ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 28 МАТЕМАТИКИ Выпуск 2
2021

В. Г. Михайлов, **В. И. Круглов** (Москва, МИ РАН им. В.А. Стеклова). **Асимптотическая нормальность числа пар цепочек с одинаковыми метками в q-ичном дереве со случайными метками вершин.**

УДК 519.212.2

Резюме: Пусть вершинам полного q-ичного дерева присвоены независимые случайные метки, являющиеся элементами некоторого конечного алфавита. В докладе рассматриваются совпадения наборов меток, соответствующих парам цепочек вершин этого дерева. Предложены теоремы о нормальной аппроксимации числа пар цепочек с совпавшими метками.

Ключевые слова: деревья с помеченными вершинами, совпадение меток, нормальная аппроксимация.

Пусть $\mathbf{Tr}$ — полное корневое q-ичное дерево высоты, достаточной для всех построений. Нам будет удобно считать, что вершины дерева занумерованы в таком порядке, что более далекие от корня вершины имеют большие номера. Корень ни номера, ни метки не имеет. Остальным вершинам сопоставлены независимые случайные метки $X_1, X_2, \ldots,$ имеющие одинаковое распределение на множестве $\{1, \ldots, d\}$ с вероятностями $\mathbf{P}\{X_i = k\} = p_b > 0,$ где $\sum_{k=1}^{d} p_k = 1.$

Путем длины s в дереве назовем последовательность s смежных ребер, началом пути будем считать ближайшую к корню инцидентную ребрам пути вершину. Каждому пути $w(s)$ из s ребер сопоставим s-цепочку $Y(w(s))$ из меток вершин, через которые он проходит (в эту цепочку не входит метка вершины, из которой начался путь). Пусть $W(H, s)$ — число s-цепочек в дереве $\mathbf{Tr},$ начало которых имеет высоту, не превосходящую заданное число $H \geqslant s.$ Нетрудно показать, что

$$W(H, s) = \frac{q^{H+1} - 1}{q - 1} \cdot q^s \asymp q^{H+s}, \quad H \to \infty.$$

Число всех пар так расположенных s-цепочек с одинаковыми метками в дереве $\mathbf{Tr}$ является случайной величиной и выражается формулой

$$\xi_2(H, s) = \sum_{1 \leqslant u < v \leqslant W(H,s)} I\{Y(u) = Y(v)\},$$

где u, v — номера путей, за которые естественно взять номера последних вершин в пути.

Интерес также представляет аналогично определяемое число $\zeta_2(H, s)$ таких пар s-цепочек с одинаковыми метками, в которых s-цепочки не имеют общих вершин. Формула для $\zeta_2(H, s)$ отличается от формулы для $\xi_2(H, s)$ лишь областью суммирования, и мы ее не приводим.

Теорема 1. *Пусть $H \to \infty,$ параметры $s = s(H)$ и $q = q(H)$ изменяются так, что $s/H \to 0,$ и найдутся числа $C > 0$ и $\varepsilon \in (0, 1]$ такие, что при всех достаточно больших H выполнено неравенство*

$$\mathbf{D}\xi_2(H, s) \geqslant C(W(T, s))^{2+\varepsilon}.$$

Тогда функция распределения и моменты случайной величины

$$\widetilde{\xi}_2(H,2) = \big(\xi_2(H,2) - \mathbf{E}\xi_2(H,2)\big)/\sqrt{\mathbf{D}\xi_2(H,2)}$$

сходятся к функции распределения и моментам стандартного нормального распределения.

Теорема 2. *Пусть $H \to \infty$, параметры $s = s(H)$ и $q = q(H)$ изменяются так, что $s/H \to 0$, и найдутся числа $C > 0$ и $\varepsilon \in (0,1]$ такие, что при всех достаточно больших H выполнено неравенство*

$$\mathbf{D}\zeta_2(H,s) \geqslant C(W(T,s))^{2+\varepsilon}.$$

Тогда функция распределения и моменты случайной величины

$$\tilde{\zeta}_2(H,2) = \big(\zeta_2(H,2) - \mathbf{E}\zeta_2(H,2)\big)/\sqrt{\mathbf{D}\zeta_2(H,2)}$$

сходятся к функции распределения и моментам стандартного нормального распределения.

