Федеральный исследовательский центр
«Карельский научный центр Российской академии наук»
Институт прикладных математических исследований
КарНЦ РАН

**X Международная
Петрозаводская конференция**

Вероятностные методы в дискретной математике

22 — 26 мая 2019

г. Петрозаводск, Россия

РАСШИРЕННЫЕ ТЕЗИСЫ

Петрозаводск 2019

Federal Research Centre
«Karelian Research Centre of the Russian Academy of Sciences»
Institute of Applied Mathematical Research of KRC RAS

X International
Petrozavodsk Conference

Probabilistic Methods in Discrete Mathematics

May 22 — 26, 2019,
Petrozavodsk, Russia

EXTENDED ABSTRACTS

Petrozavodsk 2019

УДК 519.1(063) + 519.2(063)
ББК 22.17
В35

Научный редактор *В. В. Мазалов*
Editor *V. V. Mazalov*

Ответственный редактор *Е. В. Хворостянская*
Responsible editor *E. V. Khvorostyanskaya*

Вероятностные методы в дискретной математике, X Международная
В35 Петрозаводская конференция (22 – 26 мая 2019 г., Петрозаводск, Россия): расширенные тезисы / ИПМИ КарНЦ РАН; [науч. ред. В. В. Мазалов; отв. ред. Е. В. Хворостянская]. — Петрозаводск: КарНЦ РАН, 2019. — 123 с. : ил.
ISBN 978-5-9274-0851-1

Сборник содержит расширенные тезисы докладов участников X Международной Петрозаводской конференции «Вероятностные методы в дискретной математике» (22 – 26 мая 2019 г., Петрозаводск, Россия).

Probabilistic Methods in Discrete Mathematics, X International Petrozavodsk Conference (May 22 – 26, 2019, Petrozavodsk, Russia): extended abstracts / IAMR KRC RAS; [ed. V. V. Mazalov; resp. ed. E. V. Khvorostyanskaya]. — Petrozavodsk: KRC RAS, 2019. — 123 p. : il.
ISBN 978-5-9274-0851-1

The present volume contains extended abstracts accepted for the X International Petrozavodsk Conference “Probabilistic Methods in Discrete Mathematics” (May 22 – 26, 2019, Petrozavodsk, Russia).

УДК 519.1(063) + 519.2(063)
ББК 22.17

ISBN 978-5-9274-0851-1

Main organizers

INSTITUTE OF APPLIED MATHEMATICAL RESEARCH
OF THE KARELIAN RESEARCH CENTRE RAS

STEKLOV MATHEMATICAL INSTITUTE RAS

MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

Sponsors

MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

Program committee

A.M. Zubkov (co-chair)

V.V. Mazalov (co-chair)

A.M. Raigorodskii (co-chair)

V.A. Vatutin

G.O.H. Katona

Yu.S. Kharin

A.V. Kostochka

E.V. Morozov

Yu.L. Pavlov

L.A. Petrosian

K. Szajowski

A.N. Tikhomirov

Organizing committee

chair: Vladimir Mazalov
co-chair Yury Pavlov
secretary: Marina Leri

I.A. Cheplyukova
A.V. Chirkov
E.V. Khvorostyanskaya
O.V. Lukashenko
A.N. Rettieva
A.S. Rumyantsev

Scientific program

Probability and statistical problems of discrete mathematics
Theory of random graphs and hypergraphs
Combinatorial methods in the Internet data analysis
Mathematical problems of information security
Applied statistics
Statistical modeling and simulation
Game theory and stochastic optimization problems

Invited speakers

Vatutin Vladimir A. (Russia)
Katona Gyula O.H. (Hungary)
Kostochka Aleksandr V. (USA)
Pach János (Hungary)
Simonovits Miklós (Hungary)
Tikhomirov Aleksandr N. (Russia)
Kharin Yuri S. (Belarus)

Contents

Number of trees in a random graph

N. V. Alexeev

ITMO University, St. Petersburg, Russia

E-mail: nikita.v.alexeev@gmail.com

Consider a random graph $G(n,k)$ on n vertices with k edges, that is

$$\mathbb{P}(G(n,k)=g)=\frac{1}{\binom{N}{k}},$$

where $N=\binom{n}{2}$ and g is any fixed graph on n vertices with k edges.

Let $T_m(G)$ denote the number of isolated trees on m vertices in G.

We discuss the following theorem:

Theorem 1. *For fixed m, ℓ and $k \to \infty$ as $n \to \infty$ such that there exists a finite nonzero limit $x = \lim_{n\to\infty} \frac{2k}{n}$, the following equations hold:*

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{T_m(G(n,k))}{n}\right) = \phi_m(x) = \exp(-mx)\frac{x^{m-1}m^{m-2}}{m!}; \tag{1}$$

$$\lim_{n\to\infty} \frac{\operatorname{cov}(T_m(G(n,k)), T_\ell(G(n,k))}{n} = \delta_{m,\ell}\phi_m(x) -$$

$$\phi_m(x)\phi_\ell(x)\left(m\ell x + \frac{2(m-1)(\ell-1)}{x} + 2m + 2\ell - 3m\ell\right), \tag{2}$$

where $\delta_{m,\ell}$ is the Kronecker delta.

While the result (1) is classical, the equation (2) seems to be new.

We also consider a random graph process, where each vertex i has its own propensity to make bonds p_i. The numbers p_i are randomly chosen from the uniform distribution on a standard simplex $\{(p_1, p_2, \ldots, p_n) : p_i \geq 0, \sum p_i = 1\}$ (so-called *flat Dirichlet distribution*). The process starts at a graph with n vertices and 0 edges, and on each step it adds one edge to a graph, choosing nodes i and j as the endpoints with probability $2p_ip_j$.

We call the result of this process after k steps $G_{Dir}(n,k)$. We prove the following theorem:

Theorem 2. *For fixed m, ℓ and $k \to \infty$ as $n \to \infty$ such that there exists a finite nonzero limit $x = \lim_{n\to\infty} \frac{2k}{n}$, the following equations hold:*

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{mT_m(G_{Dir}(n,k))}{n}\right) = \frac{(3m-3)!}{(m-1)!(2m-1)!}\frac{x^{m-1}}{(x+1)^{3m-2}}\,. \tag{3}$$

We note that the sequence $\frac{(3m-3)!}{(m-1)!(2m-1)!}$ $(1, 1, 3, 12, 55, \dots)$ is well-known as *Fuss–Catalan numbers* and appears in the On-Line Encyclopedia of Integer Sequences (OEIS) as the sequence `A001764`. This sequence appears in combinatorics and probability theory in many different contexts, but such an interpretation is new (to the best of our knowledge).

New bounds on Armstrong codes

P. Boyvalenkov[1], A. Sali[2]

[1] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] Rényi Alfréd Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

E-mail: sali.attila@renyi.mta.hu

An Armstrong(q,k,n) code is a q-ary code of length n of minimum distance $n-k+1$ such that for every $k-1$-subset of coordinates there exists a pair of codewords that agree exactly there. $f(q,k)$ denotes the largest n such that an Armstrong(q,k,n) code exists. Upper bounds were given on $f(q,k)$ earlier in general, and in the case when $k > k_0(q)$. In the present paper we give upper bound where k is fixed and q is large, that is for $q > q_0(k)$ that improve on the general bounds.

Keywords: Armstrong code, linear programming method, Krawtchouk polynomials

On maximum vertex degree in the configuration model

I. A. Cheplyukova

Institute of Applied Mathematical Research KRC RAS,
Petrozavodsk, Russia

E-mail: chia@krc.karelia.ru

The configuration graph where vertex degrees are independent identically distributed random variables is often used for modeling complex networks such as mobile connections, social networks, the Internet and others [1]. We consider a configuration graph with N vertices. The random variables $\xi_1, \ldots, \xi_N$ are equal to the degrees of the vertices with the numbers $1, \ldots, N$. The degrees of the vertices are drawn independently from an arbitrary given distribution. Let us know only the limit behaviour of this distribution as $k \to \infty$:

$$\mathbf{P}\{\xi_i = k\} \sim \frac{d}{k^g (\ln k)^h},$$

where $i = 1, \ldots, N, \quad d > 0, g > 1, h \geqslant 0$. These graphs were first studied in [2].

We consider a subset of this graphs under the condition that the sum of vertex degrees was bounded from above by n. Denote by $\eta_1, \ldots, \eta_N$ the random variables equal to the degrees of vertices in such a conditional random graph. It is evident that these random variables are dependent, and for natural $k_1, \ldots, k_N$ such that $k_1 + \ldots + k_N \leq n$

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\} = \mathbf{P}\{\xi_1 = k_1, \ldots, \xi_N = k_N | \xi_1 + \ldots + \xi_N \leq n\}. \quad (1)$$

The equation (1) means that for the random variables $\xi_1, \ldots, \xi_N$ and $\eta_1, \ldots, \eta_N$ the analogue of the generalized allocation scheme is valid (see [3]) and we can apply the known properties of this scheme to study conditional random graphs.

We obtained the limit distributions of the maximum vertex degree in these conditional configuration graphs for different relations between the parameters N and n tending to infinity.

Let

$$B_N = \begin{cases} (N(g-1)^h/\ln^h N)^{1/(g-1)}, & 1 < g < 3; \\ \sqrt{N \ln^{1-h} N}, & g = 3, h < 1; \\ \sqrt{N \ln\ln N}, & g = 3, h = 1; \\ \sigma\sqrt{N}, & g > 3 \quad \text{or} \quad g = 3, h > 1, \end{cases}$$

$$m = \mathbf{E}\xi_1, \qquad \sigma^2 = \mathbf{D}\xi_1.$$

The following theorem holds.

Theorem. *Let* $n, N \to \infty, r = (Nd(g-1)/(\gamma \ln^h N)^{1/(g-1)}(1+o(1)), 0 < \gamma, \varepsilon < \infty$ *and one of the following conditions is satisfied:*

1. $1 < g < 2, n/B_N \to \infty$;
2. $g = 2, h \leqslant 1, (n - dN(1+\varepsilon)\ln^{1-h} N)/B_N \geqslant -C > \infty$;
3. $g = 2, h > 1, (n - (m + (d+\varepsilon)\ln^{1-h} N)N)/B_N \geq -C > -\infty$;
4. $g > 2, (n - mN)/B_N \geq -C > -\infty$,

where C *is a positive constant. Then*

$$\mathbf{P}\{\eta_{(N)} \leqslant r\} = e^{-\gamma}(1+o(1)).$$

References

[1] *Hofstad R.* Random graphs and complex networks. Cambridge University Press, 2017, 337 p.

[2] *Pavlov Yu. L.* Conditional configuration graphs with random parameter of the power-law degree distribution // Sbornik: Mathematics. 2018. **209**(2). P. 258–275.

[3] *Chuprunov A. N., Fazecas 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). P. 101–122.

Inducibility of paths in hypergraph colorings

D. D. Cherkashin

Moscow Institute of Physics and Technology, Dolgoprudny, Russia

E-mail: `matelk@mail.ru`

Let $m(n, r)$ denote the minimal number of edges in an n-uniform hypergraph which is not r-colorable. It is known that for a fixed n one has

$$c_n r^n < m(n, r) < C_n r^n.$$

A sequence of edges $a_1, \ldots, a_r$ is an r-*chain* if $|a_i \cap a_j| = 1$ if $|i - j| = 1$ and $a_i \cap a_j = \emptyset$ otherwise. It turns out that the best current lower bounds on $m(n, r)$ are based on the evaluation of the number of r-chains. We note that r-chains are paths in the corresponding graphs; so one can apply Pippenger–Golumbic-type bounds to improve the lower bound in e times.

Limit theorems in non-homogenous generalized allocation scheme

D. E. Chickrin, A. N. Chuprunov, P. A. Kokunin

Kazan Federal University, Kazan, Russia

E-mail: dmitry.kfu@gmail.com, achuprunov@mail.ru, pkokunin@mail.ru

We say that the random variables $\eta_1, \ldots, \eta_N$ fit to the generalized allocation scheme of n particles by N cells if

$$\mathbf{P}\{\eta_{N1} = k_1, \ldots \eta_{NN} = k_N\} = \mathbf{P}\left\{\xi_{N1} = k_1, \ldots \xi_{NN} = k_N \left| \sum_{i=1}^{N} \xi_{Ni} = n \right.\right\}, \tag{1}$$

where $\xi_{N1}, \xi_{N2}, \ldots \xi_{NN}$ are independent integer valued nonnegative random variables. The aim of this paper is to study the convergence in distribution of the random variables

$$\mu_r(n, K, N) = \sum_{i=1}^{K} I_{\{\eta_{Ni}=r\}}, r = 0, 1, \ldots, \ \eta_{(K,N)} = \max_{1 \le i \le K} \eta_{Ni}, 0 < K \le N.$$

We use the following analog of Kolchin formula [1]

$$\mathbf{P}\{\mu_r(n, K, N) = k\} = \sum_{A \subset \{1,2,\ldots K\}, |A|=k} \frac{\mathbf{P}\left\{\zeta_{N,A}^{\{r\}} = n - kr\right\}}{\mathbf{P}\{\zeta_N = n\}} \times$$

$$\times \left(\prod_{i \in A} \mathbf{P}\{\xi_{Ni} = r\}\right) \left(\prod_{i \in \{1,2,\ldots K\} \setminus A} \mathbf{P}\{\xi_{Ni} \neq r\}\right), \tag{2}$$

where

$$\zeta_{N,A}^{\{r\}} = \sum_{i \in \{1,2,\ldots K\} \setminus A} \xi_{Ni}^{\{r\}} + \sum_{i=K+1}^{N} \xi_{Ni}, \quad \zeta_N = \sum_{i=1}^{N} \xi_{Ni}$$

and $\xi_{N1}^{\{r\}}, \ldots, \xi_{NK}^{\{r\}}, \xi_{N1}, \ldots, \xi_{NN}$ are independent random variables and for all $r = 0, 1, 2, \ldots$

$$\mathbf{P}\{\xi_{Ni}^{\{r\}} = j\} = \mathbf{P}\{\xi_{Ni} = j \mid \xi_{Ni} \neq r\}, \quad j = 0, 1, 2 \ldots.$$

We shell consider $\eta_{N1}, \ldots \eta_{NN}$ defined by (2.1) in which $\xi_{N1}, \xi_{N2}, \ldots \xi_{NN}$ are independent random variable with the distributions

$$\mathbf{P}\{\xi_{Ni} = k\} = \frac{b_k (x_{Ni})^k}{k! B(x_{Ni})}, \; k = 0, 1, 2, \ldots, 0 < x_{Ni} < R,$$

there $B(x) = \sum_{k=0}^{\infty} \frac{b_k}{k!} x^k$ is a sum of the series with the radius of convergence $R > 0$ and $b_0, b_1 > 0$, $b_i \geq 0$, $i \geq 2$. We will denote

$$x_{KN} = \sum_{i=1}^{K} x_{Ni}, \;\; x_N = x_{NN}, \quad x_{KN}^* = \max_{1 \leq i \leq K} x_{Ni}, \quad x_{NN}^* = x_N^*,$$

$$m_N = \sum_{i=1}^{N} \mathbf{E}\xi_{Ni}, \;\; \lambda_{rN} = \sum_{i=1}^{K} \frac{b_r (x_{Ni})^r}{r! b_0}.$$

Using (2) we obtain

Theorem 1. *Suppose $r \geq 2$, $0 \leq \lambda < \infty$ are a fixed numbers, $N, K, n \to \infty$ such that*

$$x_N \to \infty, \quad x_N^* \to 0, \quad \frac{x_{KN} x_{KN}^*}{\sqrt{x_N}} \to 0 \;\; \lambda_{rN} \to \lambda \text{ and } \frac{|n - m_N|}{\sqrt{x_N}} < C$$

for some $C > 0$. Then we have

$$\mathbf{P}\{\mu_r(n, K, N) = k\} = e^{-\lambda} \frac{\lambda^k}{k!} + o(1), \quad k = 0, 1, 2 \ldots.$$

Theorem 2. *Suppose $C_1, C > 0$, $r \geq 2$ are fixed numbers, $N, K, n \to \infty$ such that*

$$x_N \to \infty, \;\; \lambda_{rN} \to \infty, \;\; \frac{x_{KN}}{\sqrt{x_N}} \to 0, \quad x_N^* \to 0 \text{ and } \frac{|n - m_N|}{\sqrt{x_N}} < C.$$

Let $z = \frac{k - \lambda_N}{\sqrt{\lambda_N}}$. Then we have

$$\mathbf{P}\{\mu_r(n, K, N) = k\} = \frac{1}{\sqrt{2\pi \lambda_{rN}}} e^{-\frac{z^2}{2}} (1 + o(1))$$

uniformly for $0 \le k \le K$.

We use the following analog of Kolchin formula [1]

$$\mathbf{P}\{\eta_{(K,N)} \le r\} = \left(\prod_{i=1}^{K} \mathbf{P}\{\xi_{Ni} \le r\}\right) \frac{\mathbf{P}\{\zeta_N^{\{\le r\}} = n\}}{\mathbf{P}\{\zeta_N = n\}}, \ r \ge 0, \tag{3}$$

where

$$\zeta_N^{\{\le r\}} = \sum_{i=1}^{K} \xi_{Ni}^{\{\le r\}} + \sum_{i=K+1}^{N} \xi_{Ni},$$

$\xi_{N1}^{\{\le r\}}, \ldots, \xi_{NK}^{\{\le r\}}, \xi_{N1}, \ldots, \xi_{NN}$ are independent random variables and

$$\mathbf{P}\{\xi_{Ni}^{\{\le r\}} = j\} = \mathbf{P}\{\xi_{Ni} = j \mid \xi_{Ni} \le r\}, \quad j = 0, 1, 2 \ldots.$$

Using (3) we obtain

Theorem 3. *Suppose* $C > 0$, $r \in \{0, 1 \ldots\}$, $0 < \lambda < \infty$ *are fixed numbers,*

$$N, K, n \to \infty \text{ such that } x_N \to \infty, \quad x_N^* \to 0 \text{ and } \frac{|n - m_N|}{\sqrt{x_N}} < C.$$

Let one of the following conditions be valid:

(1) Let $r \ge 2$. *Let* $b_r > 0$,

$$\frac{x_{KN}}{\sqrt{x_N}} z_{KN}^* \to 0, \quad \lambda_{(r+1)N} \to \lambda,$$

(2) Let $r \in \{0, 1\}$. *Let*

$$\frac{x_{KN}}{\sqrt{x_N}} \to 0, \qquad \lambda_{(r+1)N} \to \lambda.$$

Then we have

$$\mathbf{P}\{\eta_{(K,N)} = r\} = e^{-\lambda} + o(1), \quad \mathbf{P}\{\eta_{(K,N)} = r + 1\} = 1 - e^{-\lambda} + o(1).$$

As corollaries we obtain limit theorems for a non-homogenous allocation scheme of n distingushing particles by N different cells.

References

[1] *Kolchin V. F.* Random Graphs. Cambridge University Press, 1999.

Property B_k-problem for uniform simple hypergraphs

Y. A. Demidovich

Moscow Institute of Physics and Technology (State University), Moscow, Russia

E-mail: yuradem9595@mail.ru

A *hypergraph* is a pair $H = (V, E)$, where V is a finite set whose elements are called *vertices* and E is a family of subsets of V, called the *edges*. A hypergraph is said to be *n-uniform* if each of its edges contains exactly n vertices.

One of the classical extremal problems of a hypergraph theory is the property B-problem. We say that a hypergraph has property B if there is a two-coloring of V such that no edge is monochromatic. The problem is to find the quantity $m(n)$ which is the minimum possible number of edges of an n-uniform hypergraph that does not have property B.

Let k be a natural number. The property B_k-problem is to find the value of $m_k(n)$ equal to the minimal number of edges in an n-uniform hypergraph not admitting 2-colorings of the vertex set such that every edge of the hypergraph contains at least k vertices of each color.

We say a hypergraph is *simple* if any two edges of it have no more than one vertex in the intersection.

Define a quantity $m^*_k(n)$ which is equal to minimum possible number of edges in a simple n-uniform hypergraph that does not have property B_k.

We obtain new lower and upper bounds for $m^*_k(n)$.

Theorem 1. *Let $k \geqslant 2$ and suppose that*

$$k \leqslant \sqrt{\frac{n}{\ln n}}.$$

Then for all $n \geqslant 30$,

$$m^*_k(n) \geqslant \frac{25}{18e^{52}} \frac{2^{2(n-1)}}{k(n-1)\ln(n-1)\binom{n-2}{k-1}^2}.$$

Theorem 2. *Let $k = k(n)$ be such a natural function that $k = o(n/\ln n)$. Then for large n*

$$m_k^*(n) < \frac{n^2 e^2 \ln^2 2 \cdot 2^{2n+4}}{\left(\sum_{j=0}^{k-1} \binom{n}{j}\right)^2}.$$

Theorem 3. *Let $k = k(n)$ be such a natural function that $k = o\,(n)$. Then for large n*

$$m_k^*(n) < n^7 \left\lceil \frac{n 2^{n+1} e \ln 2}{\sum_{j=0}^{k-1} \binom{n}{j}} \right\rceil^2.$$

A multiclass retrial system with coupled orbits and class-dependent service interruptions

I. Dimitriou[1], E. V. Morozov[2, 3], T. E. Morozova[3]

[1] University of Patras, Department of Mathematics, Patras, Greece
[2] Institute of Applied Mathematical Research KRC RAS, Petrozavodsk, Russia
[3] Petrozavodsk State University, Petrozavodsk, Russia

E-mail: idimit@math.upatras.gr; emorozov@karelia.ru; tiamorozova@mail.ru

In this work, we consider a multiclass retrial system with coupled orbit queues, and service interruptions. A single-server system accepts N classes of customers, which arrive according to independent Poisson streams. We assume independent and arbitrary distributed class-dependent service times.

An arriving class-i, $i = 1, \ldots, N$ customer finding the server unavailable joins the orbit queue i. We also adopt a class-dependent, queue-aware constant retrial policy. More precisely, we assume that the head blocked customer in an orbit queue attempts to connect with the server after an exponentially distributed service time, which depends both on the class of orbiting customer and on the state of the other orbit queues (i.e., whether they are idle or busy).

A distinctive feature of the model is as follows: while a customer is being served, an interruption may occur according to a Poisson process with the rate depending on the class of the interrupted customer. After such an event, a setup time follows, and its duration has general class-dependent distribution. We consider both preemptive-repeat identical, and preemptive-resume interruptions.

Such models have potential applications in the modelling of relay-assisted cooperative wireless networks.

Unlike previous works [1, 2], we now consider *combined interruptions*: the switching between the interruption modes depend on the class of the interrupted customer. This setting is motivated, for instance, by Windows and Unix-like operations system, because both of these modes are supported for multitasking, and the interruption type depends on which task has been sent to the server.

A critical point in our approach is the calculation of the Laplace-Stieltjes transform (LST) of the density of the so-called *generalized service time*, which is defined as the time elapsed from the epoch a customer succeeds to connect with the server until the epoch the server is ready to serve the next customer. To illustrate theoretical results, we simulate three-class system.

The study of EM was carried out under state order to the Karelian Research Centre of the Russian Academy of Sciences (Institute of Applied Mathematical Research KRC RAS) and supported by Russian Foundation for Basic Research, projects 18-07-00147, 18-07-00156, 19-07-00303. The research of TM is partly supported by Russian Foundation for Basic Research, projects 18-07-00147, 19-07-00303.

References

[1] *Morozov E., Morozova T., Dimitriou I.* Simulation of multiclass retrial system with coupled orbits // Proceedings of SMARTY'18 First International Conference Stochastic Modeling and Applied Research of Technology, 4-16. Petrozavodsk, Russia, September 21-25, 2018.

[2] *Dimitriou I., Morozov E., Morozova T.* A Multiclass Retrial System With Coupled Orbits And Service Interruptions // Verification of Stability Conditions (accepted). Proceedings of FRUCT2019, Moscow, April 91-12, 2019.

The domination number of the graph defined by two levels of the n-cube

G. O. H. Katona

Rényi Alfréd Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

E-mail: `ohkatona@renyi.hu`

Let $G = (V, E)$ be a simple graph. $D \subset V$ is a *dominating set* if if either $v \in D$ holds for every $v \in V$ or there is an edge $\{v, x\} \in E$ satisfying $x \in D$. The *domination number* $\gamma(G)$ of the graph G is the size of the smallest dominating set.

Consider all k-element subsets and ℓ-element subsets ($k > \ell$) of an n-element set as vertices of a bipartite graph. Two vertices are adjacent if the corresponding ℓ-element set is a subset of the corresponding k-element set. Let $G_{k,\ell}$ denote this graph. Its domination number is studied in the paper.

Theorem 1. $\gamma(G_{k,1}) = n - k + 1$.

This statement has a trivial proof. The problem becomes considerably harder when $\ell = 2$. In this case we only have an asymptotic result.

Theorem 2. $\gamma(G_{k,2}) = \frac{k+3}{2(k-1)(k+1)} n^2 + o(n^2) \quad (k \geq 3)$.

The upper estimate is proved by a random construction. We also suggest a way to find a deterministic construction, but it is completed only for $k = 3$ and 4. The lower estimate uses the *graph removal lemma*, a consequence of Szemerédi's celebrated Regularity Lemma.

The problem for $\ell = 2$ is closely related to Turán's theorem. Since the corresponding problem of Turán for 3-graphs is unsolved, our present problem seems to be difficult for $\ell \geq 3$.

Joint work with Leila Badakhshian and Zsolt Tuza.

Multi-class model with general retrials: stability analysis by simulation

R. S. Nekrasova

Institute of Applied Mathematical Research KRC RAS,
Petrozavodsk State University,
Petrozavodsk, Russia

E-mail: ruslana.nekrasova@mail.ru

We study a model of multi-class single-server retrial system with classical discipline. If a customer finds the server busy, it joins to the corresponding (virtual) infinite capacity orbit and then independently retry to attack server after random retrial time. We define renewal input process and assume, that new arrival belongs to class-i with a positive probability p_i. We consider *general* distribution of retrial time and explore the dynamics of mean orbit sizes for different configurations of considered system.

Our goal is to verify by simulation the stability of considered system and match the obtained conclusions with explicit analytical results for related system with exponential retrials. Note, that the only source of instability of presented system is infinite growth of orbit size. As the behavior of any orbit affects to other orbits, stability also means that all orbits are bounded and instability implies that all orbits go to infinity. This property significantly differs considered system from a system with constant retrial rate, presented in [1].

Necessary and sufficient stability condition for related model with classical discipline and *exponential* retrial times was presented in recent work [2]. Basing on simulation, we show, that stability requirement, that load coefficient is less than 1 (number of servers) provides stable orbits of considered system with general retrials. In particular, we present simulation results for the models with Weibull, Pareto and Uniform distributions of retrial times.

Thus, one can expect, that obtained stability criteria from [2] could be extrapolated to the general case of non-exponential retrial times.

The study was carried out under state order to the Karelian Research Centre of the Russian Academy of Sciences (Institute of Applied Mathematical Research KRC RAS). The research is partly supported by Russian Foundation for Basic Research, projects 18-07-00156, 19-07-00303.

References

[1] *Morozov E., Avrachenkov K., Nekrasova R., Steyaert B.* Stability Analysis and simulation of N-class retrial system with constant retrial rates and Poisson inputs // Asia-Pacific Journal of Operation Research. 2014. **31**(2). P. 1–18.

