Math-Net.Ru

Общероссийский математический портал

В. А. Ватутин, Асимптотические свойства числа инверсий в раскрашенных деревьях, *Матем. вопр. криптогр.*, 2019, том 10, выпуск 4, 9–24

DOI: <https://doi.org/10.4213/mvk305>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.174.149.18

26 апреля 2023 г., 15:02:49

Асимптотические свойства числа инверсий в раскрашенных деревьях

В. А. Ватутин

*Математический институт им В.А. Стеклова
Российской академии наук, Москва*

Получено 29.IV.2019

Аннотация. Рассматривается b -арное плоское корневое дерево T , вершинам которого независимо и равновероятно присвоены цвета, обозначаемые буквами алфавита $\mathcal{A} = \{A_1 < A_2 < \dots < A_m\}$. Вершина $u \in T$ является предком вершины $v \in T$ ($u \prec v$), если путь, ведущий по ребрам от корня дерева к вершине v , проходит через вершину u . Обозначим $\text{col}(u)$ цвет вершины u . Раскраска пары $u \prec v$ образует инверсию, если $\text{col}(u) > \text{col}(v)$. Исследуются вероятностные характеристики общего числа инверсий в раскрашенном b -арном плоском корневом дереве фиксированной высоты и распределения случайных величин, являющихся функционалами от числа инверсий в поддеревьях такого дерева.

Ключевые слова: b -арное корневое дерево, раскрашенное дерево, инверсия, предельные теоремы

Asymptotic properties of the inversion number in colored trees

V. A. Vatutin

Steklov Mathematical Institute of Russian Academy of Sciences, Moscow

Abstract. We consider a b -ary plane rooted tree T whose vertices are colored independently and equiprobably in m colors labelled with letters of an alphabet $\mathcal{A} = \{A_1 < A_2 < \dots < A_m\}$. A vertex $u \in T$ is an ancestor of a vertex $v \in T$ ($u \prec v$), if the path leading along the edges from the root of the tree to the vertex v passes through the vertex u . Denote $\text{col}(u)$ the color of the vertex u . The coloring of the pair $u \prec v$ forms an inversion if $\text{col}(u) > \text{col}(v)$. We study the probabilistic characteristics of the total number of inversions in a colored b -ary plane rooted tree of a fixed height and the distribution of random variables that are functionals of the number of inversions in the subtrees of such a tree.

Keywords: b -ary plane rooted tree, colored tree, inversion, limit theorems

1. Введение

В данной работе исследуются свойства некоторых случайных величин, определенных на пространстве раскрашенных плоских корневых деревьев. Напомним в связи с этим некоторые определения, характеризующие такие деревья.

Пусть T – плоское корневое дерево с конечным числом вершин и ребрами, направленными от корня.

Будем говорить, что вершина $u \in T$ является предком вершины $v \in T$, а вершина v является потомком вершины u , и писать $u \prec v$, если путь, ведущий по ребрам от корня дерева к вершине v , проходит через вершину u .

Высотой $h(u)$ вершины $u \in T$ будем называть число ребер на пути от корня к вершине u . Корень (по определению) имеет нулевую высоту. Через $|T|$ будем обозначать число вершин в дереве T .

Если $u \in T$, то $T(u)$ – корневое дерево с корнем u , вершинами, являющимися потомками вершины u , и ребрами, соединяющими эти вершины. Высотой дерева $T(u)$ будем называть величину

$$h(T(u)) = \max_{u \prec v \in T(u)} h(v).$$

Пусть имеется набор из m цветов, занумерованных буквами некоторого алфавита $\mathcal{A} = \{A_1 < A_2 < \dots < A_m\}$. Окрасим каждую вершину дерева T в один из упомянутых цветов, выбирая цвета случайно, независимо и равновероятно из множества $\mathcal{A}$. Обозначим $\text{col}(u)$ цвет вершины u . Скажем, что раскраска пары $u \prec v$ образует инверсию, если $\text{col}(u) > \text{col}(v)$.

Числом инверсий вершины $u \in T$ будем называть величину

$$Z(u) = Z(u, m) := \sum_{v \in T: u \prec v} I \{ \text{col}(u) > \text{col}(v) \},$$

где $I\{B\}$ – индикатор события B . Числом инверсий дерева T будем называть величину

$$\mathcal{I}(T, m) = \sum_{u, v \in T: u \prec v} I \{ \text{col}(u) > \text{col}(v) \} = \sum_{u \in T} Z(u).$$

Заметим, что если дерево – это отрезок длины n с ребрами единичной длины, то $\mathcal{I}(T, m)$ – это число инверсий в последовательности, порожденной последовательностью букв алфавита $\mathcal{A}$. Распределение числа инверсий в таких последовательностях изучались, например, в [1, 2].

Наша цель – исследовать свойства случайной величины $\mathcal{I}(T, m)$, численной при случайной раскраске b -арного дерева T , т. е. для дерева с конечным числом вершин, из каждой неконцевой вершины которого выходит ровно b ребер. Эта случайная величина может использоваться при анализе свойств алгоритмов поиска данных.

