N}\mathbf{P}\left\{\xi_{(N)} \leqslant uN^{1/\tau}\right\}$$
$$\times \frac{\left(\sum_{k=1}^{[uN^{1/\tau}]-1} k(k+1)p_{k+1}\right)^2}{\left(\sum_{k=1}^{[uN^{1/\tau}]} kp_k\right)^3}, \tag{9}$$

where $[x]$ is the integer part of the number x.

Using Lemma, it is not difficult to see that

$$\mathbf{P}\left\{\xi_{(N)} \leqslant uN^{1/\tau}\right\} = e^{-u^{-\tau}}(1+o(1)). \tag{10}$$

From (9) and (10) it follows that

$$C_G = \frac{e^{-u^{-\tau}}}{N} \tag{11}$$
$$\times \frac{\left(\sum_{k=1}^{[uN^{1/\tau}]-1} k(k+1)p_{k+1}\right)^2}{\left(\sum_{k=1}^{[uN^{1/\tau}]} kp_k\right)^3}(1+o(1)).$$

Using (3), we can show that

$$\sum_{k=1}^{[uN^{1/\tau}]} kp_k \tag{12}$$
$$= \sum_{k=1}^{[uN^{1/\tau}]} \frac{1}{k^\tau} - \frac{[uN^{1/\tau}]}{\left([uN^{1/\tau}]+1\right)^\tau}$$

and

$$\sum_{k=1}^{[uN^{1/\tau}]-1} k(k+1)p_{k+1} \tag{13}$$
$$= \sum_{k=2}^{[uN^{1/\tau}]} (2k-2)\frac{1}{k^\tau} - \frac{\left([uN^{1/\tau}]-1\right)^2}{\left([uN^{1/\tau}]+1\right)^\tau}$$
$$-\frac{[uN^{1/\tau}]-1}{\left([uN^{1/\tau}]+1\right)^\tau}.$$

Let $\tau = 2$. From (13) it follows that

$$\sum_{k=1}^{[uN^{1/2}]-1} k(k+1)p_{k+1} \tag{14}$$
$$= 2\sum_{k=2}^{[uN^{1/2}]} \frac{1}{k} - 2\sum_{k=2}^{[uN^{1/2}]} \frac{1}{k^2} - \frac{\left([uN^{1/2}]-1\right)^2}{\left([uN^{1/2}]+1\right)^2}$$
$$-\frac{[uN^{1/2}]-1}{\left([uN^{1/2}]+1\right)^2}.$$

Using a well-known formula

$$\sum_{k=1}^{[N]} \frac{1}{k} = \ln N + c + \varepsilon(N) \tag{15}$$

where $\varepsilon(N) \sim 1/(2N)$, we get

$$\sum_{k=2}^{[uN^{1/2}]} \frac{1}{k} = \frac{1}{2}\ln N(1+o(1) \tag{16}$$

then from (11)–(16) it follows that for $\tau = 2$

$$C_G = \frac{\exp\{-u^{-2}\}\ln^2 N}{\zeta^3(2)N}(1+o(1)). \tag{17}$$

Let $\tau \nearrow 2$. Using (13), we get

$$\sum_{k=1}^{[uN^{1/\tau}]-1} k(k+1)p_{k+1} \tag{18}$$
$$= 2\sum_{k=1}^{[uN^{1/\tau}]} \frac{1}{k^{\tau-1}}(1+o(1)) - \left[uN^{1/\tau}\right]^{2-\tau}(1+o(1)).$$

It is easy to see that

$$\sum_{k=2}^{[uN^{1/\tau}]} \frac{1}{k^{\tau-1}} = (1+o(1))N^{(2-\tau)/\tau} \int\limits_{2/N^{1/\tau}}^{u} y^{-\tau+1}dy$$
$$= (1+o(1))\frac{N^{(2-\tau)/\tau}}{2-\tau}\left(u^{2-\tau} - \frac{2^{2-\tau}}{N^{(2-\tau)/\tau}}\right)$$
$$= \frac{1}{2-\tau}\exp\left\{\frac{2-\tau}{\tau}\ln N\right\} \tag{19}$$
$$\times\left(1 - \exp\left\{-\frac{2-\tau}{\tau}\ln N\right\}\right)(1+o(1)).$$

