Compositions of Random Functions on a Finite Set

Avinash Dalal
MCS Department, Drexel University
Philadelphia, Pa. 19104
ADalal@drexel.edu

Eric Schmutz
Drexel University and Swarthmore College
Philadelphia, Pa., 19104
Eric.Jonathan.Schmutz@drexel.edu

Submitted: July 21, 2001; Accepted: July 9, 2002
MR Subject Classifications: 60C05, 60J10, 05A16, 05A05

Abstract

If we compose sufficiently many random functions on a finite set, then the composite function will be constant. We determine the number of compositions that are needed, on average. Choose random functions $f_1, f_2, f_3, \ldots$ independently and uniformly from among the n^n functions from $[n]$ into $[n]$. For $t > 1$, let $g_t = f_t \circ f_{t-1} \circ \cdots \circ f_1$ be the composition of the first t functions. Let T be the smallest t for which g_t is constant(i.e. $g_t(i) = g_t(j)$ for all i, j). We prove that $E(T) \sim 2n$ as $n \to \infty$, where $E(T)$ denotes the expected value of T.

1 Introduction

If we compose sufficiently many random functions on a finite set then the composite function is constant. We ask how long this takes, on average. More precisely, let U_n be the set of n^n functions from $[n]$ to $[n]$. Let A_n be the n element subset of U_n consisting of the constant functions: $g \in A_n$ iff $g(i) = g(j)$ for all i, j. Let $f_1, f_2, f_3, \ldots$ be a sequence of random functions chosen independently and uniformly from U_n. Let $g_1 = f_1$, and for $t > 1$ let $g_t = f_t \circ g_{t-1}$ be the composition of the first t random maps. Define $T(\langle f_i \rangle_{i=1}^{\infty})$ to be the smallest t for which $g_t \in A_n$. (If no such t exists, define $T = \infty$. It is not difficult to show that $\Pr(T = \infty) = 0$.) Our goal in this paper is to estimate $E(T)$.

It is natural to restate the problem as a question about a Markov chain. The state space is $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$. For $t > 0$ and $r \in [n]$, we are in state s_r if and only if g_t has exactly r elements in its range. With the convention that g_0 is the identity permutation, we start in state s_n at time $t = 0$. The question is how long (i.e. how many compositions) it takes to reach the absorbing state s_1.

For $m > 1$, let $\tau_m = |\{t : |Range(g_t)| = m\}|$ be the amount of time we are in state s_m. Thus $T = \sum\limits_{m=2}^{n} \tau_m$. Let $\mathcal{T}$ consist of those states that are actually visited:

for $m > 1$, $s_m \in \mathcal{T}$ iff $\tau_m > 0$. The visited states $\mathcal{T}$ are a (non-uniform) random subset of $\mathcal{S}$ that includes at least two elements, namely s_n and (with probability 1) s_1. We prove later that $\mathcal{T}$ typically contains most of the small numbered states and relatively few of the large numbered states. This observation forms the basis for our proof of

Theorem 1 $E(T) = 2n(1 + o(1))$ *as* $n \to \infty$.

We should mention that there is a standard approach to our problem using the transition matrix P and linear algebra. Let Q be the matrix that is obtained from P by striking out the first row and column of P. Then $E(T)$ is exactly the sum of the entries in the last row of $(I - Q)^{-1}$. See, for example, chapter 3 of [5]. This fact is very convenient if one wishes to compute $E(T)$ for specific small values of n. An anonymous referee conjectured that $E(T) = 2n - 3 + o(1)$ after observing that, for small values of n, $|E(T) - 2n + 3| \le 1$. This conjecture is plausible, but we are nowhere near a proof.