2. b -арные деревья

В последующем корневое дерево, из каждой неконцевой вершины которого выходит ровно $b \geq 2$ ребер и каждая концевая вершина которого имеет высоту d , будем кратко называть (d, b) -деревом. Ясно, что (d, b) -дерево имеет

$$N = N(d, b) := 1 + b + b^2 + \dots + b^d = \frac{b^{d+1} - 1}{b - 1} \quad (1)$$

вершин. В дальнейшем для краткости будем обозначать (d, b) -дерево с N вершинами символом T_N .

Асимптотическое поведение математического ожидания и дисперсии случайной величины $\mathcal{I}(T_N; m)$ описано в следующей теореме.

Теорема 1. Если b фиксировано, а $d \rightarrow \infty$, то

$$\mathbf{E}\mathcal{I}(T_N; m) \sim \frac{1}{2} \left(1 - \frac{1}{m}\right) N \log_b N,$$

$$\mathbf{D}\mathcal{I}(T_N; m) \sim \frac{1}{12} \left(1 - \frac{1}{m^2}\right) \frac{b}{b-1} N^2.$$

Теорема 1 и неравенство Чебышева влекут следующее утверждение.

Следствие 1. При $N \rightarrow \infty$

$$\frac{\mathcal{I}(T_N; m)}{\mathbf{E}\mathcal{I}(T_N; m)} \xrightarrow{p} 1,$$

где символ $\xrightarrow{p}$ обозначает сходимость по вероятности. Более того,

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbf{P} \left(\frac{|\mathcal{I}(T_N; m) - \mathbf{E}\mathcal{I}(T_N; m)|}{\sqrt{N}} > M \right) = 0.$$

Перейдем теперь к описанию свойств случайных величин $Z(u)$, $u \in T$. Если высота вершины u в (d, b) -дереве T_N равна $d - r$: $h(u) =$

$d - r$, то (r, b) -дерево $T(u)$ с корнем u имеет

$$n(r) := b \frac{b^r - 1}{b - 1} \quad (2)$$

некорневых вершин. Ясно, что распределение случайной величины $Z(u)$ зависит лишь от $h(T(u))$, но не от расстояния вершины u от корня.

Теорема 2. Если $m \geq 2$ фиксировано, а $h(T(u)) = r$, то при $r \rightarrow \infty$ распределение случайной величины $Z(u)/n(r)$ слабо сходится к равновероятному распределению на множестве $0, 1/m, 2/m, \dots, 1 - 1/m$. Если же $\min(m, r) \rightarrow \infty$, то распределение случайной величины $Z(u)/n(r)$ слабо сходится к равномерному распределению на отрезке $[0, 1]$.

Опишем теперь распределение случайной величины

$$\zeta(b, N, r) = \max_{u \in T_N: h(T(u))=r} Z(u),$$

равной максимуму числа инверсий среди всех вершин b -арного дерева T_N , находящихся на расстоянии $d - r$ от корня.

Теорема 3. Если целые числа $m \geq 2$ и $k \geq 0$ фиксированы, а параметры $r, d \rightarrow \infty$ так, что

$$n(d - r) \frac{n^k(r)}{k!} \left(1 - \frac{1}{m}\right)^{n(r)-k} \frac{1}{m^{k+1}} \rightarrow \gamma \in (0, \infty), \quad (3)$$

то

$$\begin{aligned} \lim \mathbf{P}(n(r) - \zeta(b, N, r) = k + 1) &= 1 - e^{-\gamma}, \\ \lim \mathbf{P}(n(r) - \zeta(b, N, r) = k) &= e^{-\gamma}. \end{aligned}$$

Позднее дополним теорему 3, рассмотрев случай $k/n(r) \rightarrow y \in (1 - m^{-1}, 1)$ (см. ниже теорему 5).

3. Доказательство теоремы 1

Введем для краткости обозначение

$$I(u, v) = I\{\text{col}(u) > \text{col}(v)\}$$

и займемся вычислением математического ожидания и дисперсии случайной величины $\mathcal{I}(T_N; m)$. Если $u \prec v$, то

$$\begin{aligned} \mathbf{E}I(u, v) &= \sum_{i=1}^m \mathbf{P}(\text{col}(u) > \text{col}(v) | \text{col}(u) = A_i) \mathbf{P}(\text{col}(u) = A_i) \\ &= \frac{1}{m} \sum_{i=1}^m \mathbf{P}(A_i > \text{col}(v)) = \frac{1}{m} \sum_{i=1}^m \frac{i-1}{m} = \frac{1}{2} \left(1 - \frac{1}{m}\right). \end{aligned}$$

Поскольку

$$\mathcal{I}(T_N; m) = \sum_{u, v \in T_N: u \prec v} I\{\text{col}(u) > \text{col}(v)\},$$

