А. М. Зубков, А. А. Серов, Совокупность образов подмножества конечного множества при итерациях случайных отображений, *Дискрет. матем.*, 2014, том 26, выпуск 4, 43–50

DOI: 10.4213/dm1303

Дискретная математика

том 26 выпуск 4 * 2014

УДК 519.212.2+519.213.21 DOI 10.4213/dm1303

Совокупность образов подмножества конечного множества при итерациях случайных отображений

© 2014 г. А. М. Зубков*, А. А. Серов**

Пусть $\mathcal{N}$ — множество из N элементов и $F_1, F_2, \ldots$ — последовательность случайных независимых равновероятных отображений $\mathcal{N} \to \mathcal{N}$. Для подмножества $S_0 \subset \mathcal{N}$, $|S_0| = n$, рассматриваются последовательность его образов $S_k = F_k(\ldots F_2(F_1(S_0))\ldots)$, $k = 1, 2\ldots$, и последовательность их объединений $\Psi_k = S_1 \cup \ldots \cup S_k$, $k = 1, 2\ldots$ Описан способ точного вычисления распределений $|S_k|$ и $|\Psi_k|$ при умеренных значениях N. Получены двусторонние неравенства для $\mathbf{M}|S_k|$ и $\mathbf{M}|\Psi_k|$, в которых верхние оценки асимптотически эквивалентны нижним, если $N, n, k \to \infty, nk = o(N)$. Результаты представляют интерес для анализа алгоритмов балансировки времени и памяти.

Ключевые слова: итерации случайных отображений, метод балансировки времени и памяти.

1. Введение

Одной из известных вычислительно труднорешаемых задач является поиск решения уравнения

$$G(x) = a, \tag{1}$$

где G — такое отображение конечного множества $\mathcal{N} = \{1, \ldots, N\}$ в себя, что все известные способы вычисления значения $G^{-1}(a)$ по трудоемкости сравнимы с перебором всего множества $\mathcal{N}$. Очевидным методом поиска решения уравнения (1) является последовательное вычисление значений $G(x)$ для всех $x \in \mathcal{N}$ до тех пор, пока не обнаружится решение этого уравнения. Для реализации такого метода требуется память медленно растущего при $N \to \infty$ объема (необходимого для вычисления значений функции G при любом $x \in \mathcal{N}$), но время работы (число операций) при использовании этого метода имеет порядок $O(N)$.

M. E. Hellman [10] предложил универсальный (не зависящий от вида функции G) метод поиска решений уравнения (1), позволяющий (после предварительного этапа, проводимого за время порядка $O(N)$) за счет использования памяти (по порядку меньшей $O(N)$) находить решение каждого уравнения (1) с большой вероятностью за время, по порядку меньшее $O(N)$. Этот подход был назван балансировкой

*Место работы: Математический институт им. В.А. Стеклова РАН, e-mail: zubkov@mi.ras.ru

**Место работы: Математический институт им. В.А. Стеклова РАН, e-mail: serov@mi.ras.ru

времени и памяти. На предварительном этапе метода Хеллмана и его более поздних модификаций с помощью вычисления $nt^2 = O(N)$ значений суперпозиций вида $F = R(G(x))$ (где отображения $R : \mathcal{N} \to \mathcal{N}$ выбираются тем или иным способом) составляются таблицы, содержащие в совокупности $O(N/t)$ пар вида $(x, F^t(x))$, где $F^t(x)$ — t-кратная итерация отображений вида $F = R(G(x))$; эти таблицы позволяют на основном этапе для любого $a \in \mathcal{N}$ находить решение уравнения (1) с помощью вычисления $O(N/n)$ значений вида $R(G(x))$. Если n и t имеют порядок $O(N^{1/3})$, то объем таблиц, составленных на предварительном этапе, имеет порядок $O(N^{2/3})$, а на основном этапе нахождение каждого решения имеет сложность порядка $O(N^{2/3})$ вычислений значений $R(G(x))$ (все эти оценки приведены с точностью до логарифмических множителей).

Мы рассматриваем упрощенную математическую модель процесса построения одной «радужной» таблицы (эта модель соответствует варианту метода балансировки времени и памяти, предложенному в [15]). Модель имеет следующий вид: в множестве $\mathcal{N}$ выбирается начальное подмножество S_0, $|S_0| = n$, и вычисляются его образы

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

где $F_1, \ldots, F_t$ — независимые случайные отображения множества $\mathcal{N}$ в себя, имеющие равновероятное распределение на множестве Σ_N, $|\Sigma_N| = N^N$, всех таких отображений.