2 The Transition Matrix

The $n \times n$ transition matrix P can be determined quite explicitly. Suppose g_{t-1} has i elements in its range, How many functions f have the property that $f \circ g_{t-1}$ has exactly j elements in its range? There are $\binom{n}{j}$ ways to choose the j-element range of $f \circ g_{t-1}$, and $S(i,j)j!$ ways to map the i-element range of g_{t-1} onto a given j element set. (Here $S(i,j)$ is the number of ways to partition an i element set into j disjoint subsets, a Stirling number of the second kind.) Finally, there are $n - i$ elements in the complement of the range of g_{t-1}, and n^{n-i} ways to map them into $[n]$. Thus there are $\binom{n}{j}S(i,j)j!n^{n-i}$ functions f with the desired property, and for $1 \le i,j \le n$, the transition matrix for the chain has i,j'th entry

$$P(i,j) = \binom{n}{j}\frac{S(i,j)j!}{n^i}. \tag{1}$$

The stationary distribution π assigns probability 1 to s_1. The transition matrix has some nice properties. It is lower triangular, which means the eigenvalues are just the diagonal entries: for $1 \le m \le n$,

$$\lambda_m = P(m,m) = \prod_{k=0}^{m-1}(1 - \frac{k}{n}). \tag{2}$$

For future reference we record two simple estimates for the eigenvalues, both of which follow easily from (2).

Lemma 2

$$\lambda_m = 1 - \frac{\binom{m}{2}}{n} + O(\frac{m^4}{n^2})$$

and

$$\lambda_m \le \exp(-\binom{m}{2}/n).$$

3 Lower Bound

The proof of the lower bound requires an estimate for the Stirling numbers $S(m,k)$. The literature contains many precise but complicated estimates for these numbers. Here we prove a crude inequality whose simplicity makes it convenient for our purposes.

Lemma 3 *For all positive integers m and k, $S(m,k) \le (2k)^m$.*

Proof: The proof of this lemma will be done by induction using the recurrence $S(m,k) = S(m-1,k-1) + kS(m-1,k)$. When $k=1$, we know that $S(m,1)=1$ and $(2k)^m = 2^m$. So clearly the inequality holds true for $k=1$ (for all positive integers m).

Now let ϕ_m denote the following statement: for all $k>1$, $S(m,k) \le (2k)^m$. It suffices to prove that ϕ_m is true for all m. For $m=1$, $S(1,k) = 0 \le 2k$ for all $k>1$. Now let $k>1$ and assume, inductively, that ϕ_{m-1} is true (i.e. $S(m-1,k) \le (2k)^{m-1}$ for $k>1$.) Then we have

$$S(m,k) = S(m-1,k-1) + kS(m-1,k) \le (2(k-1))^{m-1} + k(2k)^{m-1}$$
$$= (2k)^m \left\{ \frac{1}{2} + \frac{(k-1)^{m-1}}{2k^m} \right\}.$$

Realize that the quantity inside the large braces is less than one. ∎

With lemma 3 available, we can proceed with the proof that $E(T) \ge 2n(1+o(1))$. Since $T = \sum_{m=2}^{n} \tau_m$, we have

$$E(T) = \sum_{m=2}^{n} \Pr(s_m \in \mathcal{T}) E(\tau_m | s_m \in \mathcal{T}). \tag{3}$$

Obviously a lower bound is obtained by truncating this sum. To simplify notation, let $\ell = \lfloor \log\log n \rfloor$. Then

$$E(T) \ge \sum_{m=2}^{\ell} \Pr(s_m \in \mathcal{T}) E(\tau_m | s_m \in \mathcal{T}). \tag{4}$$

To estimate the second factor in each term of (4), note that

$$E(\tau_m | s_m \in \mathcal{T}) = \sum_{t=1}^{\infty} t\lambda_m^{t-1}(1-\lambda_m) = \frac{1}{1-\lambda_m}. \tag{5}$$

Applying lemma 2, we get

$$E(\tau_m | s_m \in \mathcal{T}) = \frac{n}{\binom{m}{2}}(1 + O(\frac{m^2}{n})). \tag{6}$$

To estimate the first factor of each term in (4), we make the following observation: if $s_m \notin \mathcal{T}$, then there is a transition from s_{m+d} to s_{m-j} for some positive integers d and j. Hence,

$$\Pr(s_m \notin \mathcal{T}) = \sum_{d=1}^{n-m} \sum_{j=1}^{m-1} \Pr(s_{m+d} \in \mathcal{T}) \frac{P(m+d, m-j)}{(1-\lambda_{m+d})}. \qquad (7)$$

(The factor $(1-\lambda_{m+d})^{-1} = \sum_{i=0}^{\infty} P(m+d, m+d)^i$ is there because we remain in state s_{m+d} for some number of transitions $i \geq 0$ before moving on to state s_{m-j}.)