[2] *Morozov E., Phung-Duc T.* Stability analysis of a multiclass retrial system with classical retrial policy // Performance Evaluation. 2017. **112**. P. 15–26.

On clustering coefficient of a configuration graph

Yu. L. Pavlov

Institute of Applied Mathematical Research KRC RAS,
Petrozavodsk, Russia

E-mail: pavlov@krc.karelia.ru

We consider configuration graph [1] $G = G(V, E)$, where V is a set of N vertices and E is a set of edges. Let $i, j, t \in V$ be three different vertices. The clustering coefficient [2] C_G in G is defined as

$$C_G = \frac{\Delta_G}{W_G},$$

where

$$\Delta_G = 6\Sigma_{1\leqslant i<j<t\leqslant N} I\{(ij), (jt), (tj) \in E\},$$

$I\{A\}$ is the indicator of the event A,

$$W_G = 2 \sum_{1\leqslant i<j<t\leqslant N} I\{(ij), (jt) \in E\}.$$

Let vertex degrees $\xi_1, \ldots, \xi_N$ are independent identically distributed random variables with unknown probability distribution. We assume only that

$$p_k = \mathbf{P}\{\xi_i = k\} > 0, \quad k = 1, 2, \ldots,$$

$$m = \mathbf{E}\,\xi_1 < \infty$$

and

$$p_k \sim \frac{d}{k^g (\ln k)^h}$$

as $k \to \infty, d > 0, g > 1, h \geqslant 0$.

Denote

$$\xi_{(N)} = \max_{1\leqslant s\leqslant N} \xi_s.$$

We proved the following result.

Theorem 1. *Let $N \to \infty$. Then the next asertions are hold a.a.s.*

1. *If $Var(\xi_1) < \infty$, then $C_G = O\left(\frac{1}{N}\right)$.*

2. *If $g = 3, h = 1$, then $C_G \sim \frac{(d \ln \ln N)^2}{m^3 N}$.*

3. *If $g = 3, h < 1$, then $C_G \sim \left(\frac{d}{1-h}\left(\frac{\ln N}{2}\right)^{1-h}\right)^2 \frac{1}{m^3 N}$.*

4. *If $1 < g < 3$, then there exist a constant $B > 0$ such that $N^{\frac{1}{g-1}}/B \leqslant \xi_{(N)} \leqslant BN^{\frac{1}{g-1}}$ and if $\xi_{(N)} = uN^{\frac{1}{g-1}}, 0 < u < \infty$, then*

$$C_G \sim \left(\frac{du^{3-g}}{3-g}\right)^2 \left(\frac{g-1}{\ln N}\right)^{2h} \frac{N^{\frac{7-3g}{g-1}}}{m^3}.$$

The study was carried out under state order to the Karelian Research Centre of the Russian Academy of Sciences (Institute of Applied Mathematical Research KarRC RAS).

References

[1] *Bollobas B.* A probabilistic proof of an asymptotic formulae for the number of labelled regular graphs // Eur. J. Comb. 1980. **2**. P. 311–316.

[2] *Hofstad R.* Random Graphs and Complex Networks. Vol. 1. Cambridge: Cammbridge University Press, 2017, 337 p.

Sylvester-Gallai-type theorems for vectors

A. A. Polyanskii

Moscow Institute of Physics and Technology, Dolgoprudny, Russia
CMC ASU, Maykop, Russia
IIPT RAS, Moscow, Russia

E-mail: `alexander.polyanskii@ya.ru`

The famous Sylvester-Gallai Theorem was posed as a problem by J. J. Sylvester in 1893. In 1930th it was proved by T. Gallai and others. It claims that if a finite point set $\mathcal{P}$ in the real plane is such that for any two distinct points $a, b \in \mathcal{P}$ there is a point $c \in \mathcal{P}$ lying on the line passing through a and b, then all points of $\mathcal{P}$ lie on a line.

It is known that that analogues statement is not true for the complex plane. For example, the inflection points of a cubic curve in the complex projective plane form such an example. This set of the inflection points is known as the Hesse configuration and contains nine points.

We are going to discuss the following Sylvester-Gallai-type problem.

Given positive integers m and d, a finite set $A \subset [0, 1)^d$ containing at least $m+1$ pairwise distinct vectors is called *$SG(d, m+1)$-set* if for any m distinct vectors $\mathbf{a}_1, \dots, \mathbf{a}_m \in A$ there is $\mathbf{a}_{m+1} \in A$ different from $\mathbf{a}_1, \dots, \mathbf{a}_m$ such that

$$\mathbf{a}_1 + \dots + \mathbf{a}_m + \mathbf{a}_{m+1} \in \mathbb{Z}^d.$$

Problem. Describe all SG$(d, m+1)$-sets.

We are going to discuss connections of Problem and the problem of classification of Sylvester-Gallai-type configurations lying on a cubic curve in the complex projective plane. Also, we show the solution of Problem in case $m = 2$.

Joint work with F. Nilov.

Spectra for bounded quantifier depth first-order formulae for random hypergraphs

S. N. Popova

Moscow Institute of Physics and Technology, Moscow, Russia

E-mail: popovaclaire@mail.ru

Asymptotic behavior of first-order properties probabilities of the Erdos–Renyi random graph $G(n,p)$ has been widely studied. S. Shelah and J. Spencer showed that when α is an irrational number and $p(n) = n^{-\alpha+o(1)}$ then $G(n,p)$ obeys the zero-one law (this means that for all first-order properties their limiting probabilities belong to $\{0,1\}$). In works by M.E. Zhukovskii the question when the random graph $G(n,p)$ obeys zero-one law for all first-order sentences of quantifier depth at most k was examined. In [1] the notion of spectra of first-order properties was introduced and it was shown that there exists a first-order property with infinite spectra. In [2] the notion of spectra for random graphs was studied in more detail.

In this work, we consider spectra of first-order properties in relation to random uniform hypergraphs.

Let us define the *random s-uniform hypergraph* $G^s(n,p)$. Consider the set $\Omega_n = \{G = (V_n, E)\}$ of all s-uniform hypergraphs (s-hypergraphs) with the set of vertices $V_n = \{1, 2, \dots, n\}$. The random hypergraph $G^s(n,p)$ is a random element with probability distribution

$$\Pr[G^s(n,p) = G] = p^{|E|}(1-p)^{\binom{n}{s}-|E|}.$$

For any first order property L we define two notions of its *spectra*, $S^1(L)$ and $S^2(L)$. The first considers behavior at $p = n^{-\alpha}$. $S^1(L)$ is the set of $\alpha \in (0, s-1)$ (we take the interval $(0, s-1)$ because the structure of $G^s(n, n^{-\alpha})$ for $\alpha > s-1$ is considerably simpler than that for $\alpha < s-1$ and the study of $G^s(n,p)$ with $p(n) \gg n^{-s+1}$ is a subject of a separate investigation) which do not satisfy the following property:

with $p(n) = n^{-\alpha}$, $\lim_{n\to\infty} \Pr[G^s(n,p) \models L]$ exists and is either 0 or 1.

The second considers behavior near $p = n^{-\alpha}$. $S^2(L)$ is the set of $\alpha \in (0, s-1)$ which do not satisfy the following property:

$$\text{there exists } \varepsilon > 0 \text{ so that for any } n^{-\alpha-\varepsilon} < p(n) < n^{-\alpha+\varepsilon},$$
$$\lim_{n\to\infty} \Pr[G^s(n,p) \models L] \text{ exists, is 0 or 1,}$$
$$\text{and is independent of the choice of } p(n).$$

Let $\mathcal{L}_k$ denote the set of all first-order properties which can be expressed by first-order formulae with quantifier depth at most k. Denote unions of $S^1(L)$ and $S^2(L)$ over all $L \in \mathcal{L}_k$ by S^1_k and S^2_k respectively.

We study the spectra S^1_k and S^2_k and try to answer the following questions:

- Is S^j_k, where $j \in \{1, 2\}$, finite or infinite?
- What are the maximal and the minimal numbers in S^1_k and S^2_k?
- Suppose that S^1_k and S^2_k are infinite. What are the maximal and the minimal limit points in S^1_k and S^2_k?

This work was supported by RSF 16-11-10014.

References

[1] *Spencer J. H.* Ininite spectra in the first order theory of graphs // Combinatorica. 1990. **10**(1). P. 95–102.

[2] *Spencer J. H., Zhukovskii M. E.* Bounded quantifier depth spectra for random graphs // Discrete Mathematics. 2016. **339**(6). P. 1651–1664.

Dynamic multicriteria games with random horizons

A. N. Rettieva

Institute of Applied Mathematical Research KRC RAS,
Petrozavodsk, Russia

E-mail: annaret@krc.karelia.ru

1 Introduction

Mathematical models involving more than one objective seem more adherent to real problems. Often players have more than one goal which are often not comparable. These situations are typical for game-theoretic models in economics and ecology. Models with random planning horizons in exploitation processes are most appropriate for describing reality: external random factors can cause a game breach and the participants know nothing about them a priori.

First, we construct a multicriteria Nash equilibrium applying the bargaining concept (via Nash products [3], [1]). Then, we obtain multicriteria cooperative behavior as a solution of a Nash bargaining scheme with the multicriteria Nash equilibrium payoffs playing the role of status quo points [2].

To illustrate the presented approaches a multicriteria bioresource management problem with random planning horizons is investigated.

2 Dynamic Multicriteria Game with Random Horizons

Consider a bicriteria dynamic game with two participants in discrete time. The players exploit a common resource and both wish to optimize two different criteria. The state dynamics is in the form

$$x_{t+1} = f(x_t, u_{1t}, u_{2t})\,, \ \ x_0 = x\,, \tag{1}$$

where $x_t \geq 0$ is the resource size at time $t \geq 0$, $f(x_t, u_{1t}, u_{2t})$ denotes the natural growth function, and $u_{it} \geq 0$ gives the exploitation rate of player i at time t, $i = 1, 2$.

Denote $u_t = (u_{1t}, u_{2t})$. Each player has two goals to optimize. We explore a model in which the players possess heterogeneous planning horizons. By assumption, the players stop joint exploitation at random time steps: external stochastic processes can cause a game breach.

Suppose that players 1 and 2 harvest the fish stock during n_1 and n_2 steps, respectively. Here n_1 represents a discrete random variable taking values $\{1, \dots, n\}$ with the corresponding probabilities $\{\theta_1, \dots, \theta_n\}$. Similarly, n_2 is a discrete random variable with the value set and the probabilities $\{\omega_1, \dots, \omega_n\}$. We assume that the planning horizons are independent. Therefore, during the time period $[0, n_1]$ or $[0, n_2]$ the players harvest the same stock, and the problem consists in evaluating their optimal strategies.

The payoffs of the players are determined via the expectation operator:

$$J_1^i = E\Big\{\sum_{t=1}^{n_1} \delta^t g_1^i(u_t) I_{\{n_1 \le n_2\}} +$$

$$+\Big(\sum_{t=1}^{n_2} \delta^t g_1^i(u_t) + \sum_{t=n_2+1}^{n_1} \delta^t g_1^i(u_{1t}^a)\Big) I_{\{n_1 > n_2\}}\Big\} =$$

$$= \sum_{n_1=1}^{n} \theta_{n_1} \Big[\sum_{n_2=n_1}^{n} \omega_{n_2} \sum_{t=1}^{n_1} \delta^t g_1^i(u_t) +$$

$$+\sum_{n_2=1}^{n_1-1} \omega_{n_2} \Big(\sum_{t=1}^{n_2} \delta^t g_1^i(u_t) + \sum_{t=n_2+1}^{n_1} \delta^t g_1^i(u_{1t}^a))\Big)\Big],\ i = 1, 2, \tag{2}$$

$$J_2^i = E\Big\{\sum_{t=1}^{n_2} \delta^t g_2^i(u_t) I_{\{n_2 \le n_1\}} +$$

$$+\Big(\sum_{t=1}^{n_1} \delta^t g_2^i(u_t) + \sum_{t=n_1+1}^{n_2} \delta^t g_2^i(u_{2t}^a)\Big) I_{\{n_2 > n_1\}}\Big\} =$$

$$= \sum_{n_2=1}^{n} \omega_{n_2} \Big[\sum_{n_1=n_2}^{n} \theta_{n_1} \sum_{t=1}^{n_2} \delta^t g_2^i(u_t) +$$

$$+\sum_{n_1=1}^{n_2-1} \theta_{n_1} \Big(\sum_{t=1}^{n_1} \delta^t g_2^i(u_t) + \sum_{t=n_1+1}^{n_2} \delta^t g_2^i(u_{2t}^a))\Big)\Big],\ i = 1, 2, \tag{3}$$

where, for $j = 1, 2$, u_{jt}^a specifies the strategy of player j when its partner leaves the game, $\delta \in (0, 1)$ denotes the discount.

First, we construct a multicriteria Nash equilibrium $V_i^{jN}(\tau, x)$, $i, j = 1, 2$, applying the bargaining concept ([3], [1]).

To construct multicriteria payoff functions, we adopt the Nash products. The role of the status quo points belongs to the guaranteed payoffs of the players:

$$V_1^N(\tau, x) = (V_1^{1N}(\tau, x) - G_1^1)(V_1^{2N}(\tau, x) - G_1^2)\,,$$
$$V_2^N(\tau, x) = (V_2^{1N}(\tau, x) - G_2^1)(V_2^{2N}(\tau, x) - G_2^2)\,,$$

where $V_i^{jN}(\tau, x)$ are the noncooperative payoffs as step τ occurs $(i, j = 1, 2)$.

A strategy profile $u_t^N = (u_{1t}^N, u_{2t}^N)$ is called a multicriteria Nash equilibrium [1] of the problem (1)–(3) if it maximizes $V_i^N(1, x)$, $i = 1, 2$.

Then, we obtain multicriteria cooperative behavior as a solution of a Nash bargaining scheme with the multicriteria Nash equilibrium payoffs playing the role of status quo points [2]. So, to construct the cooperative strategies it is required to solve the problem

$$\begin{gathered}(V_1^{1c}(1, x) + V_2^{1c}(1, x) - V_1^{1N}(1, x) - V_2^{1N}(1, x)) \cdot \\ (V_1^{2c}(1, x) + V_2^{2c}(1, x) - V_1^{2N}(1, x) - V_2^{2N}(1, x)) \to \max_{u_{1t}^c, u_{2t}^c}\,,\end{gathered} \tag{4}$$

where $V_i^{jc}(\tau, x)$ are the cooperative payoffs as step τ occurs $(i, j = 1, 2)$.

To illustrate the presented approaches a multicriteria bioresource management problem with random planning horizons is investigated.

References

[1] *Rettieva A. N.* Equilibria in dynamic multicriteria games // Int. Game Theory Rev. 2017. **19**(1), 1750002.

[2] *Rettieva A. N.* Dynamic multicriteria games with finite horizon // Mathematics. 2018. **6**(9). P. 156.

[3] *Rettieva A. N.* A bioresource management problem with different planning horizons // Automation and Remote Control. 2015. **76**(5). P. 919–934.

[4] *Shapley L. S.* Equilibrium points in games with vector payoffs // Naval Res. Log. Quart. 1959. **6**. P. 57–61.

Steady-state analysis of cognitive radio model by state space expansion

A. S. Rumyantsev[1], Garimella Rama Murthy[2]

[1] Institute of Applied Mathematical Research KRC RAS, Petrozavodsk, Russia
[2] Mahindra Ecole Centrale, Hyderabad, India

E-mail: [1] ar0@krc.karelia.ru

Cognitive Radio (CR) is a well-known technology used to increase spectrum utilization in wireless transmission systems and networks. The spectrum is unevenly shared between Primary and Secondary users (PU and SU, respectively), and PU has absolute priority over SU. While time sharing (interleaving of PU and SU in time) is a technically simple solution, a higher utilization may be provided by underleaving paradigm. In such a case, SU are allowed to transmit simultaneously with PU, possibly at a reduced signal strength to avoid interference.

We present a model of a single node of CR network receiving PU and SU transmissions. The PU presence in the system is itself an alternating telegraph-type process with two states: PU present (having exponentially distributed duration with rate μ_p) and PU absent (exponentially distributed with rate λ_p). The PU presence infers the transmission rate of SU as follows: transmission of SU is exponentially distributed during PU presence (absence) with rate β (μ_s). SU are served in the order of arrival driven by Poisson process with rate λ_s, and have unlimited waiting space. Finally, at each PU presence starting epoch, the SU being served (if any) returns at the first position in the queue with probability (w.p.) α, and balks w.p. $1-\alpha$.

The system state is a two-dimensional Markov process $\{X(t), J(t)\}_{t\geqslant 0}$, where $J(t) \in \{0,1\}$ is PU presence telegraph-type process, and $X(t)$ is the number of SU in the system at time $t \geqslant 0$. Such a process is the so-called Quasi-Birth-Death process with two states at each level. Recently [1] it was shown that such a process allows to obtain the steady-state distribution

$\pi_{i,j} = \mathrm{P}\{X = i, J = j\}$ explicitly. In this paper we demonstrate an alternative approach based on the so-called Complete Level Crossing Information property of a QBD process.

Consider a state i, j of the state space of a QBD process, such that the outward transitions from some states $(i+1,\cdot)$ and $(i-1,\cdot)$ are possible. Each such a state is replaced with a couple $(i, 0, j)$ and $(i, 2, j)$, where $(i, 0, j)$ receives outward transitions from states $(i-1,\cdot)$ only, and $(i, 2, j)$ receives transitions from $(i+1,\cdot)$. The inward transitions from state (i, j) are copied to the two new states, and one of the two possible destination for inter-level outward transitions from states $(i,\cdot)$ is chosen arbitrary. For completeness of the notation, a state (i_0, j_0) receiving transitions from states $(i-1,\cdot)$ only, is replaced with $(i_0, 0, j_0)$, a state (i_2, j_2) receiving transitions from states $(i+1,\cdot)$ only, is replaced with $(i_2, 0, j_2)$, and all other states (i_1, j_1) are replaced with $(i_1, 1, j_1)$.

After such a partition, the new state space allows each state (i, k, j) to receive transitions either from levels $i' \geqslant i$ or from $i' \leqslant i$ only. Thus, the infinitesimal generator $\hat{Q}$ of the new process $\{X(t), D(t), J(t)\}_{t\geqslant 0}$ has the following bidiagonal form known as LCI-complete canonical form [2]:

$$\hat{Q} = \begin{pmatrix} B_0 & B_1 & 0 & 0 & \dots \\ 0 & C & B & 0 & \dots \\ 0 & 0 & C & B & \ddots \\ 0 & 0 & \ddots & \ddots & \ddots \end{pmatrix}, \tag{1}$$

where all matrices, except B_0 and B_1, are square matrices of order m (where m is the number of states at non-boundary levels). B_0 is an $(m_0 + m) \times (m_0 + m - m_1)$ matrix, where m_0 is the number of states at boundary level and m_1 is the number of *states* at non-boundary level that do not allow downward transitions to the boundary level, while B_1 is $(m_0 + m) \times m$. Hereafter we define these matrices explicitly:

$$C = \begin{pmatrix} \mu_s & (1-\alpha)\lambda_p & -c_1 & \alpha\lambda_p \\ 0 & \beta & \mu_p & -c_2 \\ \mu_s & (1-\alpha)\lambda_p & 0 & \alpha\lambda_p \\ 0 & \beta & \mu_p & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & \lambda_s & 0 \\ 0 & 0 & 0 & \lambda_s \\ -c_1 & 0 & \lambda_s & 0 \\ d0 & -c_2 & 0 & \lambda_s \end{pmatrix},$$

$$B_{0,0} = \begin{pmatrix} -(\lambda_p+\lambda_s) & \lambda_p & \lambda_s & 0 \\ \mu_p & -(\mu_p+\lambda_s) & 0 & \lambda_s \end{pmatrix}, B_1 = \begin{pmatrix} \mathbf{0} \\ B \end{pmatrix}, B_0 = \begin{pmatrix} B_{0,0} \\ C \end{pmatrix},$$

where $c_1 = \lambda_p + \lambda_s + \mu_s$ and $c_2 = \lambda_s + \mu_s + \beta$. After some algebra, C^{-1}

can be obtained explicitly

$$C^{-1} = \begin{pmatrix} -c_3 & \frac{\alpha\lambda_p}{\mu_s c_2} & \frac{1}{\mu_s} + c_3 & -\frac{c_3}{\mu_p} - \frac{\alpha\lambda_p}{\mu_s c_2} \\ \frac{\mu_p}{\beta c_1} & 0 & -\frac{\mu_p}{\beta c_1} & \frac{1}{\beta} \\ -\frac{1}{c_1} & 0 & \frac{1}{c_1} & 0 \\ 0 & \frac{-1}{c_2} & 0 & \frac{1}{c_2} \end{pmatrix},$$

where $c_3 = \frac{(1-\alpha)\lambda_p\mu_p}{\beta\mu_s c_1}$. This allows, in particular, to obtain the steady-state solution $\hat{\pi} = [\hat{\pi}_{i,k,j}]$, where $\hat{\pi}_{i,k,j} = \mathrm{P}\{X = i, D = k, J = j\}$, in matrix-geometric form

$$\hat{\pi}_{n+1} = \hat{\pi}_n W, \quad n \geqslant 2,$$

where $W = -BC^{-1}$ can be found explicitly as

$$W = \begin{pmatrix} \frac{\lambda_s}{c_1} & 0 & -\frac{\lambda_s}{c_1} & 0 \\ 0 & \frac{\lambda_s}{c_2} & 0 & -\frac{\lambda_s}{c_2} \\ \frac{\lambda_s}{c_1} - c_3 & \frac{\alpha\lambda_p c_1}{\mu_s c_2} & \frac{c_1}{\mu_s} + c_3 - \frac{\lambda_s}{c_1} & -\frac{c_3}{\mu_p} - \frac{\alpha\lambda_p c_1}{\mu_s c_2} \\ \frac{\mu_p c_2}{\beta c_1} & \frac{\lambda_s}{c_2} & -\frac{\mu_p c_2}{\beta c_1} & \frac{c_2}{\beta} - \frac{\lambda_s}{c_2} \end{pmatrix}.$$

Finally, the following boundary value problem has to be solved in order to obtain the steady-state solution explicitly:

$$[\hat{\pi}_0 \ \ \hat{\pi}_2] \begin{pmatrix} B_0 & B_1 \\ 0 & C \end{pmatrix} = \mathbf{0}, \tag{2}$$

where π_0 is a $m_0 + m_1$–component vector, and π_2 corresponds to states $(2, \cdot, \cdot)$. Uniqueness of the solution is conditioned by $\hat{\pi}\mathbf{1} = 1$. However, the system (2) is defective, and requires $m_0 - 1$ additional equations to deliver the solution. These equations are obtained by removing the right eigenvectors r of matrix W corresponding to eigenvalues on or outside the unit disk, by equations of the type $\hat{\pi}_2 r = 0$.

References

[1] *Garimella Rama Murthy, Alexander Rumyantsev.* On an exact solution of the rate matrix of G/M/1-type Markov process with small number of phases // Journal of Parallel and Distributed Computing. 2018. **119**. P. 172–178. DOI: 10.1016/j.jpdc.2018.04.013

[2] *Beuerman S. L., Coyle E. J.* State space expansions and the limiting behavior of quasi-birth-and-death processes // Advances in Applied Probability. 1989. **21**(02), P. 284–314. https://doi.org/10/cs58tc

Multiplicative graphs and their application to the equation $n - \varphi(n) = c$

A. S. Semchankau

Lomonosov Moscow State University, Moscow, Russia

E-mail: `aliaksei.semchankau@gmail.com`

We will study the solutions to the equation $f(n) - g(n) = c$, where f and g are multiplicative functions and c is a constant. More precisely, we prove that the number of solutions does not exceed $c^{1-\epsilon}$ when f, g and solutions n satisfy some certain constraints, such as $f(n) > g(n)$ for $n > 1$.
In particular, we will prove the following estimate: the number of solutions of the equation $n - \varphi(n) = c$ (here $G(k)$ is the number of ways to represent k as a sum of two primes) is:

$$G(c+1) + O(c^{\frac{3}{4}+o(1)})$$

To obtain this we use so-called multiplicative graphs.

On the weak chromatic number of random hypergraphs

A. S. Semenov

Lomonosov Moscow State University, Moscow, Russia

E-mail: alexander.s.semenov@yandex.ru

For an integer j, a *j-independent set* in a hypergraph $H = (V, E)$ is a subset $W \subset V$ such that for every edge $e \in E : |e \cap W| \leqslant j$. A *$j$-proper coloring* of $H = (V, E)$ is a partition of the vertex set V of H into disjoint union of j-independent sets, so called colors. The *j-chromatic number* $\chi_j(H)$ of H is the minimal number of colors needed for a j-proper coloring of H. When j is greater than 1 the corresponding chromatic number usually called *weak*.

This talk is devoted to weak colorings of k-uniform random hypergraph $H(n, k, p)$. We are interested in asymptotic properties of $H(n, k, p)$ to have its j-chromatic number equal to some fixed number r. By asymptotic properties of $H(n, k, p)$ we consider n as tending to infinity, while k and r are kept constant.

It can be showed that the previously mentioned property of random hypergraph has a sharp threshold. For the case of $(k-1)$-chromatic number upper and lower bounds for that threshold are found. It should be also mentioned that the gap between this bounds is some function $O_k(1)$.

In our previous works we found very tight bounds for the case of $k - j = o(\sqrt{k})$ and $r = 2$. With this work we continue our generalization to the case when j is less than $k - 1$, but r is greater than 2. Main result is showed in a theorem below.

Theorem. *Let $H(n, k, p)$ be a random k-uniform hypergraph with $p = cn/\binom{n}{k}$. For any $r \geqslant 2$ there exists $k_0 \in \mathbb{N}$, such that if $k > k_0(r)$ and*

$$c > \frac{r^{k-1}\ln r}{kr - k + 1} - \frac{\ln r}{2} + O\left(kr^{2-k}\right)$$

then w.h.p. as n tends to infinity, $\chi_{k-2}(H(n, k, p)) > r$. Otherwise, if

$$c < \frac{r^{k-1}\ln r}{kr - k + 1} - \frac{\ln r}{2} + O(k^{-1})$$

then w.h.p. as n tends to infinity, $\chi_{k-2}\left(H(n,k,p)\right) \leqslant r$.

This is joint work with Dmitrii Shabanov. The reported study was funded by RFBR according to the research project no. 18-31-00348 and by the Russian Science Foundation under grant no. 16-11-10014.