В статье описан способ точного вычисления распределений случайных величин $\varphi_k = |S_k|$ и $\zeta_t = |S_1 \cup S_2 \cup \ldots \cup S_t|$ с помощью цепей Маркова, применимый при умеренных значениях N, и получены двусторонние оценки математических ожиданий этих случайных величин и вероятностей того, что элемент $x \in \mathcal{N}$, не зависящий от итерируемых отображений $F_1, F_2, \ldots$, принадлежит множеству S_k или множеству $S_1 \cup S_2 \cup \ldots \cup S_t$. Верхние и нижние оценки асимптотически эквивалентны при $N, n, t \to \infty$, если $nt = o(N)$. Эти результаты могут использоваться для оптимизации методов балансировки времени и памяти.

Характеристики методов балансировки времени и памяти, а также свойств итераций случайных отображений изучались в ряде работ. Перечислим некоторые полученные в них результаты.

В [10] для случайного равновероятного отображения $F : \mathcal{N} \to \mathcal{N}$, множества $S \subset \mathcal{N}$ мощности n: $|S| = n$, и случайного множества $\Phi_t = F(S) \cup F^2(S) \cup \ldots \cup F^t(S)$, где $F^k(S)$ — образ S при k-кратной итерации отображения F, $k = 1, 2, \ldots$, получены верхняя и нижняя оценки вероятности того, что $x \in \Phi_t$ для любого $x \in \mathcal{N}$, а именно:

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{t}\left(1 - \frac{it}{N}\right)^j \leqslant \mathbf{P}\left\{x \in \Phi_t\right\} \leqslant \frac{nt}{N}. \tag{2}$$

В [10] показано, что при $nt^2 \approx N$ и $n, t \gg 1$ левая часть этого неравенства близка к $0.80\,\frac{nt}{N}$; в [13] получена оценка

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{t}\left(1 - \frac{it}{N}\right)^j \geqslant \frac{nt}{N}\int\limits_0^1 \frac{1 - e^{-x}}{x}dx \approx 0.796599\,\frac{nt}{N}.$$

Приведенное в [10] доказательство неравенства (2) справедливо и для итераций независимых случайных отображений.

В [11] с помощью перехода к аппроксимирующим дифференциальным уравнениям получена другая приближенная формула:

$$\mathbf{P}\{x \in \Phi_t\} \approx 2\left(1 - \frac{1}{\tau^2}\right) \frac{e^{t/\tau} - e^{-t/\tau}}{(\tau+1)e^{t/\tau} + (\tau-1)e^{-t/\tau}}, \tag{3}$$

где $\tau = \sqrt{\frac{2N}{n}}$.

В [15] для случайного равновероятного отображения $F : \mathcal{N} \to \mathcal{N}$, последовательности $R_1, R_2, \ldots$ случайных независимых взаимно однозначных отображений $\mathcal{N} \to \mathcal{N}$, подмножества $S \subset \mathcal{N}$, $|S| = n$, последовательности его образов $S_k = R_k\left(F\left(\ldots R_2\left(F\left(R_1(F(S))\right)\right)\ldots\right)\right)$, $k = 1, 2, \ldots$, и случайного множества $\Psi_k = S_1 \cup \ldots \cup S_k$, $k = 1, 2, \ldots$, с использованием эвристических рассуждений получена приближенная формула

$$\mathbf{P}\{x \in S \cup \Psi_t\} \approx 1 - \prod_{i=1}^{t}\left(1 - \frac{n_i}{N}\right),$$

где $n_1 = n$, $n_{i+1} = N\left(1 - e^{-\frac{n_i}{N}}\right)$ при $i \geqslant 1$.

В [16] предлагается для оценивания характеристик методов балансировки времени и памяти использовать в качестве математической модели ветвящиеся процессы Гальтона-Ватсона.

