Random Compositions

Evangelos Kranakis School of Computer Science Carleton University

Outline

- 1. Random functions and iterations.
- 2. Iterated random compositions.
- 3. Markov chain for random compositions.
- 4. Waiting time until absorption:
 - (a) Lower bound.
 - (b) Upper bound.
- 5. Open Problems.

Random Functions

- For *n* integer, $[n] = \{1, 2, ..., n\}$.
- $\mathcal{F}(m,n)$ is the space of functions $f:[m] \to [n]$.

•
$$\mathcal{F}(n) := \mathcal{F}(n, n).$$

- A function f is chosen from $\mathcal{F}(n)$ randomly with the uniform distribution.
- Range of f is defined by $\operatorname{Range}(f) := \{y : \exists x \in [n]f(x) = y\}.$
- Question:

What is the expected size of the range of a random function?

4

Values of Random Functions

For random function $f:[m] \to [n]$ and given $x \in [m], y \in [n]$,

$$\Pr[f(x) = y] = \frac{n^{m-1}}{n^m}$$
$$= \frac{1}{n}.$$

For random function $f:[m] \to [n]$ and given $y \in [n]$, the size of the inverse image $f^{-1}(\{y\})$ satisfies,

$$\Pr[|f^{-1}(\{y\})| = k] = \binom{m}{k} \frac{(n-1)^{m-k}}{n^m},$$

for k = 0, 1, ..., n.

Given $f \in \mathcal{F}(n)$ the graph G(f) has [n] as its set of vertices and directed edges $(x, y) \in E \Leftrightarrow f(x) = y$. Define the k-the iterate f^k of f by f^0 = identity function and $f^k := f \circ f^{k-1}$.

The orbit (or cycle) of x is defined by

$$O_f(x) := \{x, f(x), f(f(x)), \dots, f^{n-1}(x)\}.$$

If \mathcal{L}_n is the r.v. that counts the length of an orbit it can be shown (Sachkov 1997)

$$\Pr[\mathcal{L}_n = j] = \sum_{k=j}^n \frac{(n)_k}{n^{k+1}}, \ j = 1, \dots, n$$
$$E[\mathcal{L}_n] = \sum_{j=1}^n j \sum_{k=j}^n \frac{(n)_k}{n^{k+1}}$$

Connected Components of a Random Function

The connected components of the graph G(f) consist of the orbits and trees attached to them.



The elements of a cycle are called **cyclic** elements of G(f).

The distance of an element from its cycle is called its **height**.

of Connected Components of a Random Function Let \mathcal{K}_n be the r.v. that counts the number of connected components of a random function in $\mathcal{F}(n)$.

A result of Stepanov (1966) states that

$$\Pr[\mathcal{K}_{n} = j] = \sum_{k=j}^{n} {\binom{n-1}{k-1}} \frac{S(k,j)}{n^{k}}, \ j = 1, 2, ..., n$$
$$E[\mathcal{K}_{n}] = \frac{1}{2} \log n(1+o(1))$$
$$Var[\mathcal{K}_{n}] = \frac{1}{2} \log n(1+o(1)),$$

where S(k, j) is the # of ways to partition a k element set into j disjoint subsets, a Stirling number of the 2nd kind.

of Cyclic Elements of a Random Function Let \mathcal{Z}_n be the r.v. that counts the number of distinct cyclic elements of a random function in $\mathcal{F}(n)$.

A result of Harris (1973) states that

$$\Pr[\mathcal{Z}_n = k] = \frac{k(n)_k}{n^{k+1}}, \ k = 0, 1, \dots, n$$
$$E[\mathcal{Z}_n] = \sqrt{\frac{\pi n}{2}}(1+o(1))$$
$$Var[\mathcal{Z}_n] = \left(2 - \frac{\pi}{2}\right)n(1+o(1)),$$

where
$$(n)_k = n(n-1)\cdots(n-k+1)$$
.

Expected Size of Range of a Random Function $f : [m] \to [n]$ Consider the indicator function I_i : $I_i = 1$ if $i \in \text{Range}(f)$, and $I_i = 0$, otherwise. Then we have

$$E[|\operatorname{Range}(f)|] = E\left[\sum_{i=1}^{n} I_{i}\right]$$
$$= \sum_{i=1}^{n} E[I_{i}]$$
$$= nE[I_{1}]$$
$$= nPr[I_{1} = 1]$$
$$= n(1 - \Pr[1 \notin \operatorname{Range}(f)])$$
$$= n\left(1 - \left(1 - \frac{1}{n}\right)^{m}\right).$$



What Causes Shrinkage?