то

$$\mathbf{E}\mathcal{I}(T_N; m) = \sum_{u, v \in T_N: u \prec v} \mathbf{E}I(u, v) = \frac{1}{2} \left(1 - \frac{1}{m}\right) Y(T_N),$$

где

$$\begin{aligned} Y(T_N) &= \sum_{k=1}^d \sum_{v: h(v)=k} \sum_{u \prec v} 1 \\ &= \sum_{k=1}^d k \sum_{v: h(v)=k} 1 = \sum_{k=1}^d k b^k = \frac{b(db^{d+1} - (d+1)b^d + 1)}{(b-1)^2}. \end{aligned}$$

Отсюда, используя обозначение (1), заключаем, что при фиксированном b и $d \rightarrow \infty$

$$Y(T_N) \sim \frac{db^{d+1}(b-1)}{(b-1)^2} \sim N \log_b N \quad (4)$$

и, следовательно,

$$\mathbf{E}\mathcal{I}(T_N; m) \sim \frac{1}{2} \left(1 - \frac{1}{m}\right) N \log_b N.$$

Оценим теперь дисперсию случайной величины $\mathcal{I}(T_N; m)$, используя представление

$$\mathbf{D}\mathcal{I}(T_N; m) = \sum_{u \prec v} \mathbf{D}I(u, v) + \sum_{\substack{u_1 \prec v_1, u_2 \prec v_2, \\ (u_1, v_1) \neq (u_2, v_2)}} \text{cov}(I(u_1, v_1), I(u_2, v_2)). \quad (5)$$

Ясно, что

$$\begin{aligned} S_1 &:= \sum_{u,v \in T_N: u \prec v} \mathbf{D}I(u, v) \\ &= \frac{1}{4} \left(1 - \frac{1}{m^2}\right) \sum_{u,v \in T_N: u \prec v} 1 = \frac{1}{4} \left(1 - \frac{1}{m^2}\right) Y(T_N). \end{aligned} \quad (6)$$

Запишем теперь второе слагаемое в (5) в виде суммы

$$\sum_{\substack{u_1 \prec v_1, u_2 \prec v_2, \\ (u_1, v_1) \neq (u_2, v_2)}} \text{cov}(I(u_1, v_1), I(u_2, v_2)) = S_2 + S_3 + S_4 + S_5,$$

в которой

$$\begin{aligned} S_2 &:= \sum_{v_1 \neq v_2, u \prec v_1, u \prec v_2} \text{cov}(I(u, v_1)I(u, v_2)), \\ S_3 &:= \sum_{u_1 \neq u_2, u_1 \prec v, u_2 \prec v} \text{cov}(I(u_1, v)I(u_2, v)), \\ S_4 &:= 2 \sum_{u_1 \prec v_1 = u_2 \prec v_2} \text{cov}(I(u_1, v_1)I(u_2, v_2)), \\ S_5 &:= \sum_{u_1 \prec v_1, u_2 \prec v_2, u_1, u_2, v_1, v_2 \text{ различны}} \text{cov}(I(u_1, v_1)I(u_2, v_2)). \end{aligned}$$

Для $v_1 \neq v_2, u \prec v_1, u \prec v_2$ имеем

$$\begin{aligned} &\mathbf{E}I(u, v_1)I(u, v_2) \\ &= \sum_{i=1}^m \mathbf{P}(\text{col}(u) > \text{col}(v_1), \text{col}(u) > \text{col}(v_2) | \text{col}(u) = A_i) \mathbf{P}(\text{col}(u) = A_i) \\ &= \frac{1}{m} \sum_{i=1}^m \mathbf{P}(A_i > \text{col}(v_1), A_i > \text{col}(v_2)) \\ &= \frac{1}{m} \sum_{i=1}^m \frac{(i-1)^2}{m^2} = \frac{1}{3} \left(1 - \frac{1}{m}\right) \left(1 - \frac{1}{2m}\right). \end{aligned}$$

Следовательно,

$$S_2 = \left(\frac{1}{3} \left(1 - \frac{1}{m}\right) \left(1 - \frac{1}{2m}\right) - \frac{1}{4} \left(1 - \frac{1}{m}\right)^2 \right) G_2 = \frac{1}{12} \left(1 - \frac{1}{m^2}\right) G_2,$$

где

$$G_2 := \sum_{u, v_1 \neq v_2: u \prec v_1, u \prec v_2} 1.$$

Поскольку (b, r) -дерево имеет $n(r)$ некорневых вершин (см. (2)), то

$$G_2 = \sum_{k=1}^d \sum_{u: h(u)=d-k} n(k)(n(k)-1) = \sum_{k=1}^d n(k)(n(k)-1)b^{d-k}.$$

Таким образом,

$$S_2 = \frac{1}{12} \left(1 - \frac{1}{m^2}\right) \sum_{k=1}^d n(k)(n(k)-1)b^{d-k}. \quad (7)$$

Далее,

$$|S_3| + |S_4| \leq G_3 + 2G_4,$$

где

$$G_3 := \sum_{u_1 \neq u_2, u_1 \prec v, u_2 \prec v} 1, \quad G_4 := \sum_{u_1 \prec v_1 = u_2 \prec v_2} 1 = G_3.$$

Очевидные преобразования показывают, что

$$G_3 = 2 \sum_{k=2}^d \sum_{v: h(v)=k} \sum_{u_1 \prec u_2 \prec v} 1 = \sum_{k=2}^d \sum_{v: h(v)=k} k(k-1) = \sum_{k=2}^d k(k-1)b^k.$$

Таким образом,

$$|S_3| + |S_4| \leq 3 \sum_{k=2}^d k(k-1)b^k. \quad (8)$$

Наконец,

$$S_5 = 0 \quad (9)$$

в силу независимости процедуры окраски различных вершин.