В [12] в качестве одной из моделей популяционной генетики рассматривались последовательность $F_1, F_2, \ldots$ независимых равновероятных отображений и случайная величина

$$\tau_N = \min\{t : |F_t(F_{t-1}(\ldots F_1(\mathcal{N})\ldots))| = 1\}$$

— минимальное число итераций случайных отображений $\mathcal{N}$ в себя, при котором образом $\mathcal{N}$ оказывается одноэлементное множество. В [12] отмечено, что в этой модели при $N \to \infty$ распределения случайных величин $\zeta = \frac{1}{N}\tau_N$ сходятся к распределению суммы $\xi = \sum_{j=1}^{\infty} \xi_j$, где случайные величины $\xi_1, \xi_2, \ldots$ независимы и

$$\mathbf{P}\{\xi_j \leqslant x\} = 1 - e^{-xj(j+1)/2}, \quad x \geqslant 0, \quad j = 1, 2, \ldots$$

Так как $\mathbf{E}\xi_j = 2/(j(j+1))$, то ξ имеет конечное математическое ожидание:

$$\mathbf{E}\xi = \sum_{j=1}^{\infty} \frac{2}{j(j+1)} = 2.$$

Позднее эти утверждения доказывались разными способами (см., например, [6], [2], [8]). В [14] получено обобщение этих утверждений на итерации неравновероятных отображений (когда образы разных элементов независимы и имеют одно и то же распределение на $\mathcal{N}$).

2. Основные результаты

Пусть, как и ранее, $F_1, F_2, \ldots$ — независимые случайные отображения множества $\mathcal{N} = \{1, \ldots, N\}$ в себя, $S_0 \subset \mathcal{N}$, $|S_0| = n$, $S_k = F_k(S_{k-1})$, $\Psi_k = \cup_{j=1}^{k} S_j$, $k \geqslant 1$. Положим $\varphi_0 = |S_0|$, $\zeta_0 = 0$, $\varphi_k = |S_k|$, $\zeta_k = |\Psi_k|$, $k \geqslant 1$. Так как отображения $F_1, F_2, \ldots$ независимы и одинаково распределены, то последовательности $\{\varphi_k\}_{k\geqslant 0}$ и $\{\zeta_k\}_{k\geqslant 0}$ являются цепями Маркова.

Утверждение 1. *Матрица переходных вероятностей цепи Маркова* $\{\varphi_k\}_{k\geqslant 0}$ *имеет вид*

$$P = \|p_{i,j}\|_{i,j=1}^{N},$$

$$p_{i,j} = \begin{cases} \binom{N}{j}\left(\frac{j}{N}\right)^i \sum\limits_{m=0}^{j}(-1)^m\binom{j}{m}\left(1-\frac{m}{j}\right)^i, & 1\leqslant j\leqslant i\leqslant N,\\ 0, & j>i. \end{cases}$$

Матрица переходных вероятностей цепи Маркова $\{(\varphi_k,\zeta_k)\}_{k\geqslant 0}$ *имеет вид*

$$Q = \|q_{(i,r),(j,s)}\|_{i,j,r,s=1}^{N},$$

$$q_{(i,r),(j,s)} = \begin{cases} p_{i,j}\frac{\binom{N-r}{s-r}\binom{r}{j-s+r}}{\binom{N}{j}} = \binom{N-r}{s-r}\binom{r}{j-s+r}\left(\frac{j}{N}\right)^i \sum\limits_{m=0}^{j}(-1)^m\binom{j}{m}\left(1-\frac{m}{j}\right)^i, \\ \qquad\qquad \textit{если } 1\leqslant j\leqslant i\leqslant N,\ 1\leqslant r\leqslant s\leqslant \min\{N, r+j\}, \\ 0 \qquad\qquad \textit{в остальных случаях.} \end{cases}$$

Доказательство. Согласно описанию модели вероятность перехода цепи Маркова $\{\varphi_k\}_{k\geqslant 0}$ за 1 шаг из состояния i в состояние j равна вероятности того, что при независимом равновероятном размещении i частиц по N ячейкам число занятых ячеек равно j или, что то же самое, число пустых ячеек равно $N-j$. Формула для $p_{i,j}$ в формулировке утверждения совпадает с формулами (1), (2) в [4], гл. 1, § 1.

