TWO-SIDED INEQUALITIES FOR THE AVERAGE NUMBER OF ELEMENTS IN THE UNION OF IMAGES OF FINITE SET UNDER ITERATIONS OF RANDOM EQUIPROBABLE MAPPINGS

A. A. Serov[1], A. M. Zubkov[2]
Steklov Mathematical Institute, Russian Academy of Sciences
Moscow, RUSSIA
e-mail: [1]`serov@mi.ras.ru`, [2]`zubkov@mi.ras.ru`

Abstract

Let $\mathcal{N}$ be a set of N elements and $F_1, F_2, \ldots$ be a sequence of random independent equiprobable mappings $\mathcal{N} \to \mathcal{N}$. For a subset $S_0 \subset \mathcal{N}$, $|S_0| = n$, we consider a sequence of its images $S_k = F_k(\ldots F_2(F_1(S_0))\ldots)$, $k = 1, 2\ldots$, and a sequence of their unions $\Psi_k = S_1 \cup \ldots \cup S_k$, $k = 1, 2\ldots$ An approach to the exact computation of distribution of $|S_k|$ and $|\Psi_k|$ for moderate values of N is described. Two-sided inequalities for $\mathbf{M}|S_k|$ and $\mathbf{M}|\Psi_k|$ such that upper bound are asymptotically equivalent to lower ones for $N, n, k \to \infty, nk = o(N)$ are derived. The results are of interest for the analysis of time-memory tradeoff algorithms.

This work was supported by RFBR, grant 14-01-00318.

1 Introduction

One of the well-known time-consuming task is the search for solution of the equation

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

where G be a mapping of the finite set $\mathcal{N} = \{1, \ldots, N\}$ to itself such that the complexity of any known method to compute the value $G^{-1}(a)$ is comparable with exhaustive search over the entire set $\mathcal{N}$. The trivial method of searching the solution of the equation (1) is the sequential computation of values $G(x)$ for all $x \in \mathcal{N}$ until the solution of (1) will be found. The implementation of such method requires a memory of slowly growing size for $N \to \infty$ (necessary to calculate a value of the function G for any $x \in \mathcal{N}$), but the time (number of operations) needed this method has the order $O(N)$.

M. E. Hellman [2] proposed the universal (independent of the type of function G) method for searching the solutions of the equation (1), permitting (after the preliminary stage of the complexity $O(N)$) to find the solution of equation (1) with a high probability for a time in order less than $O(N)$ by means of tables having volume less than $O(N)$. This approach has been called the time-memory tradeoff.

We consider a simplified mathematical model of the “rainbow” table construction (this model corresponds to the version of the time-memory tradeoff method that has been proposed in [6]). The model is as follows: an initial subset $S_0 \subset \mathcal{N}$, $|S_0| = n$, is chosen and its images

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

are calculated, where $F_1, \ldots, F_t$ are independent random mappings of the set $\mathcal{N}$ to itself having uniform distribution on the set Σ_N, $|\Sigma_N| = N^N$, of all such mappings.

We propose the method to compute distributions of random variables $\varphi_k = |S_k|$ and $\zeta_t = |S_1 \cup S_2 \cup \ldots \cup S_t|$ by means of Markov chains, applicable for moderate values of N, and obtain two-sided estimates for the expectation of these random variables and for the probabilities that an element $x \in \mathcal{N}$, independent of the iterated mappings $F_1, F_2, \ldots$, belongs to the set S_k or to the set $S_1 \cup S_2 \cup \ldots \cup S_t$. Upper and lower bounds are asymptotically equivalent for $N, n, t \to \infty$, if $nt = o(N)$. These results may be used to optimize the methods of the time-memory tradeoff.