Consider a random function $f : [m] \to [n]$ (with $m \le n$): shrinkage is caused by collisions among the elements $f(1), f(2), \ldots, f(m)$, i.e., f(x) = f(y), for some $x \ne y$.

$$\begin{aligned} \Pr[\text{collision}] &= & \Pr[\exists x \neq y(f(x) = f(y))] \\ &= & 1 - \Pr[\forall x \neq y(f(x) \neq f(y))] \\ &= & 1 - \prod_{i=0}^{m-1} \frac{n-i}{n} \text{ (Birthday paradox)} \\ &= & 1 - \prod_{i=0}^{m-1} \left(1 - \frac{i}{n}\right) \end{aligned}$$

Hence, the bigger the m the higher the probability of a collision! We see later that this causes the "Markov Chain" to skip large states! Variance of |Range(f)| for a Random Function $f : [m] \to [n]$ Let X be the r.v. that counts the size of Range(f) and U = n - X. Consider the indicator function $I'_i: I'_i = 1$ if $i \notin \text{Range}(f)$, and $I'_i = 1$, otherwise. Observe that

$$\begin{aligned} Var(X) &= Var(U) \\ &= E[U^2] - E[U]^2 \\ &= \sum_{i \neq j} E[I'_i I'_j] + \sum_i E[I'_i] - E[U]^2 \\ &= n(n-1) \left(1 - \frac{2}{n}\right)^m + n \left(1 - \frac{1}{n}\right)^m - n^2 \left(1 - \frac{1}{n}\right)^{2m} \\ &= n^2 \left(\left(1 - \frac{2}{n}\right)^m - \left(1 - \frac{1}{n}\right)^{2m} \right) \\ &+ n \left(\left(1 - \frac{1}{n}\right)^m - \left(1 - \frac{2}{n}\right)^m \right). \end{aligned}$$

Compositions of Random Functions

- k functions f_1, f_2, \ldots, f_k are chosen from $\mathcal{F}(n)$ randomly and independently with the uniform distribution.
- Let $f^{(k)} := f_k \circ f_{k-1} \circ \cdots \circ f_1$.
- Convention: $f^{(0)} =$ identity function on [n].
- $f^{(k)}$ is called a random composition.

• Question:

Since the expected size of the range of a random function is a constant fraction of the size of its domain, how long does it take for a random composition to become constant?

Iterations of Random Functions

Model proposed by (Diaconis & Freedman 1999)

- There is a state space S.
- There is a family of functions \mathcal{F} such that each $F \in \mathcal{F}$ maps the state space into itself $F: S \to S$.
- There is a probability distribution μ on \mathcal{F} .
- If the chain is at state $s \in S$ it moves to state F(s) by choosing $F \in \mathcal{F}$ at random.
- The process starts with F_0 and inductively defines $X_{t+1} = F(X_t)$ where F is a random function $F \in \mathcal{F}$.

Example I: Linear Affine Functions

- The state space S is the real line.
- $\mathcal{F} = \{F_+, F_-\}$ has just two functions defined as follows

$$F_{+} : \mathbf{R} \to \mathbf{R} : x \to F_{+}(x) = ax + 1$$
$$F_{-} : \mathbf{R} \to \mathbf{R} : x \to F_{-}(x) = ax - 1,$$

where 0 < a < 1.

- The probability distribution μ on \mathcal{F} is $\mu(F_+) = \mu(F_-) = 1/2$. Let $\xi_i = \pm 1$ with probability 1/2, respectively.
- The process starts with ξ_0 and inductively defines $X_{t+1} = F(X_t)$ where F is a random function $F \in \mathcal{F}$.
- Clearly, $X_{t+1} = aX_t + \xi_t$ and the stationary distribution $X_{\infty} = \xi_1 + a\xi_2 + a^2\xi_3 + \cdots$ converges since a < 1.

Example II: *d*-dimensional Affine Functions

- The state space S is the d-dimensional space \mathbf{R}^d .
- \mathcal{F} contains a set of functions of the form

$$F: \mathbf{R}^d \to \mathbf{R}^d: x \to F(x) = Ax + B,$$

A is an $d \times d$ matrix and B is a $d \times 1$ vector.

- \mathcal{F} can be identified with a set of pairs (A, B) of matrices and we have a probability distribution μ on \mathcal{F} .
- The basic chain is $X_{t+1} = A_t X_t + B_t$, where A_t is an $d \times d$ matrix and B_t a $d \times 1$ vector, and (A_t, B_t) are idependent and identically distributed.

This has applications in fractal geometry.