Переход цепи $(\varphi_k,\zeta_k) = (|S_k|,|\Psi_k|)$ из состояния (i,r) в состояния (j,s) с помощью отображения F_{k+1} можно разбить на два этапа: на первом этапе происходит переход от множества S_k, $|S_k| = i$, к множеству S_{k+1} с $|S_{k+1}| = j$ (этот переход не зависит от Ψ_k и имеет такую же вероятность $p_{i,j}$, как в цепи Маркова $\{\varphi_k\}$), а на втором этапе строится множество $\Psi_{k+1} = \Psi_k \cup S_{k+1}$; при этом в силу равновероятности отображения F_{k+1} и его независимости от $F_1,\dots,F_k$ множество S_{k+1} с $|S_{k+1}| = j$ имеет равновероятное распределение на совокупности всех j-элементных подмножеств множества $\mathcal{N}$; поэтому

$$\mathbf{P}\left\{|\Psi_{k+1}| = s \,\middle|\, |\Psi_k| = r,\, |S_{k+1}| = j\right\} = \frac{\binom{N-r}{s-r}\binom{r}{j-s+r}}{\binom{N}{j}}.$$

Утверждение доказано.

Вероятности переходов цепи Маркова $\{\varphi_k\}_{k\geqslant 0}$ за k шагов образуют матрицу $P^{(k)} = \|p_{(i,j)}^{(k)}\|_{i,j=1}^{N} = P^k$. Таким образом, наборы чисел $\{p_{(n,j)}^{(k)} = \mathbf{P}\{\varphi_k = j \,|\, \varphi_0 = n\},\ j = 1,\dots,N\}$ задают распределения φ_k, что позволяет находить численные значения характеристик распределения φ_k при умеренных значениях N.

Двусторонние оценки величин $\mathbf{P}\{x\in S_k \,|\, \varphi_0 = n\}$, $\mathbf{P}\{x\in \Psi_k \,|\, \varphi_0 = n\}$ и первых моментов случайных величин φ_k, ζ_k содержатся в следующей теореме.

Теорема 1. *Пусть* $F_1, F_2, \dots$ *— независимые равновероятные отображения множества* $\mathcal{N} = \{1,\dots,N\}$ *в себя,* $S_0\subseteq\mathcal{N}$, $|S_0| = n$, $S_k = F_k(\dots(F_1(S_0))\dots)$, $k\geqslant 1$. *Для любого элемента* $x\in\mathcal{N}$, *не зависящего от отображений* F_1, $F_2,\dots$, *при любых* $1\leqslant k$, $n\leqslant N$ *справедливы неравенства*

$$\begin{gathered} \tfrac{n}{N} - C_n^2\tfrac{k}{N^2} \leqslant \mathbf{P}\{x\in S_k \,|\, \varphi_0 = n\} < \tfrac{n}{N} - C_n^2\tfrac{k}{N^2} + \tfrac{n^3k^2}{4N^3}, \\ \tfrac{nt}{N} - C_{t+1}^2\tfrac{3n^2}{2N^2} < \mathbf{P}\{x\in\Psi_t \,|\, \varphi_0 = n\} < \tfrac{nt}{N} - C_n^2C_{t+1}^2\tfrac{1}{N^2} + \tfrac{n^3(t+1)^3}{12N^3}. \end{gathered} \tag{4}$$

Справедливы также следующие оценки:

$$n - C_n^2 \tfrac{k}{N} \leqslant \mathbf{M}\{\varphi_k \,|\, \varphi_0 = n\} < n - C_n^2 \tfrac{k}{N} + \tfrac{n^3 k^2}{4N^2}\,, \tag{5}$$

$$nt - C_{t+1}^2 \tfrac{3n^2}{2N} < \mathbf{M}\,\{\zeta_t \mid \varphi_0 = n\} < nt - C_n^2 C_{t+1}^2 \tfrac{1}{N} + \tfrac{n^3 (t+1)^3}{12N^2}\,,$$

$$\mathbf{D}\{\varphi_k \,|\, \varphi_0 = n\} < \tfrac{kn^3}{N}\left(1 + \tfrac{(n+2)k}{4nN}\right). \tag{6}$$

