4. *Caranti A., Volta F., and Sala M.* An application of the O'Nan-Scott theorem to the group generated by the round functions of an AES-like cipher // Des. Codes Cryptogr. 2009. V. 52. P. 293–301.
5. *Caranti A., Volta F., and Sala M.* On some block ciphers and imprimitive groups // Appl. Algebra Eng. Commun. Comput. 2009. V. 20. P. 339–350.
6. *Leander G., Abdelraheem M. A., AlKhzaimi H., and Zenner E.* A cryptanalysis of PRINTcipher: The invariant subspace attack // LNCS. 2011. V. 6841. P. 206–221.
7. *Трифонов Д. И., Фомин Д. Б.* Об инвариантных подпространствах в XSL-шифрах // Прикладная дискретная математика. 2021. № 54. С 58–76.
8. *Todo Y., Leander G., and Sasaki Y.* Nonlinear invariant attack: practical attack on full SCREAM, iSCREAM, and Midori64 // ASIACRYPT 2016. LNCS. 2016. V. 10032. P. 3–33.
9. *Буров Д. А.* О существовании нелинейных инвариантов специального вида для раундовых преобразований XSL-алгоритмов // Дискретная математика. 2021. Т. 33. № 2. С. 31–45.
10. *Mattarei S.* Inverse-closed additive subgroups of fields // Israel J. Math. 2007. V. 159. P. 343–347.
11. *Goldstein D., Guralnick R., Small L., and Zelmanov E.* Inversion-invariant additive subgroups of division rings // Pacific J. Math. 2006. V. 227. P. 287–294.
12. *Nyberg K.* Differentially uniform mappings for cryptography // LNCS. 1994. V. 765. P. 55–64.
13. *Carlet C.* Open questions on nonlinearity and on APN Functions // LNCS. 2015. V. 9061. P. 83–107.
14. *Hua L.-K.* Some properties of a sfield // Proc. NAS USA. 1949. V. 35. P. 533–537.

УДК 519.214 DOI 10.17223/2226308X/15/2

ОБ АСИМПТОТИЧЕСКОЙ НОРМАЛЬНОСТИ ЧИСЛА КРАТНЫХ СОВПАДЕНИЙ ЦЕПОЧЕК В ПОЛНЫХ q-ИЧНЫХ ДЕРЕВЬЯХ И ЛЕСАХ СО СЛУЧАЙНЫМИ МЕТКАМИ

В. Г. Михайлов, В. И. Круглов

Рассматриваются полные q-ичные корневые деревья высоты H, каждой вершине которых независимо от остальных вершин присвоена случайная метка, выбираемая из множества $\{1, 2, \ldots, N\}$. Исследуются случайные величины, равные числу наборов из $r \geqslant 2$ путей одинаковой длины s, у которых совпадают соответствующие s-цепочки меток вершин. Представлена теорема о достаточных условиях асимптотической нормальности рассматриваемых случайных величин при неограниченном увеличении высоты дерева. При исследовании повторений цепочек в лесе деревьев предполагается, что имеется r деревьев, которые могут иметь разные высоты $H_1, \ldots, H_r$ и вершинам которых аналогичным образом поставлены в соответствие независимые в совокупности случайные метки. Изучается число наборов из r путей длины s, в которые входит по одному пути с каждого дерева, для которых совпадают соответствующие цепочки меток вершин, для этой случайной величины также получены достаточные условия асимптотической нормальности.

Ключевые слова: *деревья с метками, цепочки меток на дереве, повторения цепочек, условия асимптотической нормальности.*

Введение

Повторения событий могут свидетельствовать о наличии закономерностей, при анализе которых возникают задачи о вычислении или оценке значений вероятностей повторений событий в наборах независимых случайных величин. В исследованиях по этой тематике первоначально рассматривались задачи о повторениях цепочек случайных символов [1 – 4], естественным продолжением этих исследований стали работы, связанные с повторениями паттернов в деревьях со случайно помеченными вершинами. Распределения числа вхождений заданного поддерева в случайное дерево рассматривались в [5, 6], задачи такого рода возникают в компьютерных науках [7, 8] при анализе алгоритмов или, например, в связи с древовидной структурой XML-документов, которые используются, в частности, на портале Госуслуг. Подобные задачи также могут возникать в связи с построением статистических критериев и анализом генетических последовательностей.

Предельные пуассоновские теоремы для числа совпадений меток цепочек в двоичном или q-ичном дереве, метки вершин которого независимы и имеют равновероятное распределение на конечном алфавите, получены в [9, 10], предельная пуассоновская теорема для числа совпадений паттернов в q-ичном дереве с равновероятными метками вершин доказана в [4].