Из соотношений (5)–(9) вытекает, что

$$\mathbf{DI}(T_N; m) = \frac{1}{4} \left(1 - \frac{1}{m^2}\right) Y(T_N) + \frac{1}{12} \left(1 - \frac{1}{m^2}\right) G_2 + O(G_3).$$

Отталкиваясь от этого представления, исследуем асимптотику дисперсии случайной величины $\mathcal{I}(T_N; m)$ при фиксированном b и $N \rightarrow \infty$.

Имеем

$$\begin{aligned}
 G_2 &= \sum_{k=1}^d b \frac{b^k - 1}{b - 1} \left(b \frac{b^k - 1}{b - 1} - 1 \right) b^{d-k} \\
 &= \frac{b}{(b - 1)^2} \sum_{k=1}^d (b^d - b^{d-k}) (b^{k+1} - 2b + 1) \\
 &= \frac{b^{d+1}}{(b - 1)^2} \sum_{k=1}^d (b^{k+1} - 2b + 1) - \frac{b}{(b - 1)^2} \sum_{k=1}^d (b^{d+1} - 2b^{d-k+1} + b^{d-k}).
 \end{aligned}$$

Поскольку при $d \rightarrow \infty$

$$\sum_{k=1}^d (b^{k+1} - 2b + 1) = b^2 \frac{b^d - 1}{b - 1} - 2db + d \sim bN$$

и

$$\begin{aligned}
 \sum_{k=1}^d (b^{d+1} - 2b^{d-k+1} + b^{d-k}) &= db^{d+1} - (2b - 1)b \frac{b^d - 1}{b - 1} \\
 &\sim bN \log_b N - (2b - 1)N \sim bN \log_b N,
 \end{aligned}$$

то

$$G_2 \sim \frac{b^{d+1}}{(b - 1)^2} bN - \frac{b}{(b - 1)^2} (b - 1)N \log_b N \sim \frac{b}{b - 1} N^2.$$

Далее, при $d \rightarrow \infty$

$$\begin{aligned}
 G_3 &= \sum_{k=2}^d k(k - 1)b^k = b^2 \frac{d^2}{dx^2} \left(\frac{x^{d+1} - 1}{x - 1} \right)_{x=b} \\
 &= b^2 \frac{d(d + 1)b^{d-1}(b - 1)^2 - 2(db^{d+1} - (d + 1)b^d + 1)}{(b - 1)^3} \\
 &\sim \frac{d(d + 1)b^{d+1}}{b - 1} \sim N \log_b^2 N = o(G_2).
 \end{aligned}$$

Объединяя полученные оценки с (4), заключаем, что

$$\mathbf{DI}(T_N; m) \sim \frac{1}{12} \left(1 - \frac{1}{m^2} \right) G_2 \sim \frac{1}{12} \left(1 - \frac{1}{m^2} \right) \frac{b}{b - 1} N^2.$$

Теорема 1 доказана.

4. Свойства случайной величины $Z(u)$

Рассмотрим теперь распределение числа инверсий

$$Z(u) = \sum_{v \in T_N: u \prec v} I(u, v),$$

порождаемых вершиной u b -арного дерева T_N , имеющего высоту d . Формула полной вероятности дает

$$\begin{aligned} \mathbf{P}(Z(u) = k) &= \sum_{i=1}^m \mathbf{P}(Z(u) = k, \text{col}(u) = A_i) \\ &= \frac{1}{m} \sum_{i=1}^m \mathbf{P}(Z(u) = k | \text{col}(u) = A_i). \end{aligned}$$

Всюду ниже мы рассматриваем лишь такие вершины u (d, b) -арного дерева T_N , высота $h(u)$ каждой из которых равна $d - r$, что, в свою очередь, означает, что высота $h(T(u))$ b -арного дерева $T(u)$ с корнем в вершине u равна r .