Доказательство. Будем использовать обозначение $F_{k\ldots1}(x)$ для $F_k(\ldots(F_1(x))\ldots)$. Из равновероятности отображений F_j, $j \geqslant 1$, следует, что величина $\mathbf{P}\{x \in S_k \,|\, |S_0| = n\}$ не зависит от $x \in \mathcal{N}$ как при фиксированном x, так и при случайном x, принимающем значения в $\mathcal{N}$ и не зависящем от $F_1, F_2, \ldots$ Поэтому неравенства (4) будем доказывать для $x = 1$. Так как

$$\{1 \in S_k\} = \bigcup_{x \in S_0} \{F_{k\ldots1}(x) = 1\},$$

то в силу неравенств Бонферрони

$$\sum_{x \in S_0} \mathbf{P}\{F_{k\ldots1}(x) = 1\} - \sum_{\substack{x,y \in S_0 \\ x<y}} \mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = 1\} \leqslant$$

$$\leqslant \mathbf{P}\{1 \in S_k \,|\, \varphi_0 = n\} \leqslant \tag{7}$$

$$\leqslant \sum_{x \in S_0} \mathbf{P}\{F_{k\ldots1}(x) = 1\} - \sum_{\substack{x,y \in S_0 \\ x<y}} \mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = 1\} +$$

$$+ \sum_{\substack{x,y,z \in S_0 \\ x<y<z}} \mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = F_{k\ldots1}(z) = 1\}.$$

Очевидно,

$$\mathbf{P}\{F_{k\ldots1}(x) = 1\} = \tfrac{1}{N} \text{ для любого } x \in \mathcal{N}. \tag{8}$$

Далее, при любых $x, y \in \mathcal{N}, x \neq y$,

$$\mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = 1\} =$$

$$= \mathbf{P}\{F_{k\ldots1}(x) = 1\}\mathbf{P}\{F_{k\ldots1}(y) = F_{k\ldots1}(x) \,|\, F_{k\ldots1}(x) = 1\} =$$

$$= \tfrac{1}{N}\,\mathbf{P}\left\{\bigcup_{j=1}^{k} [\min\{i :\, F_{i\ldots1}(y) = F_{i\ldots1}(x)\} = j]\right\}, \tag{9}$$

поскольку в силу равновероятности отображений F_j условная вероятность не зависит от значения $F_{k\ldots1}(x)$. События в правой части (9) несовместны, и в силу независимости и равновероятности отображений $F_1, F_2, \ldots$ при любом $j = 1, \ldots, k$

$$\mathbf{P}\left\{\{\min\{i :\, F_{k\ldots1}(y) = F_{k\ldots1}(x)\} = j\}\right\} = \tfrac{1}{N}\left(1 - \tfrac{1}{N}\right)^{j-1},$$

т. е.

$$\mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = 1\} =$$

$$= \tfrac{1}{N}\,\mathbf{P}\left\{\bigcup_{j=1}^{k} \{[\min\{i :\, F_{i\ldots1}(y) = F_{i\ldots1}(x)\} = j]\right\} = \tag{10}$$

$$= \tfrac{1}{N}\sum_{j=1}^{k} \tfrac{1}{N}\left(1 - \tfrac{1}{N}\right)^{j-1} = \tfrac{1}{N}\left(1 - \left(1 - \tfrac{1}{N}\right)^k\right) \in \left[\frac{k}{N} - C_k^2 \frac{1}{N^2}, \frac{k}{N}\right].$$

Аналогично для любых попарно различных $x, y, z \in \mathcal{N}$

$$\mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = F_{k\ldots1}(z) = 1\} =$$
$$= \mathbf{P}\{F_{k\ldots1}(x) = 1\}\mathbf{P}\{F_{k\ldots1}(z) = F_{k\ldots1}(y) = F_{k\ldots1}(x) \,|\, F_{k\ldots1}(x) = 1\} =$$
$$= \tfrac{1}{N}\,\mathbf{P}\Big\{ \bigcup_{j,m=1}^{k} \big[\min\{i :\, F_{i\ldots1}(y) = F_{i\ldots1}(x)\} = j,$$
$$\min\{r :\, F_{r\ldots1}(z) \in \{F_{r\ldots1}(x), F_{r\ldots1}(y)\}\} = m\big]\Big\}.$$

При разных парах (j, m) события в правой части несовместны и

$$\mathbf{P}\big\{ \min\{i :\, F_{i\ldots1}(y) = F_{i\ldots1}(x)\} = j,$$
$$\min\{r :\, F_{r\ldots1}(z) \in \{F_{r\ldots1}(x), F_{r\ldots1}(y)\}\} = m\big\} =$$
$$= \begin{cases} \frac{1}{N}\left(1-\frac{1}{N}\right)^{j-1} \frac{2}{N}\left(1-\frac{2}{N}\right)^{m-1}, & m < j, \\ \frac{1}{N}\left(1-\frac{1}{N}\right)^{j-1} \frac{1}{N}\left(1-\frac{2}{N}\right)^{j-1}\left(1-\frac{1}{N}\right)^{m-j}, & m \geqslant j. \end{cases}$$