Example III: Random Compositions

- States specify the "size of the range" of a random composition and these states form the state space S.
- A family of functions $\{\circ_f : f \in \mathcal{F}(n)\}$ maps the state space into itself as follows: Given a function $g \in \mathcal{F}(n)$ already in state s,

 $s \to \circ_f(s) :=$ size of range of $f \circ g$.

- The probability distribution on $\{\circ_f : f \in \mathcal{F}(n)\}$ is uniform.
- If the chain is at state $s \in S$ it moves to state $\circ_f(s)$ by choosing f at random.
- The process starts with f_0 (identity function) and inductively defines $X_{t+1} = \circ_f(X_t)$ where f is a random function $f \in \mathcal{F}(n)$.

Example IV: Random Compositions of Hashes

- States specify the "size of the range" of a random composition and these states form the state space S.
- $\mathcal{H}(n)$ is the set of functions $f: [n] \to [n/2]$.
- A family of functions $\{\circ_h : h \in \mathcal{H}(n)\}$ maps the state space into itself as follows: Given a function $g \in \mathcal{H}(n)$ already in state s,

 $s \to \circ_h(s) :=$ size of range of $h \circ g$.

- The probability distribution on $\{\circ_h : h \in \mathcal{H}(n)\}$ is uniform.
- If the chain is at state $s \in S$ it moves to state $\circ_f(s)$ by choosing f at random.
- The process starts with a given function h_0 and inductively defines $X_{t+1} = \circ_h(X_t)$ where h is a random function $h \in \mathcal{H}(n)$.

Waiting Time until Absorption

- For t > 0, we are in state s_r iff $|\text{Range}(f^{(t)})| = r$.
- $\tau_r = |\{t : |\operatorname{Range}(f^{(t)})| = r\}|$ is the amount of time in state s_r .
- State s_r is visited if $\tau_r > 0$. and \mathcal{T} is the set of states that are actually visited.
- Let T be the smallest t for which $f^{(t)}$ is constant, i.e.,



How is T Computed

- The Markov chain starts with the identity function $f^{(0)}$ at time 0 in state s_n .
- By the nature of the problem, states are visited in non-increasing order.
- It is possible that states may be "skipped".
- Eventually it reaches s_1 , the absorbing state.
- T is really the time it takes to reach the absorbing state s_1 .
- The main result is the following
 Theorem: E[T] = 2n(1 + o(1)), as n → ∞.

Transition Probabilities

For $j \leq i$, what is the probability $f^{(t)} \in s_j$ given that $f^{(t-1)} \in s_i$? Given that $f^{(t-1)} \in s_i$, how many functions f are there such that $f \circ f^{(t-1)}$ has j elements in its range?

- The are $\binom{n}{i}$ ways to choose the range of $f \circ f^{(t-1)}$,
- S(i, j)j! ways to map the *i*-element range of f^(t-1) onto a given j element set, where S(i, j) is the # of ways to partition a i element set into j disjoint subsets, a Stirling number of the 2nd kind, and
- n^{n-i} ways to map them into [n].

It follows that

$$p(i,j) := \Pr[f^{(t)} \in s_j | f^{(t-1)} \in s_i] = \binom{n}{j} \frac{S(i,j)j!}{n^i},$$

Upper Bound on Stirling Numbers S(i, j)

S(i, j) = # of ways to partition a *i* element set into *j* subsets.

- Prove by induction $S(i,j) \leq (2j)^i$.
- For i = 1: $S(1, j) \le 2j$
- We have that

$$\begin{split} S(i,j) &= S(i-1,j-1) + jS(i-1,j) \text{ (Identity)} \\ &\leq (2(j-1))^{i-1} + j(2j)^{i-1} \text{ (Induction)} \\ &= (2j)^i \left(\frac{(j-1)^{i-1}}{2j^i} + \frac{1}{2} \right) \\ &\leq (2j)^i. \end{split}$$

Eigenvalues of the Transition Matrix

The transition matrix $P := (p(i, j)_{i,j})$ is lower diagonal.

Eigenvalues are the diagonal elements of the matrix, i.e.,

$$\lambda_r = p(r, r)$$

$$= \binom{n}{r} \frac{S(r, r)r!}{n^r}$$

$$= \prod_{i=1}^{r-1} \left(1 - \frac{i}{n}\right)$$

$$= 1 - \frac{\binom{r}{2}}{n} + O\left(\frac{r^4}{n^2}\right)$$

Note: $1 > 1 - \frac{1}{n} = \lambda_2 \ge \cdots \ge \lambda_n > 0$ and $\frac{1}{1 - \lambda_{r+d}} \le n - 1$