Probability thresholds for coloring properties of random graphs and hypergraphs

D. A. Shabanov

Moscow Institute of Physics and Technology, Dolgoprudny, Russia

E-mail: dmitry.shabanov@phystech.edu

The talk deals with the problems concerning colorings of random graphs and hypergraphs. Let $H(n,k,p)$ denote the classical binomial model of a random k-uniform hypergraph: every edge of a complete k-uniform hypergraph on n vertices is included into $H(n,k,p)$ independently with probability $p \in (0,1)$. Recall also the definition of the random graph $G(n,p) = H(n,2,p)$. It is well-known that every decreasing monotone property $Q = (Q_n, n \in \mathbb{N})$ of hypergraphs has *a probability threshold* in the binomial model, i.e. there exists a function $\widehat{p} = \widehat{p}(n)$ such that

$$\lim_{n\to\infty} \mathsf{P}(H(n,k,p) \text{ has } Q_n) = \begin{cases} 1, & \text{if } p = p(n) = o(\widehat{p}); \\ 0, & \text{if } p = p(n) = \omega(\widehat{p}). \end{cases}$$

In the talk we concentrate on coloring properties of hypergraphs. Recall that for given $r \geq 2$, a hypergraph H is said to be *r-colorable* if there exists a coloring of the vertices of H with r colors such that every edge is not monochromatic under this coloring. The property of r-colorability is decreasing, but moreover, its probability threshold is sharp in the $H(n,k,p)$ model: if $(r,k) \neq (2,2)$ then there exists a function $\widehat{p}_{r,k} = \widehat{p}_{r,k}(n)$ such that for any $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \mathsf{P}(H(n,k,p) \text{ is } r\text{-colorable}) = \begin{cases} 1, & \text{if } p = p(n) \leq (1-\varepsilon)\widehat{p}_{r,k}; \\ 0, & \text{if } p = p(n) \geq (1+\varepsilon)\widehat{p}_{r,k}. \end{cases}$$

The main problem is to *find or estimate the sharp probability function* $\widehat{p}_{r,k}$.

We will survey the recent advances concerning estimating the probability thresholds for coloring properties of random graphs and hypergraphs including:

- panchromatic colorings (every edge should contain the vertices of all the colors);
- strong colorings (every two vertices of the same color cannot share a common edge).

We also give new results concerning the concentration in few values of the chromatic number of a random graph $G(n,p)$ in the non-sparse case ($np \to +\infty$ slowly enough).

The work is supported by the Russian Science Foundation under grant N 16-11-10014.

Stability methods in extremal graph theory

M. Simonovits

Rényi Alfréd Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

E-mail: `miki@renyi.hu`

My two lectures, the first one at MIPT, the second one in Petrozavodsk, are strongly related to each other, yet, they describe different features of the same area, namely, of extremal graph theory. They are self-contained.

- . -

Extremal graph theory is one of the fastest developing areas of Graph Theory. Recently Endre Szemerédi and I have published a long survey (in the volume of the conference Building Bridges II, on the occasion of the 70th birthday of L. Lovász) on results and methods in extremal graph theory, and on embedding small or large subgraphs into large graphs. Here "large" means very large: we do not care of the small cases. We arrive at such problems typically when for small values of n the situation is chaotic but for large values the extremal structures are simple. (Another case where we get such results is when we apply the Szemerédi Regularity Lemma, or its variants.)

Extremal graph theory investigates problems in various settings, where a universe of objects (graphs, multigraph, digraphs, hypergraphs, random graphs) is fixed and also a family $\mathcal{L}$ of excluded graphs (from this universe) and we try to determine the maximum number of edges such a graph (say on n vertices) can have without containing forbidden subgraphs.

The extremal objects are the ones not containing forbidden substructures and having the maximum number of edges (for a fixed number of vertices).

One reason why extremal problems are important is that they led to extremely important new methods. They are strongly connected to Ramsey-Turán problems, Anti-Ramsey problems, to Dual Anti-Ramsey problems, Supersaturated graphs,...

I shall discuss an important subcase of these problems: the application of stability methods. Stability methods are applicable mostly when we conjecture that the extremal structures are fairly simple and also the nearly extremal objects have this simple structure. Then the stability method provides exact extremal structures in many cases, for n sufficiently large.

I shall discuss old and new extremal results and open problems in this and related areas. I shall speak of three particular stability methods: (a) Progressive Induction, (b) Separating the almost extremal cases from the fuzzy cases, (c) application of the Removal Lemma.

Invariants of graph drawings in the plane

A. B. Skopenkov

Moscow Institute of Physics and Technology, Independent University of Moscow, Moscow, Russia

E-mail: `skopenko@mccme.ru`
`http://www.mccme.ru/~skopenko`.

We present a simplified exposition of some classical and modern results on graph drawings in the plane. These results are chosen so that they illustrate some spectacular recent higher-dimensional results on the border of geometry, combinatorics and topology [2]. We define a mod2-valued *self-intersection invariant* (i.e. the van Kampen number) and its generalizations. We present elementary formulations and arguments, so we do not require any knowledge of algebraic topology. This talk will be accessible to mathematicians not specialized in any of the areas discussed. It may serve as an introduction into algebraic topology motivated by algorithmic, combinatorial and geometric problems. See [2].

Supported in part by the Simons-IUM fellowship and Russian Foundation for Basic Research Grant No. 19-01-00169.

References

[1] *Skopenkov A.* A user's guide to the topological Tverberg Conjecture // Russian Math. Surveys. 2018. **73**(2). P. 323–353. Earlier version: arXiv:1605.05141v4.

[2] *Skopenkov A.* Invariants of graph drawings in the plane, submitted, arXiv:1805.10237.

Angles of random polytopes

D. N. Zaporozhets

St. Petersburg Department of Steklov Institute, St. Petersburg, Russia

E-mail: zap1979@gmail.com

We will consider some problems on calculating the average angles of random polytopes. Some of them are open.

Large deviations in single-server retrial queues

K. A. Zhukova

Institute of Applied Mathematical Research KRC RAS,
Petrozavodsk, Russia

E-mail: kalininaksenia90@gmail.com

We consider a single-server retrial system with a renewal input of customers arriving at instants $t_n,\ n \geqslant 1$, with independent identically distributed (iid) interarrival times $\tau_n := t_{n+1} - t_n,\ n \geqslant 1,\ t_1 = 0$, and with the iid service times S_n, $n \geqslant 1$. We denote a renewal input with rate $\lambda := 1/\mathsf{E}\tau \in (0, \infty)$, and the service rate of the system is $\mu := 1/\mathsf{E}S \in (0, \infty)$.

In a retrial system, if a new customer finds the server busy he joins an infinite-capacity virtual orbit and attempts to occupy server after an exponentially distributed time with rate γ. We consider a constant retrial rate system. It means that retrial intensity of attempts equals γ and stays constant regardless of the orbit size. In this case, for convenience, we treat the orbit as a FIFO queue in which only the top (oldest) orbital customer makes attempts to enter server [1].

Denote K_0 the index of the first costumer which meets an empty system upon arrival, and K_N – the index of the first costumer which reaches the level N within busy cycle. We consider an overflow probability $\mathsf{P}(K_N < K_0)$ that the number of customers in the system reaches a (high) level $N > 1$ during busy cycle.

There are two basic assumptions required for the large deviation analysis: the system is in stationary regime and the possibility of an arbitrary (large) value of the queue. The sufficient stability condition for retrial system is [1]

$$\rho := \lambda/\mu < 1, \tag{1}$$

and coincides with the stability criterion of classic buffered system $GI/G/1$. Also we assume that

$$\mathsf{P}(\tau < S) > 0, \tag{2}$$

so the arbitrary large value of the queue is possible.

For a random variable X, we introduce the log moment generating function,

$$\Lambda_X(\theta) := \log \mathsf{E}[e^{\theta X}]. \tag{3}$$

We assume that $\Lambda_S(\theta)$ exists for some $\theta > 0$. Denote

$$\hat{\theta} = \max(\theta > 0 : \mathsf{E}e^{\theta S} < \infty). \tag{4}$$

Define

$$\theta_* = \sup(\theta \in (0,\, \hat{\theta}) : \Lambda_\tau(-\theta) + \Lambda_S(\theta) \leqslant 0) \tag{5}$$

and

$$\theta^* = \sup(\theta \in (0,\, \min(\hat{\theta}, \gamma)) : \Lambda_\tau(-\theta) + \Lambda_S(\theta) + \log \frac{\gamma}{\gamma - \theta} \leqslant 0). \tag{6}$$

Theorem 1. *Assume that conditions*

$$\lambda < \frac{\gamma\mu}{\gamma + \mu} \tag{7}$$

and (2) *hold. Then the decay rate of the overflow probability in single-server constant rate retrial system satisfies*

$$\Lambda_\tau(-\theta_*) \leqslant \limsup_{N\to\infty} \frac{1}{N} \ln \mathsf{P}(K_N < K_0) \leqslant \Lambda_\tau(-\theta^*), \tag{8}$$

where θ_* and θ^* are defined in (5) *and* (6)*, respectively. The similar statement holds with all limsups replaced by liminf.*

To prove the statement presented in Theorem 1 we transform original retrial system to a classic buffered system with special type of dependence between service times

$$\bar{S} = S + \mathbb{I} \cdot \exp(\gamma), \tag{9}$$

where $\exp(\gamma)$ is a random variable (r.v.) exponentially distributed with parameter γ, and $\mathbb{I}$ is an indicator function

$$\mathbb{I} = \begin{cases} 1, & \text{if server is busy upon an arrival;} \\ 0, & \text{otherwise.} \end{cases}$$

Service time $\bar{S}$ (9) the retrial time. Such classic system is equivalent to original retiral system from the point of view stability. To obtain a lower asymptotic bound, we construct a minorant classic buffered system $GI/G/1$

with the minimal service times $\tilde{S} = S$ and use a monotonicity of the queue-size process. To obtain an upper bound, we also construct a dominating classic buffered system with the service time $\hat{S} = S + \exp(\gamma)$. This approach allows to obtain exponential decay rate of the overflow probability as $N \to \infty$, with different exponents in the asymptotic lower and upper bounds (5) and (6), respectively.

To simplify the analysis of a retrial system it can be useful to compare it to a classic system. In the retrial system customer goes to orbit if server is busy upon the arrival of this customer. Occupation of server can be expressed by means of load coefficient $\rho = \lambda/\mu$. Hence, retrial system can be approximated by classic buffered system with service times

$$S_c = S + \mathbf{I} \cdot \exp(\gamma), \qquad (10)$$

where

$$\mathbf{I} = \begin{cases} 1, \text{with probability} & \rho = \lambda/\mu; \\ 0, \text{with probability} & 1 - \rho. \end{cases}$$

Simulation results shows that the estimated probabilities both in the retrial and classic systems are indeed located between the bounds. Moreover they are quite close to each other. It means that more simple classic buffered system can be analysed instead of the origin retrial system.

References

[1] *Avrachenkov K., Morozov E., Nekrasova R., Steyaert B.* Stability analysis and simulation of N-class retrial system with constant retrial rates and Poisson inputs // Asia-Pacific Journal of Operational Research. 2014. **31**(2).

[2] *Buijsrogge A., De Boer P.-T., Rosen K., Scheinhardt W.* Large deviations for the total queue size in non-Markovian tandem queues. // Queueing Systems. 2017. **85**. P. 305–312.

[3] *Kim J., Kim B.* Tail asymptotics for the queue size distribution in the MAP/G/1 retrial queue // Queueing Systems. 2010. **66**. P.79–94.

[4] *Kim J., Kim B., Ko S.-S.* Tail asymptotics for the queue size distribution in an M/G/1 retrial queue
J. Appl. Probab. 2007. **44**. P. 1111–1118.

[5] *Sadowsky J. S.* Large deviations theory and efficient simulation of excessive backlogs in a GI/GI/m queue // IEEE Trans. Autom. Control. 1991. **36**(12). P. 1383–1394.

Об оптимальном управлении параллельной приоритетной очередью

Е. А. Аксёнова, А. В. Соколов

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводск, Россия

E-mail: aksenova@krc.karelia.ru, sokavs@gmail.com

Во многих приложениях, например, в операционных системах и сетях, требуется структура данных, в которой основными операциями являются операции вставки нового элемента и удаления элемента с наибольшим значением ключа (приоритета). Такую структуру данных называют приоритетной очередью. Основными методами реализации таких структур данных являются упорядоченные и неупорядоченные списки и массивы, пирамиды, различные варианты бинарных деревьев поиска и другие методы. Все эти методы имеют разные оценки сложности по времени и памяти [1]. В этой работе анализируется способ представления приоритетной очереди в виде нескольких последовательных циклических FIFO-очередей. Такой метод представления используется в сетевых приложениях [2].

В [3–5] мы анализировали модели работы с FIFO-очередями и приоритетными очередями в случае последовательного выполнения операций с очередями. В настоящее время многоядерные процессоры и параллельные вычисления все шире используются в различных приложениях, в том числе и в сетевых.

В [6] предложены и анализируются модели работы с параллельными FIFO-очередями, а в [7–9] – модели работы с work-stealing деками. Эти структуры данных используются при реализации work-stealing балансировщиков параллельных вычислений [10]. Отметим, что «расслабленные» модификации приоритетной очереди тоже могут применяться в модифицированных реализациях work-stealing балансировщиков, и на ряде тестов, например, на алгоритме нахождения кратчайшего пути Дейкстры, такие реализации показывают лучшие результаты [11], чем при реализации с деками.

В этой работе мы предлагаем модель работы с двухприоритетной очередью, в которой допускаются параллельные операции.

Рассмотрим очередь, имеющую два приоритета, расположенную в ограниченной памяти. Очередь представлена в виде двух FIFO-очередей первого и второго приоритета.

Работа с очередями происходит в дискретном времени. На каждом шаге могут произойти с заданными вероятностями параллельные операции с очередями. Исключение из приоритетной очереди происходит по наивысшему приоритету. Это значит, что пока вторая очередь не пустая исключение происходит из нее. Как только она станет пустой, исключения будут происходить из первой очереди.

В работе предложена математическая модель процесса в виде случайного блуждания по целым точкам на плоскости и решается задача нахождения оптимальных размеров очередей, когда в качестве критерия оптимальности рассмотрено максимальное среднее время до переполнения памяти. Для решения поставленной задачи разработана имитационная модель. В докладе будут представлены результаты численных экспериментов с разработанной моделью.

Работа выполнена при финансовой поддержке РФФИ, грант 18-01-00125-а.

Список литературы

[1] *Knuth D.* The Art of Computer Programming. Vol. 1. Addison-Wesley, 2001.

[2] *Олифер В. Г., Олифер Н. А.* Компьютерные сети. Принципы, технологии, протоколы: учебник для вузов. 4-е изд. СПб.: Питер, 2010.

[3] *Sokolov A. V., Drac A. V.* The Linked List Representation of n LIFO-Stacks and/or FIFO-Queues in the Single-Level Memory // Information Processing Letters. 2013. **13**. P. 832–835.

[4] *Aksenova E. A., Sokolov A. V.* The optimal implementation of two FIFO-queues in single-level memory // Applied Mathematics. 2011. **2**(10). P. 1297–1302.

[5] *Аксёнова Е. А., Соколов А. В.* Анализ некоторых методов реализации приоритетной очереди // Межвуз. сб. «Стохастическая оптимизация в информатике» Изд-во С.-Петербургского университета. 2008. **4**. С. 61–71.

[6] *Аксёнова Е. А., Соколов А. В.* Об оптимальном параллельном управлении n FIFO-очередями в общей памяти // Обозрение прикладной и промышленной математики. 2018. **25**(1).

[7] *Sokolov A. V., Barkovsky E. A.* The Mathematical Model and The Problem of Optimal Partitioning of Shared Memory for Work-Stealing Deques // Lecture Notes in Computer Science. 2015. **9251**. P. 102–106.

[8] *Aksenova E. A., Sokolov A. V.* Modeling of the Memory Management Process for Dynamic Work-Stealing Schedulers // IEEE Proceedings: Proceedings of 2017 Ivannikov ISPRAS Open Conference (ISPRAS). 2018. P. 12–15. DOI: 10.1109/ISPRAS.2017.00009

[9] *Барковский Е. А., Лазутина А. А., Соколов А. В.* Построение и анализ модели процесса работы с двумя деками, двигающимися друг за другом в общей памяти // Программные системы: теория и приложения. 2019. **10:1**(40), С. 3–17. DOI: 10.25209/2079-3316-2019-10-1-3-17. URL: http://psta.psiras.ru/read/psta2019_1_3-17.pdf

[10] *Kuchumov R., Sokolov A., Korkhov V.* Staccato: Cache-Aware Work-Stealing Task Scheduler for Shared-Memory Systems // ICCSA 2018. Lecture Notes in Computer Science, **10963**. Springer, Cham. 2018. P. 91–102. DOI: 10.1007/978-3-319-95171-3_8

[11] *Wimmer M., Versaci F., Cederman D., Tráff J. L., Tsigas F.* Data structures for task-based priority. (Poster Paper) In the Proceedings of the 19th ACM SIGPLAN symposium on Principles and practice of parallel programming (PPoPP 2014). ACM press 2014. P. 379–380.

Апостериорная оценка погрешности дискретизации на ансамбле решений и эффект концентрации меры

А. К. Алексеев

Московский физико-технический институт, Долгопрудный, Россия

E-mail: aleksey.k.alekseev@gmail.com

Известно, что эффект концентрации меры [1–3] имеет существенное значение в статистической термодинамике, как системе очень многих переменных. Сеточные функции, используемые при дискретизации многомерных уравнений в частных производных (например, уравнений Эйлера, описывающих невязкое течение газа и рассмотренных в докладе), также принадлежат пространству большой размерности (от $N \sim 10^6$ и выше). Соответственно, интересно, сказывается ли (и как) эффект концентрации меры в задачах такого типа, в частности, при апостериорной оценке погрешности дискретизации.

Влияние эффекта концентрации зависит как от геометрии множества, занимаемого решением (или его погрешностью) в многомерном пространстве, так и от алгоритма выбора векторов на этом множестве (степени произвольности (независимости) выбора). Численные решения уравнений аэрогазодинамики, полученные разными методами, нельзя рассматривать как независимые вектора, так как они описывают единственное решение и, поэтому, должны проявлять высокую степень корреляции. В то же время легко показать, что локальная погрешность аппроксимации (ошибка усечения) в случае алгоритмов разного порядка точности является линейно независимой. Степень независимости погрешности дискретизации для решений, полученных разными алгоритмами определяется, как ошибкой усечения так и искусственными монотонизаторами, применяемыми в современных алгоритмах, *a приори* не ясна и обсуждается ниже. Систему уравнений обозначим как $A\tilde{u} = f$, численное решение, полученное k-*м* алгоритмом определяется дискретным оператором $A_h u^{(k)} = f_h$. Ошибка усечения $\delta u^{(k)}$ получается из разложения в ряд Тейлора численного решения (сеточной функции) $u^{(k)}$ и подстановки результата в основную систему уравнений. Лагранжева форма ошибки усечения $\delta u^{(k)}(\alpha_n) = Ch^m \cdot \partial^{m+1} u(x_n + \alpha_n)/\partial x^{m+1}$ (n номер узла сетки) содержит неизвестный параметр $\alpha_n \in [0,1]$ [4]. В связи с этим и тем, что $\Delta u^{(k)} = (u^{(k)} - \tilde{u}) = A_h^{-1} \delta u^{(k)}$, ошибка дискретизации обычно воспринимается как детерминированная величина. Однако, существует ряд численных работ (например, [5]), которые основаны на том, что компоненты ошибки дискретизации $\Delta u_i^{(k)}$ являются случайной нормально распределенной величиной. Другие численные эксперименты [4] указывают на наличие достаточно универсальной формы плотности распределения лагранжева параметра $P(\alpha_n)$ для уравнения теплопроводности. Данные расчетов, представленные в данном докладе, указывают на то, что для ансамбля численных решений, полученных независимыми методами, максимум расстояния между решениями $\|du_{\max}\|$ может служить верхней оценкой погрешности дискретизации с индексом эффективности $\frac{\|du_{\max}\|}{\|\Delta u\|} \sim 0.6 \div 3$.

Таким образом, имеется противоречие между случайным (наблюдаемым) и детерминированным (теоретически обоснованным) подходами к оценке погрешности дискретизации. Возможно, этот парадокс можно разрешить с помощью теории концентрации меры, связывающей вероятность и геометрию многомерных пространств. Этой надежде способствует то, что, как погрешность усечения $\delta u^{(i)} \in R^N$, так и погрешность дискретизации $\Delta u^{(i)} = u^{(i)} - \tilde{u} \in R^N$, принадлежат пространству очень высокой размерности.

Для конечно-разностных схем несовпадающего порядка аппроксимации k погрешность $\delta u^{(k)}$ линейно независима, так как содержит производные разного порядка. Что касается погрешностей дискретизации $\Delta u^{(k)}$, то для таких схем не существует алгоритма, трансформирующего их друг в друга, поэтому естественным выглядит предположение, что они являются алгоритмически независимыми (случайными по Колмогорову). Рассмотрим ансамбль из M численных решений, полученных на одной сетке с помощью независимых разностных схем (в простейшем случае схем разного порядка). Если предположить, что алгоритмическая независимость погрешности дискретизации $\Delta u^{(k)}$, обеспечивает произвольность выбора, то принадлежность погрешности к пространствам очень большой размерности дает нестандартные возможности для определения нормы погрешности и положения истинного решения. Известно, что в пространствах достаточно большой размерности N расстояние $d_{1,2}$ между двумя произвольно выбранными векторами $v^{(1)} \in R^N$ и $v^{(2)} \in R^N$ “с вероятностью 1” больше длины этих векторов $\left\|v^{(i)}\right\| \leq d_{1,2}$. Это вызвано тем обстоятельством, что хорда в этих пространствах “почти всегда больше радиуса” [3], а два произвольно выбранных единичных вектора с большой вероятностью ортогональны [1], а именно $P\{(v^{(1)}, v^{(2)}) > \delta\} < \sqrt{\pi/2}e^{-\delta^2 N/2}$.

Применительно к нашему случаю, учтем, что разность между численными решениями равна разности между погрешностями дискретизации этих решений $u^{(1)} - u^{(2)} = u^{(1)} - \tilde{u} - u^{(2)} + \tilde{u} = \Delta u^{(1)} - \Delta u^{(2)}$. Считаем также, что норма погрешностей ограничена $\|\Delta u^{(i)}\| \le R$ (они принадлежат некоторой гиперсфере радиуса R с центром в нуле). В таком случае, если погрешности дискретизации выбраны произвольно, расстояние $d_{1,2} = \|u^{(1)} - u^{(2)}\| = \|\Delta u^{(1)} - \Delta u^{(2)}\|$ между двумя численными решениями $u^{(1)} \in R^N$ и $u^{(2)} \in R^N$ "с вероятностью 1" больше расстояния от точного решения до численного $d(u^{(i)}, \tilde{u}) = \|\tilde{u} - u^{(i)}\| \le d_{1,2} = \|\Delta u^{(1)} - \Delta u^{(2)}\|$.

В итоге, для двух численных решений, полученных независимыми методами, расстояние между ними "с вероятностью 1" определяет верхнюю грань нормы погрешности дискретизации (расчета).

Таким образом, обоснованием апостериорной оценки погрешности дискретизации на ансамбле решений может служить "благословение размерности", вследствие которого два произвольно выбранных вектора в пространстве большой размерности "почти всегда" ортогональны. Близкий результат получается из других геометрических соображений, а именно, с помощью неравенства треугольника [6].

Список литературы

[1] *Зорич В. А.* Многомерная геометрия, функции очень многих переменных и вероятность // ТВП. 2014. **59**(3). С. 436–451.

[2] *Gorban A. N., Tyukin I. Y.* Blessing of dimensionality: mathematical foundations of the statistical physics of data. 2018. arXiv:1801.03421v1.

[3] *Sidiropoulos P.* N-sphere chord length distribution. 2014. arXiv:1411.5639v1.

[4] *Alekseev A. K., Makhnev I. N.* On using the Lagrange coefficients for a posteriori error estimation // Numer. Analys. Appl. 2009. **2**(4). P. 302–313.

[5] *Rauser F., Marotzke J., Korn P.* Ensemble-type numerical uncertainty quantification from single model integrations // Journal Comp. Physics. 2015. **292**(C). P. 30–42.

[6] *Alekseev A. K., Bondarev A. E., Navon I. M.* On Triangle Inequality Based Approximation Error Estimation // 2017. arXiv:1708.04604.

Применение метода расщепления для оценивания характеристик графа большой размерности

К. Е. Авраченков[1], А. В. Бородина[2,3], Н. А. Нечепаренко[3]

[1] Inria, Sophia Antipolis, France
[2] Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия
[3] Петрозаводский государственный университет, Петрозаводск, Россия

E-mail: k.avrachenkov@sophia.inria.fr, borodina@krc.karelia.ru, nechepor2880@gmail.com

Одним из методов для оценивания характеристик современных инфокоммуникационных сетей и социальных сетей является стандартное случайное блуждание в графе большой размерности. Для сетей большого объема характерно, что вся сеть целиком недоступна для исследования, либо время на обход всей структуры слишком велико для эффективного оценивания характеристик сети.

В данной работе предлагается использовать технику расщепления для ускорения процедуры случайного блуждания. В частности, такой подход будет применен для построения оценки вероятности достижения некоторого узла сети (или области сети) до момента возвращения в стартовый узел (или множество).

Рассмотрим модель сети в виде неориентированного графа $G = (V, E)$, со стартовым множеством узлов $I_n \subset V$, $|I_n| = n$, в которое входят узлы из всех компонент связности графа G. Для некоторого узла $b \in V$ необходимо построить оценку для вероятности $p_{esc(a\to b)}$, что стартуя из узла a случайное блуждание достигнет узла b до момента возвращения в узел a:

$$p_{esc(a\to b)} = \mathbb{P}\{X_j = b,\ X_i \neq a,\ 0 \leq i \leq j < \infty | X_0 = a\}.$$

Построение оценок для такого типа характеристик актуально для задач вычисления эффективного сопротивления в электрических цепях (*escape probability* [3]), а также в задачах, связанных с редкими события (*first entrance probability* [4]). В веб-графах большой размерности такая вероятность не всегда является достаточно малой (кроме ситуаций с притягивающими узлами), чтобы использовать технику ускоренного моделирования. Тем не менее, время оценивания даже таких вероятностей (порядка $0,1$ - $0,5$) с помощью случаного блуждания достаточно велико. Ускоренное построение циклов случайного блуждания позволит строить оценку с большей точностью.

Основным требованием для использования метода случайного блуждания является связность графа. Для построения связного графа $G^{'} = (V^{'}, E^{'})$, эквивалентного исходному с точки зрения случайного блуждания используется метод введения "суперузла, предложенного в работах [1,2], а также метод частичного случайного блуждания, который позволяет строить несмещенные и состоятельные оценки характеристик узлов на основе известного подграфа графа G. Далее, считаем граф G связным с учетом необходимых преобразований.

Пусть $X = \{X_t,\ t \geq 0\}$ - цепь Маркова, соотвествующая случайному блуждания в графе G со счетным пространством состояний χ. Определим k-ую траекторию как последовательность узлов, посещенных случайным блужданием со стартом в узле a до момента возврата обратно

$$X_1^{(k)}, \dots, X_{\xi_k}^{(k)},\ k \geq 1, \tag{1}$$

где дискретные моменты ξ_k являются моментами первого перехода в a. Последовательность $\{\xi_k\}_{k\geq 1}$ определяет процесс восстановления, а независимые траектории можно рассматривать как циклы регенерации. Условие связности графа гарантирует конечность цикла регенерации.

Далее регенеративная постановка дает возможность применить метод расщепления, где в качестве функции значимости $S : \chi \to \mathbb{R}$ используется кратчайшее расстояние по алгоритму Дейкстры от текущей вершины x до целевой вершины b. Таким образом, на каждом шаге случайного блуждания вершина b становится либо ближе, либо дальше. Расщепление траектории случайного блуждания происходит при приближении к вершине b.