Так как

$$\sum_{j=2}^{k}\sum_{m=1}^{j-1} \tfrac{1}{N}\left(1-\tfrac{1}{N}\right)^{j-1} \tfrac{2}{N}\left(1-\tfrac{2}{N}\right)^{m-1} < \tfrac{2}{N^2}\sum_{j=1}^{k-1} j = \tfrac{k(k-1)}{N^2},$$
$$\sum_{m=1}^{k}\sum_{j=1}^{m} \tfrac{1}{N}\left(1-\tfrac{1}{N}\right)^{m-1} \tfrac{1}{N}\left(1-\tfrac{2}{N}\right)^{j-1} < \tfrac{1}{N^2}\sum_{m=1}^{k} m = \tfrac{k(k+1)}{2N^2},$$

то

$$\mathbf{P}\{F_{k\ldots1}(x) = F_{k\ldots1}(y) = F_{k\ldots1}(z) = 1\} < \tfrac{k(k-1)}{N^2} + \tfrac{k(k+1)}{2N^2} < \tfrac{3k^2}{2N^3}\,. \tag{11}$$

Из (7), (8), (10), (11) следует, что

$$\mathbf{P}\{1 \in S_k \,|\, \varphi_0 = n\} \geqslant \tfrac{n}{N} - C_n^2 \tfrac{k}{N^2},$$
$$\mathbf{P}\{1 \in S_k \,|\, \varphi_0 = n\} \leqslant \tfrac{n}{N} - C_n^2 \tfrac{1}{N}\left(1 - \left(1-\tfrac{1}{N}\right)^k\right) + C_n^3 \tfrac{3k^2}{2N^3} \leqslant$$
$$\leqslant \tfrac{n}{N} - C_n^2 \tfrac{k}{N^2} + C_n^2 C_k^2 \tfrac{1}{N^3} + C_n^3 \tfrac{3k^2}{2N^3} < \tfrac{n}{N} - C_n^2 \tfrac{k}{N^2} + \tfrac{n^3 k^2}{4N^3}\,.$$

Тем самым первое неравенство в (4) доказано.

Доказательство второго неравенства тоже проводится с помощью неравенств Бонферрони

$$\sum_{k=1}^{t} \mathbf{P}\{1 \in S_k \,|\, \varphi_0 = n\} - \sum_{1 \leqslant k < m \leqslant t} \mathbf{P}\{1 \in S_k,\, 1 \in S_m \,|\, \varphi_0 = n\} \leqslant$$
$$\leqslant \mathbf{P}\left\{1 \in \Psi_t = \bigcup_{k=1}^{t} S_k \,\middle|\, \varphi_0 = n\right\} = \mathbf{P}\left\{\bigcup_{k=1}^{t}\{1 \in S_k\} \,\middle|\, \varphi_0 = n\right\} \leqslant$$
$$\leqslant \sum_{k=1}^{t} \mathbf{P}\{1 \in S_k \,|\, |S_0| = n\}. \tag{12}$$

Используя первые неравенства в (4), находим:

$$\frac{nt}{N} - C_n^2 C_{t+1}^2 \tfrac{1}{N^2} = \sum_{k=1}^{t} \left(\tfrac{n}{N} - C_n^2 \tfrac{k}{N^2}\right) \leqslant \sum_{k=1}^{t} \mathbf{P}\{1 \in S_k \,|\, \varphi_0 = n\} \leqslant$$

$$\leqslant \sum_{k=1}^{t} \left(\tfrac{n}{N} - C_n^2 \tfrac{k}{N^2} + \tfrac{n^3 k^2}{4N^3}\right) < \tfrac{nt}{N} - C_n^2 C_{t+1}^2 \tfrac{1}{N^2} + \tfrac{n^3 (t+1)^3}{12N^3}\,. \tag{13}$$