Так как распределение случайной величины $Z(u)$ зависит лишь от $h(T(u))$, но не от высоты $h(u)$ вершины u , а событие $\{Z(u) = k, \text{col}(u) = A_i\}$ означает, что в раскрашенном b -арном дереве с корнем u ровно k некорневых вершин покрашены в цвета $A_j < A_i$, то

$$\mathbf{P}(Z(u) = k | \text{col}(u) = A_i) = C_{n(r)}^k \left(\frac{i-1}{m} \right)^k \left(\frac{m-i+1}{m} \right)^{n(r)-k}$$

и, таким образом,

$$\mathbf{P}(Z(u) = k) = \frac{1}{m} \sum_{i=1}^m C_{n(r)}^k \left(\frac{i-1}{m} \right)^k \left(\frac{m-i+1}{m} \right)^{n(r)-k}, \quad (10)$$

что приводит к соотношению

$$\begin{aligned} F_r(s) &:= \mathbf{E}[s^{Z(u)}] \\ &= \frac{1}{m} \sum_{k=0}^{n(r)} \sum_{i=1}^m C_{n(r)}^k \left(\frac{i-1}{m} \right)^k \left(\frac{m-i+1}{m} \right)^{n(r)-k} s^k \\ &= \frac{1}{m} \sum_{i=1}^m \sum_{k=0}^{n(r)} C_{n(r)}^k \left(\frac{i-1}{m} \right)^k \left(\frac{m-i+1}{m} \right)^{n(r)-k} s^k \\ &= \frac{1}{m} \sum_{i=1}^m \left(\frac{m-i+1}{m} + \frac{i-1}{m} s \right)^{n(r)} = \frac{1}{m} \sum_{j=0}^{m-1} \left(1 - (1-s) \frac{j}{m} \right)^{n(r)}. \end{aligned}$$

Теперь все готово для доказательства теоремы 2.

Доказательство теоремы 2.

Положим

$$1 - s = \frac{\lambda}{n(r)}, \quad \lambda > 0.$$

Поскольку

$$e^{-(1+\varepsilon)x} \leq e^{-\frac{x}{1-\varepsilon}} \leq 1 - x \leq e^{-x} \quad (11)$$

при $x \in (0, \frac{\varepsilon}{1+\varepsilon})$, то для достаточно больших значений $n(r)$ имеем

$$e^{-(1+\varepsilon)\frac{\lambda}{n(r)}\frac{j}{m}} \leq 1 - \frac{\lambda}{n(r)}\frac{j}{m} \leq e^{-\frac{\lambda}{n(r)}\frac{j}{m}}.$$

Таким образом,

$$\frac{1}{m} \sum_{j=0}^{m-1} \left(1 - \frac{\lambda}{n(r)}\frac{j}{m}\right)^{n(r)} \leq \frac{1}{m} \sum_{j=0}^{m-1} e^{-\lambda\frac{j}{m}} = \frac{1}{m} \frac{1 - e^{-\lambda}}{1 - e^{-\lambda/m}} \quad (12)$$

и

$$\frac{1}{m} \sum_{j=0}^{m-1} \left(1 - \frac{\lambda}{n(r)}\frac{j}{m}\right)^{n(r)} \geq \frac{1}{m} \frac{1 - e^{-(1+\varepsilon)\lambda}}{1 - e^{-(1+\varepsilon)\lambda/m}}. \quad (13)$$

Заметим теперь, что в силу (11) для достаточно малых значений $x > 0$ справедливы оценки

$$(1 - \varepsilon)x \leq 1 - e^{-x} \leq x.$$

Отсюда и из произвольности выбора $\varepsilon > 0$ следует, что если m фиксировано, то

$$\begin{aligned} \lim_{r \rightarrow \infty} \mathbf{E} [e^{-\lambda Z(u)/n(r)}] &= \lim_{r \rightarrow \infty} F_r(e^{-\lambda/n(r)}) \\ &= \frac{1}{m} \sum_{j=0}^{m-1} \lim_{n(r) \rightarrow \infty} \left(1 - (1 - e^{-\lambda/n(r)})\frac{j}{m}\right)^{n(r)} \\ &= \frac{1}{m} \sum_{j=0}^{m-1} \lim_{n(r) \rightarrow \infty} \left(1 - \frac{\lambda j}{n(r)m}\right)^{n(r)} = \frac{1}{m} \sum_{j=0}^{m-1} e^{-j/m}, \end{aligned}$$

что доказывает первое утверждение теоремы 2, поскольку правая часть этого соотношения является преобразованием Лапласа равновероятного распределения на множестве $\{0, 1/m, 2/m, \dots, 1 - 1/m\}$.

Если же $m \rightarrow \infty$, то для достаточно больших значений m

$$\frac{1}{m} \frac{1 - e^{-\lambda}}{1 - e^{-\lambda/m}} \leq \frac{1 + \varepsilon}{1 - \varepsilon} \frac{1}{\lambda} (1 - e^{-\lambda})$$

и

$$\frac{1}{m} \frac{1 - e^{-(1+\varepsilon)\lambda}}{1 - e^{-(1+\varepsilon)\lambda/m}} \geq \frac{1 - \varepsilon}{1 + \varepsilon} \frac{1}{\lambda} (1 - e^{-(1+\varepsilon)\lambda}).$$

Отсюда и из (12)–(13) следует, что

$$\lim_{m, r \rightarrow \infty} \mathbf{E} [e^{-\lambda Z(r)/n(r)}] = \frac{1 - e^{-\lambda}}{\lambda} = \int_0^1 e^{-\lambda x} dx,$$

Это доказывает второе утверждение теоремы 2. поскольку интеграл в последнем соотношении является преобразованием Лапласа равномерного распределения на отрезке $[0, 1]$.