Доказательства этих теорем используют результаты из [1] и [2]. Другие применения этих результатов см. в [3].

Как оказалось, изучение асимптотических свойств дисперсий величин $\xi_2(H,2)$ и $\zeta_2(H,2)$ требует громоздких вычислений, и в общем случае эту задачу решить пока не удалось. При небольших значениях параметра s асимптотику дисперсий можно найти, рассматривая все возможные случаи взаимного расположения цепочек длины s. Таким способом для $s = 2$ удалось показать, что в случае, когда распределение случайных величин, используемых в качестве меток в вершинах дерева, отличается от равновероятного, $\mathbf{D}\xi_2(H,2) \asymp q^{3H}$ и $\mathbf{D}\zeta_2(H,2) \asymp q^{3H}$ при $H \to \infty$. Поэтому из теорем 1 и 2 следует, что в этих условиях функции распределения и моменты случайных величин $\widetilde{\xi}_2(H,2)$ и $\tilde{\zeta}_2(H,2)$ сходятся к функции распределения и моментам стандартного нормального распределения.

Другие постановки задач о деревьях со случайными метками в вершинах можно найти в [4], [5], [6], [7].

СПИСОК ЛИТЕРАТУРЫ

1. *Janson S.* Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. — Ann. Probab., 1988, v. 16, №1, c. 306–312.
2. *Михайлов В. Г.* Об одной теореме Янсона. — Теория вероятн. и ее примен., 1991, т. 36, в. 1, с. 168–170. *Mikhailov V. G.* On a theorem of Janson. — Theory Probab. Appl., 1991, v. 36, is. 1, p. 173–176.
3. *Mikhailov V. G.*, *Mezhennaya N. M.* Normal aproximation for U - and V-statistics of a stationary absolutely regular sequence. — Siberian Electronic Mathematical Reports, 2020, v. 17, p. 672–682.
4. *Hoffmann C. M.*, *O'Donnell M. J.* Pattern matching in trees. — J. ACM, 1982, v. 29, is. 1, p. 68–95.
5. *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. *Zubkov A. M.*, *Kruglov V. I.* On coincidences of tuples in a binary tree with random labels of vertices. — Discrete Math. Appl., 2016, v. 26, is. 3, p. 145–153.
6. *Kruglov V.*, *Zubkov A.* Number of Pairs of Template Matchings in q-ary Tree with Randomly Marked Vertices. — Analytical and Computational Methods in Probability Theory (LNCS, Springer), 2017, v. 10684, p. 336–346.

7. *Зубков А. М.*, *Круглов В. И.* Непродолжаемые к корню повторения цепочек на q-ичном дереве со случайными метками вершин. — Обозрение прикл. и промышл. матем., 2018, т. 25, в. 3, с. 249–251. *Zubkov A. M.*, *Kruglov V. I.* Non-Extendable to the Root Coincedences of Marks of Chains in q-ary Tree with Randomly Marked Vertices. — Review of Applied and Industrial Mathematics, 2018, v. 25, is. 3, p. 249–251. (In Russian.)

UDC 519.212.2

Mikhailov V. G., ***Kruglov V. I.*** (Moscow, Steklov Mathematical Institute of RAS).
Asymptotical normality of number of pairs of identically labeled chains in q-ary tree with randomly labeled vertices.

Abstract: Let the vertices of a complete q-ary tree be assigned independent random labels of some finite alphabet. We consider pairs of identically labeled chains of vertices and propose theorems for normal approximation for the number of such pairs.

Keywords: complete tree, labeled vertices, pattern matching, normal aproximation.