Далее, с учетом того, что при условии вида $\varphi_k = i$, $\varphi_m = j$ множества S_k и S_m независимы, не зависят от предыстории и что последовательность φ_r, $r \geqslant 0$, образует цепь Маркова с невозрастающими траекториями, находим, что при $1 \leqslant k < m$

$$\mathbf{P}\{1 \in S_k,\, 1 \in S_m \,|\, \varphi_0 = n\} =$$

$$= \sum_{n \geqslant i \geqslant j \geqslant 1} \mathbf{P}\{1 \in S_k,\, 1 \in S_m,\, \varphi_k = i,\, \varphi_m = j \,|\, \varphi_0 = n\} = \tag{14}$$

$$= \sum_{i=1}^{n} \mathbf{P}\{\varphi_k = i \,|\, \varphi_0 = n\} \tfrac{i}{N} \sum_{j=1}^{i} \mathbf{P}\{\varphi_m = j \,|\, \varphi_k = i\} \tfrac{j}{N} \leqslant$$

$$\leqslant \sum_{i=1}^{n} \mathbf{P}\{\varphi_k = i \,|\, \varphi_0 = n\} \tfrac{i}{N}\, \tfrac{i}{N} \leqslant \tfrac{1}{N^2} \mathbf{E}\{\varphi_k^2 \,|\, \varphi_0 = n\} \leqslant \tfrac{n^2}{N^2}\,.$$

Из (12), (13), (14) следует второе неравенство в (4):

$$\tfrac{nt}{N} - C_{t+1}^2 \tfrac{3n^2}{2N^2} < \tfrac{nt}{N} - C_n^2 C_{t+1}^2 \tfrac{1}{N^2} - C_t^2 \tfrac{n^2}{N^2} \leqslant \mathbf{P}\left\{1 \in \Psi_t \,|\, \varphi_0 = n\right\} <$$

$$< \tfrac{nt}{N} - C_n^2 C_{t+1}^2 \tfrac{1}{N^2} + \tfrac{n^3 (t+1)^3}{12N^3}.$$

Неравенства для математических ожиданий непосредственно следуют из (4), так как

$$\varphi_k = \sum_{j=1}^{N} \mathbb{I}\{j \in S_k\}\,, \quad \zeta_t = \sum_{j=1}^{N} \mathbb{I}\{j \in \Psi_t\}\,. \tag{15}$$

Для оценки $\mathbf{D}\{\varphi_k \,|\, \varphi_0 = n\}$ используем (15), а также независимость и равновероятность отображений F_k, $k \geqslant 1$:

$$\mathbf{D}\{\varphi_k \,|\, \varphi_0 = n\} = \mathbf{M}\left\{\varphi_k^2 \,|\, \varphi_0 = n\right\} - \left(\mathbf{M}\{\varphi_k \,|\, \varphi_0 = n\}\right)^2 =$$

$$= \mathbf{M}\left\{ \sum_{i,j=1}^{N} \mathbb{I}\{i, j \in S_k\} \,\middle|\, \varphi_0 = n \right\} - \left(\mathbf{M}\{\varphi_k \,|\, \varphi_0 = n\}\right)^2 =$$

$$= \mathbf{M}\{\varphi_k \,|\, \varphi_0 = n\} + N(N-1)\mathbf{P}\{1, 2 \in S_k \,|\, \varphi_0 = n\} - \left(\mathbf{M}\{\varphi_k \,|\, \varphi_0 = n\}\right)^2\,. \tag{16}$$

Оценки для $\mathbf{M}\{\varphi_k \,|\, \varphi_0 = n\}$ уже получены. Далее, в силу равновероятности и независимости отображений F_k, $k \geqslant 1$,

$$\mathbf{P}\left\{1,\, 2 \in S_k \,|\, \varphi_0 = n\right\} = \mathbf{P}\left\{ \bigcup_{x,y=1}^{n} \{F_{k\ldots 1}(x) = 1,\, F_{k\ldots 1}(y) = 2\} \right\} \leqslant$$

$$\leqslant \sum_{x,y=1}^{n} \mathbf{P}\left\{F_{t\ldots 1}(x) \neq F_{t\ldots 1}(y)\,(1 \leqslant t < k), F_{k\ldots 1}(x) = 1,\, F_{k\ldots 1}(y) = 2\right\} =$$