Доказательство теоремы 3. Обозначим $X_{n,p}$ случайную величину, имеющую биномиальное распределение с параметрами (n, p) :

$$\mathbf{P}(X_{n,p} \geq n - k) = \sum_{0 \leq j \leq k} C_n^{n-j} p^{n-j} (1-p)^j = \sum_{0 \leq j \leq k} \frac{n^{[j]}}{j!} p^{n-j} (1-p)^j,$$

где $n^{[j]} = n(n-1) \cdots (n-j+1)$. В силу (10) и предположения $h(u) = d-r$ имеем

$$\begin{aligned} \mathbf{P}(\zeta(b, N, r) \leq k) &= \mathbf{P}\left(\max_{u \in T_N: h(T(u))=r} Z(u) < n(r) - k\right) \\ &= \mathbf{P}^{n(d-r)}(Z(u) < n(r) - k) = (1 - \mathbf{P}(Z(u) \geq n(r) - k))^{n(d-r)} \\ &= \left(1 - \frac{1}{m} \sum_{i=1}^{m-1} \mathbf{P}(X_{n(r), i/m} \geq n(r) - k)\right)^{n(d-r)}. \end{aligned}$$

Согласно закону больших чисел для фиксированного k и любого $i \in \{1, 2, \dots, m-1\}$

$$\mathbf{P}(X_{n(r), i/m} \geq n(r) - k) \rightarrow 0$$

при $n(r) \rightarrow \infty$. Кроме того, ввиду равенства

$$\mathbf{P}(X_{n(r), i/m} \geq n(r) - k) = \sum_{0 < j \leq k} \frac{(n(r))^{[j]}}{j!} \left(\frac{i}{m}\right)^{n(r)-j} \left(1 - \frac{i}{m}\right)^j$$

оценка

$$\mathbf{P} \left(X_{n(r),i/m} \geq n(r) - k \right) = o \left(\mathbf{P} \left(X_{n(r),(m-1)/m} \geq n(r) - k \right) \right)$$

справедлива для всех $1 \leq i \leq m - 2$. Поэтому в условиях теоремы 3

$$\begin{aligned} & \mathbf{P} (\zeta(b, N, r) < n(r) - k) \\ &= \exp \left\{ -\frac{1+o(1)}{m} n(d-r) \mathbf{P} \left(X_{n(r),i/m} \geq n(r) - k \right) \right\} \\ &= \exp \left\{ -\frac{1+o(1)}{m} n(d-r) \sum_{j=0}^k \frac{(n(r))^{[j]}}{j!} \left(1 - \frac{1}{m}\right)^{n(r)-j} \frac{1}{m^j} \right\} \\ &= \exp \left\{ -\frac{1+o(1)}{m} n(d-r) \frac{n^k(r)}{k!} \left(1 - \frac{1}{m}\right)^{n(r)-k} \frac{1}{m^k} \right\} \rightarrow e^{-\gamma}, \end{aligned}$$

в то время как

$$\begin{aligned} & \mathbf{P} (\zeta(b, N, r) < n(r) - k + 1) \\ &= \exp \left\{ -\frac{1+o(1)}{m} n(d-r) \sum_{j=0}^{k-1} \frac{(n(r))^{[j]}}{j!} \left(1 - \frac{1}{m}\right)^{n(r)-j} \frac{1}{m^j} \right\} \rightarrow 1 \end{aligned}$$

в силу оценки

$$\begin{aligned} & n(d-r) \sum_{j=0}^{k-1} \frac{(n(r))^{[j]}}{j!} \left(1 - \frac{1}{m}\right)^{n(r)-j} \frac{1}{m^j} \\ & \leq e^{1/m} n(d-r) (n(r))^{[k-1]} \left(1 - \frac{1}{m}\right)^{n(r)-k} \\ & = o \left(n(d-r) n^k(r) \left(1 - \frac{1}{m}\right)^{n(r)-k} \right) = o(1), \end{aligned}$$

следующей из (3) (здесь по определению $\sum_{j=0}^{-1} \cdot = 0$), и асимптотических соотношений

$$\begin{aligned} & \mathbf{P} (\zeta(b, N, r) < n(r) - k - 1) \\ &= \exp \left\{ -\frac{1+o(1)}{m} n(d-r) \sum_{j=0}^{k+1} \frac{(n(r))^{[j]}}{j!} \left(1 - \frac{1}{m}\right)^{n(r)-j} \frac{1}{m^j} \right\} \\ &= \exp \left\{ -\frac{1+o(1)}{m} n(d-r) \frac{(n(r))^{[k+1]}}{(k+1)!} \left(1 - \frac{1}{m}\right)^{n(r)-k-1} \frac{1}{m^{k+1}} \right\} \rightarrow 0, \end{aligned}$$