This relation together with (11), (12) and (18) yields

$$C_G = \frac{\exp\{-u^{-2}\}}{\zeta^3(2)N}$$
$$\times\left(\frac{2(\exp\{\frac{2-\tau}{\tau}\ln N\}-1)(1+o(1))}{2-\tau}\right.$$
$$\left.-[uN^{1/\tau}]^{2-\tau}(1+o(1))\right)^2$$
$$= \frac{\exp\{-u^{-2}\}(1+o(1))}{\zeta^3(2)N}\left(2\frac{\exp\{\frac{2-\tau}{\tau}\ln N\}-1}{2-\tau}\right.$$

$$- \exp\left\{\frac{2-\tau}{\tau}\ln N\right\}\Bigg)^2$$

$$= \frac{\exp\{-u^{-2}\}\left(-2+\tau\exp\left\{\frac{2-\tau}{\tau}\ln N\right\}\right)^2}{(2-\tau)^2\zeta^3(2)N}(1+o(1))$$

$$= \frac{4e^{-u^{-2}}(N^{1-\tau/2}-2/\tau)^2}{(2-\tau)^2\zeta^3(2)N}(1+o(1)). \qquad (20)$$

Let $1 < c_4 \leqslant \tau \leqslant c_5 < 2$. Using relations (13), (18), and the first equality of (19) we get that

$$\sum_{k=1}^{[uN^{1/\tau}]-1} k(k+1)p_{k+1}$$

$$= \frac{\tau}{2-\tau}u^{2-\tau}N^{(2-\tau)/\tau}(1+o(1)).$$

From here and (11) it is easy to see that

$$C_G = \left(e^{-u^{-\tau}}\zeta^3(\tau)N\right)^{-1}\Bigg(2N^{(2-\tau)/\tau}\frac{u^{2-\tau}}{(2-\tau)}$$

$$-(uN^{1/\tau})^{2-\tau}\Bigg)^2(1+o(1)) \qquad (21)$$

$$= \left(e^{-u^{-\tau}}\zeta^3(\tau)\right)^{-1}N^{(4-3\tau)/\tau}\left(\frac{u^{2-\tau}\tau}{2-\tau}\right)^2(1+o(1)).$$

Let $\tau \searrow 1$. In the same way as in the previous case, using the known expansion of the zeta-function about point 1 (7) we get that

$$C_G = \left(e^{-u^{-\tau}}\zeta^3(\tau)N\right)^{-1}$$

$$\times\left(2N^{(2-\tau)/\tau}\frac{u^{2-\tau}}{(2-\tau)} - (uN^{1/\tau})^{2-\tau}\right)^2(1+o(1))$$

$$= \frac{e^{-u^{-1}}(\tau-1)^3}{N}\left(u^{2-\tau}N^{(2-\tau)/\tau}\right)^2(1+o(1))$$

$$= e^{-u^{-1}}(\tau-1)^3u^2N^{(4-3\tau)/\tau}(1+o(1)). \qquad (22)$$

Let $\tau = 1$. From (12) and (15) we get that

$$\sum_{k=1}^{[uN]} kp_k = \sum_{k=1}^{[uN]}\frac{1}{k} - 1 + o(1)$$

$$= (1+o(1))\ln N. \qquad (23)$$

Using (3) and (15) we can find that

$$\sum_{k=1}^{[uN]-1} k(k+1)p_{k+1} = \sum_{k=1}^{[uN]-1}\left(1-\frac{2}{k+2}\right)$$