$$= n(n-1)\left(1 - \tfrac{1}{N}\right)^{k-1} \tfrac{1}{N^2}\,. \tag{17}$$

Из (5), (16), (17) и неравенства $\left(1-\frac{1}{N}\right)^k \leqslant 1-\frac{k}{N}+C_k^2\frac{1}{N^2}$ следует, что

$$\begin{gathered}\mathbf{D}\{\varphi_k \,|\, \varphi_0 = n\} < \\ < n - C_n^2\frac{k}{N} + \frac{n^3k^2}{4N^2} + \frac{(N-1)n(n-1)}{N}\left(1-\frac{1}{N}\right)^{k-1} - \left(n - C_n^2\frac{k}{N}\right)^2 = \\ = \frac{n^3k^2}{4N^2} + n(n-1)\left(\left(1-\frac{1}{N}\right)^k - \left(1-\frac{k(n-1)}{2N}\right)\left(1-\frac{kn}{2N}\right)\right) \leqslant \\ \leqslant \frac{n^3k^2}{4N^2} + n(n-1)\left(\left(1-\frac{k}{N}+\frac{k(k-1)}{2N^2}\right) - \left(1-\frac{k(2n-1)}{2N}+\frac{kn(n-1)}{4N^2}\right)\right) = \\ = \frac{n^3k^2}{4N^2} + n(n-1)\left(\frac{k(2n-1)-2k}{2N} - \frac{kn(n-1)-2k(k-1)}{4N^2}\right) < \frac{kn^3}{N}\left(1+\frac{(n+2)k}{4nN}\right).\end{gathered}$$

Тем самым неравенство (6) и теорема 1 доказаны.

Неравенства теоремы можно уточнять, если использовать больше членов в неравенствах Бонферрони.

Список литературы

1. Гульден Я., Джексон Д., *Перечислительная комбинаторика*, М. : Наука. Физматлит, 1990, 503 с.
2. Зубков А. М., Шибанов О. К., “Время до объединения всех частиц при равновероятных размещениях по последовательности слоев ячеек”, *Матем. заметки*, **85**:3 (2009), 373–381.
3. Колчин В. Ф., *Случайные отображения*, М. : Наука, 1984, 208 с.
4. Колчин В. Ф., Севастьянов Б. А., Чистяков В. П., *Случайные размещения*, М. : Наука, 1976, 224 с.
5. Степанов В. Е., “О распределении числа вершин в слоях случайного дерева”, *Теория вероятн. и примен.*, **14**:1 (1969), 64–77.
6. Dalal A., Schmutz E. “Compositions of random functions on a finite set”, *Electr. J. Comb.*, **9**:R26 (2002).
7. Flajolet P., Odlyzko A. M., “Random mapping statistics”, Advances in Cryptology, Proc. Eurocrypt’89, *Lect. Notes Comput. Sci.*, **434**, 1990, 329–354.
8. Goh W. M. Y., Hitczenko P., Schmutz E., “Iterating random functions on a finite set”, 2014, 7 pp., arXiv:math/0207276v2.
9. Harris B., “Probability distributions related to random mappings”, *Ann. Math. Statist.*, **31**:2 (1960), 1045–1062.
10. Hellman M.E., “A cryptanalytic time–memory trade-off”, *IEEE Trans. Inf. Theory*, 1980, 401–406.
11. Hong J., Ma D., “Success probability of the Hellman trade-off”, *Inf. Process. Lett.*, **109**:7 (2009), 347–351.
12. Kingman J. F. C., “The coalescent”, *Stoch.Proc. Appl.*, **13** (1982), 235–248.
13. Kusuda K., Matsumoto T., “Optimization of time-memory trade-off cryptanalysis and its application to DES”, *IEICE Trans. on Fundamentals*, **1**:E-79A (1996), 35–48.
14. McSweeney J. K., Pittel B. G., “Expected coalescence time for a nonuniform allocation process”, *Adv. Appl. Probab.*, **40**:4 (2008), 1002–1032.
15. Oechslin P., “Making a faster cryptanalytic time-memory trade-off”, *Lect. Notes Comput. Sci.*, **2729** (2003), 617–630.
16. 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*, **5**:2 (2014), 103–108.
17. Rubin H., Sitgreaves R., *Probability distributions related to random transformations of a finite set*, Tech. report. №19A, Appl. math. and statist. lab., Stanford Univ., 1954.

Статья поступила 20.06.2014.