вытекающих из условия $n(r) \rightarrow \infty$ при $r \rightarrow \infty$ и неравенств

$$\begin{aligned} n(d-r) \frac{(n(r))^{[k+1]}}{(k+1)!} \left(1 - \frac{1}{m}\right)^{n(r)-k-1} \frac{1}{m^k} \\ \geq \frac{(n(r)-k)}{(k+1)} \times n(d-r) \frac{(n(r))^{[k]}}{k!} \left(1 - \frac{1}{m}\right)^{n(r)-k} \frac{1}{m^{k+1}} \\ \geq \frac{(n(r)-k)}{(k+1)} \frac{\gamma}{2}, \end{aligned}$$

справедливых для достаточно больших значений d и r .

Теорема 3 доказана.

Для более детального изучения распределения случайной величины $\zeta(b, N, r)$ нам понадобится следующий результат, установленный в работе [3]. Пусть

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

– функция распределения стандартного нормального закона, а функция $H(x, p)$ для параметров x и p , принадлежащих интервалу $(0, 1)$, задана соотношением

$$H(x, p) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}.$$

Положим $C_{n,p}(0) = (1-p)^n$, $C_{n,p}(n) = 1 - p^n$,

$$C_{n,p}(k) = \Phi \left(\operatorname{sgn} \left(\frac{k}{n} - p \right) \sqrt{2nH \left(\frac{k}{n}, p \right)} \right), 1 \leq k \leq n-1,$$

где $\operatorname{sgn}(x) = x/|x|$ при $x \neq 0$ и $\operatorname{sgn}(0) = 0$.

Теорема 4 (см. [3]). *При любом $k = 0, 1, \dots, n-1$*

$$C_{n,p}(k) \leq \mathbf{P}(X_{n,p} \leq k) \leq C_{n,p}(k+1). \quad (14)$$

Опираясь на это утверждение, докажем следующую теорему.

Теорема 5. *Если параметры b и N меняются при $r \rightarrow \infty$ так, что*

$$\frac{k}{n(r)} \rightarrow y \in \left(\frac{m-1}{m}, 1 \right) \quad (15)$$

u

$$\frac{n(d-r)}{m\sqrt{2n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)}\sqrt{2\pi}} e^{-n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)} \rightarrow g(y, m)$$

для некоторой функции $g(y, m)$, то при указанном изменении параметров

$$\begin{aligned} \frac{y}{(1-y)(m-1)} e^{-g(y, m)} &\leq \liminf \mathbf{P} \left(\frac{\zeta(b, N, r)}{n(r)} \leq y \right) \\ &\leq \limsup \mathbf{P} \left(\frac{\zeta(b, N, r)}{n(r)} \leq y \right) \leq e^{-g(y, m)}. \end{aligned}$$

Доказательство теоремы 5. Поскольку $h(u) = d - r$, то

$$\mathbf{P} \left(\max_{u \in T_N: h(T(u))=r} Z(u) \leq k \right) = \mathbf{P}^{n(d-r)} (Z(u) \leq k).$$

В силу асимптотического представления

$$\bar{\Phi}(x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad (16)$$

справедливого при $x \rightarrow \infty$, и правого из неравенств в (14) для

$$k \geq \frac{(m-1)n(r)}{m} \quad (17)$$

имеем

$$\begin{aligned} \mathbf{P} \left(\frac{\zeta(b, N, r)}{n(r)} \leq \frac{k}{n(r)} \right) &\leq \left(\frac{1}{m} \sum_{i=1}^m \Phi \left(\sqrt{2n(r)H\left(\frac{k}{n(r)}, \frac{i}{m}\right)} \right) \right)^{n(d-r)} \\ &\leq \left(1 - \frac{1}{m} \bar{\Phi} \left(\sqrt{2n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)} \right) \right)^{n(d-r)} \\ &= \left(1 - \frac{1 + \varepsilon_1(k, r)}{m\sqrt{2n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)}\sqrt{2\pi}} e^{-n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)} \right)^{n(d-r)} \\ &= \exp \left\{ -\frac{(1 + \varepsilon_2(k, r))n(d-r)}{m\sqrt{2n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)}\sqrt{2\pi}} e^{-n(r)H\left(\frac{k}{n(r)}, \frac{m-1}{m}\right)} \right\}, \end{aligned}$$

где $|\varepsilon_1(k, r)| + |\varepsilon_2(k, r)| \rightarrow 0$ при $r \rightarrow \infty$ равномерно по значениям k , удовлетворяющих условию (17).