Let $\sigma := \sum_{d=1}^{n-m} \sum_{j=1}^{m-1} \frac{S(m+d,m-j)}{n^{j+d}} \frac{\lambda_{m-j}}{1-\lambda_{m+d}}$. Putting (1) and $\Pr(s_{m+d} \in \mathcal{T}) \leq 1$ into (7), we get

$$\Pr(s_m \notin \mathcal{T}) \leq \sum_{d=1}^{n-m} \sum_{j=1}^{m-1} 1 \cdot \binom{n}{m-j} \frac{S(m+d, m-j)(m-j)!}{n^{m+d}(1-\lambda_{m+d})} = \sigma. \qquad (8)$$

A first step in bounding σ is to note that $1 > (1-\frac{1}{n}) = \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \ldots \geq \lambda_n > 0$, and therefore

$$\frac{\lambda_{m-j}}{1-\lambda_{m+d}} \leq \frac{1}{1-\lambda_{m+d}} \leq \frac{1}{1-\lambda_2} = n-1.$$

Hence

$$\sigma \leq (n-1) \sum_{d=1}^{n-m} \frac{1}{n^d} \sum_{j=1}^{m-1} \frac{S(m+d, m-j)}{n^j}.$$

Applying lemma 3 to each term of the inside sum, we get

$$\sum_{j=1}^{m-1} \frac{S(m+d, m-j)}{n^j} \leq \sum_{j=1}^{m-1} \frac{(2(m-j))^{m+d}}{n^j}$$

$$\leq \frac{m(2m-2)^{m+d}}{n} < \frac{\ell(2\ell)^{\ell+d}}{n}.$$

Hence

$$\sigma \leq (n-1) \frac{\ell(2\ell)^{\ell}}{n} \sum_{d=1}^{n-m} (\frac{2\ell}{n})^d = O(\frac{(2\ell)^{\ell+2}}{n}) = o(1).$$

Thus $\Pr(s_m \in \mathcal{T}) \geq 1 - o(1)$ for all $m \leq \ell$, Putting this and (6) back into (4), and using the fact that $\sum_{m=2}^{\ell} \frac{1}{\binom{m}{2}} = \sum_{m=2}^{\ell} (\frac{2}{m-1} - \frac{2}{m}) = 2 - \frac{2}{\ell}$, we get the lower bound $E(T) \geq 2n(1+o(1))$. ∎

4 Upper Bound

If $|Range(g_{t-1})| = m$, then the restriction of f_t to $Range(g_{t-1})$ is a random function from an m element set to $[n]$. Before proving that $E(T) \leq 2n(1+o(1))$, we gather a simple lemma about the size of the size of the range for such random maps.

Lemma 4 *Suppose $h : [m] \to [n]$ is selected uniformly at random from among the n^m functions from $[m]$ into $[n]$, and let R be the cardinality of the range of h. Then the mean and variance of R are respectively $E(R) = n - n(1-\frac{1}{n})^m$ and $Var(R) = n^2\{(1-\frac{2}{n})^m - (1-\frac{1}{n})^{2m}\} + n\{(1-\frac{1}{n})^m - (1-\frac{2}{n})^m\}$.*

Proof: Let $U = n - R = \sum_{i=1}^{n} I_i$, where I_i is 1 if i is not in the range of h, and otherwise I_i is zero. Then $E(R) = n - E(U)$, and $Var(R) = Var(U)$.

$$E(U) = nE(I_1) = n(1 - \frac{1}{n})^m. \tag{9}$$

$$E(U^2) = \sum_{i \neq j} E(I_i I_j) + E(U)$$

$$= n(n-1)(1 - \frac{2}{n})^m + E(U).$$

Therefore

$$Var(U) = n^2 \Big\{ (1 - \frac{2}{n})^m - (1 - \frac{1}{n})^{2m} \Big\} + n \Big\{ (1 - \frac{1}{n})^m - (1 - \frac{2}{n})^m \Big\}.$$

■

The next corollary shows that there are gaps between the large states in $\mathcal{T}$. Let $\xi_2 = \lfloor \frac{n}{\log^2 n} \rfloor$, and let $\beta = \beta(n) = \frac{1}{2}(\xi_2 - n + n(1 - \frac{1}{n})^{\xi_2})$. Although β is quite large ($\beta \gg \frac{n}{\log^4 n}$) all we really need for our purposes is that $\beta \to \infty$ as $n \to \infty$.