В настоящей работе рассматриваются полные q-ичные корневые деревья и леса, составленные из таких корневых деревьев. Некорневым вершинам деревьев присвоены случайные метки, выбранные независимо из множества $\{1, 2, \ldots, N\}$ в соответствии с некоторым вероятностным распределением.

Изучается число наборов по $r \geqslant 2$ путей длины s на одном или нескольких деревьях, для которых совпадают соответствующие цепочки меток вершин. Получены достаточные условия асимптотической нормальности этой случайной величины при неограниченном увеличении высоты деревьев.

1. Повторения цепочек на дереве

Пусть $Tr(H)$ — полное q-ичное корневое дерево высоты H и вершинам этого дерева присвоены незавимые в совокупности случайные метки, выбираемые из конечного множества $\{1, 2, \ldots, N\}$ в соответствии с положительными вероятностями $p_1, \ldots, p_N$, где $p_1 + \ldots + p_N = 1$.

Пусть $s < H$. Через $W(H, s)$ будем обозначать множество цепочек длины s в дереве $Tr(H)$, начало которых имеет высоту, не превосходящую $H - s$. Нетрудно показать [11], что

$$|W(H, s)| = \frac{q^{H-s+1} - 1}{q - 1}\, q^s = \frac{q^{H+1} - q^s}{q - 1}.$$

Определим случайную величину $\xi_r(H, s)$, которая равна числу всех таких наборов из r различных путей длины s в дереве $Tr(H)$, для которых совпадают соответствующие s-цепочки меток вершин, составляющих эти пути. Для этого занумеруем элементы множества $W(H, s)$ числами от 1 до $|W(H, s)|$, путь с номером u, где $1 \leqslant u \leqslant |W(H, s)|$, будем обозначать ω_u, а цепочку меток вершин на этом пути — $Y(\omega_u)$. Тогда

$$\xi_r(H, s) = \sum_{1 \leqslant u_1 < \ldots < u_r \leqslant |(W(H,s)|} I\{Y(\omega_{u_1}) = \ldots = Y(\omega_{u_r})\}.$$

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

больших H выполнено неравенство

$$\mathbf{D}\xi_r(H,s) \geqslant C\left(\frac{q^{H+1}-q^s}{q-1}\right)^{2(r-1)+\varepsilon}. \tag{1}$$

Тогда функции распределения и моменты случайной величины

$$\tilde{\xi}_r(H,s) = \frac{\xi_r(H,s) - \mathbf{E}\xi_r(H,s)}{\sqrt{\mathbf{D}\xi_r(H,s)}}$$

сходятся к функции распределения и моментам стандартного нормального распределения.

Доказательство теоремы 1 для случая $r = 2$ опубликовано в [11].

Можно отметить, что при $s = 1$ величина $\xi_r(H,1)$ совпадает по распределению с числом ξ_r наборов по r одинаковых исходов в последовательности из $|W(H,1)| - 1$ независимых случайных величин X_i, принимающих значения на множестве $\{1,\ldots,N\}$ с вероятностями $\mathsf{P}[X_i = k] = p_k > 0$, $k = 1,\ldots,N$, $\sum\limits_{k=1}^{N} p_k = 1$. Свойства распределения величины ξ_r известны: для неё условие (1) выполняется для любых неравновероятных распределений величин X_i и не выполняется, если $p_1 = \ldots = p_N = 1/N$.

2. Повторения цепочек в лесах

Рассмотрим набор из r полных q-ичных корневых деревьев $Tr_1(H_1),\ldots,Tr_r(H_r)$ высот $H_1,\ldots,H_r$ соответственно, и пусть вершинам этих деревьев присвоены независимые в совокупности случайные метки, выбираемые из множества $\{1,2,\ldots,N\}$ в соответствии с положительными вероятностями $p_1,p_2,\ldots,p_N$, где $p_1+p_2+\ldots+p_N = 1$.

Пусть случайная величина $\xi_{(r)}(H_1,\ldots,H_r;s)$ равна числу таких наборов из r путей длины s, что в эти наборы входит по одному пути из каждого из деревьев $Tr_1(H_1),\ldots,Tr_r(H_r)$ и для этих путей совпадают соответствующие s-цепочки меток вершин. Тогда

$$\xi_{(r)}(H_1,\ldots,H_r;s) = \sum_{\omega_{u_1}\in W(H_1,s)} \ldots \sum_{\omega_{u_r}\in W(H_r,s)} I\{Y(\omega_{u_1}) = \ldots = Y(\omega_{u_r})\}.$$

Минимальную высоту деревьев $Tr_1(H_1),\ldots,Tr_r(H_r)$ будем обозначать через $H_{\min} = \min\{H_1,\ldots,H_r\}$.