Согласно функции $S(x)$ определяются уровни, на которых происходит расщепление и последовательность вложенных событий

$$l = l_0 \geq l_1 \geq \cdots \geq l_M = 0,\ \ l := S(a);\ \ E_M \subset E_{M-1} \subset \cdots \subset E_0,$$

где $T_k = \inf\{t > 0 : S(X_t) \leq l_k\}$, $E_k = \{T_k < \tau_A\}$, $\tau_A = \inf\{t \geq 0 : X_{t-1} \notin A \ \&\ X_t \in A\}$, $A = \{x \in \chi : S(x) = l\}$. Множество A может состоять как из одного узла a, так и из нескольких. Тогда

$$p_{esc(a\to b)} = \prod_{k=1}^{M} p_k = \mathbb{P}[E_M], \tag{2}$$

где $p_k = \mathbb{P}[E_k|E_{k-1}],\ k = 1,\ldots,M.$ — вероятности перехода на уровень k при условии достижения предыдущих $k-1$ уровней.

Тогда оценка для требуемой вероятности методом расщепления определяется формулой:

$$\widehat{p}_{esc}[SP] = \prod_{k=1}^{M} \widehat{p}_k, \quad \widehat{p}_k = \frac{N_k}{r_{k-1}N_{k-1}}, \tag{3}$$

где r_k – число расщеплений на уровне l_k, N_k — число достижений уровня k, а $r_{k-1}N_{k-1}$ —общее число траекторий, стартующих с уровня $k-1$.

Численные эксперименты показали преимущество предлагаемого метода по сравнению со стандартным методом Монте-Карло и численным решением системы линейных уравнений, построенной по матрице Лапласа.

Исследование выполнено по госзаказу в Карельском научном центре РАН (Институт прикладных математических исследований КарНЦ РАН) и при поддержке Российского фонда фундаментальных исследований, проекты 18-07-00187, 18-07-00147.

Список литературы

[1] *Avrachenkov K., Ribeiro B., Towsley D.*, Improving Random Walk Estimation Accuracy with Uniform Restarts // Proceedings of WAW, 2010, P. 98–109.

[2] *Avrachenkov K., Ribeiro B., Sreedharan J. K.* Inference in OSNs via Lightweight Partial Crawls // Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science. 2016. P. 165–177.

[3] *Vos V. S. S.* Methods for determining the effective resistance. Master's thesis, 20 December 2016.

[4] *Dean T., Dupuis P.* Splitting for rare event simulation: A large deviation approach to design and analysis // Stochastic processes and their applications. 2009. **119**(2). P. 562–587.

Реализуемость графов с вращениями на ориентируемых поверхностях

Т. В. Березин

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

E-mail: `timurovadia@gmail.com`

Граф называется реализуемым на поверхности, если его можно на ней расположить (нарисовать) без самопересечений.

Проблема реализуемости графов на плоскости, торе, листе Мёбиуса и других поверхностях хороша известна и является одной из основных в топологической теории графов [4, 7].

Реализуемость графов на поверхности является проблемой не только теоретической, но и практической. Во многих современных инженерных задачах (например, проектирование печатных плат сверхбольших интегральных схем, проектирование транспортных развязок) важно выяснить, возможно ли нарисовать граф на поверхности так, чтобы его изображение удовлетворяло определенным требованиям. Эти задачи решаются путем применения компьютерных алгоритмов [1].

Хорошо известен критерий реализуемости графа на плоскости — теорема Понтрягина – Куратовского: *граф планарен тогда и только тогда, когда он не содержит подграфов, гомеоморфных* K_5 *или* $K_{3,3}$ [6].

Получение критериев реализуемости графов на двумерном торе может быть полезным с алгоритмической точки зрения и представляет самостоятельный теоретический интерес.

Для распознавания реализуемости графов будем пользоваться дополнительной конструкцией **графа с вращениями** (другие интерпретации: ориентированное утолщение, ленточный граф, R-граф).

Графом с вращениями называется граф вместе с указанием для его вершин ориентированного циклического порядка выходящих из нее полуребер или концов петель.

Будем рассматривать мультиграфы с петлями и кратными ребрами. Договоримся сокращать «мультиграф» до «граф».

Пусть G — граф с одной вершиной и конечным числом n петель.

Иероглифом называется граф G с указанием для его единственной вершины ориентированного циклического порядка выходящих из нее концов петель.

Обозначим петли G буквами алфавита в порядке перечисления выходящих из вершины концов петель при обходе вокруг вершины по часовой стрелке. В таком случае иероглиф запишется как слово-строка длины $2n$ из n букв, в котором каждая буква встречается дважды, с точностью до переименования букв и циклического сдвига. Будем сокращать «иероглиф, соответствующий слову x» до «иероглиф x».

Графом петель иероглифа называется граф, вершинами которого являются петли иероглифа. Две вершины графа петель соединены ребром, если соответствующие две петли образуют иероглиф $(abab)$ (т.е. пересекаются).

Редукция иероглифа — композиция некоторого числа следующих преобразований:

(D) Удаление некоторой изолированной петли, т.е. такой петли, которая в циклическом порядке задается двумя подряд идущими буквами $(aa...)$;

(R) Замена двух 'параллельных' петель a и a', т.е. петель, соответствующие которым буквы в циклическом порядке находятся на соседних местах и не чередуются $(aa'...a'a...)$ на одну.

Основным результатом работы является доказательство следующего утверждения.

Теорема. *Следующие условия на иероглиф эквивалентны:*

(A) *Иероглиф реализуем на торе.*

(B) *Иероглиф не содержит ни одного из иероглифов:* $(ababcdcd)$, $(abcdabcd)$, $(abacdcbd)$, $(abcadbdc)$.

(C) *Граф петель иероглифа является объединением изолированных вершин и полного дву- или трехдольного графа.*

(D) *Иероглиф редуцируется до одного из иероглифов:* $()$, $(abab)$, $(abcabc)$.

Критерий (B) теоремы получен из [3], где на рисунке 2 страницы 9 изображены иероглифы не реализуемые на торе. Доказательство этого результата может было и известно, но нигде ранее не опубликовано. В [5] А. Б. Скопенковым, опираясь на результаты [2], были сформулированы без доказательства остальные критерии теоремы. Однако для критерия (C) в первоначальной формулировке был найден контрпример и критерий уточнен в работе.

Список литературы

[1] *Агеева Т. и др.* Алгоритмы автотрассировки проводников на поверхности печатной платы // Тезисы конференции. Университет ИТМО, 2018. https://openbooks.ifmo.ru/ru/file/7259/7259.pdf

[2] *Culler M.* Using surfaces to solve equations in free groups // Topology. 1981. **20**. P. 133–145.

[3] *Каибханов А., Пермяков Д., Скопенков А.* Реализуемость графов с вращениями. https://www.turgor.ru/lktg/2005/3/rotrus.ps

[4] *Mohar B., Thomassen C.* Graphs on Surfaces. Baltimore, MD: Johns Hopkins University Press, 2001.

[5] *Скопенков А. Б.* Алгебраическая топология с геометрической точки зрения. М.: МЦНМО, 2015. 272 с. // Обновляемая электронная версия. https://www.mccme.ru/circles/oim/obstruct.pdf

[6] *Thomassen C.* Kuratowski's theorem // J. Graph. Theory. 1981. **5**, P. 225–242.

[7] *Звонкин А. К., Ландо С. К.* Графы на поверхностях и их приложения. М.: МЦНМО, 2010. 480 с.

Исследование $MMPP$-потока в условиях предельно частых и предельно редких изменений состояний

О. С. Бобкова, С. П. Моисеева, А. А. Назаров

Томский государственный университет, Томск, Россия

E-mail: osia153@yandex.ru

Значительная часть исследований по теории массового обслуживания проводится в предположении, что входящий поток - пуассоновский. Он является достаточно простым для соответствующих математических моделей, а также адекватным для широкого круга прикладных задач. Наблюдаемые потоки часто являются суммами большого числа независимых потоков, а они, в достаточно общих предположениях хорошо аппроксимируются пуассоновским процессом. При описании реальных объектов возникают модели, в которых первоначальный поток, проходя ряд последовательных систем, теряет часть своих требований. Такой редеющий поток также будет приближаться к пуассоновскому. Возросший в последнее время интерес к более сложным входящим потокам, в частности, к модулированным пуассоновским потокам ($MMPP$, MAP), в первую очередь связан с существованием практически важных задач, для которых предположение о том, что входящий поток пуассоновский, приводит к искажению операционных характеристик системы, например, занижению среднего числа находящихся в ней требований [1, 2]. Получение явного вида характеристик систем обслуживания со сложными потоками (такими как MAP) возможно лишь для узкого класса моделей, поэтому исследование, как правило, ведется в двух направлениях: построение оценок для этих характеристик и изучение условий высокой загрузки, когда в системе скапливается большое число требований, и малой, когда прибор значительную долю времени простаивает. Эта общность делает достаточно затруднительным применение полученных результатов к конкретным моделям, так как приходится проверять большое число условий и интерпретировать их в терминах, удобных для приложений.

MAP-поток (от Markovian Arrival Process), предложенный Ньютсом и Лукантони [3], широко применяется для моделирования потоков информации в спутниковых, компьютерных, беспроводных и мобильных сетях связи и представляет один из видов коррелированных потоков событий [1,2]. Известны два эквивалентных способа задания MAP [5], в данной работе используется модель, описанная в [5].

Пусть цепь Маркова $k(t)$ с конечным числом состояний $k = 1, 2, ..., K$ задана матрицей инфинитезимальных характеристик $\mathbf{Q}$ с элементами $q_{\nu k}$. Также задан набор неотрицательных чисел $\lambda_k \geq 0$, имеющим смысл условной интенсивности наступления события в потоке, кода цепь Маркова находится в k-м состоянии. Известно [6], что распределение вероятностей числа событий в высокоинтенсивном $MMPP$-потоке, наступивших за время t, можно аппроксимировать нормальным. Обозначим $n(t)$ – число событий определяемого случайного потока, наступивших за время t.

Рассмотрим поведение потока при других предельных условиях, а именно в условии предельно частых и предельно редких изменениях состояний управляющей потоком цепи Маркова.

Под условием предельно частых изменений состояния потока будем понимать такие значения параметров потока, при которых время пребывания управляющей цепи Маркова в каждом состоянии достаточно малое, то есть стремится к нулю.

Длина интервала времени, когда цепь Маркова находится в k-м состоянии, распределена по экспоненциальному закону с параметром $-q_{kk}$, то есть среднее время пребывания равно $M\{\tau\} = \frac{1}{-q_{kk}}$. Таким образом, условие предельно частых изменений состояния потока можно записать как $q_{kk} \to \infty$.

Показано, что производящая функция числа наступивших событий в потоке при асимптотическом условии предельно частых изменений состояний управляющей цепи Маркова имеет вид

$$f(z,t) = M\{z^{n(t)}\} = \{(z-1)\kappa t\},$$

где $\mathbf{r\Lambda e} = \kappa$.

То есть $MMPP$ поток при условии предельно частых изменениях управляющей цепи Маркова является пуассоновским с параметром $\kappa = \mathbf{r\Lambda e}$.

Под условием предельно редких изменений состояния потока будем понимать такие значения параметров потока, при которых время пребывания управляющей цепи Маркова в каждом состоянии достаточно долгое, то есть стремится к бесконечности.

Длина интервала времени, когда цепь Маркова находится в k-ом состоянии, распределена по экспоненциальному закону с параметром $-q_{kk}$, то есть среднее время пребывания равно $M\{\tau\} = \frac{1}{-q_{kk}}$. Таким образом, условие предельно редких изменений состояния потока можно записать как $q_{kk} \to 0$. Тогда производящая функция числа наступивших событий в потоке при асимптотическом условии предельно редких изменений состояний управляющей цепи Маркова вид средневзвешенного пуассоновского потока:

$$f(z,t) = \sum_{k=1}^{N} \mathbf{r}(k) exp\{(z-1)\lambda_k\}.$$

Такой поток будем называть гиперпуассоновский поток.

Аналогчным образом в указнных условиях проведены исследования систем массового обслуживания с входящим MMPP потоком и неограниченным числом обслуживающих приборов.

Список литературы

[1] *Kang S. H., Kim Y. H., Sung D. K., Choi B. D.* An application of Markovian Arrival Process to modeling superposed ATM cell streams // IEEE Transactions on Communications. 2002. **50**(4). P. 633–642.

[2] *Klemm A., Lindermann C., Lohmann M.* Modelling IP traffic using the batch Markovian arrival process // Performance Evaluation. 2003. **54**. P. 149–173.

[3] *Lucantoni D. M.* New results on single server queue with a batch Markovian arrival process // Stochastic Models. 1991. **7**. P. 1–46.

[4] *Назаров А. А., Моисеева С. П., Моисеев А. Н.* Способы задания MAP-потока и их эквивалентность // Теория вероятностей, математическая статистика и их приложения: сб. науч. ст. / под ред. Н.Н. Труша, Г.А. Медведева, Ю.С. Харина. – Минск: РИВШ. 2014. С. 160–164.

[5] *Назаров А. А., Моисеева С. П.* Метод асимптотического анализа в теории массового обслуживания. Томск: Изд-во НТЛ, 2006. 112 с.

[6] *Моисеев А. Н., Назаров А. А.* Бесконечнолинейные системы и сети массового обслуживания. Томск: Изд-во НТЛ, 2015. 240 с.

Об использовании бета- и гамма-распределений в численных рандомизированных моделях

Т. Е. Булгакова, А. В. Войтишек

Новосибирский государственный университет, Новосибирск, Россия

E-mail: vav@osmf.sscc.ru

С развитием вычислительной техники возрастает роль численных (реализуемых на компьютере) методов решения прикладных задач, в частности, методов численного статистического моделирования (или методов Монте-Карло); см., например, [1].

Применение методов Монте-Карло обусловлено наличием в соответствующей математической модели случайных параметров и (или) функций. Введение (выбор) этих параметров и функций называется *рандомизацией* модели. Данный доклад посвящен проблемам выбора одномерного случайного параметра ξ и соответствующей плотности распределения $f_\xi(u)$ (здесь $u, \xi \in (a,b) \subseteq \mathbb{R}$) при рандомизации численных моделей.

Существует достаточно устойчивое (и в целом – ошибочное) мнение ученых, занимающихся вопросами конструирования прикладных вероятностных моделей (в том числе компьютерных), о том, что если требуется ввести одномерный случайный параметр ξ, распределенный *в конечном интервале* (a,b); $-\infty < a < b < +\infty$, то (с точностью до линейного преобразования) следует использовать случайную величину $\xi = \beta^{(\mu,\nu)}$, имеющую плотность *бета-распределения*

$$f_\beta^{(\mu,\nu)}(u) = \frac{u^{\mu-1}\,(1-u)^{\nu-1}}{B(\mu,\nu)}, \quad 0 < u < 1 \tag{1}$$

со специально подобранными положительными параметрами μ и ν; здесь $B(\mu,\nu) = \int_0^1 t^{\mu-1}\,(1-t)^{\nu-1}\,dt$ – бета-функция.

Если же требуется ввести случайный параметр ξ, распределенный *в полубесконечном интервале (полуоси)* (a,b) (здесь $-\infty < a < b = +\infty$

или $-\infty = a < b < +\infty$), то (с точностью до линейного преобразования) рассматривается (и тоже – не вполне оправдано) случайная величина $\xi = \gamma^{(\lambda,\nu)}$, имеющая плотность *гамма-распределения*

$$f_\gamma^{(\lambda,\nu)}(u) = \frac{\lambda^\nu\, u^{\nu-1}\, e^{-\lambda\, u}}{\Gamma(\nu)}, \quad u > 0 \tag{2}$$

со специально подобранными положительными параметрами λ и ν; здесь $\Gamma(\nu) = \int_0^{+\infty} t^{\nu-1}\, e^{-t}\, dt$ – гамма-функция.

Наконец, если требуется использовать случайный параметр ξ, распределенный *на всей числовой прямой* $(-\infty, +\infty)$, то чаще всего используется (с точностью до линейного преобразования) случайная величина $\xi = \xi^{(0,1)}$, имеющая плотность *стандартного нормального (гауссовского) распределения*

$$f_\xi^{(0,1)}(u) = \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}, \quad -\infty < u < +\infty. \tag{3}$$

Широкое применение распределения (3) вполне оправдано при использовании *усредненных* величин (средних ошибок, средних затрат ресурсов и т. п.) по соображениям, связанным с применением центральной предельной теоремы.

Необходимость применения распределений (1) и (2) для соответствующих интервалов распределения вводимого параметра ξ не столь очевидна. Считается, что бета-распределение (1) неплохо описывает процессы, обладающих естественными нижними и верхними пределами, а также различные доли (доля минералов, находящихся в горной породе; доля радиации, поглощенной материалом; продолжительность работ по проекту и т. п.). В свою очередь, гамма-распределение (2) используется в теории надежности (описывает срок службы изделия, время достижения изделием предельного состояния при коррозии, время наработки до k-го отказа и т. п.), в медицинских статистических исследованиях (описывает продолжительность жизни больного при хронических заболеваниях, время достижения лечебного эффекта от лекарств и т. п.), в экономических математических конструкциях (описание спроса и т. п.) и в других областях.

Случайные величины с распределениями (1) и (2) обладают рядом важных свойств. В частности, гамма-распределение (2) обладает *свойством безграничной делимости*, основанном на том, что если случайные величины $\gamma^{(\lambda,\nu_1)}$ и $\gamma^{(\lambda,\nu_2)}$ имеют гамма-распределение (2) и независимы, то величина $\gamma^{(\lambda,\nu_1)} + \gamma^{(\lambda,\nu_2)} = \gamma^{(\lambda,\nu_1+\nu_2)}$ тоже имеет гамма-распределение (2) с параметром $\nu = \nu_1 + \nu_2$; равенство означает здесь совпадение распределений соответствующих случайных величин. Кроме того, для связанных с распределениями (1) и (2) бета- и гамма-функций имеется множество полезных формул, которые позволяют получить аналитические соотношения (рекуррентные, предельные) для математических конструкций, в которых применяются распределения (1) и (2). Весьма распространенным среди прикладников является тезис (весьма спорный) о том, что формулы (1) и (2) дают возможность (с помощью варьирования параметров μ, ν и λ) получать весьма широкий спектр форм графиков плотностей на соответствующих интервалах.

Однако, с точки зрения численного моделирования, все приведенные выше полезные свойства распределений (1) и (2) нивелируются следующим обстоятельством.

ЗАМЕЧАНИЕ. *Практически все алгоритмы численного получения выборочных значений* $\beta_0^{(\mu,\nu)}$ *и* $\gamma_0^{(\lambda,\nu)}$ *случайных величин* $\beta^{(\mu,\nu)}$ *и* $\gamma^{(\lambda,\nu)}$*, имеющих бета- и гамма-распределение, соответственно, являются неэффективными (неэкономичными), поэтому по возможности следует избегать использования распределений* (1) *и* (2) *в расчетах по методу Монте-Карло.*

Обоснованию этого замечания посвящена первая часть доклада. Во второй части доклада в качестве альтернативы применения распределений (1) и (2) предложено либо выбирать $f_\xi(u)$ из некоторого «банка» элементарных (эффективно моделируемых) вероятностных плотностей, либо использовать кусочно-постоянные версии плотности $f_\xi(u)$ (гистограммы или приближения сложно моделируемых плотностей).

Список литературы

[1] *Михайлов Г. А., Войтишек А. В.* Численное статистическое моделирование. Методы Монте-Карло. М.: Издательство «Юрайт», 2018.

Исследование зон модальности распределения Неймана типов A, B и C.

К. А. Власова, В. В. Корников
Санкт-Петербургский ГУ, Санкт-Петербург, Россия
E-mail: kseniia.vlasova.910@gmail.com

Распределение Неймана относится к классу так называемых контагиозных распределений. Этот термин введён Нейманом в 1939 году [6] в статье, описывающей процессы кластеризации или распространения инфекций.

Изначально распределение Неймана описывало биологическую модель распределения личинок [5]. В настоящее время распределение Неймана можно применять и в других областях, в том числе для решения логистических задач. Для этого нужно быть уверенным в унимодальности распределения.

Само распределение Неймана (типа A) задаётся формулами:

$$x = \sum_{i=1}^{N} \xi_i, \quad \xi_i \in Poisson(\phi), \quad N \in Poisson(\lambda),$$

$$G_A(z) = \exp(\lambda(e^{\phi(z-1)} - 1)).$$

Также существуют распределения Неймана типов B и C, задаваемые производящими функциями вероятностей:

$$G_B(z) = \exp \frac{\lambda(e^{\phi(z-1)} - 1)}{\phi(z-1) - 1}, \quad G_C(z) = \exp \frac{\lambda(2e^{\phi(z-1)} - \phi(z-1) - 1)}{(\phi(z-1))^2 - 1}.$$

Целью данной работы является проведение исследования модальности распределения Неймана типа A, сравнение полученных результатов с результатами предыдущих исследований модальности распределения Неймана типа A [2,3] и проведение аналогичного исследования для распределений Неймана типов B и C.

В статье «Neyman Type A Distribution Revisited» [3] было проведено аналитическое исследование модальности Неймана типа A и получены три зоны модальности.

В статье «Анализ обобщённого контагиозного распределения Неймана на многовершинность» [2] было проведено альтернативное исследование модальности Неймана типа A и получены более широкие множества (ϕ, λ), в которых распределение унимодально.

Ввиду сложности распределений Неймана типов B и C нахождение зон модальности аналитическим способом является затруднительным. Поэтому была разработана программа в среде matlab, подбирающая значения параметров распределения ϕ, λ, при которых оно унимодально. Для вычислений были использованы следующие рекуррентные формулы:
Распределение Неймана типа A:

$$P_0 = \exp(\lambda(e^{-\phi} - 1)), \quad P_x = (\lambda\phi/x)e^{-\phi}\sum_{j=0}^{x-1}(\lambda^j/j!)P_{x-j-1}.$$

Распределение Неймана типа B:

$$P_0 = \exp(\lambda/\phi(1 - \phi - e^{\phi})),\ P_x = \frac{\lambda}{x\phi}\sum_{j=0}^{x-1}(j+1)(1 - e^{-\phi}\sum_{i=0}^{j+1}\frac{\phi^i}{i!})P_{x-j-1}.$$

Распределение Неймана типа C:

$$P_0 = \exp(2\frac{\lambda}{\phi^2}(e^{-\phi} - 1 - \frac{\phi^2}{2} + \phi)),$$

$$P_x = \frac{2\lambda e^{\phi}}{x\phi^2}\sum_{j=0}^{x-1}(j+1)(\phi(e^{\phi} - \sum_{i=0}^{j}\frac{\phi^i}{i!}) - (j+2)(e^{\phi} - \sum_{i=0}^{j+1}\frac{\phi^i}{i!}))P_{x-j-1}.$$

По результатам работы программы были получены множества пар ϕ, λ, в которых сменяется модальность соответствующего распределения, и найдены три зоны модальности для распределения Неймана типа A на интервале $\lambda \in [0, 20]$ с шагом 0.1 и $\phi \in [0, 50]$ с шагом 0.1: в двух зонах распределение унимодально, в одной — мультимодально. Они представлены ниже на Рис. 1.

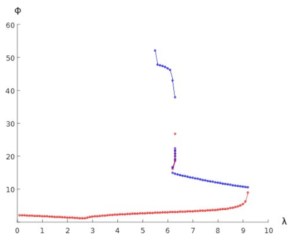

Рис. 1: распр. Неймана типа A.

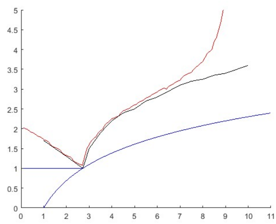

Рис. 2: Сравнение.

Область под красной линией соответствует значениям (ϕ, λ), при которых распределение Неймана типа A унимодально. Область между красной и синей линиями соответствует значениям, при которых распределение мультимодально. При значениях (ϕ, λ), расположенных выше синей линии, распределение снова унимодально. Стоит обратить внимание на поведение распределений Неймана типа A при $\lambda = 6.2$. Их унимодальность сильно зависит от значения ϕ.

Результаты, полученные в статье «Анализ обобщённого контагиозного распределения Неймана на многовершинность» [2], подтвердились. Они также согласуются с выводами статьи «Neyman Type A Distribution Revisited» [3]: полученное множество (ϕ, λ), при которых распределение Неймана типа A унимодально, шире, чем множество (ϕ, λ), полученное аналитически в статье (Рис. 2).

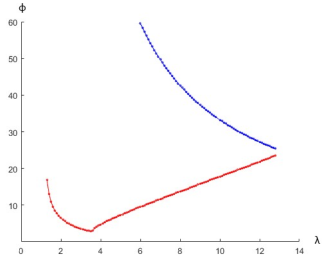

Рис. 3: распр. Неймана типа C.

Метод был применён для определения зон модальности распеределений Неймана типов B и C. Получено множество пар (ϕ, λ), в которых сменяется модальность распределений. В случае многомодальности графики обоих распределений имеют характерную форму, и при наложении дополнительного ограничения на нижнюю границу возможных значений параметра λ, распределения можно считать унимодальными. График зон модальности для распределения Неймана типа C представлен выше на Рис. 3.

Список литературы

[1] *Tripathi Ram C.* Neyman's type A, B, C Distributions. 2004.

[2] *Ордин Л. И.*Анализ обобщённого контагиозного распределения Неймана на многовершинность. 2017.

[3] *Claude Masse J., Theodorescu R.* Neyman Type A distribution revisited. 2004.

[4] *Beall G.* The fit and significance of contagious distributions when applied to observations on larval insects. 1940.

[5] *Gurland J.* A genetalized class of contagious distributions. 2014.

[6] *Neyman J.* On a new class of contagious distributions, applicable in entomology and bacteriology. 1939.

Двух-шаговая игра ценообразования на графе

В. В. Гусев

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: gusev@krc.karelia.ru

Пусть $G =< V, E >$ неориентированный граф, V — множество вершин, E — множество ребер. На графе G рассмотрим игру ценообразования. Пусть N множество игроков. Стратегией игрока является выбор подмножества вершин V и установление цен в выбранных вершинах.

Введем обозначения:

n^v — число потребителей в вершине $v, v \in V$;

c_i^v — затраты игрока $i, i \in N$ на реализацию единицы товара в вершине $v, v \in V, c_i^v > 0$;

p_i^v — цена, по которой игрок i продает единицу товара в вершине $v, p_i^v \geq c_i^v$;

a_i^v — коэффициент качества товара игрока i в вершине $v, a_i^v > 0$

$n^v \cdot e^{-a_i^v p_i^v}$ — доля потребителей в вершине v, которые готовы приобрести товар игрока i, при условии, что игрок i единственный на рынке.

Пусть v некоторая фиксированная вершина графа G. K — множество игроков, которые выбрали вершину v. Перенумеруем игроков в вершине v как 1, 2, ..., $|K|$. Рассмотрим игру

$$\Gamma^v = \langle K, Q^v, \left\{h_i^v(p^v(K))\right\}_{i \in K} \rangle,$$

где $Q^v = \prod_{i \in K} Q_i^v$ декартово произведение множеств $Q_i^v = [c_i^v; +\infty)$. Функции выигрышей определяются по формуле 2.