$$= uN(1+o(1)).$$

From this and (11), (23) it follows that

$$C_G = e^{-1/u}\frac{u^2N}{\ln^3 N}(1+o(1)). \qquad (24)$$

Let us consider the last case where $0 < c_1 \leqslant \tau < 1$. It is not hard to see that in this case

$$\sum_{k=1}^{[uN^{1/\tau}]}\frac{1}{k^\tau} = N^{-(\tau-1)/\tau}\int\limits_{1/N^{1/\tau}}^{u}\frac{1}{y^\tau}dy(1+o(1))$$

$$= \frac{u^{1-\tau}N^{(1-\tau)/\tau}}{1-\tau}(1+o(1)). \qquad (25)$$

From this and (12) we get that

$$\sum_{k=1}^{[uN^{1/\tau}]} kp_k = \sum_{k=1}^{[uN^{1/\tau}]}\frac{1}{k^\tau} - \frac{[uN^{1/\tau}]}{\left([uN^{1/\tau}]+1\right)^\tau}$$

$$= u^{1-\tau}N^{(1-\tau)/\tau}\frac{\tau}{1-\tau}(1+o(1)). \qquad (26)$$

Using (13), the second equality of (19), and (25) we get that

$$\sum_{k=1}^{[uN^{1/\tau}]} k(k+1)p_{k+1} = 2\sum_{k=2}^{[uN^{1/\tau}]}\frac{1}{k^{\tau-1}} - 2\sum_{k=2}^{[uN^{1/\tau}]}\frac{1}{k^\tau}$$

$$-\left(uN^{1/\tau}\right)^{2-\tau}(1+o(1))$$

$$= u^{2-\tau}N^{(2-\tau)/\tau}\frac{\tau}{2-\tau}(1+o(1)).$$

From (11), (26), and the previous relation we get

$$C_G = \frac{\exp\{-u^{-\tau}\}(1-\tau)^3u^{1+\tau}}{\tau(2-\tau)^2}N^{1/\tau}(1+o(1)).$$

Taking into account the above considerations on the maximum vertex degree, the assertion of Theorem 2 follows from the reasoning here and (17), (20)–(22), (24).

References

1. *Bateman H., Erdelyi A.* Higher transcendental functions. New York-Toronto-London: MC GRAW-HILL Book Company INC; 1953.

2. *Bianconi G., Barabasi A.-L.* Bose-Einstein condensation in complex networks. *Phys. Rev. Lett.* 2001;86:5632–5635. doi: 10.1103/PhysRevLett.86.5632

3. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Comb.* 1980;1(4):311–316. doi: 10.1016/S0195-6698(80)80030-8

4. *Durrett R.* Random graph dynamics. Cambridge: Cambridge University Press; 2007. 223 p. doi: 10.1017/CBO9780511546594

5. *Faloutsos C., Faloutsos P., Faloutsos M.* On power-law relationships of the internet topology. *Comput. Commun. Rev.* 1999;29(1):251–262. doi: 10.1145/316194.316229

6. *Gnedenko B.* Sur la distribution limite du terme maximum d'une série alétorie. *Ann. Math.* 1943;44(3):423–453. doi: 10.2307/1968974

7. *Hofstad R.* Random graphs and complex networks. Vol. 1. Cambridge: Cambridge University Press; 2017. 337 p. doi: 10.1017/9781316779422

8. *Newman M. E. J.* The structure and function of complex networks. *SIAM Rev.* 2003;45(2):167–256. doi: 10.1137/S003614450342480

9. *Pavlov Yu. L.* On the asymptotics of the clustering coefficient in a configuration graph with unknown distribution of vertex degrees. *Informatics and Applications.* 2019;13(3):9–13. (In Russ.). doi: 10.14357/19922264190302

