О Б О З Р Е Н И Е
П Р И К Л А Д Н О Й И П Р О М Ы Ш Л Е Н Н О Й
Т о м 25 М А Т Е М А Т И К И В ы п у с к 3
2018

А. М. З у б к о в, **В. И. К р у г л о в** (Москва, Математический институт им. В. А. Стеклова Российской академии наук)**. Непродолжаемые к корню повторения цепочек на q-ичном дереве со случайными метками вершин.**

Пусть T_q^n — полное q-ичное дерево высоты n, состоящее из корня $*$ и n слоев вершин; q^k элементов множества $I^{(k)}$ вершин k-го ($k = 1, 2, \ldots, n$) слоя можно занумеровать q-ичными наборами $\mathbf{i} = (i_1, i_2, \ldots, i_k) \in \{0, 1, \ldots, q-1\}^k$. На множестве вершин дерева T_q^n зададим естественное лексикографическое упорядочение, считая, что $\mathbf{i} = (i_1, \ldots, i_k) \prec \mathbf{j} = (j_1, \ldots, j_h)$, если либо $\mathbf{i} = *$, $\mathbf{j} \neq *$, либо $1 \leqslant k < h$, либо $1 \leqslant k = h$ и $\sum_{m=1}^{k} i_m q^{k-m} < \sum_{m=1}^{k} j_m q^{k-m}$.

Цепочкой $C_\mathbf{i}$ длины l, начинающейся в вершине $\mathbf{i} = (i_1, i_2, \ldots, i_k) \in I^{(v_\mathbf{i})}, 0 \leqslant v_\mathbf{i} \leqslant n - l + 1$, будем называть любую последовательность l вершин

$$(i_1, i_2, \ldots, i_k), (i_1, i_2, \ldots, i_k, i_{k+1}), \ldots, (i_1, i_2, \ldots, i_k, i_{k+1}, \ldots, i_{k+l-1}).$$

Обозначим эти вершины цепочки $C_\mathbf{i}$ соответственно $C_\mathbf{i}[0], C_\mathbf{i}[1], \ldots, C_\mathbf{i}[l-1]$ и определим отношение порядка на множестве цепочек фиксированной длины l: $C_\mathbf{i} \prec C_\mathbf{j}$ тогда и только тогда, когда $C_\mathbf{i}[l-1] \prec C_\mathbf{j}[l-1]$. Множество упорядоченных пар цепочек $(C_\mathbf{i}, C_\mathbf{j}), \mathbf{i} \prec \mathbf{j}$, длины l обозначим $\mathcal{P}_{n,l}$.

Пусть каждой вершине $\mathbf{i}$ дерева T_q^n присвоена случайная метка $m(\mathbf{i})$, имеющая равновероятное распределение на множестве $\{1, \ldots, d\}$, причем значения $m(\mathbf{i})$, $\mathbf{i} \in T_q^n$, независимы в совокупности. Тогда каждой цепочке $C_\mathbf{i}$ длины l соответствует случайная последовательность меток $M(C_\mathbf{i}) = (m(C_\mathbf{i}[0]), m(C_\mathbf{i}[1]), \ldots, m(C_\mathbf{i}[l-1]))$, и $\mathbf{P}\{M(C_\mathbf{i}) = M(C_\mathbf{j})\} = \frac{1}{d^l}$ для любых двух непересекающихся цепочек $C_\mathbf{i}$ и $C_\mathbf{j}$.

Для каждой отличной от корня дерева вершины $\mathbf{i} = (i_1, \ldots, i_k)$ мы можем рассмотреть предшествующую вершину $\mathbf{i}^- = (i_1, \ldots, i_{k-1})$. Повторения последовательностей меток $M(C_\mathbf{i})$ и $M(C_\mathbf{j})$, для которых $m(\mathbf{i}^-) \neq m(\mathbf{j}^-)$ и, следовательно, $M(C_{\mathbf{i}^-}) \neq M(C_{\mathbf{j}^-})$, будем называть непродолжаемыми к корню. Для пары непересекающихся цепочек $(C_\mathbf{i}, C_\mathbf{j}) \in \mathcal{P}_{n,l}$ положим

$$X_{C_\mathbf{i} C_\mathbf{j}} = \mathbb{I}\{M(C_\mathbf{i}) = M(C_\mathbf{j}), m(\mathbf{i}^-) \neq m(\mathbf{j}^-)\},$$

в случае $\mathbf{i} = *$ будем считать условие $m(\mathbf{i}^-) \neq m(\mathbf{j}^-)$ выполненным; рассмотрим случайную величину

$$V_{n,l} = \sum\nolimits_{(C_\mathbf{i}, C_\mathbf{j}) \in \mathcal{P}_{n,l} \colon C_\mathbf{i} \cap C_\mathbf{j} = \varnothing} X_{C_\mathbf{i} C_\mathbf{j}}.$$

Значение d мы будем считать фиксированным при изменении n и l.