Corollary 5 $\Pr(s_{m-\delta} \notin \mathcal{T}$ for $1 \leq \delta \leq \beta \mid s_m \in \mathcal{T}) = 1 - o(1)$ *uniformly for* $\xi_2 \leq m \leq n$.

Proof: Suppose we are in state s_m at time $t-1$ and select the next function f_t. Let h be the restriction of f_t to the range of g_{t-1}, and let R be the cardinality of the range of h, and let $B = m - R$. Observe that if $B > \beta$ then the next β states are missed: $s_{m-\delta} \notin \mathcal{T}$ for $1 \leq \delta \leq \beta$. Note that $E(B) = m - n + n(1 - \frac{1}{n})^m > 2\beta$. Applying Chebyshev's inequality to the random variable B, we get

$$\Pr(B \leq \beta) \leq \Pr(B \leq \frac{1}{2}E(B)) \leq \frac{4Var(B)}{(E(B))^2}. \tag{10}$$

For $\xi_2 \leq m \leq n$, we have $E(B) = m - n + n(1 - \frac{1}{n})^m \geq \xi_2 - n + n(1 - \frac{1}{n})^{\xi_2} \gg \frac{n}{\log^4 n}$. (A calculus exercise shows that $E(B)$ is an increasing function of m.) To bound $Var(B)$ note that,

$$(1 - \frac{2}{n})^m - (1 - \frac{1}{n})^{2m} = O(\frac{m}{n^2}).$$

Therefore (10) yields

$$\Pr(B \leq \beta) = O(\frac{m \log^8 n}{n^2}) = o(1).$$

■

Now we proceed with the proof of the upper bound $E(T) \leq 2n(1 + o(1))$. Split the sum (3) into three separate sums as follows. Let $\xi_1 = \lfloor \sqrt{\frac{n}{\log n}} \rfloor$, and let $\xi_2 = \lfloor \frac{n}{\log^2 n} \rfloor$, so that (3) becomes

$$E(T) = \sum_{m=2}^{\xi_1} + \sum_{m=\xi_1+1}^{\xi_2} + \sum_{m=\xi_2+1}^{n} \tag{11}$$

The first sum in (11) is estimated using (5), lemma 2, and the fact that $\Pr(s_m \in \mathcal{T}) \leq 1$:

$$\begin{aligned}
\sum_{m=2}^{\xi_1} \Pr(s_m \in \mathcal{T}) E(\tau_m | s_m \in \mathcal{T}) &\leq \sum_{m=2}^{\xi_1} \frac{1}{1-\lambda_m} \\
&= \sum_{m=2}^{\xi_1} \frac{1}{\frac{\binom{m}{2}}{n} + O(\frac{m^4}{n^2})} \\
&= (1 + O(\frac{\xi_1^2}{n}))n \sum_{m=2}^{\xi_1} \frac{1}{\binom{m}{2}} = 2n(1 + o(1)).
\end{aligned}$$

The second sum in (11) is estimated using a crude bound on the eigenvalues. For $\xi_1 < m \leq \xi_2$, we have $\lambda_m \leq \lambda_{\xi_1} = 1 - \frac{1}{2 \log n} + O(\frac{1}{\sqrt{n \log n}})$. Hence the second sum in (11) is at most

$$\begin{aligned}
\sum_{m=\xi_1+1}^{\xi_2} \frac{1}{1-\lambda_m} &\leq \frac{1}{1-\lambda_{\xi_1}} \sum_{m=\xi_1}^{\xi_2} 1 \\
&= O(\xi_2 \log n) = O(\frac{n}{\log n}).
\end{aligned}$$