Таким образом, если выполнены условия доказываемой теоремы, то

$$\limsup \mathbf{P} \left(\frac{\zeta(b, N, r)}{n(r)} \leq \frac{k}{n(r)} \right) \leq e^{-g(y, m)}.$$

Доказательство заявленной оценки снизу требует бóльших усилий. Поскольку

$$\frac{\partial H(x, p)}{\partial p} = \frac{p - x}{p(1 - p)} < 0, \quad \frac{\partial H(x, p)}{\partial x} = \log \frac{x}{1 - x} - \log \frac{p}{1 - p} > 0 \quad (18)$$

при $0 < p < x < 1$, то при $x \in (1 - 1/m, 1)$

$$\begin{aligned} & \inf_{1 - m^{-1} \leq x < 1} \min_{1 \leq i \leq m-2} \left(H \left(x, \frac{i}{m} \right) - H \left(x, 1 - \frac{1}{m} \right) \right) \\ &= \inf_{1 - m^{-1} \leq x < 1} \left(H \left(x, 1 - \frac{2}{m} \right) - H \left(x, 1 - \frac{1}{m} \right) \right) \\ &= H \left(1 - \frac{1}{m}, 1 - \frac{2}{m} \right) - H \left(1 - \frac{1}{m}, 1 - \frac{1}{m} \right) \\ &= \left(1 - \frac{1}{m} \right) \log \frac{m-1}{m-2} - \frac{1}{m} \log 2 > 0. \end{aligned}$$

Отсюда и из (16) следует, что при $r \rightarrow \infty$

$$\bar{\Phi} \left(\sqrt{2n(r)H \left(\frac{k-1}{n(r)}, \frac{i}{m} \right)} \right) = o \left(\bar{\Phi} \left(\sqrt{2n(r)H \left(\frac{k-1}{n(r)}, 1 - \frac{1}{m} \right)} \right) \right)$$

для $1 \leq i \leq m-2$ и значений k , удовлетворяющих неравенству (17). Таким образом, при $r \rightarrow \infty$

$$\sum_{i=1}^m \bar{\Phi} \left(\sqrt{2n(r)H \left(\frac{k-1}{n(r)}, \frac{i}{m} \right)} \right) \sim \bar{\Phi} \left(\sqrt{2n(r)H \left(\frac{k-1}{n(r)}, 1 - \frac{1}{m} \right)} \right).$$

Объединяя эти оценки с левым неравенством в (14), заключаем, что

$$\begin{aligned} \mathbf{P} \left(\frac{\zeta(b, N, r)}{n(r)} \leq \frac{k}{n(r)} \right) &\geq \left(\frac{1}{m} \sum_{i=1}^m \Phi \left(\sqrt{2n(r)H \left(\frac{k-1}{n(r)}, \frac{i}{m} \right)} \right) \right)^{n(d-r)} \\ &= \left(1 - \frac{1 + \varepsilon_3(k, r)}{m} \bar{\Phi} \left(\sqrt{2n(r)H \left(\frac{k-1}{n(r)}, 1 - \frac{1}{m} \right)} \right) \right)^{n(d-r)} \\ &= \left(1 - \frac{1 + \varepsilon_4(k, r)}{m \sqrt{2n(r)H \left(\frac{k-1}{n(r)}, \frac{m-1}{m} \right)} \sqrt{2\pi}} e^{-n(r)H \left(\frac{k-1}{n(r)}, \frac{m-1}{m} \right)} \right)^{n(d-r)}, \end{aligned}$$

где $|\varepsilon_3(k, r)| + |\varepsilon_4(k, r)| \rightarrow 0$ при $r \rightarrow \infty$ равномерно по значениям k , подчиняющимся условию (17).

Используя второе соотношение в (18) и вспоминая (15), выводим, что при $r \rightarrow \infty$

$$\begin{aligned} &n(r) \left(H \left(\frac{k-1}{n(r)}, \frac{m-1}{m} \right) - H \left(\frac{k}{n(r)}, \frac{m-1}{m} \right) \right) \\ &= (1 + \varepsilon_5(k, r)) \left(\log \frac{\frac{k-1}{n(r)}}{1 - \frac{k-1}{n(r)}} - \log(m-1) \right) \\ &= (1 + \varepsilon_6(k, r)) \log \frac{y}{(1-y)(m-1)}, \end{aligned}$$

где $|\varepsilon_5(k, r)| + |\varepsilon_6(k, r)| \rightarrow 0$. Отсюда следует, что

$$\liminf \mathbf{P} \left(\frac{\zeta(b, N, r)}{n(r)} \leq \frac{k}{n(r)} \right) \geq \frac{y}{(1-y)(m-1)} e^{-g(y, m)}.$$

Теорема 5 доказана.

Список литературы

- [1] Ватутин В. А., “Предельные теоремы для числа отрезков возрастания в случайных перестановках, порождаемых алгоритмами сортировки”, *Дискретная математика*, **6:1** (1994), 83–99.
- [2] Ватутин В. А., Михайлов В. Г., “О числе чтений случайных неравновероятных файлов при устойчивой сортировке”, *Дискретная математика*, **8:2** (1996), 14–30.
- [3] Зубков А. М., Серов А. А., “Полное доказательство универсальных неравенств для функции распределения биномиального закона”, *Теория вероятн. и ее примен.*, **57:3** (2012), 597–602.