В игре Γ^v нас будут интересовать равновесные цены.

$$h_i^v(p^v(K)) = (p_i^v - c_i^v) e^{-a_i^v p_i^v} f_i(p^v(K \setminus \{i\})),$$

Теорема 1. *Игра Γ^v является порядковой потенциальной игрой с потенциальной функцией*

$$P(p^v(K)) = \prod_{k \in K} (p_k^v - c_k^v) e^{-a_k^v p_k^v}$$

и точкой равновесия $\bar{p}^v(K) = \left(\bar{p}_1^v, ..., \bar{p}_{|K|}^v\right)$, *где*

$$\bar{p}_i^v = c_i^v + \frac{1}{a_i^v} \; \forall i \in K.$$

Пусть $N = \{1, 2, 3\}$, вершину v выбрали только игроки 1 и 2. Функции выигрышей игроков и ordinal potential в игре Γ^v имеют вид:

$$h_1^v(p_1^v, p_2^v) = n^v(p_1^v - c_1^v)e^{-a_1^v p_1^v}\left(\frac{1}{2}e^{-a_2^v p_2^v} + (1 - e^{-a_2^v p_2^v})\right);$$

$$h_2^v(p_1^v, p_2^v) = n^v(p_2^v - c_2^v)e^{-a_2^v p_2^v}\left(\frac{1}{2}e^{-a_1^v p_1^v} + (1 - e^{-a_1^v p_1^v})\right);$$

$$P(p_1^v, p_2^v) = (p_1^v - c_1^v)(p_2^v - c_2^v)e^{-a_1 p_1^v - a_2 p_2^v}.$$

Пусть $b^v = 5, a_1^v = 0.1, a_2^v = 0.2$. Из (3) следует что $c_1^v = 50, c_2^v = 25$. График функции $P(p_1^v, p_2^v)$ и точка равновесия $\bar{p}^v(K) = \bar{p}^v(\{1, 2\}) = (\bar{p}_1^v, \bar{p}_2^v) = (60, 30)$ изображены на рисунке 1.

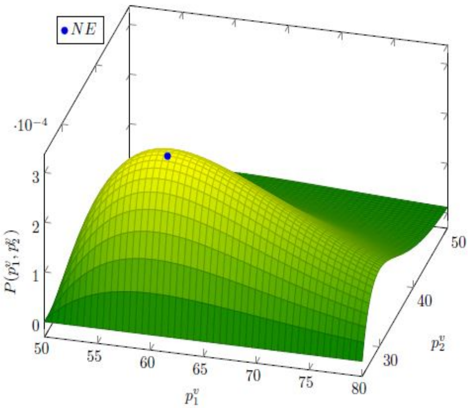

Рис. 1. График
$P(p_1^v, p_2^v) = (p_1^v - 50)(p_2^v - 25)e^{-0.1p_1^v - 0.2p_2^v}, p_1^v \in [50; 80], p_2^v \in [25; 50]$.

Вероятностный метод оценки гафниана на примере одного класса тёплицевых матриц

Д. Б. Ефимов

Физико-математический институт Коми НЦ УрО РАН, Сыктывкар, Россия

E-mail: dmefim@mail.ru

Пусть $A = (a_{ij})$ — симметричная матрица четного порядка n над коммутативным ассоциативным кольцом. Ее гафниан определяется как

$$\mathrm{Hf}(A) = \sum_{(i_1 i_2|\ldots|i_{n-1}i_n)} a_{i_1 i_2} \ldots a_{i_{n-1} i_n},$$

где суммирование ведется по всем разбиениям множества $\{1, 2, \ldots, n\}$ на непересекающиеся пары $(i_1 i_2), \ldots, (i_{n-1} i_n)$ с точностью до порядка пар и порядка элементов в каждой паре. Так, если $n = 4$, то $\mathrm{Hf}(A) = a_{12}a_{34} + a_{13}a_{24} + a_{14}a_{23}$. Понятие гафниана ввел итальянский физик Э.Р. Каяньелло как удобный математический аппарат для описания некоторых величин в квантовой теории поля [1]. Гафниан обладает также интересным комбинаторным свойством, связанным с решением важной задачи в теории графов. А именно, гафниан матрицы смежности неориентированного графа четного порядка равен общему числу совершенных паросочетаний данного графа.

С вычислительной точки зрения нахождение гафниана является труднорешаемой проблемой. Так, наиболее быстрый на данный момент алгоритм для точного вычисления гафниана $n \times n$ комплексной матрицы работает за время $O(n^3 2^{n/2})$ [2]. Для преодоления проблемы большой временной сложности вычисления гафниана рассматриваются различные подходы: нахождение быстрых полиномиальных алгоритмов для более узких классов матриц, нахождение быстровычислимых детерминированных приближенных оценок, нахождение быстровычислимых вероятностных оценок.

В работе [3] А. Барвинок предложил следующую вероятностную оценку гафниана произвольной симметричной матрицы $A = (a_{ij})$ порядка n с неотрицательными элементами. Пусть дано множество $\{u_{ij}, 1 \leq i < j \leq n\}$ независимых вещественных случайных величин со стандартным нормальным распределением. Рассмотрим кососимметрическую матрицу $B = (b_{ij})$ следующего вида:

$$b_{ij} = \begin{cases} u_{ij}\sqrt{a_{ij}} & \text{, если } 1 \leq i < j \leq n, \\ -u_{ij}\sqrt{a_{ij}} & \text{, если } 1 \leq j < i \leq n, \\ 0 & \text{, если } i = j. \end{cases}$$

Тогда $E(\det B) = \mathrm{Hf}(A)$. Так как определитель матрицы размера $n \times n$ можно вычислить за время $O(n^3)$, то $\det B$ является быстровычислимой несмещенной оценкой гафниана.

Пусть a, b — элементы произвольного ассоциативного коммутативного кольца с единицей. Рассмотрим двупараметрическое семейство симметричных тёплицевых матриц $T_{a,b}$ порядка $2m$ с нулевой диагональю, у которых элементы с индексами вида $(i, i+2)$ и $(i, i-2)$ равны a, а все остальные элементы равны b. В данной работе мы получаем следующую аналитическую формулу для точного вычисления гафниана данных матриц:

$$\mathrm{Hf}(T_{a,b}) = \sum_{k=0}^{m} (a-b)^{m-k} b^k \frac{(2k)!}{k!2^k} \left(\sum_{i=\max\left(0, \left\lceil \frac{m-3k}{2} \right\rceil\right)}^{\left\lfloor \frac{m-k}{2} \right\rfloor} C_{2k+i}^{m-k-i} \cdot C_{m-k-i}^{i} \right).$$

Полученный точный результат мы сравниваем с упомянутой выше вероятностной оценкой.

Список литературы

[1] *Caianiello E. R.* On quantum field theory - I: Explicit solution of Dyson's equation in electrodynamics without use of Feynman graphs // IL Nuovo Cimento. 1953. **10**(12). P. 1634–1652.

[2] *Björklund A., Gupt B., Quesada N.* A faster hafnian formula for complex matrices and its benchmarking on the Titan supercomputer // arXiv:1805.12498v2 [cs.DS] 25 Sep2018.

[3] *Barvinok A.* Polynomial Time Algorithms to Approximate Permanents and Mixed Discriminants Within a Simply Exponential Factor // Random Structures Algorithms. 1999. **14**. P. 29–61.

Предельные теоремы, связанные с алгоритмами выбора случайных элементов групп

А. М. Зубков

Математический институт им. В. А. Стеклова РАН, Москва, Россия

E-mail: zubkov@mi.ras.ru

Исследования «типичных» свойств элементов конечных групп G (например, их порядков) можно проводить, анализируя свойства случайно выбираемых элементов этих групп. В общем случае разработка алгоритмов построения независимых равновероятных элементов групп с большим числом элементов и сложной структурой является нетривиальной задачей.

В докладе приводится обзор небольшой части результатов, связанных с исследованием предложенного в [2] итерационного алгоритма мультипликативных замен, позволяющего строить случайные элементы любой наперед заданной конечной группы, имеющих распределение, сколь угодно близкое к равномерному. Обзор результатов, связанных с алгоритмом мультипликативных замен, содержится в статье [3]; в ней список литературы содержит более 90 работ, из которых более 10 непосредственно связаны с этим алгоритмом.

Алгоритм мультипликативных замен состоит в следующем. Пусть G — конечная группа, $\mathcal{N}_k(G)$ — множество всех *порождающих k-элементных наборов*, т. е. таких k-элементных наборов $(g) = (g_1, \ldots, g_k)$ элементов группы G, что $\langle g_1, \ldots, g_k \rangle = G$. Чтобы сделать одношаговое преобразование данного порождающего набора $(g) = (g_1, \ldots, g_k)$, нужно случайно равновероятно выбрать пару (i, j), $1 \leqslant i \neq j \leqslant k$, и случайно, равновероятно, независимо от пары (i, j) применить одну из 4 операций

$$R^{\pm}_{i,j} : (g_1, \ldots, g_i, \ldots, g_k) \to (g_1, \ldots, g_i \cdot g_j^{\pm 1}, \ldots, g_k),$$
$$L^{\pm}_{i,j} : (g_1, \ldots, g_i, \ldots, g_k) \to (g_1, \ldots, g_j^{\pm 1} \cdot g_i, \ldots, g_k).$$

Эти операции переводят порождающий k-элементный набор в порождающий k-элементный набор. После t таких независимых случайных переходов можно выбрать случайный элемент из этого набора в качестве случайного элемента группы G.

В [2] предполагалось, что групповые операции умножения, обращения и сравнения с единичным элементом e имеют единичную сложность (или реализуются обращением к оракулу) и что в начальном значении набора $(g_1, \ldots, g_k)$ первые l элементов $g_1, \ldots, g_l$ порождают группу G, а $g_{l+1} = \ldots = g_k = e$. В [2] было доказано, что алгоритм мультипликативных замен работает, если $k > 2\log_2 |G|$, а число итераций t достаточно велико; однако практически он работает и при меньших значениях k.

Алгоритм мультипликативных замен порождает случайные элементы группы, распределение которых близко к равномерному, но может и отличаться от равномерного.

Пусть G — конечная группа и Q_k, Q_k^* — распределения первого и случайно выбранного элементов порождающего k-элементного набора $g \in \mathbb{N}_k(G)$. Из симметрии следует, что $Q_k = Q_k^*$. В [1] показано, что если σ — первый элемент равновероятного случайного k-элементного набора, порождающего группу A_n, то

$$\mathbf{P}\{\sigma(1) = 1\} = \frac{1}{n} - \frac{1}{n^k} + O\left(\frac{1}{n^{2k-1}}\right).$$

В [1] приведены примеры, в которых неравновероятность элементов равновероятных порождающих k-наборов можно обнаруживать статистически.

Рассмотрим простой случай, когда конечная группа G абелева. Тогда элементы возникающих при итерациях алгоритма мультипликативных замен наборов порождающих элементов $(g)^{(t)} = (g_1^{(t)}, \ldots, g_k^{(t)})$ можно представить в виде произведений степеней элементов исходного порождающего набора $(g)^{(0)} = (g_1^{(0)}, \ldots, g_k^{(0)}) = (g_1, \ldots, g_k)$. Если $k = 2$ и на каждом шаге нзависимо от предыстории с равными вероятностями используются только операции двух видов: $(h_1, h_2) \to (h_1, h_1 h_2)$ и $(h_1, h_2) \to (h_1 h_2, h_2)$, то $(g)^{(t)} = (g_1^{S_1(t)}, g_2^{S_2(t)})$, где

$$\begin{aligned}\mathbf{P}\{(S_1(t+1), S_2(t+1)) = (S_1(t), S_1(t) + S_2(t))\} = \\ = \mathbf{P}\{(S_1(t+1), S_2(t+1)) = (S_1(t) + S_2(t), S_2(t))\} = \frac{1}{2}.\end{aligned}$$

Таким образом, последовательность $(S_1(t), S_2(t))$ образует двумерную цепь Маркова с начальным состоянием $(1, 1)$.

Эквивалентной интерпретацией этой цепи является следующая урновая схема. Пусть в урне находятся шары белого и черного цветов. За 1 шаг с вероятностями, равными $\frac{1}{2}$, либо число белых шаров увеличивается на число черных шаров, либо число черных шаров увеличивается на число белых шаров. В [4] получены формулы для первых двух моментов общего числа шаров в урне и доказано, что предел функции распределения доли числа белых шаров в урне совпадает с теоретико-числовой монотонно возрастающей функцией Минковского, соответствующей сингулярному распределению на [0, 1].

Аналогичные результаты получены для обобщения этой урновой схемы на случай урны с шарами n цветов, когда на каждом шаге число шаров одного случайно выбранного цвета увеличивается на число шаров другого случайно выбранного цвета. При слабых предположениях о распределении выбираемых цветов показано, что вектор долей чисел шаров разных цветов имеет невырожденное предельное распределение.

Список литературы

[1] *Babai L., Pak I.* Strong bias of group generators: an obstacle to the "product replacement algorithm-// J. Algorithms. 2004. **50**(2). P. 215–231.

[2] *Celler F., Leedham-Green C. R., Murray S., Niemeyer A., O'Brien E. A.* Generating random elements of a finite group // Comm. Algebra. 1995. **23**. P. 4931–4948.

[3] *Pak I.* What do we know about the product replacement algorithm? // Groups and Computation. 2001. **3**. P. 301–347.

[4] *Зубков А. М., Колесникова К. А.* Цепь Маркова с теоретико-числовым предельным распределением // Дискретная математика. 2015. **27**(3). С. 17–24.

[5] *Зубков А. М., Колесникова К. А.* Предельная теорема для процесса аддитивных замен // Теория вероятностей и ее применения. 2017. **62**(2). С. 393–404.

Мощность образа подмножества при композиции случайных отображений конечного множества

А. М. Зубков, А. А. Серов

Математический институт им. В.А. Стеклова РАН, Москва, Россия

E-mail: zubkov@mi.ras.ru, serov@mi.ras.ru

Пусть $F_1, F_2, \ldots$ — независимые случайные равновероятные отображения множества $\mathfrak{X}_N = \{X_1, \ldots, X_N\}$ в себя , $S_0 \subseteq \mathfrak{X}_N$ — начальное подмножество $\mathfrak{X}_N$, $|S_0| = m$, тогда $F_t\left(\ldots F_2\left(F_1\left(S_0\right)\right)\ldots\right)$ — образ подмножества S_0 при композиции t отображений. Рассматриваются образы S_0 при последовательном действии отображений $F_1, \ldots, F_t$:

$$S_1 = F_1(S_0),\ S_2 = F_2(F_1(S_0)), \ldots, S_t = F_t(F_{t-1}(\ldots(F_1(S_0))\ldots)).$$

Очевидно, последовательности $\{S_j\}_{j=0}^t$ и $\{|S_j|\}_{j=0}^t$ являются цепями Маркова и траектории цепи $\{|S_j|\}_{j=0}^t$ не возрастают.

Оценкам размеров образов $\{S_k\}_{k=0}^t$ посвящены многочисленные работы (см., например, [1, 2, 4–7]). Публикации, посвящённые исследованию распределений размеров образов $\{S_k\}_{k=0}^t$, авторам неизвестны.

В теореме 1 указаны условия на m, t и N, при которых распределение размеров образов S_t асимптотически нормально.

Теорема 1. *Если m, n, $N \to \infty$ так, что m имеет порядок $N^{1/4}$ и $n = o(m)$, то для любого фиксированного $x \in \mathbb{R}$ при целых t вида*

$$t = t(N, m, n) = 2N\left(\frac{1}{n} - \frac{1}{m}\right) + x\,\frac{2N}{\sqrt{3}}\sqrt{\left(\frac{1}{n^3} - \frac{1}{m^3}\right)}$$

справедливо соотношение

$$\mathbf{P}\left\{|S_t| \leq \frac{N}{N/m + t/2}\right\} \to \Phi(x)\,,$$

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

Этот результат соответствует полученному ранее в [3] с помощью эвристических рассуждений приближению $\frac{|S_t|}{N} \approx \frac{1}{N/|S_0| + t/2}$.

Пусть при $m \geq n$, $m, n \in \mathbb{N}$,

$$p_{m,n}(t, N) = \mathbf{P}\left\{|F_{t\ldots 1}(X_1, X_2, \ldots, X_m)| = n\right\} = \mathbf{P}\left\{|S_t| = n \big| |S_0| = m\right\}, \quad (1)$$

где $F_{t\ldots 1}(\cdot) = F_t(\ldots(F_1(\cdot))\ldots)$ и $X_1, X_2, \ldots, X_m \in \mathcal{X}_N$ попарно различны.

Из [8, Утверждение 1] следует, что

$$p_{n,n}(t,N) = \left(\prod_{i=1}^{n-1}\left(1-\frac{i}{N}\right)\right)^t, \tag{2}$$

$$p_{n+1,n}(t,N) = \frac{n+1}{2}\left(1-\left(1-\frac{n}{N}\right)^t\right)\left(\prod_{i=1}^{n-1}\left(1-\frac{i}{N}\right)\right)^t, \tag{3}$$

$$p_{m,n}(1,N) = \sum_{2\leq j_1<j_2<\ldots<j_{m-n}\leq m}\prod_{i=1}^{n-1}\left(1-\frac{i}{N}\right)\prod_{v=1}^{m-n}\frac{j_v-v}{N}.$$

В Утверждении 1 представлен способ точного рекуррентного вычисления распределения $|S_t|$.

Утверждение 1. *При $m > n$, $m,n > 1$, $t \in \mathbb{N}$ справедливо рекуррентное тождество*

$$p_{m,n}(t,N) = \sum_{i=1}^{t}\left(\frac{N^{[m]}}{N^m}\right)^{i-1}\sum_{j=n}^{m-1}p_{j,n}(t-i,N)\left(\tfrac{j}{N}\right)^m\tbinom{N}{j}\mathbf{P}\left\{\mu_0(m,j)=0\right\},$$

где $p_{n,n}(0,N) = 1$, $p_{j,n}(0) = 0$, $j = n+1, n+2, \ldots, m$, $p_{n,n}(t,N)$ определены в (2), $p_{n+1,n}(t,N)$ — в (3).

Получены двусторонние неравенства для $\mathbf{M}\{|S_t| \mid |S_0| = m\}$, в которых разность между верхней и нижней оценками имеет порядок $o(m)$, если $m,t,N\to\infty$, $mt = o(N)$.

Теорема 2. *Для любого элемента $x \in \mathcal{X}_N$, не зависящего от отображений F_1, $F_2, \ldots$, при любых $1 \leqslant m \leqslant N$, $t \geq 1$ справедливы неравенства*

$$\frac{1}{N}\sum_{k=1}^{4}(-1)^{k-1}\binom{m}{k}p_k(t,N) \leqslant \mathbf{P}\{x\in S_t \mid |S_0| = m\}$$
$$\leqslant \frac{1}{N}\sum_{k=1}^{5}(-1)^{k-1}\binom{m}{k}p_k(t,N),$$

где $p_1(t,N) = 1$, $p_2(t,N) = 1-\left(\frac{N^{[2]}}{N^2}\right)^t$, $p_3(t,N) = 1-\frac{3}{2}\left(\frac{N^{[2]}}{N^2}\right)^t + \frac{1}{2}\left(\frac{N^{[3]}}{N^3}\right)^t$, $p_4(t,N) = 1-\frac{9N-11}{5N-6}\left(\frac{N^{[2]}}{N^2}\right)^t+\left(\frac{N^{[3]}}{N^3}\right)^t-\frac{N-1}{5N-6}\left(\frac{N^{[4]}}{N^4}\right)^t$, $p_5(t,N) = 1-\frac{5(4N-5)}{2(5N-6)}\left(\frac{N^{[2]}}{N^2}\right)^t+\frac{5(4N-7)}{2(7N-12)}\left(\frac{N^{[3]}}{N^3}\right)^t-\frac{5(N-1)}{2(5N-6)}\left(\frac{N^{[4]}}{N^4}\right)^t+\frac{N-1}{2(7N-12)}\left(\frac{N^{[5]}}{N^5}\right)^t$.

Справедливы также оценки

$$\sum_{k=1}^{4}(-1)^{k-1}\binom{m}{k}p_k(t,N) \leqslant \mathbf{M}\{|S_t| \mid |S_0|=m\}$$

$$< \sum_{k=1}^{5}(-1)^{k-1}\binom{m}{k}p_k(t,N).$$

Результаты представляют интерес для анализа алгоритмов балансировки времени и памяти.

Список литературы

[1] *Flajolet P., Odlyzko A. M.* Random mapping statistics // Eurocrypt'89 Lect. Notes Comput. Sci. 1990. **434** P. 329–354.

[2] *Oechslin P.* Making a faster cryptanalytic time-memory trade-off // Lect. Notes Comput. Sci. 2003. **2729**. P. 617–630.

[3] *Hong J.* The cost of false alarms in Hellman and rainbow tradeoffs // Designs, Codes and Cryptography. 2010. **57**(3). P. 293–327.

[4] *Hong J., Moon S.* A comparison of cryptanalytic tradeoff algorithms // J. Cryptology. 2013.**26**. P. 559–637.

[5] *Pilshchikov D. V.* Estimation of the characteristics of time-memory-data tradeoff methods via generating functions of the number of particles and the total number of particles in the Galton-Watson process // Математические вопросы криптографии. 2014. **5**(2). P. 103–108.

[6] *Зубков А. М., Серов А. А.* Совокупность образов подмножества конечного множества при итерациях случайных отображений // Дискретная математика. 2014. **26**(4). С. 43–50. англ. пер.: *Zubkov A. M., Serov A. A.* Images of subset of finite set under iterations of random mappings // Discrete Math. Appl. 2015. **25**(3). P. 179–185.

[7] *Серов А. А.* Образы конечного множества при итерациях двух случайных зависимых отображений // Дискретная математика. 2015. **27**(4). С. 133–140. англ. пер.: *Serov A. A.* Images of a finite set under iterations of two random dependent mappings // Discrete Math. Appl. 2016. **26**(3). P. 175–181.

[8] *Зубков А. М., Серов А. А.* Предельная теорема для мощности образа подмножества при композиции случайных отображений // Дискретная математика. 2017. **29**(1). С. 17–26. англ. пер.: *Zubkov A. M., Serov A. A.* Limit theorem for the size of an image of subset under compositions of random mappings // Discrete Math. Appl. 2018. **28**(2). P. 131–138.

О приближении вероятностных мер мерами с конечными носителями

А. В. Иванов

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: alvlivanov@krc.karelia.ru

Пусть $P(X)$ – пространство вероятностных мер на компактном метрическом пространстве (X, ρ), наделенное метрикой Канторовича-Рубинштейна ρ_P, и $P_n(X)$ – подпространство $P(X)$, состоящее из мер, носитель которых имеет не более n точек. Пусть $\varepsilon > 0$ и $\mu \in P(X)$. Положим $N(\mu, \varepsilon) = \min\{n : \rho_P(\mu, P_n(X)) \leq \varepsilon\}$. Число $N(\mu, \varepsilon)$ есть наименьшая мощность носителя вероятностной меры, дающей ε-приближение меры μ. Если носитель μ бесконечен, то $N(\mu, \varepsilon)$ неограниченно возрастает при $\varepsilon \to 0$. Скорость возрастания $N(\mu, \varepsilon)$ характеризуют верхний $\overline{ord}(\mu)$ и нижний $\underline{ord}(\mu)$ *порядки метрической аппроксимации* меры μ:

$$\overline{ord}(\mu) = \overline{\lim}_{\varepsilon \to 0} \frac{\ln(N(\mu, \varepsilon))}{\ln 1/\varepsilon}, \; \underline{ord}(\mu) = \underline{\lim}_{\varepsilon \to 0} \frac{\ln(N(\mu, \varepsilon))}{\ln 1/\varepsilon}. \tag{1}$$

Величины $\overline{ord}(\mu)$ и $\underline{ord}(\mu)$ определены в рамках общей функториальной схемы (см. [1]), по которой могут быть введены, в частности, емкостные размерности подмножеств компакта X, которые находят приложение в теории динамических систем.

Экспонентой $\exp(X)$ метрического компакта (X, ρ) называется пространство непустых замкнутых подмножеств X с метрикой Хаусдорфа ρ_H. Для каждого $n \in N$ определено подпространство $\exp_n(X) = \{F \in \exp(X) : |F| \leq n\} \subset \exp(X)$. Для $F \in \exp(X)$ и $\varepsilon > 0$ $N(F, \varepsilon) = \min\{n : \rho_H(F, \exp_n(X)) \leq \varepsilon\}$. Верхняя и нижняя емкостные размерности множества F могут быть определены по формулам, аналогичным (1):

$$\overline{\dim}_B F = \overline{\lim}_{\varepsilon \to 0} \frac{\ln(N(F, \varepsilon))}{\ln 1/\varepsilon}, \; \underline{\dim}_B F = \underline{\lim}_{\varepsilon \to 0} \frac{\ln(N(F, \varepsilon))}{\ln 1/\varepsilon}.$$

Справедливы следующие утверждения:

1. Для любой меры $\mu \in P(X)$ выполнены неравенства:

$$\overline{ord}(\mu) \leq \overline{\dim}_B(supp(\mu)), \ \underline{ord}(\mu) \leq \underline{\dim}_B(supp(\mu)).$$

При этом порядки метрической аппроксимации вероятностной меры могут существенно отличаться от емкостных размерностей ее носителя:

2. Для любого бесконечного метрического компакта (X, ρ) существует мера $\mu \in P(X)$ такая, что $supp(\mu) = X$ и $\overline{ord}(\mu) = \underline{ord}(\mu) = 0$.

В теории динамических систем известны понятия верхней и нижней емкостных размерностей меры $\mu \in P(X)$. По определению,

$$\underline{\dim}_B\mu = \lim_{\varepsilon \to 0} \inf\{\underline{\dim}_B F : \mu(F) > 1 - \varepsilon\},$$

$$\overline{\dim}_B\mu = \lim_{\varepsilon \to 0} \inf\{\overline{\dim}_B F : \mu(F) > 1 - \varepsilon\}.$$

Вопрос о совпадении порядков метрической аппроксимации меры и ее емкостных размерностей решается отрицательно:

3. Существует вероятностная мера μ на счетном метрическом компакте, для которой $\underline{ord}(\mu) > 0$. При этом μ имеет нулевые емкостные размерности.

Метод приближения меры последовательностью мер с конечными носителями можно использовать для определения понятия равномерно распределенной вероятностной меры на метрическом компакте (X, ρ).

Последовательность $\mu_n = \frac{1}{|X_n|} \sum_{x \in X_n} \delta_x$ вероятностных мер с конечными носителями $X_n = supp(\mu_n) \subset X$ будем называть *допустимой*, если $|X_n| = N(X, 1/n)$ и $\rho_H(X_n, X) \leq 1/n$. Будем говорить, что метрический компакт (X, ρ) является *стабильным*, если любая допустимая последовательность имеет предел. Если компакт стабилен, то все его допустимые последовательности сходятся к одной мере, которую обозначим через μ_X^u. Если $supp(\mu_X^u) = X$, то меру μ_X^u будем называть *равномерно распределенной* на X и говорить, что метрический компакт (X, ρ) допускает равномерно распределенную меру.