Transition Probabilities for Affine Matrices

- Consider $d \times d$ matrices over, say, the finite field Z_p^* and let $A^{(0)} := I$ be the identity matrix.
- $A^{(t)} = A_t A^{(t-1)}$, where A_t is a random matrix.
- For t > 0 we are in state r iff $rank(A^{(t)}) = r$.
- $\tau_r = |\{t : \operatorname{rank}(A^{(t)}) = r\}|$ is the amount of time in state s_r .
- **Open Question:** For $j \leq i$, compute the transition probabilities

$$p(i,j) := \Pr[\operatorname{rank}(A^{(t)}) = j \mid \operatorname{rank}(A^{(t-1)}) = i]$$

This is equivalent to computing the number of matrices B such that rank(BA) = j, given that rank(A) = i.

Lower Bound on E[T]

We have the identity

$$E[T] = E\left[\sum_{r=2}^{n} \tau_{r}\right]$$
$$= \sum_{r=2}^{n} E[\tau_{r}]$$
$$= \sum_{r=2}^{n} E[\tau_{r}|\tau_{r} > 0] \cdot \Pr[\tau_{r} > 0]$$

It remains

- to compute $E[\tau_r | \tau_r > 0]$, and
- give a lower bound on $\Pr[\tau_r > 0]$.

Computing $E[\tau_r | \tau_r > 0]$

This is the expected amount of time you stay in state τ_r , given that you visit it?

Given $\tau_r > 0$, τ_r follows the geometric distribution, with probability of success $p(r, r) = \lambda_r$.

$$E[\tau_r | \tau_r > 0] = \sum_{t=1}^{\infty} t \lambda_r^{t-1} (1 - \lambda_r)$$
$$= \frac{1}{1 - \lambda_r}$$
$$= \frac{n}{\binom{r}{2}} \left(1 + O\left(\frac{r^2}{n}\right) \right)$$

Estimating $\Pr[\tau_r = 0]$

We give an upper bound on $\Pr[\tau_r = 0]$.

- If $\tau_r = 0$ then the state s_r is never visited.
- Therefore there must exist a transition

$$s_{r+d} \rightarrow s_{r-j},$$

for some positive integers d, j.

- In fact, before moving to state s_{r-j} it may stay in state s_{r+d} a number of times $t = 0, 1, 2, 3, \ldots$
- Therefore we must take into account how long we stay in state s_{r+d} given that this state is visited.

Upper Bound on $\Pr[\tau_r = 0]$

We can show that

$$\Pr[\tau_r = 0] = \sum_{d=1}^{n-r} \sum_{j=1}^{r-1} \Pr[\tau_{r+d} > 0] \sum_{t=0}^{\infty} p(r+d, r-j) p(r+d, r+d)^t$$
$$= \sum_{d=1}^{n-r} \sum_{j=1}^{r-1} \Pr[\tau_{r+d} > 0] \frac{p(r+d, r-j)}{1 - \lambda_{r+d}}$$
$$\leq \sum_{d=1}^{n-r} \sum_{j=1}^{r-1} \binom{n}{r-j} \frac{S(r+d, r-j)(r-j)!}{n^{r+d}(1 - \lambda_{r+d})}$$
$$\leq (n-1) \sum_{d=1}^{n-r} \frac{1}{n^d} \sum_{j=1}^{r-1} \frac{S(r+d, r-j)}{n^j}$$
$$\operatorname{Recall: 1)} \Pr[\tau_{r+d} > 0] \leq 1, 2) \ 1 > 1 - \frac{1}{n} = \lambda_2 \geq \cdots \geq \lambda_n > 0 \text{ and}$$
$$\frac{1}{1 - \lambda_{r+d}} \leq n - 1$$

Lower Bound on $\Pr[\tau_r > 0]$ Hence we obtain for $r \leq |\log \log n|$ $\Pr[\tau_r > 0] \ge 1 - (n-1) \sum_{r=1}^{n-r} \frac{1}{n^d} \sum_{r=1}^{r-1} \frac{S(r+d, r-j)}{n^j}$ $\geq 1 - (n-1) \sum_{d=1}^{n-r} \frac{1}{n^d} \sum_{j=1}^{r-1} \frac{(2(r-j))^{r+d}}{n^j}$ $\geq 1 - (n-1) \sum_{d=1}^{n-r} \frac{1}{n^d} \frac{r(2r)^{r+d}}{n}$ $\geq 1 - O\left(\frac{(2\ell)^{\ell+2}}{n}\right)$ = 1 - o(1).