10. *Pavlov Yu. L., Cheplyukova I. A.* On the asymptotics of degree structure of configuration graphs with bounded number of edges. *Discrete Mathematics and Applications.* 2019;29(4):219–232. doi: 10.1515/dma-2019-0020

11. *Reittu H., Norros I.* On the power-law random graph model of massive data networks. *Perform. Eval.* 2004;55(1-2):3–23. doi: 10.1016/S0166-5316(03)00097-X

Поступила в редакцию / received: 13.03.2025; принята к публикации / accepted: 14.05.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Чеплюкова Ирина Александровна
канд. физ.-мат. наук, старший научный сотрудник

e-mail: chia@krc.karelia.ru

CONTRIBUTOR:

Cheplyukova, Irina
Cand. Sci. (Phys.-Math.), Senior Researcher

Труды Карельского научного центра РАН. 2025. № 4. С. 114
Transactions of the Karelian Research Centre RAS. 2025. No. 4. P. 114

ПАМЯТИ СЕРГЕЯ ВЯЧЕСЛАВОВИЧА СТАФЕЕВА (1970–2025)

9 мая 2025 г. на 55-м году жизни скончался высококвалифицированный специалист в области теории вероятностей, математической и прикладной статистики Сергей Вячеславович Стафеев.

Он родился 30 октября 1970 г. в Петрозаводске. После окончания школы № 40 поступил на механико-математический факультет МГУ, который окончил в 1993 г. по специальности «теория вероятностей и математическая статистика». С 1993 г. работал в Карельском научном центре РАН и до конца жизни был научным сотрудником лаборатории теории вероятностей и компьютерной статистики Института прикладных математических исследований. В 2008 г. защитил диссертацию на тему «Идентифицируемость и обучение гауссовских графовых моделей с латентными переменными», и ему была присуждена ученая степень кандидата физико-математических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ.

Им опубликовано более 40 работ. Основные научные интересы С. В. Стафеева были связаны с построением и исследованием графовых моделей, позволяющих выдвигать и проверять гипотезы о структуре сложных объектов и о влиянии разнообразных взаимодействующих факторов на поведение таких объектов. Получен ряд новых результатов о моделях факторного анализа с зависимыми остатками, а также с латентными (скрытыми, не наблюдаемыми) переменными. Решалась проблема параметрической идентифицируемости таких моделей, состоящая в ответе на вопрос о возможности однозначного определения параметров по совместному распределению наблюдаемых случайных величин и, как следствие, о получении состоятельных оценок параметров.

Большое внимание С. В. Стафеев уделял использованию методов математической статистики для решения прикладных задач, возникающих в различных научных исследованиях. Совместно с профильными специалистами он был соавтором работ по медицине, экономике, зоологии и лесному хозяйству. Наиболее заметны в этом ряду результаты по археологии, опубликованные в ведущих международных журналах. В этих работах предложен статистический метод сравнения стадий обработки каменных орудий из различных стоянок эпохи неолита. Метод основан на сравнении фрактальных размерностей распределения отщепов готовых изделий. Было проведено статистическое сравнение стоянок западного побережья Онежского озера.

С. В. Стафеев отличался доброжелательностью к коллегам, был интеллигентным, скромным и порядочным человеком. Светлая память о нем надолго останется в сердцах знавших его людей.

Коллеги из ИПМИ КарНЦ РАН

ПРАВИЛА ДЛЯ АВТОРОВ

Серия «Математическое моделирование и информационные технологии»

(требования к работам, представляемым к публикации в «Трудах Карельского научного центра Российской академии наук»)

«Труды Карельского научного центра Российской академии наук» (далее – Труды КарНЦ РАН) публикуют результаты завершенных оригинальных исследований в различных областях современной науки: теоретические и обзорные статьи, сообщения, материалы о научных мероприятиях (симпозиумах, конференциях и др.), персоналии (юбилеи и даты, утраты науки), статьи по истории науки. Представляемые работы должны содержать новые, ранее не публиковавшиеся данные.