Справедливы следующие утверждения:

4. Любой бесконечный метрический компакт (X, ρ) с единственной предельной точкой a стабилен, и $\mu_X^u = \delta_a$.

5. Существует метрический компакт (Y, ρ) с двумя неизолированными точками, не являющийся стабильным.

6. Стандартное канторово совершенное множество Π, лежащее на отрезке $[0, 1]$, допускает равномерно распределенную меру.

7. Существует совместимая с топологией метрика ρ_1 на канторовом совершенном множестве Π такая, что (Π, ρ_1) не является стабильным.

8. Существует совместимая метрика ρ_2 на канторовом совершенном множестве Π такая, что (Π, ρ_2) стабилен, но не допускает равномерно распределенной меры.

9. Если бесконечный компакт (X, ρ) допускает равномерно распределенную меру, то X не имеет изолированных точек.

10. Если X допускает равномерно распределенную меру, то для любого замкнутого подмножества $F \subset X$ с непустой внутренностью имеют место равенства $\underline{\dim}_B F = \underline{\dim}_B X$, $\overline{\dim}_B F = \overline{\dim}_B X$.

11. Ограниченное канонически замкнутое измеримое по Жордану множество X в (R^n, ρ) с метрикой ρ, определяемой по формуле $\rho(x, y) = \max\{|x_i - y_i| : i = 1, \ldots, n\}$, допускает равномерно распределенную меру, причем $\mu_X^u = \frac{1}{\mu(X)}(\mu|_X)$, где μ – мера Лебега в R^n, являющаяся продолжением меры Жордана.

Список литературы

[1] *Ivanov A. V.* On metric order in spaces of the form $\mathcal{F}(X)$ // Topology and its Applications. 2017. **221**. С. 107–113.

Игра наилучшего выбора двух объектов с неполной информацией

А. А. Ивашко

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: aivashko@krc.karelia.ru

В работе исследована следующая теоретико-игровая модель наилучшего выбора с неполной информацией, связанная с телевизионной игрой «Голос». Пусть несколько игроков одновременно просматривают некоторую последовательность объектов. Каждому игроку необходимо набрать команду из двух объектов. Качество каждого объекта определяется двумя случайными характеристиками. Первая характеристика игрокам известна, а вторая скрыта от них. Каждый из игроков стремится выбрать объект в свою команду, суммарное значение качества которого больше, чем у остальных игроков. При этом выбор осуществляется только на основе значения известной характеристики качества.

Теоретико-игровые модели с неполной информацией, в которых каждому игроку необходимо выбрать только один объект, рассматривались в работах [1, 2]. В данной работе найдены оптимальные пороговые стратегии и выигрыши игроков для игры с равноправными участниками и возможностью выбрать двух объектов в команду.

Список литературы

[1] *Konovalchikova E. N., Mazalov V. V.* A Game-Theoretic Model of TV Show "The Voice" // Automation and Remote Control. 2016. **77**(8). P. 1468–1479.

[2] *Mazalov V. V., Ivashko A. A., Konovalchikova E. N.* Optimal Strategies in Best-Choice Game with Incomplete Information — The Voice Show // International Game Theory Review. 2016. **18**(2). art. no. 164001.

Универсальные стратегии в задаче об одноруком бандите и оптимизация обработки больших данных

А. В. Колногоров

Новгородский государственный университет им. Ярослава Мудрого, Великий Новгород, Россия

E-mail: `Alexander.Kolnogorov@novsu.ru`

Рассматривается задача об одноруком бандите, т.е. о двуруком бандите с известным распределением дохода, соответсвующим выбору первого действия. Для бернуллиевских доходов она была рассмотрена в [1, 2]. Формально однорукий бандит есть управляемый случайный процесс ξ_n, $n = 1, 2, \ldots, N$, значения которого, интерпретируемые как доходы, зависят только от текущих выбираемых действий. Известное математическое ожидание дохода за выбор первого действия m_1 без ограничения общности считаем равным нулю (в противном случае можно рассмотреть процесс $\xi_n - m_1$, $n = 1, 2, \ldots, N$). Распределение дохода, соответствующее выбору второго действия, предполагается нормальным (гауссовским) с плотностью распределения $f_D(x|m) = (2\pi D)^{-1/2} \exp(-(x-m)^2/(2D))$, причем математическое ожидание m неизвестно, а дисперсия D – известна (в дальнейшем предположение об известности D можно снять). Однорукий бандит удобно описать параметром $\theta = (0, m)$, допустимое множество значений параметра выбрано следующим $\Theta = \{\theta : |m| \le C\}$, $0 < C < \infty$. Гауссовские распределения доходов возникают при пакетной обработке данных, если для обработки имеются два метода (действия), причем эффективность одного из них априори известна. В качестве дохода можно рассматривать, например, количество успешно обработанных данных. В этом случае все данные в каждом пакете обрабатываются одним и тем же методом, а для управления используются суммарные доходы в пакетах. В силу центральной предельной теоремы эти доходы при широких предположениях имеют приблизительно нормальные распределения независимо от исходных распределений доходов, и это определяет универсальность постановки задачи.

Стратегия управления σ в каждый момент времени n обеспечивает выбор действия в зависимости от всей известной предыстории процесса. Цель управления формулируется в минимаксной постановке. Для этого определяется функция потерь

$$L_N(\sigma,\theta) = N\max(0,m) - \mathbf{E}_{\sigma,\theta}\left(\sum_{n=1}^{N}\xi_n\right),$$

характеризующая потери математического ожидания полного дохода за применение стратегии σ в сравнении с максимально возможным доходом $N\max(0,m)$ в случае известного параметра. Обозначим через $\lambda(m)$ априорную плотность распределения на множестве параметров. По функции потерь определяются минимаксный и байесовский риски

$$R_N^M(\Theta) = \inf_{\{\sigma\}}\sup_{\Theta} L_N(\sigma,\theta), \quad R_N^B(\lambda) = \inf_{\{\sigma\}}\int_{\Theta} L_N(\sigma,\theta)\lambda(\theta)d\theta,$$

соответствующие оптимальные стратегии называются минимаксной и байесовской стратегиями. Преимущество байесовского подхода состоит в том, что для любого априорного распределения он позволяет искать байесовские стратегию и риск методом динамического программирования с помощью решения рекуррентного уравнения. Однако решение зависит как от самого априорного распределения, так и от распределений доходов, т.е. не является универсальным. Минимаксный подход не использует априорное распределение параметра, и это также определяет универсальность получаемых стратегий.

Установлены следующие результаты [3, 4].

В соответствии с основной теоремой теории игр минимаксные стратегия и риск могут быть найдены как байесовские, соответствующие наихудшему априорному распределению. Для их нахождения получено рекуррентное интегро-разностное уравнение. В предельном случае при $N \to \infty$ получено дифференциальное уравнение в частных производных второго порядка. Установлено, что $R_N^M(\Theta) \sim r(DN)^{1/2}$ при больших N, где $r \approx 0{,}37$.

Для большого числа исходных данных минимаксный риск мало зависит от количества пакетов, на которые разбиты эти данные, если количество пакетов достаточно велико. Например, минимаксный риск, соответствующий пакетной обработке данных, разбитых на 50 пакетов, всего лишь на 3% больше предельного значения, соответствующего обработке данных, разбитых на сколь угодно большое число пакетов. Если исходных данных достаточно много, это означает, что предложенные стратегии обеспечивают близкие значения минимаксного риска для широкого класса процессов с одинаковыми математическими ожиданиями и дисперсиями при сравнительно небольшом числе обрабатываемых пакетов.

Полученные стратегия и минимаксный риск мало меняются при существенном изменении дисперсии (например, при изменении дисперсии на 10%). Если исходных данных достаточно много, это означает, что можно сделать оценку дисперсии D для некоторой начальной группы данных, а затем использовать полученную оценку для управления. Таким образом, если обрабатываемых данных много. то предположение о том, что значение D известно, можно опустить.

Для бернуллиевского однорукого бандита получено рекуррентное уравнение для нахождения байесовских стратегии и риска, соответствующих оптимальной обработке данных по-одному и его предельное описание дифференциальным уравнением в частных производных второго порядка. При этом дифференциальные уравнения в гауссовском и бернуллиевском случае получились одинаковыми. Это означает, что при обработке больших данных с бернуллиевскими доходами обработка по-одному не позволяет снизить минимаксный риск, соответствующий пакетной обработке. Можно предположить, что это свойство пакетной обработки сохранится и для других распределений доходов.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект номер 18-29-16223-мк).

Список литературы

[1] *Bradt R. N., Johnson S. M., Karlin S.* On Sequential Designs for Maximizing the Sum of n Observations // Ann. Math. Statist. 1956. **27**. P. 1060–1074.

[2] *Chernoff H., Ray S. N.* A Bayes Sequential Sampling Inspection Plan // Ann. Math. Statist. 1965. **36**. P. 1387–1407.

[3] *Колногоров А. В.* Задача об одноруком бандите для систем с параллельной обработкой данных // Пробл. передачи информ. 2015. **51**(2). С. 99–113.

[4] *Колногоров А. В.* Гауссовский двурукий бандит и оптимизация групповой обработки данных // Пробл. передачи информ. 2018. **54**(1). С. 27–44.

Некоторые аспекты развития обобщенной схемы размещения

А. В. Колчин

Московский автомобильно-дорожный государственный технический университет, Москва, Россия

E-mail: akolchin@madi.ru

Памяти академика Академии криптографии РФ Колчина Валентина Федоровича посвящается

Комбинаторика сыграла важную роль в начале развития теории вероятностей, и эти два раздела математики продолжают развиваться в тесном взаимодействии. В настоящее время теория вероятностей, предлагая новые подходы к решению задач дискретной математики, как бы отдает долги комбинаторике. Среди этих новых подходов отметим хорошо развитые в теории вероятностей методы асимптотического анализа, которые успешно используются при решении сложных комбинаторных задач. Комбинаторные задачи и методы занимают значительное место в исследованиях по теории вероятностей. Среди многочисленных работ в этой области можно выделить несколько направлений: комбинаторные задачи в теории случайных процессов, задачи, связанные со случайными отображениями и случайными графами, задачи размещения частиц по ячейкам.

Для решения широкого круга подобных комбинаторных задач весьма плодотворным оказывается *вероятностный подход*. Если распределение вероятностей задано на множестве рассматриваемых комбинаторных структур, то числовые характеристики этих структур можно рассматривать как случайные величины и анализировать их вероятностными методами. При таком вероятностном подходе мы автоматически ограничиваемся рассмотрением типичных структур, которые составляют основную массу рассматриваемого множества, и исключаем из рассмотрения небольшую долю структур с нестандартными свойствами.

Вероятностный подход впервые был использован в почти современном виде В. Л. Гончаровым, применившим его к изучению множества S_n

всех подстановок степени n и серий в случайных $(0,1)$ -последовательностях. Среди тех, трудами которых развивалась вероятностная комбинаторика в России, были С. Н. Бернштейн, Н. В. Смирнов, В. Е. Степанов, ее успехи тесно связаны с блестящей Российской вероятностной школой, школой А. А. Маркова, П. Л. Чебышёва, А. М. Ляпунова, А. Я. Хинчина, А. Н. Колмогорова, Ю. В. Прохорова.

При анализе случайных структур успешно применялись разнообразные вероятностные методы; в частности, в вероятностной комбинаторике находит успешное применение *обобщенная схема размещения*, позволяющая сводить ряд комбинаторных задач к задачам о суммах независимых случайных величин, классическому объекту изучения в теории вероятностей. Обобщенная схема размещения была введена в 1968 г. В. Ф. Колчиным и заняла заметное место в асимптотических исследованиях в вероятностной комбинаторике. Свое название эта схема получила в связи с тем, что она является обобщением классической задачи о случайном размещении частиц по ячейкам. Она оказалась удобным средством исследования таких интереснейших процессов, как эволюция случайных графов, случайных лесов, систем линейных уравнений со случайными коэффициентами, случайных подстановок, в том числе в связи с построением и анализом вычислительных алгоритмов. В настоящее время активные исследования асимптотического поведения различных комбинаторных объектов с использованием обобщенной схемы размещения ведутся, в частности, Ю. Л. Павловым в Карельском научном центре РАН, А. Н. Чупруновым в Казанском федеральном университете и И. Фазекашем в Дебреценском университете.

Напомним, что в обобщенной схеме размещения частиц распределение заполнений ячеек представимо как условное распределение *независимых* случайных величин при условии, что их сумма принимает фиксированное значение. Пусть $\eta_1, \ldots, \eta_N$ — неотрицательные целочисленные случайные величины, рассматриваемые как некоторые числовые характеристики комбинаторной структуры из N компонент, состоящей из n элементов, такие, что $\eta_1 + \cdots + \eta_N = n$. Если существуют независимые случайные величины $\xi_1, \ldots, \xi_N$ такие, что совместное распределение $\eta_1, \ldots, \eta_N$ допускает представление

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\} \\ = \mathbf{P}\{\xi_1 = k_1, \ldots, \xi_N = k_N \mid \xi_1 + \cdots + \xi_N = n\}, \quad (1)$$

где $k_1, \ldots, k_N$ — произвольные целые числа, то говорят, что $\eta_1, \ldots, \eta_N$ образуют обобщенную схему размещения с параметрами n и N и независимыми случайными величинами $\xi_1, \ldots, \xi_N$. Случайные величины

$\eta_1, \ldots, \eta_N$ интерпретируются как заполнения ячеек.

А. Н. Чупруновым активно исследуются обобщенные схемы размещения *случайного числа* K частиц по N ячейкам. Ее частными случаями являются обобщенная схема размещения с неполным комплектом частиц и другие аналоги обобщенной схемы размещения.

И. Фазекашем исследуются расширения обобщенных схем размещения, где в N ячеек размещаются либо *по крайней мере* n частиц, либо *не более* n частиц.

Изучение многих характеристик обобщенной схемы размещения сводится к задачам о суммах независимых случайных величин. В случае, когда распределения слагаемых одинаковы и фиксированы (не зависят от числа слагаемых), можно пользоваться хорошо развитой теорией суммирования независимых случайных величин. Однако во многих применениях обобщенной схемы возникает необходимость в *локальных предельных теоремах в схеме серий.*

Как правило, распределение $\xi_1, \ldots, \xi_N$ допускает представление

$$\mathsf{P}\{\xi_1 = k\} = \frac{b_k \theta^k}{k!\, B(\theta)}, \tag{2}$$

где $b_0, b_1, b_2, \ldots$ — некоторая последовательность неотрицательных чисел, $B(\theta) = \sum_{k=0}^{\infty} b_k \theta^k / k!$, и θ — параметр, принимающий положительные значения из области сходимости ряда $B(\theta)$. Для изучения характеристик обобщенной схемы размещения требуются локальные предельные теоремы при всех значениях параметра θ. Автором рассмотрены основные случаи возможных областей изменения N и θ в схеме, в которой выполнено либо *условие* A_r, $r \geq 2$: $b_0 > 0$, $b_1 = \cdots = b_{r-1} = 0$, $b_r > 0$, и максимальный шаг распределения (2) равен 1; либо *условие* A_1: $b_0, b_1 > 0$, очевидно, являющейся частным случаем общей схемы.

Автором получен ряд предельных, в том числе *локальных*, теорем для сумм вида $S_N = \sum_{k=1}^{N} \xi_k$ при $N \to \infty$ и различных вариантах поведения параметра $\theta = \theta(N)$. В *простейшем случае* значения θ отделены от 0 и не приближаются к значению радиуса сходимости $B(\theta)$. При $N \to \infty$ и $\theta \to 0$ имеет место *переход распределения* S_N *с одной решетки на другую*: именно, на решетке целых чисел имеет место сходимость к нормальному распределению, в то время как на решетке целых неотрицательных чисел с шагом r имеет место сходимость к распределению Пуассона. В случае, когда значения параметра θ приближаются к границе сходимости ряда $B(\theta)$, как можно показать на примерах, могут появляться и другие предельные распределения.

О применении схем разделения секрета в распределенных системах хранения информации

С. А. Круглик

Сколковский институт науки и технологий, Москва, Россия

E-mail: stanislav.kruglik@skoltech.ru

Значительный рост объема информации, хранимой человечеством, рождает новые вызовы к задаче конструирования распределенных систем облачного хранения информации ввиду их повсеместного распространения вследствие удобства использования. Одним из таких вызовов, полностью нерешенным в данный момент, является задача восстановления информации в случае отказа малого числа узлов. Несмотря на то, что случаи одновременного отказа множества узлов также возможны, отказ малого числа является наиболее частым сценарием и задача построения кодов, ее оптимизирующих, является достаточно актуальной и практически мотивированной.

Известно два класса кодов, ее решающих, в первом случае, когда речь идет о так называемых восстанавливающих кодах, оптимизируется суммарный объем информации, используемый при восстановлении одного узла. Во втором случае мы оптимизируем число серверов, участвующих в данном процессе, и называем такие коды кодами с локальным восстановлением. Несмотря на большой прогресс, достигнутый в построении и практическом применении данных кодов они имеют существенный недостаток связанный с игнорированием вопросов сохранения конфиденциальности пользовательской информации.

В данной работе рассмотрена задача сохранения конфиденциальности информации, хранимой в распределенной системе, при наличии злоумышленника, имеющего доступ к небольшому числу серверов. В частности рассмотрена связь данной задачи с задачей разделения секрета и возможность совмещения свойств локального восстановления с ней.

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 18-07-01427, 18-37-00459, 19-01-00364, 19-37-80006.

Неравенства для ранга случайной двоичной матрицы с независимыми строками заданных весов

В. И. Круглов, В. Г. Михайлов

Математический институт им. В.А. Стеклова РАН, Москва, Россия

E-mail: kruglov@mi-ras.ru, mikhail@mi-ras.ru

1. Введение

Матрицы над конечными полями активно используются в задачах теории кодирования, криптографии и в других направлениях дискретной математики (см., например, [7], [6], [3]). Во многих из этих задач интерес представляют свойства ранга матрицы при случайном выборе ее элементов.

Наиболее изученной является задача о ранге матрицы, элементы которой выбраны независимо и равновероятно из множества элементов рассматриваемого поля. Изучение предельного поведения распределения ранга таких матриц дает, к примеру, следующий результат.

Пусть число строк n и число столбцов m матрицы A, элементы которой независимы в совокупности и распределены на множестве элементов поля $GF(q)$ равновероятно, связаны соотношением $m - n = k$, где k — фиксированное целое число. Тогда (см. [4], [5]) при любом неотрицательном целом l, для которого $l + k \geq 0$,

$$\lim_{n\to\infty} \mathbf{P}\{\mathrm{rank}(A) = n - l\} = q^{l(k+l)} \prod_{i=l+1}^{\infty} \left(1 - \frac{1}{q^i}\right) \prod_{j=1}^{k+l} \left(1 - \frac{1}{q^j}\right)^{-1}.$$

Понимая под весом вектора число его ненулевых элементов, рассмотрим класс матриц, состоящих из n независимых случайных строк, которые выбираются равновероятным способом из множества всех m-мерных ($m > n$) двоичных векторов заданных весов s_i, $i = 1, \ldots, n$.

В настоящей работе для функций распределения ранга таких случайных матриц строятся явные оценки сверху в форме неравенств.

2. Формулировки результатов

Теорема 1. *Пусть B — двоичная матрица размера $n\times m$, $n < m$, строки $b_1, \dots, b_n$ которой выбраны независимо и равновероятно из множества всех m-мерных двоичных векторов заданных весов $s_1, \dots, s_n$ соответственно. Тогда при любом $l \in \{1, \dots, n-1\}$*

$$\mathbf{P}\{\operatorname{rank}(B) \le n-l\} \le$$
$$\le \frac{1}{l2^m} \sum_{t=0}^{m} C_m^t \left[\prod_{i=1}^{n} \left(1 + \frac{K_{s_i}^m(t)}{C_m^{s_i}} \right) - \sum_{i=1}^{n} \frac{K_{s_i}^m(t)}{C_m^{s_i}} - 1 \right],$$

где $K_s^m(t) = \sum\limits_{j=0}^{\min(s,t)} (-1)^j C_t^j C_{m-t}^{s-j}$ — многочлены Кравчука, задаваемые соотношением

$$\sum_{s=0}^{\infty} K_s^m(t) z^s = (1-z)^m (1+z)^{m-t}.$$

Отметим, что $\sum\limits_{j=0}^{\min(s,t)} C_t^j C_{m-t}^{s-j} = C_m^s$, и, следовательно, $\left| \frac{K_s^m(t)}{C_m^s} \right| < 1$. Положим

$$Q(s_1, \dots, s_n) = \sum_{i=1}^{n} \frac{1}{\sqrt{C_m^{s_i}}}.$$

Теорема 2. *Пусть выполнены условия теоремы 1 и числа $s_1, \dots, s_n$ таковы, что найдется натуральное число $t_0 \le \frac{m}{2}$, для которого*

$$C_m^{t_0} \ge 2^m Q^2(s_1, \dots, s_n).$$

Тогда при любом $l \in \{1, \dots, n-1\}$

$$l\,\mathbf{P}\{\operatorname{rank}(B) \le n-l\} \le 2^{n-m+1} \sum_{t=0}^{t_0} C_m^t + e(m-2t_0+1) Q^2(s_1, \dots, s_n),$$

в частности,

$$\mathbf{P}\{\operatorname{rank}(B) < n\} \le 2^{n-m+1} \sum_{t=0}^{t_0} C_m^t + e(m-2t_0+1) Q^2(s_1, \dots, s_n).$$

Теоремы 1 и 2 получены в продолжение исследований работ [1], [2] и [8] о свойствах случайных линейных кодов над простыми конечными полями.

Список литературы

[1] *Зубков А. М., Круглов В. И.* Моментные характеристики весов векторов в случайных двоичных линейных кодах // Матем. вопр. криптогр. 2014. **3**(4). С. 55–70.

[2] *Зубков А. М., Круглов В. И.* Статистические характеристики весовых спектров случайных линейных кодов над GF(p) // Матем. вопр. криптогр. 2014. **5**(1). С. 27–38.

[3] *Коблиц Н.* Курс теории чисел и крипографии. М: Научное изд-во ТВП, 2001.

[4] *Коваленко И. Н., Левитская А. А., Савчук М. Н.* Избранные задачи вероятностной комбинаторики. Киев: Наукова думка, 1986.

[5] *Колчин В. Ф.* Случайные графы. Москва, ФИЗМАТЛИТ, 2000.

[6] *Лидл Р., Нидеррайтер Г.* Конечные поля. **2**. М: Мир, 1988.

[7] *Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А.* Теория кодов, исправляющих ошибки. М: Связь, 1979.

[8] *Zubkov A. M., Kruglov V. I.* On quantiles of minimal codeword weights of random linear codes over F_p // Матем. вопр. криптогр. 2018. **9**(2). P. 99–102

Об оптимальном управлении деками в двухуровневой памяти

А. А. Лазутина[1], А. В. Соколов[2]

[1] Московский государственный университет имени М.В.Ломоносова, Москва, Россия
[2] Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: alazutina@yandex.ru, sokavs@gmail.com

Для динамической балансировки параллельных вычислений в общей памяти часто используют стратегию «Work-stealing», при которой пустые ядра забирают задачи у других ядер [1–3]. У каждого ядра есть дек, "с ограниченным входом" [4]. «Work-stealing» используется в таких системах как Cilk [5], Cilk++, TPL [6], TBB, X10 [7].

В работах [8, 9] мы предлагали модели для оптимального управления деками в памяти одного уровня. Как показал наш опыт программной реализации, большое значение имеет оптимизация работы с кэш-памятью в параллельных балансировщиках задач, работающих по стратегии work-stealing. Например, в реализации [10] за счет этого удалось уменьшить время выполнения тестовых задач в 2.5 раза и кеш-промахи до 30% по сравнению с work-stealing балансировщиками Intel TBB и Intel/MIT Cilk. Поэтому важно исследовать и специальные методы работы с деками в двухуровневой памяти, а не только пытаться работать с универсальными реализациями кэш-памяти.

Для наиболее часто используемых структур данных можно предложить свои отличные от универсальных методы работы в двухуровневой памяти, например, регистры – оперативная память.

В случае переполнения стека в быстрой памяти, в медленную память надо переносить элементы из основания стека, так как они не понадобятся в ближайшем будущем. В случае переполнения очереди, в медленную память следует переносить элементы из более новой части очереди. В работах [11] и [12] решались задачи оптимального управления одним и двумя стеками, в [13] – FIFO-очередью в двухуровневой памяти. На практике в разных архитектурах был аппаратно реализован ряд методов управления стеками в двухуровневой памяти, как альтернатива классической кэш-памяти [14]. В этой работе мы рассмотрим модель для оптимального управления деком в двухуровневой памяти.

Рассмотрим классический последовательный циклический метод представления дека в памяти. В случае переполнения дека в быстрой памяти, в медленную память мы переносим элементы из средней части структуры данных, так как данные с концов дека могут понадобиться раньше.

Пусть мы работаем с несколькими деками параллельно. Рассмотрим модель и постановку нашей задачи для одного из деков системы. Можно считать, что мы разделили файл регистров между деками и нам надо найти оптимальный метод работы с каждым деком.

Пусть заданы вероятности возможных параллельных операций с деком. Наша задача найти количества элементов с двух сторон дека, которые мы оставляем в быстрой памяти, чтобы среднее время до попадания в состояния, когда требуется провести перераспределение памяти, было максимальным.

Рассмотрены две постановки задачи, в зависимости от используемой стратегии кражи задач.

1. Кражи элементов из деков происходят по k элементов.
2. Кражи элементов из деков происходят по половине элементов.

Для решения поставленной задачи в общем случае разработана имитационная модель. В некоторых частных случаях построены математические модели с использованием аппарата разностных уравнений.

В докладе будут представлены результаты численных экспериментов с разработанными моделями.

Работа выполнена при финансовой поддержке РФФИ, грант 18-01-00125-а.

Список литературы

[1] *Blumofe R. D., Leiserson C. E.* Scheduling Multithreaded Computations by Work Stealing // Journal of the ACM. 1999. **46**(5). P. 720–748

[2] *Herlihy M., Shavit N.* The Art of Multiprocessor Programming. Elsevier, 2008. 536 p.

[3] *Hendler D., Shavit N.* Non-blocking Steal-half Work Queues // ACM PODC. 2002. P. 280–289.

[4] *Knuth D.* The Art of Computer Programming. Vol.1. Addison-Wesley Professional, 1997. 672 p.

[5] *Blumofe R. D., Joerg C. F., Kuszmaul B. C., Leiserson C. E., Randall K. H., Zhou Y.* Cilk: An Efficient Multithreaded Runtime System // Journal of Parallel and Distributed Computing. 1996. **37**(1). P. 55–69.

[6] *Leijen D., Schulte W., Burckhardt S.* The Design of a Task Parallel Library // ACM SIGPLAN. 2009.**44**(10). P. 227–242.

[7] *Tardieu O., Wang H., Lin H.* A work-stealing scheduler for X10's task parallelism with suspension // ACM SIGPLAN. 2012. **47**(8). P. 267–276.

[8] *Sokolov A. V., Barkovsky E. A.* The Mathematical Model and The Problem of Optimal Partitioning of Shared Memory for Work-Stealing Deques // Lecture Notes in Computer Science. 2015. **9251**. P. 102–106.