Лемма. *Если* $n, l \to \infty$ *так, что* $n - 2l \to \infty$, *то*

$$\mathbf{E} V_{n,l} = \frac{d-1}{d^{l+1}} \cdot \frac{q^{2n+2}}{2(q-1)^2} (1 + o(1)).$$

Обозначим $\pi_k = \frac{\mathbf{E} V_{n,l}}{k} \mathbf{P}\big\{ \sum_{(C'_\mathbf{i}, C'_\mathbf{j}) \in \mathcal{P}_{n,l}} X_{C'_\mathbf{i} C'_\mathbf{j}} = k \,\big|\, X_{C_\mathbf{i} C_\mathbf{j}} = 1 \big\}, k \in \mathbb{N}$. Сложное распределение Пуассона $CP(\pi)$ можно определить как распределение случайной величины $\Xi_\pi = \sum_{k=1}^{\infty} k\xi_k$, где $\xi_1, \xi_2, \ldots$ независимы и при каждом k случайная величина ξ_k имеет распределение Пуассона с параметром π_k. Несложно показать, что $\mathbf{E}\Xi_\pi = \sum_{k=1}^{\infty} k\pi_k = \mathbf{E}V_{n,l}$. Укажем достаточные условия, при которых расстояние по вариации между распределением $\mathcal{L}(V_{n,l})$ случайной величины $V_{n,l}$ и распределением $CP(\pi)$

$$d_{\mathrm{tv}}(\mathcal{L}(V_{n,l}), CP(\pi)) = \frac{1}{2}\sum_{k=0}^{\infty} |\mathbf{P}\{V_{n,l} = k\} - \mathbf{P}\{\Xi_\pi = k\}|$$

стремится к нулю.

Теорема. *Если* $n, l \to \infty$ *так, что* $n - 2l \to \infty$ *и* $\mathbf{E}V_{n,l} \to \lambda \in (0, \infty)$, *то существуют такие* $\varepsilon(l, n)$, *что* $\varepsilon(l, n) \to 0$ *и*

$$d_{\mathrm{tv}}(\mathcal{L}(V_{n,l}), CP(\pi)) \leqslant \frac{16}{q(q-1)} H_1(\pi) \left(\mathbf{E}V_{n,l}\right)^2 \frac{q^{2l}}{q^n} \left(1 + \varepsilon(l, n)\right)) + \\ + \frac{1}{d^{\,l}} \frac{q^{n+l}}{q-1} + \frac{d-1}{d^{l+1}} \frac{q^{n+2}l}{(q-1)^2} \to 0,$$

где $H_1(\pi) \leqslant \min\left(1, \dfrac{1}{\pi_1}\right) \cdot \exp\left(\sum\limits_{k=1}^{\infty} \pi_k\right)$.

Явные формулы для величин $\varepsilon(l, n)$ здесь не приводятся из-за их громоздкости.

Задачи о повторениях цепочек меток на вершинах двоичного дерева были рассмотрены в [1], в [2] были исследованы аналогичные задачи для повторения меток на пучках цепочек в q-ичных деревьях. В данных тезисах представлено обобщение некоторых результатов работы [1] на случай цепочек знаков на вершинах q-ичных деревьев.

Исследование выполнено за счет гранта Российского научного фонда (проект № 14-50-00005).

СПИСОК ЛИТЕРАТУРЫ

1. *Зубков А. М.*, *Круглов В. И.* Повторения цепочек на бинарном дереве со случайными метками вершин. — Дискретн. матем., 2015, т. 27, в. 4, с. 38–48.
2. *Zubkov A.*, *Kruglov V.* Number of Pairs of Template Matchings in q-ary Tree with Randomly Marked Vertices. — Anal. and Comp. Methods in Prob. Theory, LNCS, 2017, v. 10684, p. 336–346.