Статьи проходят обязательное рецензирование. Решение о публикации принимается редакционной коллегией серии или тематического выпуска Трудов КарНЦ РАН после рецензирования, с учетом научной значимости и актуальности представленных материалов. Редколлегии серий и отдельных выпусков Трудов КарНЦ РАН оставляют за собой право возвращать без регистрации рукописи, не отвечающие настоящим правилам.

При получении редакцией рукопись регистрируется (в случае выполнения авторами основных правил ее оформления) и направляется на отзыв рецензентам. Отзыв состоит из ответов на типовые вопросы анкеты и может содержать дополнительные расширенные комментарии. Кроме того, рецензент может вносить замечания и правки в текст рукописи. Авторам высылается электронная версия анкеты и комментарии рецензентов. Доработанный экземпляр автор должен вернуть в редакцию вместе с первоначальным экземпляром и ответом на все вопросы рецензента не позднее чем через месяц после получения рецензии. Перед опубликованием авторам высылается электронная версия статьи, которую авторы вычитывают и заверяют.

Журнал имеет систему электронной редакции на базе Open Journal System (OJS), позволяющую вести представление и редактирование рукописи, общение автора с редколлегиями серий и рецензентами в электронном формате и обеспечивающую прозрачность процесса рецензирования при сохранении анонимности рецензентов (http://journals.krc.karelia.ru/).

Содержание выпусков Трудов КарНЦ РАН, аннотации и полнотекстовые электронные версии статей, а также другая полезная информация, включая настоящие Правила, доступны на сайтах – http://transactions.krc.karelia.ru; http://journals.krc.karelia.ru

Почтовый адрес редакции: 185910, г. Петрозаводск, ул. Пушкинская, 11, КарНЦ РАН, редакция Трудов КарНЦ РАН. Телефон: (8142) 762018.

ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСИ

Статьи публикуются на русском или английском языке. Рукописи должны быть тщательно выверены и отредактированы авторами.

Объем рукописи (включая таблицы, список литературы, подписи к рисункам, рисунки) не должен превышать: для обзорных статей – 30 страниц, для оригинальных – 25, для сообщений – 15, для хроники и рецензий – 5–6. Объем рисунков не должен превышать 1/4 объема статьи. Рукописи большего объема (в исключительных случаях) принимаются при достаточном обосновании по согласованию с ответственным редактором.

При оформлении рукописи применяется полуторный межстрочный интервал, шрифт Times New Roman, кегль 12, выравнивание по обоим краям. Размер полей страницы – 2,5 см со всех сторон. Все страницы, включая список литературы и подписи к рисункам, должны иметь сплошную нумерацию в правом нижнем углу. Страницы с рисунками не нумеруются.

Рукописи подаются в электронном виде в систему электронной редакции на сайте http://journals.krc.karelia.ru либо высылаются на e-mail: trudy@krc.karelia.ru, или же представляются в редакцию лично (г. Петрозаводск, ул. Пушкинская, 11, каб. 502).

Для публикации в выпусках серии «Математическое моделирование и информационные технологии» рукописи принимаются в формате .tex (LaTex 2є) с использованием стилевого файла, который находится по адресу: http://transactions.krc.karelia.ru/section.php?id=755. Статья в файле с расширением .pdf загружается на сайт журнала http://journals.krc.karelia.ru. Исходный файл с расширением .tex и необходимые рисунки загружаются на 4-м шаге «Загрузка дополнительных файлов».

Обязательные элементы рукописи располагаются в следующем порядке:

УДК курсивом в левом верхнем углу первой страницы; заглавие статьи на русском языке полужирным шрифтом; инициалы и фамилии авторов на русском языке полужирным шрифтом; полное название и полный почтовый адрес организации – места работы каждого автора в именительном падеже на русском языке курсивом (если авторов несколько и работают они в разных учреждениях, следует отметить арабскими цифрами соответствие фамилий авторов аффилированным организациям; автора, ответственного за переписку, следует отметить звездочкой и указать в аффилиации его электронный адрес); аннотация на русском языке; ключевые слова на русском языке; указание источников финансирования выполненных исследований на русском языке.