[9] *Aksenova E. A., Sokolov A. V.* Modeling of the Memory Management Process for Dynamic Work-Stealing Schedulers // IEEE Proceedings: Proceedings of 2017 Ivannikov ISPRAS Open Conference (ISPRAS). 2018. P. 12–15. DOI: 10.1109/ISPRAS.2017.00009

[10] *Kuchumov R., Sokolov A., Korkhov V.* Staccato: Cache-Aware Work-Stealing Task Scheduler for Shared-Memory Systems // ICCSA 2018. Lecture Notes in Computer Science, Vol. 10963. Springer, Cham. 2018. P. 91–102. DOI: 10.1007/978-3-319-95171-3_8

[11] *Aksenova E. A., Lazutina A. A., Sokolov A. V.* Study of a Non-Markovian Stack Management Model in a Two-Level Memory // Programming and Computer Software. 2004. **30**(1). P. 25–33.

[12] *Аксенова Е. А., Соколов А. В.* Оптимальное управление двумя параллельными стеками в двухуровневой памяти // Дискретная математика. 2007. **19**(1). С. 67–75.

[13] *Соколов А. В.* Об оптимальном кешировании FIFO-очередей // Стохастическая оптимизация в информатике. 2013. **9**(2). С. 72–88.

[14] *Koopman P. J.* Stack Computers: The New Wave. Ellis Horwood Ltd., 1989. 502 p.

О сообществах в коммуникационных графах

В. В. Мазалов, Н. Н. Никитина, А. А. Печников

Институт прикладных математических исследований КарНЦ РАН,
Петрозаводск, Россия

E-mail: nikitina@krc.karelia.ru

Веб-граф фрагмента Веба — это ориентированный граф без петель и кратных дуг, множеством вершин которого является целевое множество сайтов, а множество дуг строится следующим образом: две вершины связаны дугой, если есть хотя бы одна гиперссылка, связывающая соответствующие сайты. Коммуникационный граф веб-графа — это неориентированный граф, имеющий то же самое множество вершин, что и веб-граф, получаемый из веб-графа путём замены встречных дуг на ребра. Коммуникационный граф, построенный таким образом, может иметь изолированные вершины и/или несколько компонент связности. В этом случае мы исключаем изолированные вершины, поскольку они не влияют на связность, и изучаем компоненты связности каждую по отдельности, начиная с максимальной.

В работе исследуется вопрос о структуре сообществ коммуникационных графов для трёх случаев на примерах сайтов институтов РАН (академический Веб), сайтов российских университетов (университетский Веб) и объединенного множества сайтов университетов и институтов, который назовем научно-образовательным Вебом.

Авторские базы данных позволяют построить веб-графы для трех указанных фрагментов Веба, имеющие следующие характеристики: академический граф – 588 вершин, 2617 дуг; университетский граф – 279 вершин, 1242 дуги; научно-образовательный граф – 867 вершин, 5030 дуг. Построенные для этих веб-графов коммуникационные графы имеют следующие характеристики: академический коммуникационный граф – 231 вершина, 341 ребро; университетский коммуникационный граф – 35 вершин, 37 рёбер; научно-образовательный коммуникационный граф – 313 вершин, 468 рёбер.

Для поиска сообществ в каждом из графов использовались программы, реализованные в системе Wolfram Mathematica на основе двух различных подходов. В первом случае это известный алгоритм иерархического разбиения Гирвана-Ньюмана [1], реализованный в пакете IGraph/M [2]. Второй подход, описанный в работе [3], и основанный на использовании метода максимального правдоподобия, реализован в Wolfram Mathematica.

Для первого подхода число сообществ не было задано априори и определялось автоматически для наибольшего значения коэффициента модулярности. Во втором подходе полученное число сообществ использовалось в качестве параметра алгоритма.

Разбиения, полученные двумя алгоритмами, оказались различными для всех трех графов, но достаточно близкими по таким характеристикам, как численные меры сходства двух разбиений NMI (Normalized Mutual Information)и ARI (Adjusted RandIndex).

В то же время, такие характеристики, как коэффициент модулярности, средняя плотность связей внутри сообществ и средняя плотность связей между сообществами получились различными.

Содержательная интерпретация результатов дает основания полагать, что полученные разбиения на сообщества качественно близки во всех трех случаях.

Список литературы

[1] *Newman M. E., Girvan M.* Finding and evaluating community structure in networks // Physical Review E. 2004. **69**(2). P. 026113.

[2] IGraph/M | Zenodo. https://zenodo.org/record/2349281.

[3] *Мазалов В. В., Никитина Н. Н.* Метод максимального правдоподобия для выделения сообществ в коммуникационных сетях // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2018. **14**(3). С. 200–214.

Признаки распределения степеней вершин теоретико-графовых моделей текстов

Н. Д. Москин, А. А. Рогов

Петрозаводский государственный университет, Петрозаводск, Россия

E-mail: moskin@petrsu.ru, rogov@petrsu.ru

Атрибуция текстов в настоящее время является одной из актуальных задач в филологии. Например, уже многие годы исследуется коллекция, содержащая около 500 неатрибутированных текстов из журналов "Время"(1861-1863), "Эпоха"(1864-1865) и еженедельника "Гражданин"(1873-1874), часть из которых предположительно могла принадлежать перу Ф. М. Достоевского [2]. Другой задачей является разграничение фольклорных текстов и текстов, стилизованных под фольклор (которые сочинили такие известные поэты как Н. А. Клюев, А. К. Толстой и С. А. Есенин) [3].

Задача установления авторства текстов давно стала междисциплинарной, часто она решается с применением математических методов [2]. Один из подходов подразумевает семантический анализ текста, в результате которого текст представляется в виде графа. Схожесть текстов означает близость их графов. При этом анализ осуществляют разными методами. Один из методов заключается в том, чтобы сопоставить каждой вершине графа ее степень и отсортировать их в порядке убывания. Примеры двух таких диаграмм, полученных при анализе 48 бесёдных песен Заонежья XIX – начала XX века, изображены на рис. 1. Однако, применение многомерных методов (например, деревьев решений) требует представление диаграммы одним числом. В данной работе авторы показывают на эмпирическом материале возможность использовать для такой оценки один из параметров гиперболической регрессии, т. е. кривой вида $y = \frac{a}{x} + b$. Этой характеристикой будет коэффициент a (как показали эксперименты, в задаче разграничения народных песен и песен, стилизованных под фольклор, этот коэффициент среди восьми показателей является определяющим уже на третьем узле проверки в дереве решений [3]).

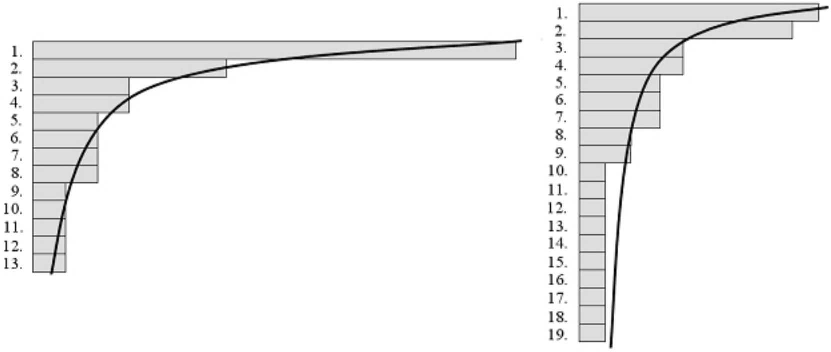

Рис. 1: Распределение степеней вершин по убыванию.

Фрагмент таблицы с исходными данными представлен на Табл. 1 (параметры гиперболической регрессии a и b были найдены при помощи метода наименьших квадратов). Здесь z_{ij} – это степень вершины с номером j в отсортированном списке для i текста ($i = 1, \ldots, 48$), если $j \leq 3$, или сумма степеней оставшихся вершин, если $j = 4$.

Таблица 1: Фрагмент таблицы с исходными данными

№	Текст	a_i	b_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}
1	Все мужья до жен добры	23,268	1,037	10	5	3	22
2	Девушка в горенке сидела	30,998	-0,389	17	14	4	25
3	Уж ты Ванюша, Иван	14,48	2,744	5	4	3	22
4	Широкая борода!	34,41	-1,126	14	4	3	17
5	Не огонь горит	23,593	1,448	8	7	6	17
6	Ты, отеческая дочь!	20,012	1,833	9	7	4	26
7	Мальчик ты, мальчик	39,855	-1,851	12	9	3	10
8	Невызревшей рябинушки	16,551	2,023	10	8	7	37
9	Как вила серая утка	21,318	0,813	10	5	3	26
10	Ах ты, вьюнчик	42,305	0,681	5	3	2	2
…	…	…	…	…	…	…	…

Для проверки гипотезы найдем коэффициент корреляции Пирсона между модулем разности коэффициентов регрессии $d_{ij}^{(1)} = |a_i - a_j|$ и расстоянием χ^2 между диаграммами. Расстояние вычислялось по формуле $d_{ij}^{(2)} = n_{ij}(\sum\limits_{k=1}^{4}(\frac{z_{ik}^2}{z_i(z_{ik}+z_{jk})} + \frac{z_{jk}^2}{z_j(z_{ik}+z_{jk})}) - 1)$, где $n_{ij} = \sum\limits_{k=1}^{4}(z_{ik} + z_{jk})$, $z_i = \sum\limits_{k=1}^{4} z_{ik}$ - сумма степеней по строкам ($i, j = 1, \ldots, 48$) [1].

Как показали расчеты, коэффициент корреляции Пирсона составил

$r \approx 0,76$. Согласно шкале Чеддока такую связь считают высокой. Облако рассеивания между соответствующими значениями $d_{ij}^{(1)}$ и $d_{ij}^{(2)}$ представлено на рис. 2. Если рассматривать второй коэффициент b, то корреляция между $d_{ij}^{(2)}$ и $d_{ij}^{(3)} = |b_i - b_j|$ равна $r \approx 0,38$ (умеренная связь).

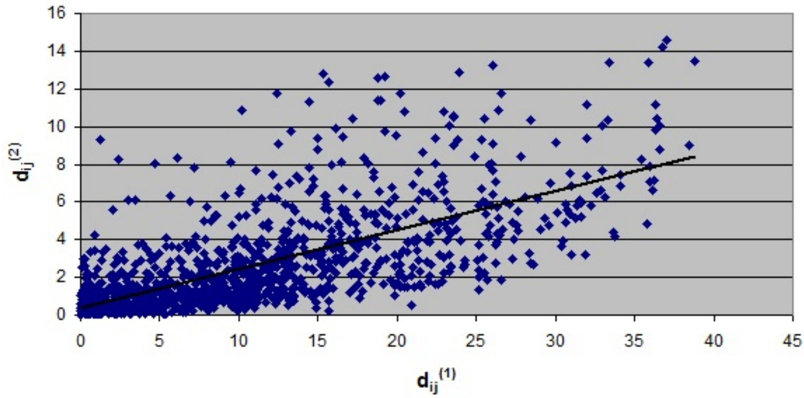

Рис. 2: Облако рассеивания $d_{ij}^{(1)}$ от $d_{ij}^{(2)}$.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-012-90026.

Список литературы

[1] *Айвазян С. А., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное издание. М.: Финансы и статистика, 1983. 471 с.

[2] *Рогов А. А., Москин Н. Д., Сидоров Ю. В., Суровцова Т. Г.* Математические методы атрибуции анонимных статей // Современные методы прикладной математики, теории управления и компьютерных технологий: сборник трудов XI международной научной конференции ПМТУКТ-2018 (Воронеж, 18-24 сентября 2018 г.). Воронеж: Изд-во «Научная книга», 2018. С. 227–230.

[3] *Shchegoleva L., Lebedev A., Moskin N.* Recognition of Folklore Texts and Author's Poems Using Classification Trees and Neural Networks // Proceedings of the 22st Conference of Open Innovations Association FRUCT. Jyvaskyla, Finland, 2018. P. 418–420.

Потенциал в игре заполнения для виртуального скрининга лекарств на базе Desktop Grid

Н. Н. Никитина

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: nikitina@krc.karelia.ru

Виртуальный скрининг лекарств. Создание новых лекарств представляет собой трудоемкую задачу и связано со значительными временными затратами на поиск веществ, обладающих терапевтическим действием для белка-мишени и соответствующих всем критериям безопасности для человека. Один из первых этапов процесса разработки лекарства состоит в отборе множества химических соединений-хитов, которые потенциально имеют требуемую биохимическую активность по отношению к мишени. Хиты выбираются из множества лигандов – соединений, способных создавать устойчивые комплексы с мишенью. На следующих этапах проводится отбор и оптимизация хитов. Метод автоматического роботизированного поиска хитов в лабораторных условиях называется высокопроизводительным скринингом. С развитием высокоточных методов компьютерного моделирования, виртуальный скрининг (ВС) стал *«in silico»* альтернативой высокопроизводительному. В ходе ВС проводится компьютерное моделирование взаимодействия лигандов и мишени и оценивается вероятность образования устойчивых молекулярных комплексов. Лиганды с высокими оценками такой вероятности становятся хитами. На ранних этапах ВС требуется скорейшее нахождение как можно более химически разнообразных хитов для тестирования в лаборатории.

Системы Desktop Grid. Большие объемы библиотек моделей лигандов и мишеней приводят к необходимости привлечения инструментов высокопроизводительных вычислений для проведения ВС. Одним из эффективных вычислительных инструментов для проведения ВС являются вычислительные системы Desktop Grid, объединяющие территориально распределенные неспециализированные вычислители, связанные с управляющим узлом сетью передачи данных. В работе предложена и исследована математическая модель управления заданиями в

Desktop Grid для проведения ВС, нацеленная на скорейшее нахождение как можно более химически разнообразных хитов.

Математическая модель. Рассмотрим Desktop Grid с N вычислительными узлами ($N \geq 2$), или игроками, $C_1, \ldots, C_N$. Узел C_i имеет производительность $w_i > 0$ – число вычислительных операций, которое он выполняет за шаг времени. Будем говорить, что игрок C_i имеет вес w_i. Обозначим произведение весов всех игроков как $W = \prod_{i=1}^{N} w_i$.

Разобьем множество вычислительных заданий T на R непересекающихся подмножеств $T = T_1 \cup \ldots \cup T_R$. На каждом шаге игрок выбирает в точности одно подмножество заданий $r \in \mathcal{R} = \{1, \ldots, R\}$.

Определим приоритет подмножества r следующим образом:

$$\sigma_r = \begin{cases} \frac{p_r}{p_1 + \ldots + p_R}, & p_r > 0, \\ \epsilon > 0, & \text{иначе}, \end{cases} \tag{1}$$

где p_r – ожидаемая доля хитов в подмножестве r.

После того как вычислительный узел выполнил задания, он отсылает результаты серверу и запрашивает следующие задания. Пусть выигрыш игрока C_i на данном шаге выражает количество выполненной им полезной работы. Оно зависит от числа выполненных заданий, приоритета подмножества и от числа других игроков, выбравших данное подмножество. Чем меньше игроков на данном шаге выполняют подмножество заданий r, тем более ценна работа каждого из них, поскольку найденные хиты ВС являются химически разнообразными.

Обозначим как n_r число игроков, выбравших подмножество r на данном шаге, а $\delta(n_r)$ – коэффициент заполнения подмножества:

$$\delta(n_r) = \frac{N + 1 - n_r}{N}. \tag{2}$$

Выигрыш узла C_i, выбравшего подмножество r, составляет

$$U_{ir} = w_i \left(\alpha \delta(n_r) + (1 - \alpha) \sigma_r \right). \tag{3}$$

Здесь $\alpha \in [0, 1]$ – коэффициент, выражающий характер игрока. При $\alpha \to 0$ игрок стремится быть «добытчиком» (получать как можно больше хитов), при $\alpha \to 1$ – «разведчиком» (исследовать незаполненные подмножества заданий).

Таким образом, имеется игра заполнения $\Gamma = \langle C, \mathcal{R}, U \rangle$, где C — множество игроков (вычислительных узлов), $\mathcal{R}$ – множество стратегий, U — множество функций выигрыша. Профиль стратегий в данной игре – вектор-расписание $\bar{s} = (s_1, \ldots, s_N)$, где компонента $s_i = r$ равна номеру подмножества r, выбранного игроком C_i. Профиль стратегий является равновесием по Нэшу, если ни один вычислительный узел не может увеличить количество своей полезной работы, в одностороннем порядке отклонившись от данного расписания.

Для невзвешенных игр заполнения ($w_1 = \ldots = w_N$) известно, что игра Γ имеет равновесие по Нэшу в чистых стратегиях [1]. Известно также, что динамика наилучших ответов сходится к равновесию за полиномиальное время [2]. Покажем, что известные результаты верны и для рассматриваемой нами игры. Воспользуемся тем фактом, что функции выигрыша являются мультипликативно-сепарабельными [3].

Рассмотрим функцию

$$\Phi(\bar{s}) = \prod_{r \in R} \prod_{k=1}^{n_r} W\left(\alpha \frac{N+1-k}{N} + (1-\alpha)\sigma_r\right) = \prod_{r \in R} \phi_r. \tag{4}$$

Заметим, что если $\sigma_r > 0$, то $\phi_r > 0 \ \forall r, \forall \bar{s}$.

При отклонении игрока k от выбранной стратегии r на стратегию t изменение его выигрыша составит

$$\Delta U_k = w_k \left((1-\alpha)(\sigma_t - \sigma_r) + \frac{\alpha}{N}(n_r - n_t - 1)\right), \tag{5}$$

а изменение значения функции Φ составит

$$\Delta \Phi = W \phi'_r \phi_t \left((1-\alpha)(\sigma_t - \sigma_r) + \frac{\alpha}{N}(n_r - n_t - 1)\right), \tag{6}$$

где ϕ'_r вычислено в новом профиле стратегий, а ϕ_t – в старом.

Знаки выражений (5) и (6) совпадают. Таким образом, функция Φ является ординальным потенциалом игры Γ, а сама игра – потенциальной. В ней всегда существует равновесие по Нэшу. Действуя по алгоритму наилучших ответов, игроки пошагово оптимизируют единую глобальную функцию – потенциал (4), и в вычислительной системе Desktop Grid децентрализованно достигается равновесие по Нэшу.

Работа выполнена при финансовой поддержке РФФИ в рамках научных проектов №18-07-00628 и №18-37-00094.

Список литературы

[1] *Rosenthal R.* A Class of Games Possessing Pure-strategy Nash Equilibria // International Journal of Game Theory. 1973. **2**(1). P. 65–67.

[2] *Ieong S. et al.* Fast and Compact: A Simple Class Of Congestion Games // AIII Proceedings. 2005. P. 1–6.

[3] *Milchtaich I.* Weighted congestion games with separable preferences // Games and Economic Behavior. 2009. **67**(2). P. 750–757.

Закон больших чисел для количества непоявившихся непересекающихся цепочек в последовательности испытаний Бернулли

М. П. Савелов

МГУ им. М.В. Ломоносова, Московский физико-технический институт, Москва, Россия

E-mail: savelovmp@gmail.com

Рассмотрим последовательность $X_1, X_2, \ldots$ независимых одинаково распределенных бернуллиевских случайных величин: $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$. Будем считать, что у нас n независимых цепочек наблюдений, в каждой из которых $k(n)$ наблюдений, причем $n \to \infty$, $k = k(n) \to \infty$. Обозначим через μ_0 число цепочек, не появившихся в такой выборке ни разу.

В докладе будет обсуждаться статистика

$$\hat{H} = \frac{\ln(2^k - \mu_0)}{k},$$

которая является «грубой оценкой» энтропии $H(p) = -p \ln p - (1 - p) \ln(1 - p)$: будет пояснено, почему данная оценка естественна, а также будет приведен ряд теорем. Наиболее важным является следующий результат.

Положим $f(p) = \log_{\frac{1-p}{p}} 2(1 - p)$ при $p \in (0, \frac{1}{2}) \bigcup (\frac{1}{2}, 1)$ и $f(p) = p$ при $p \in \{0, \frac{1}{2}, 1\}$. Функция $f(p)$ непрерывна, строго возрастает, принимает значения в диапазоне от 0 до 1 и удовлетворяет соотношению $f(1-p) = 1-f(p)$. Мы предполагаем, что n и k стремятся к бесконечности, причем по техническим причинам нам будет удобнее вместо зависимости $k(n)$ рассматривать зависимость $n = n(k) \to \infty$ при $k \to \infty$.

Теорема 1. *Пусть $n = q_k 2^k$ и $\lim_{k\to\infty} \frac{\ln q_k}{k} = 0$. Тогда $\hat{H}$ является сильно состоятельной оценкой величины $H(f(p))$, причем $H(f(p)) \geq H(p)$ и равенство достигается только при $p = 0, \frac{1}{2}, 1$.*

Пример 1. *Если $n = q_k 2^k$ и $\frac{1}{k^{100}} \leq q_k \leq k^{100}$, то условия теоремы 1 выполнены.*

Структурные и стохастические свойства специальных графов

Л. Я. Савельев

Новосибирский государственный университет, ИМ СО РАН, Новосибирск, Россия

E-mail: savelev@math.nsc.ru

Рассматриваются некоторые частные случаи графов, ребрами которых являются ребра конечных прямоугольных решеток и диагонали ячеек. Каждая нумерация элементов решетки определяет двоичный номер графа, составленный из индикаторов его ребер. Для описания структурных свойств графа используются характеристики серий единиц и серий нулей двоичного номера: общие длины таких серий, их числа и максимумы длин. Они позволяют преобразовывать графы в более простые. Такие задачи возникают при повышении надежности автоматического чтения.

Стохастические свойства графов при последовательном выборе ребер связываются с распределениями таких характеристик серий в двоичных марковских последовательностях.

Двоичные номера каждого конечного семейства рассматриваемых графов составляют прямоугольную двоичную матрицу. Нумерация ребер решетки накладывает ограничения на характеристики серий единиц и серий нулей в строках и столбцах такой матрицы. Возникает задача о числе матриц, удовлетворяющих этим ограничениям.

1. Прямоугольные графы цифр. Электронные цифры можно рассматривать как графы с ребрами на прямоугольной решетке, у которой 6 узлов и 7 ребер. Существует $7! = 5040$ различных нумераций ребер такой решетки. В качестве стандартной можно принять нумерацию снизу вверх и слева направо: левые вертикальные 1, 2; горизонтальные 3, 4, 5; правые вертикальные 6, 7. Остальные получаются перестановками. Структура цифр определяет следующие связи элементов $b[i,j,t]$ их двоичных номеров $b[i,t]$ цифр i при перестановке t: $b[1,t[6],t] = b[1,t[7],t] = 1$; $b[2,t[2],t] = b[2,t[6],t] = 0$; $b[3,t[1],t] = b[3,t[2],t] = 0$; $b[4,t[1],t] = b[4,t[3],t] = b[4,t[5],t] = 0$; $b[5,t[1],t] = b[1,t[7],t] = 0$; $b[6,t[7],t] = 0$; $b[7,t[5],t] = b[7,t[6],t] = b[7,t[7],t] = 1$; $b[8,j,t] = 1$; $b[9,t[1],t] = 0$; $b[0,t[4],t] = 0$. Неуказанные позиции дополняются противоположными значениями. Двоичный номер $b[i,t]$ является i-й строкой нумерационной матрицы $B[t]$, получающейся из стандартной перестановкой столбцов t. Стандартная матрица имеет строки 0000011, 1011101, 0011111, 0101011, 0111110, 1111110, 0000111, 11111111, 0111111, 1110111.

Серийные номера $v[i,t] = \{x[i,t], y[i,t], z[i,t]\}$ цифры i при перестановке t ($x[i,t]$ – число 1-ц, $y[i,t]$ – число 1-серий, $z[i,t]$ – максимум длин 1-серий) вычисляются по формулам $x[i,t] = \sum_{j=1}^{7} b[i,j,t]$,
$y[i,t] = b[i,1,t] + \sum_{j=2}^{7} c[i,j,t]$,
$z[i,t] = \max \sum_{j=1}^{7} b[i,j,t] \prod_{k=j}^{m} b[i,k,t]$; $c[i,j,t] = (1 - b[i,j-1,t])b[i,j,t]$. Введем также краткие серийные номера $w[i,t] = \{y[i,t], z[i,t]\}$ только с числом и максимумом длин серий единиц. При некоторых перестановках они различают прямоугольные цифры.

Простая программа направленного перебора позволила найти все такие перестановки. Их оказалось 18. Использовалось то, что графы цифр 6, 9, 0 не имеют ровно по одному разному ребру и позиции j, k, m единственных нулей в их строках нумерационной матрицы определяют соответствующие три ее столбца. Оба нуля номера цифры 5 находятся на местах j, k. Один из нулей номера цифры 3 находится на месте k нуля номера графа цифры 9. Второй занимает новое место n, совпадающее с одним из нулей номера цифры 2 и не совпадающее с местами всех предыдущих нулей. Второй нуль номера цифры 2 занимает новое место p. Единицы номера цифры 4 занимают места m, n, p и новое место q. Единицы номера цифры 1 занимает места j и q. Две единицы номера цифры 7 занимают места j, q, а место третьей единицы совпадает с общим местом s нулей в полученных номерах цифр 1 и 4.

Найти допустимые перестановки для рассматриваемой простой решетки можно и полным перебором всех перестановок. Направленность нужна в более сложных случаях для состоящих из цифр комбинаций натуральных чисел.

Методы решения получаемых систем нелинейных двоичных уравнений могут иметь специфику. Перестановка $t = 2453671$ дает пример различных кратких серийных номеров цифр $1, \ldots, 9, 0$: 21,23,33,32,15, 16,22,17,25,24. Простой визуализацией могут служить прямые углы, стороны которых имеют соответствующие числа звеньев. Различение их направлений по четности цифр усилит контраст.

2. Графы с диагоналями. Цифры на индексах почтовых конвертов можно рассматривать как графы с ребрами на решетке, полученной из прежней добавлением двух диагоналей в клетках. Существует $9! = 362880$ различных нумераций ребер такой решетки. В стандартную нумерацию добавим номера 8 и 9 для нижней и верхней диагоналей. Остальные нумерации дают перестановки номеров. Связи элементов двоичных номеров цифр нетрудно выписать.

Кроме прежних двух и трехзначных рассматриваются пятизначные серийные номера цифр при перестановках, полученные добавлением числа серий нулей и максимума их длин, которые вычисляются по аналогичным формулам. Ставятся задачи о классах допустимых нумераций, различающих цифры по выбранным серийным номерам, и о мощностях этих классов. Приводятся примеры допустимых нумераций.

Стилизованные буквы можно рассматривать как графы с ребрами на решетке из четырех квадратов с диагоналями, у которой 9 узлов и 20 ребер. Существует 20! различных нумераций ребер такой решетки. Ставятся аналогичные задачи о классах допустимых нумераций и их мощностях. Приводятся примеры допустимых нумераций выбранных букв.

3. Стохастические свойства нумераций. В книге В.Ф. Колчина [1] описываются случайные графы с независимыми ребрами и матрицы с независимыми элементами. Несомненный интерес представляет марковская зависимость между ребрами и элементами. Стохастические свойства серийных номеров выражаются совместными распределениями характеристик серий в двоичных марковских цепях. Такие распределения, случайные соответствия, двоичные матрицы и серии, подробно описаны в статье [2]. Там выведены формулы для рациональных производящих функций распределений и коэффициентов их разложений в степенные ряды, дающих вероятности. Рассматривается частный случай последовательностей независимых значений. Формулы удобны для компьютерных вычислений.