For the last sum in (11), we can no longer get away with the trivial estimate $\Pr(s_m \in \mathcal{T}) \leq 1$. However now the size of the eigenvalues can be handled less carefully:

$$\sum_{m=\xi_2+1}^{n} \Pr(s_m \in \mathcal{T}) \frac{1}{1-\lambda_m} \leq \Big(\max_{m \geq \xi_2} \frac{1}{1-\lambda_m}\Big)\Big(\sum_{m=\xi_2}^{n} \Pr(s_m \in \mathcal{T})\Big). \tag{12}$$

The first factor in (12) is easily estimated using (2):

$$\max_{m \geq \xi_2} \frac{1}{1-\lambda_m} = \frac{1}{1-\lambda_{\xi_2}} \leq \frac{1}{1-\exp(-\binom{\xi_2}{2}/n)} \leq 2$$

for all sufficiently large n.

To deal with the second factor in (12) we use Corollary 5. The idea is that there cannot be too many "hits"(visited states) simply because every time there is a hit it is followed by β "misses". To make this precise, define $V = \sum_{m=\xi_2}^{n} \chi_m$, where χ_m is 1 if $s_m \in \mathcal{T}$ and 0 otherwise. Thus the second factor in (12) is just $E(V)$. Also count large numbered states that are *not* in $\mathcal{T}$ with $W = \sum_{m=\xi_2}^{n} (1 - \chi_m)$ so that $W + V = n + 1 - \xi_2$ and $E(V) = n + 1 - \xi_2 - E(W)$. If a state s_m is in $\mathcal{T}$, and if the next β possible states $s_{m-1}, s_{m-2}, \ldots, s_{m-\beta}$ are *not* in $\mathcal{T}$, then those β missed states together contribute exactly β to W.

If we let $J_m = \chi_m \cdot \prod_{\delta=1}^{\beta} (1-\chi_{m-\delta})$, then $W \geq \beta \sum_{m \geq \xi_2} J_m$. But then

$$E(W) \geq \beta \sum_{m \geq \xi_2} E(J_m) =$$

$$\beta \sum_{m \geq \xi_2} \Pr(s_m \in \mathcal{T}) \Pr(s_{m-1}, s_{m-2}, \ldots s_{m-\beta} \notin \mathcal{T} | s_m \in \mathcal{T}).$$

By Corollary 5,

$$\Pr(s_{m-1}, s_{m-2}, \ldots s_{m-\beta} \notin \mathcal{T} | s_m \in \mathcal{T}) = 1 - o(1).$$

Hence

$$E(W) \geq \beta(1+o(1)) \sum_{m=\xi_2}^{n} \Pr(s_m \in \mathcal{T}) = (1+o(1))\beta E(V).$$

But then

$$E(V) = n + 1 - \xi_2 - E(W) \leq n + 1 - \xi_2 - \beta(1+o(1)))E(V),$$

which implies that

$$E(V) \leq \frac{n+1-\xi_2}{1+\beta(1+o(1))} = O(\log^4 n).$$

Thus the second factor of (12) is $o(n)$, which means that the third sum in (11) is negligible. ■

References

[1] D.Aldous and J.Fill, Reversible Markov Chains and Random Walks on Graphs" http://stat.berkeley.edu/users/aldous.

[2] P.Diaconis and D.Freedman, Iterated Random Functions, *SIAM Review* **41** No. 1, p 45–76.

[3] J.C.Hansen and J.Jaworski, Large Components of Random Mappings, *Random Structures and Algorithms* **17** (2000) 317–342.

[4] J.Kemeny, J.L.Snell, and A.W.Knapp, Denumerable Markov Chains, Van Nostrand Co., 1966.

[5] J.G.Kemeny, J.L.Snell, Finite Markov Chains, Springer Verlag, 1976.

[6] J.Jaworski, A Random Bipartite Mapping, *Annals of Discrete Math.,* **28** 137–158 (1985).

[7] V.F.Kolchin, Random Mappings, Optimization Software, 1986.

[8] V.F.Kolchin, B.A.Sevastyanov, and V.P.Chistaykov, Random Allocations, Winston, 1978.

[9] J. S. Rosenthal, Convergence Rates for Markov Chains, *SIAM Review* **37** 387–405.