Далее располагаются все вышеуказанные элементы на английском языке.

Текст статьи (статьи экспериментального характера, как правило, должны иметь разделы: **Введение. Материалы и методы. Результаты и обсуждение. Выводы** либо **Заключение**); благодарности; списки литературы на языке оригинала (**Литература**) и на английском языке (**References**); таблицы на русском и английском языках (на отдельных листах); рисунки (на отдельных листах); подписи к рисункам на русском и английском языках (на отдельном листе).

На отдельном листе дополнительные сведения об авторах: фамилии, имена, отчества всех авторов полностью на русском и английском языке; должности, ученые звания, ученые степени авторов; адрес электронной почты каждого автора; можно указать телефон для контакта редакции с авторами статьи.

ЗАГЛАВИЕ СТАТЬИ должно точно отражать ее содержание и состоять из 8–10 значащих слов.

АННОТАЦИЯ должна быть лишена вводных фраз, создавать возможно полное представление о содержании статьи и иметь объем не менее 200 слов. Рукопись с недостаточно раскрывающей содержание аннотацией может быть отклонена.

Отдельной строкой приводится перечень КЛЮЧЕВЫХ СЛОВ (как правило, не менее пяти). Ключевые слова или словосочетания отделяются друг от друга точкой с запятой, в конце точка не ставится.

Раздел «Материалы и методы» должен содержать сведения об объекте исследования с обязательным указанием латинских названий и сводок, по которым они приводятся, авторов классификаций и пр. Транскрипция географических названий должна соответствовать атласу последнего года издания. Единицы физических величин приводятся по Международной системе СИ. Желательна статистическая обработка всех количественных данных. Необходимо возможно точнее обозначать местонахождения (в идеале – с точным указанием географических координат).

Изложение результатов должно заключаться не в пересказе содержания таблиц и графиков, а в выявлении следующих из них закономерностей. Автор должен сравнить полученную им информацию с имеющейся в литературе и показать, в чем заключается ее новизна. На табличный и иллюстративный материал следует ссылаться так: на рисунки, фотографии и таблицы в тексте (рис. 1, рис. 2, табл. 1, табл. 2 и т.д.), фотографии, помещаемые на вклейках (рис. I, рис. II). Обсуждение завершается формулировкой в разделе «Заключение» основного вывода, которая должна содержать конкретный ответ на вопрос, поставленный во «Введении». Ссылки на литературу в работах **серии «Математическое моделирование и информационные технологии»** даются цифрами.

ТАБЛИЦЫ нумеруются в порядке упоминания их в тексте, каждая таблица имеет свой заголовок. Заголовки таблиц, заголовки и содержание столбцов, строк, а также примечания приводятся на русском и английском языках. Диаграммы и графики не должны дублировать таблицы. Материал таблиц должен быть понятен без дополнительного обращения к тексту. Все сокращения, использованные в таблице, поясняются в Примечании, расположенном под ней. При повторении цифр в столбцах нужно их повторять, при повторении слов – в столбцах ставить кавычки. Таблицы могут быть книжной или альбомной ориентации.

РИСУНКИ представляются отдельными файлами с расширением TIFF (*.TIF) или JPG. При первичной подаче материала в редакцию рисунки вставляются в общий текстовый файл. При сдаче материала, принятого в печать, все рисунки должны быть представлены в виде отдельных файлов в вышеуказанном формате. Графические материалы могут быть снабжены указанием желательного размера рисунка, пожеланиями и требованиями к конкретным иллюстрациям. На каждый рисунок должна быть как минимум одна ссылка в тексте.