Для любого $l \in \mathbb{N}$ определим величину $P_l = \sum\limits_{k=1}^{N} p_k^l$, которая равна вероятности того, что l различных вершин, лежащих в одном или нескольких деревьях, имеют одинаковые метки.

Теорема 2. Пусть $H_1,\ldots,H_r \to \infty$ и параметры $s = s(H_1,\ldots,H_r)$ и $q = q(H_1,\ldots,H_r)$ изменяются так, что $s/H_{\min} \to 0$. Пусть существуют такие числа $C > 0$ и $\varepsilon \in (0,1]$, что при всех достаточно больших $H_{\min}$ выполнено неравенство

$$\mathbf{D}\xi_{(r)}(H_1,\ldots,H_r;s) \geqslant Cq^{2(H_1+\ldots+H_r)-(2-\varepsilon)H_{\min}}.$$

Тогда функции распределения и моменты случайной величины

$$\tilde{\xi}_{(r)}(H_1,\ldots,H_r;s) = \frac{\xi_{(r)}(H_1,\ldots,H_r;s) - P_r^s \prod\limits_{k=1}^{r} \dfrac{q^{H_k+1}-q^s}{q-1}}{\sqrt{\mathbf{D}\xi_{(r)}(H_1,\ldots,H_r;s)}}$$

сходятся к функции распределения и моментам стандартного нормального распределения.

Доказательства теорем 1 и 2 основаны на модификации метода Янсона [12], предложенной в работе В. Г. Михайлова [13].

ЛИТЕРАТУРА

1. *Guibas L. J. and Odlyzko A. M.* Long repetitive patterns in random sequences // Z. Wahrscheinlichkeitstheorie verw. Geb. 1980. No. 1. P. 241–262.
2. *Зубков А. М., Михайлов В. Г.* Предельные распределения случайных величин, связанных с длинными повторениями в последовательности независимых испытаний // Теория вероятн. и ее примен. 1974. Т. 19. № 1. С. 173–181.
3. *Михайлов В. Г.* Оценка точности сложной пуассоновской аппроксимации для распределения числа совпадающих цепочек // Теория вероятн. и ее примен. 2001. Т. 46. № 4. С. 713–723.
4. *Kruglov V. and Zubkov A.* Number of pairs of template matchings in q-ary tree with randomly marked vertices // LNCS. 2017. V. 10684. P. 336–346.
5. *Hoffmann C. M. and O'Donnell M. J.* Pattern matching in trees // J. ACM. 1982. V. 29. No. 1. P. 68–95.
6. *Steyaert J.-M. and Flajolet P.* Patterns and pattern-matching in trees: an analysis // Inf. & Control. 1983. V. 58. No. 1. P. 19–58.
7. *Singh G., Smolka S. A., and Ramakrishnan I. V.* Distributed algorithms for tree pattern matching // LNCS. 1988. V. 312. P. 92–107.
8. *Tahraoui M. A., Pinel-Sauvagnat K., Laitang C., et al.* A survey on tree matching and XML retrieval // Computer Science Rev. 2013. No. 8. P. 1–23.
9. *Зубков А. М., Круглов В. И.* Повторения цепочек на бинарном дереве со случайными метками вершин // Дискретная математика. 2015. Т. 27. № 4. С. 38–48.
10. *Kruglov V. I.* On coincidences of tuples in a q-ary tree with random labels of vertices // Discr. Math. Appl. 2018. V. 28. No. 5. P. 293–307.
11. *Михайлов В. Г., Круглов В. И.* Об асимптотической нормальности в задаче о повторениях цепочек в помеченном полном дереве // Матем. вопр. криптогр. 2021. Т. 12. № 4. С. 59–64.
12. *Janson S.* Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs // Ann. Probab. 1988. V. 16. No. 1. P. 306–312.
13. *Михайлов В. Г.* Об одной теореме Янсона // Теория вероятн. и ее примен. 1991. Т. 36. № 1. С. 168–170.

УДК 519.214 DOI 10.17223/2226308X/15/3

О ТОЧНОСТИ НОРМАЛЬНОЙ АППРОКСИМАЦИИ ДЛЯ РАСПРЕДЕЛЕНИЯ ЧИСЛА КРАТНЫХ ПОВТОРЕНИЙ ЗНАКОВ В СТАЦИОНАРНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

В. Г. Михайлов, Н. М. Меженная

Изучается задача об асимптотической нормальности числа r-кратных повторений знаков в отрезке длины n стационарной в узком смысле случайной последовательности со значениями в конечном множестве, удовлетворяющей условию равномерно сильного перемешивания. Показано, что если существует такое число $\alpha > 0$, что коэффициент равномерно сильного перемешивания $\varphi(t)$ убывает