В связи с двоичными матрицами рассматривается задача о пакетном обслуживании заявок специальной линией автоматов, которая является модификацией предложенной Л. Н. Королевым задачи об оценке эффективности блочной организации памяти компьютера. Полученные распределения характеристик серий в случайных последовательностях можно использовать для оценки эффективности рассматриваемого процесса пакетного обслуживания. Представляют интерес его графические интерпретации.

Список литературы

[1] *Колчин В. Ф.* Случайные графы. М.: ФИЗМАТ ЛИТ, 2004. 256 с.

[2] *Савельев Л. Я.* Случайные соответствия, двоичные матрицы и серии // Дискретная математика. 1999. **11**(4). С. 3–26.

Параметрическая идентифицируемость модели структурных уравнений с латентными переменными

С. В. Стафеев

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: stafeev@krc.karelia.ru

Пусть $G = (V, E)$ – смешанный граф с множеством вершин $V = \{1, ..., n\}$ и множеством ребер E. Множество ребер E графа G состоит из ориентированных $(i \to j)$ и двуориентированных $(i \leftrightarrow j)$ ребер.

Рассмотрим следующую, связанную с графом G, систему структурных уравнений с латентными переменными:

$$X = B_G^t X + A^t H + \varepsilon, \tag{1}$$

где $X = (X_1, ..., X_n)^t$ – вектор наблюдаемых случайных величин с матрицей ковариаций $\Sigma = (\sigma_{ij})$; $H = (H_1, ..., H_k)^t$ – вектор независимых нормально распределенных латентных (скрытых) случайных величин с $\mathbf{M}(H_j) = 0$, и $\mathbf{M}(H_j)^2 = 1$; $A = (a_{ij}) \in \mathbb{R}^{k \times n}$; $B_G = (b_{ij}) \in \mathbb{R}^{n \times n}$, причем $b_{ij} = 0$, если $i \to j \notin E$. Вектор остатков $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^t$ имеет нормальное распределение с нулевым вектором математических ожиданий и неизвестной положительно определенной матрицей ковариаций $\Omega_G = (\omega_{ij})$, причем $\omega_{ij} = 0$, если $i \leftrightarrow j \notin E$. Мы будем предполагать, что вектора H и ε являются независимыми. Таким образом, с помощью графа G постулируются некоторые нули в матрицах B_G и Ω_G. Вектор параметров $\theta_G = (A, B_G, \Omega_G)$ модели (1) состоит из матрицы A и ненулевых элементов матриц B_G и Ω_G. Пусть $\mathbf{PD}_V$ – множество $n \times n$ симметричных положительно определенных матриц, $\mathbf{B}_G$ – множество матриц, для которых матрица $I - B_G$ обратима. Определим параметрическое множество модели (1):

$$\Omega_G = \{\theta_G : A = (a_{ij}) \in \mathbb{R}^{k \times n}, B_G \in \mathbf{B}_G, \Omega_G \in \mathbf{PD}_V\}.$$

Пусть $\theta_G \in \Theta_G$. Тогда, матрица ковариаций наблюдаемых случайных величин имеет вид:

$$\Sigma = (I - B_G)^{-t}(\Omega + AA^t)(I - B_G)^{-1}. \tag{2}$$

Пусть $\mathbf{M}_G$ – множество всех матриц, допускающих разложение (2). Таким образом, граф G задает семейство нормальных распределений

$$\mathbb{N}_G = \{N(0, \Sigma) : \Sigma \in \mathbf{M}_G\}. \tag{3}$$

Одной из наиболее важных проблем, связанных с моделями, содержащими латентные переменные (в частности с моделью (1)), является проблема параметрической идентифицируемости [1,2]. Данная проблема заключается в ответе на вопрос о возможности однозначного определения неизвестных параметров по имеющимся данным.

Рассмотрим вероятностно-статистическую модель $\mathcal{M}(\Theta)$, заданную семейством распределений $\{\mathbb{P}_\theta, \theta \in \Theta\}$. Модель $\mathcal{M}(\Theta)$ называется идентифицируемой, если любым различным $\theta \in \Theta$ и $\theta^{'} \in \Theta$ соответствуют различные распределения $\mathbb{P}_\theta$ и $\mathbb{P}_{\theta'}$.

Модель $\mathcal{M}(\Theta)$ называется локально идентифицируемой, для любого $\theta \in \Theta$ найдется конечное число таких $\theta^{'} \in \Theta$, $\theta^{'} \notin \theta$, что $\mathbb{P}_{\theta'} = \mathbb{P}_\theta$.

Модель $\mathcal{M}(\Theta)$ называется почти всюду (п.в.) локально идентифицируемой, если она является п.в. локально идентифицируемой при $\theta \in \Theta^{'} \subseteq \Theta$, а множество $\Theta \setminus \Theta^{'}$ имеет меру нуль.

В нашем случае, модель задается семейством распределений (3). Получаем, что модель (1) будет идентифицируемой, если по матрице Σ матрица A определяется с точностью до ортогонального преобразования, а элементы матриц Ω_G и B_G определяются однозначно.

Пусть $\overline{G} = (V, \overline{E})$ – дополнительный к G граф. Образуем граф $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, где $\mathbf{V} = \{\mathbf{i} = (i_1, ..., i_k), 1 \leqslant i_1 < i_2 < ... < i_k \leqslant k\}$, а $\mathbf{E} = \{(\mathbf{i}, \mathbf{j}) | \mathbf{i} = (i_1, ..., i_k), \mathbf{j} = (j_1, ..., j_k), (i_s, j_l) \in \overline{E}, s, l = 1, ..., k\}$. (Заметим, что граф $\mathcal{G}$ использовался в работе [3] для решения вопроса об идентифицируемости модели факторного анализа с зависимыми остатками.)

В работе [4] получены достаточные условия п.в. локальной параметрической идентифицируемости модели (1) в случае, когда G является простым ациклическим графом (т.е. графом, в котором любые две различные вершины могут быть соединены только одним ребром и он не содержит ориентированных циклов) и $k = 1$. В следующей теореме получены условия п.в. локальной идентифицируемости для произвольного k.

Теорема 1. *Пусть G простой ациклический граф. Модель* (1) *будет п.в. локально идентифицируемой, если каждая компонента связности графа $\mathcal{G}$ содержит по крайней мере один нечетный цикл.*

Список литературы

[1] *Maathuis M., Drton M., Lauritzen S., Wainwright M.* (Eds) Handbook of Graphical Models. Chapman & Hall/CRC., 2018. 536 p.

[2] *Drton M.* Algebraic problems in structural equation modeling // The 50th Anniversary of Grobner Bases, Mathematical Society of Japan, Tokyo, Japan, 2018. P. 35–86, doi:10.2969/aspm/07710035

[3] *Стафеев С. В.* Об условиях глобальной идентифицируемости для моделей факторного анализа // Труды Карельского научного центра Российской академии наук. 2011. **5**. С. 111–114.

[4] *Leung D., Drton M., Hara H.* Identifiability of directed Gaussian graphical models with one latent source // Electron. J. Statist. 2016. **10**. P. 394–422. doi:10.1214/16-EJS1111

Вероятностно-статистический анализ дискретных временных рядов: модели и методы

Ю. С. Харин

НИИ прикладных проблем математики и информатики БГУ,
Минск, Беларусь

E-mail: kharin@bsu.by

Адекватной моделью динамики многих реальных стохастических явлений является временной ряд (случайный процесс) $x_t \in A$ на вероятностном пространстве (Ω, F, P), где t — время, x_t — регистрируемая в момент t величина, A — пространство (алфавит) состояний исследуемого явления. Вероятностно-статистический анализ временных рядов в настоящее время глубоко разработан «для непрерывного случая» [1], когда A — множество ненулевой меры Лебега, а время t — дискретно или непрерывно.

В последние годы (в том числе по причине «цифровизации общества») растет интерес к дискретным временным рядам, для которых теория вероятностно-статистического анализа не разработана. Дискретный временной ряд (ДВР) — это случайный процесс x_t с дискретным временем $t \in Z = \{\ldots, -1, 0, 1, \ldots\}$ и дискретным множеством состояний A мощности $|A| = N$, $2 \leq N \leq +\infty$; на практике часто N — конечно. Именно этот случай рассматривается в данном докладе; без потери общности полагаем $A = \{0, 1, \ldots, N-1\}$.

ДВР адекватно описывает реальные явления в генетике, экономике, социологии, метеорологии, защите информации [2 - 4].

В докладе представлены результаты вероятностно-статистического анализа для следующих моделей ДВР.

1. Модель Джекобса-Льюиса [4]:

$$x_t = \mu_t x_{t-\gamma_t} + (1 - \mu_t)\xi_t,$$

где $\xi_t \in A$, $\mu_t \in V = \{0, 1\}$, $\gamma_t \in \{1, 2, \ldots, s\}$ — независимые в совокупности случайные величины.

2. Обобщенная МТД-модель Рафтери [4]:

$$p_{i_1,\ldots,i_s,i_{s+1}} ::= P\{x_{t+1} = i_{s+1} | x_t = i_s, \ldots, x_{i-s+1} = i_1\} = \\ = \sum_{j=1}^{s} \lambda_j (Q^{(j)})_{i_j, i_{s+1}}, i_1, \ldots, i_{s+1} \in A,$$

где $\{\lambda_j\}, \{Q^{(j)}\}$ — параметры модели.

3. Дискретная авторегрессия порядка s [3]:

$$x_t = (\theta_1 x_{t-1} + \ldots + \theta_s x_{t-s} + \xi_t) \mod N,$$

где $\theta_1, \ldots, \theta_s \in A$, $\theta_s \neq 0$, — параметры модели.

4. Цепь Маркова порядка s с r частичными связями [4]:

$$p_{i_1,\ldots,i_s,i_{s+1}} = (Q)_{i_{m_1},\ldots,i_{m_r},i_{s+1}},$$

где $s, 1 \leq r \leq s$, Q, $M = (m_1, \ldots, m_r)$ — параметры модели.

5. Цепь Маркова условного порядка:

$$p_{i_1,\ldots,i_s,i_{s+1}} = \sum_{k=0}^{K} I\{< (i_{s-L+1}, \ldots, i_1) >= k\} (Q)^{(m_k)})_{i_{b_k}, i_{s+1}},$$

где $s, K, L, \{b_k\}, \{m_k\}$, $\{Q^{(m_k)}\}$ — параметры модели, $I\{\cdot\}$ — индикатор.

6. Бернуллиевская условно нелинейная авторегрессионная модель $(N = 2)$ [5]:

$$p_{i_1,\ldots,i_s,i_{s+1}} = \begin{cases} F(\sum_{j=1}^{m} a_j \psi_j(i_1, \ldots, i_s)), & \text{если } i_{s+1} = 1 \ , \\ 1 - F(\sum_{j=1}^{m} a_j \psi_j(i_1, \ldots, i_s)), & \text{если } i_{s+1} = 0, \end{cases}$$

где $F(\cdot)$ — функция распределения, $\{\psi_j(\cdot)\}$ — базис, $m, \{a_j\}$ — параметры.

7. Биномиальная условно нелинейная авторегрессионная модель [6]:

$$p_{i_1,\ldots,i_s,i_{s+1}} = C_N^{i_{s+1}} \Theta^{i_{s+1}} (1 - \Theta)^{N - i_{s+1}},$$

$$\Theta = \Theta(i_1, \ldots, i_s) = F(\sum_{j=1}^{m} a_j \psi_j(i_1, \ldots, i_s)).$$

Все рассматриваемые модели являются малопараметрическими моделями цепей Маркова высокого порядка. Используя марковские свойства удается построить состоятельные оценки параметров модели и статистические тесты проверки гипотез о значениях параметров модели.

Теоретические результаты иллюстрируются компьютерными экспериментами.

Список литературы

[1] *Андерсон Т. В.* Статистический анализ временных рядов. М.: Наука, 1976.

[2] *Weiss C. H.* An Introduction to Discrete-Valued Time Series. N.Y.: Wiley, 2018.

[3] *Харин Ю. С., Агиевич С. В., Васильев Д. В., Матвеев Г. В.* Криптология. Минск: БГУ, 2013.

[4] *Kharin Yu.* Robustness in Statistical Forecasting. N.Y.: Springer, 2013.

[5] *Kharin Yu., Voloshko V., Medved E.* Statistical estimation of parameters for binary conditionally nonlinear autoregressive time series // Mathematical Methods of Statistics. 2018. **27** (2).

[6] *Харин Ю. С., Волошко В. А.* Биномиальная условно нелинейная авторегрессионная модель дискретного временного ряда и ее вероятностно-статистические свойства // Труды Института математики НАН Беларуси. 2019. **26**(1).

О вторых степенях вершин конфигурационного графа

Е. В. Хворостянская

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: `cher@krc.karelia.ru`

Рассматривается конфигурационный граф, содержащий N вершин, занумерованных числами от 1 до N. Степени вершин конфигурационного графа являются независимыми одинаково распределенными случайными величинами. Эти случайные величины задают число выходящих из каждой вершины полуребер, т.е. ребер, для которых еще не определены инцидентные им вторые вершины. После определения степеней вершин полуребра равновероятно соединяются, образуя ребра. В случае, когда сумма степеней всех вершин нечетна, добавляется еще одна вершина степени степени 1, что не влияет на асимптотические свойства графа. Из построения графа следует, что он может содержать петли и кратные ребра.

Под второй степенью произвольной вершины A конфигурационного графа понимается случайная величина $\eta^{(2)}$, равная сумме степеней вершин, смежных с вершиной A, без учета ребер, идущих к A.

Для конфигурационного графа, степени вершин $\eta_1, \ldots, \eta_N$ которого имеют распределение Пуассона с параметром $\lambda > 0$

$$\mathbf{P}\{\eta_i = k\} = \frac{\lambda^k}{k!}e^{-\lambda},\ k = 0, 1, 2, \ldots,\ i = 1, \ldots, N,$$

получено следующее утверждение.

Теорема 1. *При $N \to \infty$ производящая функция $F(z)$ случайной величины $\eta^{(2)}$ равна*

$$F(z) = e^{\lambda\left(e^{-\lambda}-1\right)} + e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}s_k(\lambda)z^k,$$

где

$$s_1(\lambda) = \lambda e^{-\lambda} e^{\lambda e^{-\lambda}}, \quad s_k = \lambda e^{-\lambda} \frac{ds_{k-1}(\lambda)}{d\left(\lambda e^{-\lambda}\right)}, \quad k = 2, 3, \ldots, \tag{1}$$

а предельное распределение случайной величины $\eta^{(2)}$ *задается равенствами*

$$\mathbf{P}\left\{\eta^{(2)} = 0\right\} = e^{\lambda\left(e^{-\lambda}-1\right)}, \quad \mathbf{P}\left\{\eta^{(2)} = k\right\} = \frac{\lambda^k e^{-\lambda}}{k!} s_k(\lambda), \ k \geqslant 1,$$

где $s_k(\lambda)$ *определены в* (1).

В работе [1] изучались конфигурационные графы, в которых степени вершин $\eta_1, \ldots, \eta_N$ имеют распределение

$$\mathbf{P}\left\{\eta_i = k\right\} = p_k > 0, \quad k = 1, 2, \ldots, \quad i = 1, \ldots, N \tag{2}$$

такое, что при $k \to \infty$

$$p_k \sim \frac{d}{k^g (\ln k)^h}, \ h \geqslant 0, \ g \geqslant 1, \ h + g > 1, \ d > 0.$$

Для таких конфигурационных графов справедливо следующее утверждение.

Теорема 2. *При* $N \to \infty$ *для производящей функции* $F(z)$ *второй степени вершины конфигурационного графа с распределением степеней вершин* (2) *при* $g > 7/3$ *выполнено равенство*

$$F(z) = F_1\left(\frac{F_1'(z)}{F_1'(1)}\right),$$

где $F_1(z) = \sum\limits_{k=1}^{\infty} p_k z^k$, $F_1'(z) = \sum\limits_{k=1}^{\infty} k p_k z^{k-1}$.

Список литературы

[1] *Павлов Ю. Л.* Условные конфигурационные графы со случайным параметром степенного распределения степеней // Математический сборник. 2018. **209**(2). С. 120–137.

Оптимальный поиск с вероятностью ошибки распознавания

И. А. Чернов, Е. Е. Ивашко

Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия

E-mail: `chernov@krc.karelia.ru`, `ivashko@krc.karelia.ru`

В докладе обсуждаются различные постановки следующей задачи дискретного ненадежного поиска. Имеется бесконечное число пронумерованных урн, содержащих суммарно конечное число шаров, некоторые из которых помечены. Извлекаемый из урны шар изучается; непомеченные шары распознаются верно и перемещаются следующую по счету урну. Затраты на исследование шара известны и могут зависеть от номера шара и/или урны; в простейшем случае они равны единице. Помеченный шар может быть не опознан и ошибочно перемещен в следующую урну. Вероятность ошибки известна и может зависеть от шара и/или урны. Правильно распознанный помеченный шар откладывается в сторону. Цель — найти все помеченные шары с достаточно высокой вероятностью и с минимальными в среднем затратами, выбирая урну для извлечения очередного шара на основе анализа распределения шаров по урнам. Кроме того, требуется оценить средние затраты на обнаружение заданного количества помеченных шаров.

Перечислим частные случаи задачи:

- Стоимость изучения шара равна единице, помеченный шар единственный, вероятность ошибки постоянна.
- Те же условия, но помеченных шаров некоторое известное число.
- Количество помеченных шаров неизвестно, требуется найти все с вероятностью не ниже заданного порога.
- Вероятность ошибки зависит от урны; в частном случае число урн конечно и в последней ошибка исключена.
- Трудоемкость изучения шара зависит от его номера и от урны, в которой он находится.

Мы доказываем оптимальность стратегии, состоящей в исчерпании непустой урны с минимальным номером, независимо от распределения шаров, для простейшей постановки, а также выявляем условия, при которых эта стратегия остается оптимальной в более сложных вариантах. Так, она оптимальна также для задачи, в которой из выбранной урны шар извлекается неслучайно, а стоимость его изучения возрастает по отношению к номеру шара и убывает по номеру урны не слишком быстро. Оптимальна она и для конечного числа урн при условии, что оно достаточно велико. Показано, что возможны «парадоксальные» результаты, например, оптимальность стратегии «изучать всегда один и тот же шар» при быстром убывании трудоемкости изучения как функции номера урны (однако полная трудоемкость при этом конечна).

Получены аналитические выражения для оптимального среднего числа изученных шаров и, в более общем случае, трудоемкости поиска. Отдельно рассмотрен практически важный случай, когда изначально все шары сосредоточены в первой урне — тогда при оптимальном поиске непусты могут быть лишь первые две урны. Приведены результаты численных экспериментов, в том числе сравнивающие оптимальную стратегию с альтернативными.

Обсуждаются также приложения задачи, прежде всего в области распределенных вычислений.

О необходимых условиях алгебраической аппроксимации случайных величин на конечном множестве

А. Д. Яшунский

Институт прикладной математики им. М. В. Келдыша РАН, Москва, Россия

E-mail: yashunsky@keldysh.ru

В математической кибернетике достаточно давно рассматривается задача выражения дискретных случайных величин в виде алгебраических комбинаций случайных величин с заданными распределениями. Так, например, Р.Л.Схиртладзе показал [1], что, подставляя независимые одинаково распределенные бернуллиевские случайные величины в булевы функции конъюнкции, дизъюнкции и отрицания, можно построить случайную величину, распределение которой сколь угодно близко к любому наперед заданному бернуллиевскому распределению. Настоящая работа посвящена дальнейшему исследованию условий, при которых система алгебраических операций над случайными величинами позволяет аппроксимировать случайную величину с любым распределением.

Обозначим $E_k = \{0, 1, \ldots, k-1\}$. Будем рассматривать случайные величины со значениями в множестве E_k. Распределение такой случайной величины — вектор $\mathbf{p} = (p_0, \ldots, p_{k-1})$, удовлетворяющий условиям $\sum p_i = 1$, $p_i \geqslant 0$ при всех $i = 0, \ldots, k-1$. Всевозможные распределения образуют в $\mathbb{R}^k$ симплекс, будем обозначать его $\mathbf{S}^{(k)}$, а его вершины — $\mathbf{e}^0, \ldots, \mathbf{e}^{k-1}$.

Положим $P_k(n) = \{f(x_1, \ldots, x_n) \colon E_k^n \to E_k\}$ и $P_k = \bigcup\limits_{n=1}^{\infty} P_k(n)$.

Пусть $X_1, \ldots, X_n$ — независимые в совокупности случайные величины со значениями в E_k с распределениями $\mathbf{p}^1, \ldots, \mathbf{p}^n$ соответственно, и пусть $f(x_1, \dots, x_n) \in P_k(n)$. Тогда $Y = f(X_1, \dots, X_n)$ — случайная величина с распределением $\mathbf{q}$, компоненты которого могут быть вычислены следующим образом:

$$\mathbf{q}_i = \sum_{\substack{\sigma_1, \dots, \sigma_n \in E_k: \\ f(\sigma_1, \dots, \sigma_n) = i}} p^1_{\sigma_1} \cdots p^n_{\sigma_n}.$$

Эти равенства определяют вектор-функцию $\hat{f}\colon (\mathbf{S}^{(k)})^n \to \mathbf{S}^{(k)}$, задающую преобразование распределений $\mathbf{q} = \hat{f}(\mathbf{p}^1, \dots, \mathbf{p}^n)$. Для множества $B \subseteq P_k$ положим $\hat{B} = \{\hat{f} \mid f \in B\}$.

Пусть $G \subseteq \mathbf{S}^{(k)}$. Обозначим через $W_B(G)$ наименьшее по включению подмножество множества $\mathbf{S}^{(k)}$, содержащее множество G, замкнутое относительно применения операций из $\hat{B}$ и содержащее все свои предельные точки. Содержательно $W_B(G)$ состоит из всех распределений, которые могут быть сколь угодно точно аппроксимированы путем применения (возможно многократного) функций из множества B к независимым случайным величинам с распределениями из множества G.

Будем говорить, что система функций B *аппроксимационно полна*, если найдется такое множество $G \subset \mathbf{S}^{(k)} \setminus \{\mathbf{e}^0, \dots, \mathbf{e}^{k-1}\}$, $|G| < \infty$, что $W_B(G) = \mathbf{S}^{(k)}$. В данной работе устанавливаются некоторые необходимые условия аппроксимационной полноты, обобщающие найденные ранее условия для систем функций из P_2 (см. [2, 3]).

Далее используются следующие обозначения. Множество функций, выразимых суперпозициями над системой B, обозначается $[B]$ (подробнее см. [4]). Функции из $P_k(1)$, равные константам $0, 1, \dots, k-1$ обозначаются через $c_0, \dots, c_{k-1}$ соответственно. Через S_k обозначается множество функций из $P_k(1)$, являющихся перестановками на множестве E_k. Для $G \subseteq \mathbf{S}^{(k)}$ через $Conv(G)$ обозначается выпуклая оболочка множества G.

Для функции $f(x_1, \dots, x_n) \in P_k$ положим

$$U_i(f) = \{f(\alpha_1, \dots, \alpha_{i-1}, x, \alpha_i, \dots, \alpha_{n-1}) \mid (\alpha_1, \dots, \alpha_{n-1}) \in E_k^{n-1}\}.$$

Очевидно, что $U_i(f) \subseteq P_k(1)$. Отображение $\sigma\colon P_k \to \mathbb{N}$ будем называть *индексным*, если для любой функции $f(x_1, \dots, x_n) \in P_k$ выполнено $\sigma(f) \in \{1, \dots, n\}$, т.е. σ сопоставляет каждой функции какой-то номер ее переменной.

Пусть $B \subseteq P_k$ и σ — некоторое индексное отображение. Положим:

$$\mathfrak{U}_\sigma(B) = \left[\bigcup_{f \in B} U_{\sigma(f)}(f) \right].$$

Имеет место следующая теорема.

Теорема 1. *Пусть $G \subseteq \mathbf{S}^{(k)}$, $B \subseteq P_k$ и σ — некоторое индексное отображение. Положим $H = Conv(\{\widehat{\varphi}(\mathbf{g}) \mid \varphi \in \mathfrak{U}_\sigma(B), \mathbf{g} \in G\})$, тогда $W_B(H) = H$.*

С помощью теоремы 1 доказывается следующее необходимое условие аппроксимационной полноты.

Теорема 2. *Пусть $B \subseteq P_k$ аппроксимационно полна. Тогда для любого индексного отображения σ выполнено $\mathfrak{U}_\sigma(B) \ni c_0, \ldots, c_{k-1}$.*

Из этого общего условия можно выводить более специальные условия, рассматривая различные подмножества $P_k(1)$, замкнутые по суперпозиции. В частности выполнено следующее утверждение.

Следствие 1. *Пусть $B \subseteq P_k$ аппроксимационно полна. Тогда найдется такая функция $f(x_1, \ldots, x_n) \in B$, что $U_i(f) \nsubseteq S_k$ при $i = 1, \ldots, n$.*

Используя установленные условия, можно привести ранее неизвестный пример системы функций из P_2 заведомо не являющейся аппроксимационно полной.

Следствие 2. *Пусть каждая функция $f(x_1, \ldots, x_n) \in B \subset P_2$ представляется в виде $x_i + f'(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \bmod 2$ для некоторого $i = 1, \ldots, n$. Тогда B не является аппроксимационно полной.*

Исследование выполнено при финансовой поддержке РФФИ, проект №18–01–00337.

Список литературы

[1] *Схиртладзе Р. Л.* О методе построения булевой случайной величины с заданным распределением вероятностей // Дискретный анализ. Вып. 7. Сб. науч. тр. Новосибирск: Ин-т матем. СО АН СССР. 1966. С. 71–80.

[2] *Яшунский А. Д.* О преобразованиях вероятности бесповторными булевыми формулами // Мат-лы XVI Международной школы-семинара «Синтез и сложность управляющих систем» (Санкт-Петербург, 26–30 июня 2006 г.) М.: Изд-во механико-математического факультета МГУ. 2006. С. 150–155.

[3] *Yashunsky A. D.* Clone-induced approximation algebras of Bernoulli distributions // Algebra Universalis. 2019. **80**(5). doi:10.1007/s00012-019-0578-4

[4] *Яблонский С. В.* Введение в дискретную математику. М.: Высш. шк., 2001.

Научное издание

X Международная Петрозаводская конференция

ВЕРОЯТНОСТНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

22 — 26 мая 2019 г.
Петрозаводск, Россия

РАСШИРЕННЫЕ ТЕЗИСЫ

Печатается по решению Ученого совета Института прикладных математических исследований Карельского научного центра РАН

Издается в авторской редакции
Оригинал-макет Е. В. Хворостянская

Подписано в печать 06.05.2019.
Формат 60x84 1/16. Бумага офсетная.
Уч.-изд. л. 5. Усл. печ. л. 7,21. Тираж 80 экз. Заказ № 554

Отпечатано с готового оригинал-макета

Федеральный исследовательский центр
«Карельский научный центр Российской академии наук»
Редакционно-издательский отдел
185003, г. Петрозаводск, пр. А. Невского, 50