ПОДПИСИ К РИСУНКАМ приводятся на русском и английском языках, должны содержать достаточную информацию для того, чтобы приводимые данные могли быть понятны без обращения к тексту (если эта информация уже не дана в другой иллюстрации). Аббревиации расшифровываются в подрисуночных подписях, детали на рисунках следует обозначать цифрами или буквами, значение которых также приводится в подписях.

СОКРАЩЕНИЯ. Разрешаются лишь общепринятые сокращения – названия мер, физических, химических и математических величин и терминов и т. п. Все прочие сокращения должны быть расшифрованы, за исключением небольшого числа общеупотребительных.

БЛАГОДАРНОСТИ. Располагаются после основного текста статьи отдельным абзацем, в котором авторы выражают признательность частным лицам, сотрудникам учреждений и организациям, оказавшим содействие в проведении исследований и подготовке статьи.

ИНФОРМАЦИЯ О КОНФЛИКТЕ ИНТЕРЕСОВ. При подаче статьи авторы должны раскрыть потенциальные конфликты интересов, которые могут быть восприняты как оказавшие влияние на результаты или выводы, представленные в работе. Если конфликт интересов отсутствует, следует об этом сообщить в отдельной формулировке.

СПИСОК ЛИТЕРАТУРЫ оформляется по ГОСТ Р 7.0.5-2008. Источники располагаются в алфавитном порядке. Все ссылки даются на языке оригинала (названия на японском, китайском и других языках, использующих нелатинский шрифт, пишутся в русской транскрипции). Сначала приводится список работ на русском языке и на языках с близким алфавитом (украинский, болгарский и др.), а затем – работы на языках с латинским алфавитом. В списке литературы между инициалами авторов ставится пробел.

REFERENCES. Приводится отдельным списком, повторяя все позиции основного списка литературы. Библиографические записи источников оформляются согласно стилю Vancouver (см. примеры в ГОСТ Р 7.0.7–2021 и образцы ниже). Заголовки русскоязычных работ приводятся на английском языке; для журналов и сборников, в которых размещены цитируемые работы, указывается параллельное английское наименование (при его наличии) либо русскоязычное наименование приводится в латинской транслитерации (вариант BSI) с переводом на английский язык. Прочие элементы библиографической записи приводятся на английском языке (русскоязычное название издательства транслитерируется). При наличии переводной версии источника в References желательно указать ее. Библиографические описания прочих работ приводятся на языке оригинала.

Для каждого источника обязательно указание DOI при его наличии; если приводится адрес интернет-страницы источника (URL), нужно указать дату обращения к ней.

Transactions of the Karelian Research Centre of the Russian Academy of Sciences
No. 4, 2025
MATHEMATICAL MODELING AND INFORMATION TECHNOLOGIES

TABLE OF CONTENTS

Научный журнал

Труды Карельского научного центра
Российской академии наук
№ 4, 2025

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Печатается по решению Ученого совета
Федерального исследовательского центра
«Карельский научный центр Российской академии наук»

Выходит 8 раз в год

Издание зарегистрировано Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций
Регистрационная запись ПИ № ФС 77-72429 от 28.02.2018 г.

Редактор А. И. Мокеева
Компьютерная верстка Л. Э. Бюркланд

Подписано в печать 24.06.2025. Дата выхода 30.06.2025. Формат 60х84 1/8.
Печать офсетная. Уч.-изд. л. 10,0. Усл. печ. л. 13,8.
Тираж 100 экз. Заказ 856. Цена свободная

Учредитель и издатель: Федеральное государственное бюджетное учреждение науки
Федеральный исследовательский центр «Карельский научный центр Российской академии наук»
185910, г. Петрозаводск, ул. Пушкинская, 11

Оригинал-макет: Редакция научного издания «Труды КарНЦ РАН»

Типография: Редакционно-издательский отдел КарНЦ РАН
185030, г. Петрозаводск, пр. А. Невского, 50