2 Basic results

Let, as before, $F_1, F_2, \ldots$ be independent random mappings of the set $\mathcal{N} = \{1, \ldots, N\}$ to itself, $S_0 \subset \mathcal{N}$, $|S_0| = n$, $S_k = F_k(S_{k-1})$, $\Psi_k = \cup_{j=1}^k S_j$, $k \geqslant 1$. Let $\varphi_0 = |S_0|$, $\zeta_0 = 0$, $\varphi_k = |S_k|$, $\zeta_k = |\Psi_k|$, $k \geqslant 1$. Since the mappings $F_1, F_2, \ldots$ are independent and identically distributed, the sequences $\{\varphi_k\}_{k\geqslant 0}$ and $\{\zeta_k\}_{k\geqslant 0}$ form the Markov chains.

Assertion 1. The transition probability matrix of the Markov chain $\{\varphi_k\}_{k\geqslant 0}$ has the form

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

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

The transition probability matrix of the Markov chain $\{(\varphi_k, \zeta_k)\}_{k\geqslant 0}$ has the form

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

$$q_{(i,r),(j,s)} = \begin{cases} p_{i,j}\dfrac{\binom{N-r}{s-r}\binom{r}{j-s+r}}{\binom{N}{j}} = \binom{N-r}{s-r}\binom{r}{j-s+r}\left(\frac{j}{N}\right)^i \sum\limits_{m=0}^{j}(-1)^m\binom{j}{m}\left(1-\frac{m}{j}\right)^i, \\ \qquad \text{if } 1 \leqslant j \leqslant i \leqslant N,\ 1 \leqslant r \leqslant s \leqslant \min\{N, r+j\}, \\ 0 \qquad \text{otherwise.} \end{cases}$$

The transition probabilities of the Markov chain $\{\varphi_k\}_{k\geqslant 0}$ for k steps form the matrix $P^{(k)} = \|p_{(i,j)}^{(k)}\|_{i,j=1}^N = P^k$. Thus the collections of numbers $\{p_{(n,j)}^{(k)} = \mathbf{P}\{\varphi_k = j \mid \varphi_0 = n\}$, $j = 1, \ldots, N\}$ define the distributions of φ_k that allows to find the numerical values of the distribution characteristics of φ_k for the moderate values of N.

The two-sided estimates of $\mathbf{P}\{x \in S_k \mid \varphi_0 = n\}$, $\mathbf{P}\{x \in \Psi_k \mid \varphi_0 = n\}$ and the first moments of the random variables φ_k, ζ_k are contained in the following Theorem.

Theorem 1. *Let $F_1, F_2, \ldots$ be the independent equiprobable mappings of the set $\mathcal{N} = \{1, \ldots, N\}$ to itself, $S_0 \subseteq \mathcal{N}$, $|S_0| = n$,$S_k = F_k(\ldots(F_1(S_0))\ldots)$, $k \geqslant 1$. For any element $x \in \mathcal{N}$, which does not depend on F_1, $F_2, \ldots$, for all $1 \leqslant k$, $n \leqslant N$ we have*

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

The following inequalities are also valid:

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

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

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

References

[1] Harris B. (1960). Probability distributions related to random mappings. *Ann. Math. Statist.* Vol. **31**, No. 2, pp. 1045–1062.

[2] Hellman M.E. (1980). A cryptanalytic time–memory trade-off. *IEEE Trans. Inf. Theory.* Vol. **26**, pp. 401–406.

[3] Flajolet P., Odlyzko A.M. (1990). Random Mapping Statistics *Advances in Cryptology — Proc. Eurocrypt'89, J-J. Quisquater Ed., Lect. Notes Comp. Sci.* Vol. **434**, pp. 329–354.

[4] Kolchin V.F., Sevastyanov B.A., Chistyakov V.P. (1978). Random allocations. *Scripta Series in Math.* V. H. Winston & Sons, Washington, pp. 262.

[5] Kolchin V.F. (1986). Random mappings. *Trans. Ser. in Math. and Eng.*, Optimization Software Inc. Publications Division, New York, pp. 207.

[6] Oechslin P. (2003). Making a faster cryptanalytic time-memory trade-off. *Lect. Notes Comput. Sci.* Vol. 2729, pp. 617–630.

[7] Zubkov A.M., Serov A.A. (2014). Images of subset of finite set under iterations of random mappings. *Discrete Math. Appl.* Vol. 2015, No. 3, pp. 179–185.