Proving the Lower Bound: $E[T] \ge 2n(1 + o(1))$ We have the idequality

$$E[T] \geq \sum_{r=1}^{\ell} E[\tau_r | \tau_r > 0] \cdot \Pr[\tau_r > 0]$$

$$\geq \sum_{r=2}^{\ell} \frac{n}{\binom{r}{2}}$$

$$= 2n \sum_{r=2}^{\ell} \frac{1}{r(r-1)}$$

$$= 2n \sum_{r=2}^{\ell} \left(\frac{1}{r-1} - \frac{1}{r}\right)$$

$$= 2n \left(1 - \frac{1}{\ell}\right)$$

This completes the proof of the lower bound.



$$\xi_1 = \left\lfloor \sqrt{\frac{n}{\log n}} \right\rfloor$$
$$\xi_2 = \left\lfloor \frac{n}{\log^2 n} \right\rfloor$$

and make upper bound estimates on each of them.

1st Sum:
$$= \sum_{m=2}^{\xi_1} E[\tau_m | \tau_m > 0] \cdot \Pr[\tau_m > 0]$$

Observe that

$$1 \text{st Sum} \leq \sum_{m=2}^{\xi_1} \frac{1}{1 - \lambda_m}$$
$$= \sum_{m=2}^{\xi_1} \frac{1}{\frac{\binom{m}{2}}{n} + O(m^4/n^2)}$$
$$= \sum_{m=2}^{\xi_1} \frac{n}{\binom{m}{2} + O(m^4/n)}$$
$$\leq 2n \sum_{m=2}^{\xi_1} \frac{1}{m(m-1)}$$
$$= 2n(1 - 1/\xi_1)$$
$$= 2n(1 + o(1))$$

$$\begin{aligned} \mathbf{2nd} \ \mathbf{Sum:} &= \sum_{m=\xi_1+1}^{\xi_2} E\left[\tau_m | \tau_m > 0\right] \cdot \Pr[\tau_m > 0] \\ \text{Observe that } \lambda_{\xi_1} &= 1 - \frac{1}{2\log n} + O(1/\log^2 n). \end{aligned}$$

$$\begin{aligned} 2nd \ \text{Sum} &\leq \sum_{m=\xi_1+1}^{\xi_2} \frac{1}{1 - \lambda_m} \\ &= \frac{1}{1 - \lambda_{\xi_1}} \sum_{m=\xi_1+1}^{\xi_2} 1 \\ &= O(\xi_2 \log n) \\ &= O(n/\log n) \\ &= o(1). \end{aligned}$$

$$\begin{aligned} \mathbf{3rd} \ \mathbf{Sum:} &= \sum_{m=\xi_2+1}^n E\left[\tau_m | \tau_m > 0\right] \cdot \Pr[\tau_m > 0] \\ \text{Observe that } \max_{m>\xi_2} \frac{1}{1-\lambda_m} &= \frac{1}{1-\lambda_{\xi_2+1}}. \text{ Hence,} \\ \text{3rd Sum} &\leq \sum_{m=\xi_2+1}^n \frac{1}{1-\lambda_m} \Pr[\tau_m > 0] \\ &\leq \frac{1}{1-\lambda_{\xi_2+1}} \left(\sum_{m=\xi_2+1}^n \Pr[\tau_m > 0]\right) \\ &\leq \frac{1}{1-\exp\left(-\left(\frac{\xi_2}{2}\right)/n\right)} \left(\sum_{m=\xi_2+1}^n \Pr[\tau_m > 0]\right) \\ &\leq 2\sum_{m=\xi_2+1}^n \Pr[\tau_m > 0], \end{aligned}$$

for n large enough. It remains to bound the RHS above.

On Skipping Large States

To deal with $\sum_{m=\xi_2+1}^{n} \Pr[\tau_m > 0]$ we will show that "every hit (i.e., $\tau_m > 0$)" is followed by "many (i.e., β) misses (i.e., $\tau_{m-\delta} = 0$, forall δ such that $1 \leq \delta \leq \beta$)"

Let us define

$$\beta := \beta(n) = \frac{1}{2} \left(\xi_2 - n + n \left(1 - \frac{1}{n} \right)^{\xi_2} \right)$$

Observe that

$$\beta(n) >> \frac{n}{\log^4 n}.$$

Suppose we are in state s_m at time t - 1, i.e., $|\text{Range}(f^{(t-1)})| = m$, and select the next function f_t at random. Claim 1: If $B > \beta$ then $\tau_{m-\delta} = 0$, for $1 \le \delta \le \beta$. Let h be the restriction of f_t to the range of $f^{(t-1)}$. Let R be the cardinality of the range of h, and B = m - R. To prove the claim notice that

$$B > \beta \implies m - R > \beta$$
$$\Rightarrow R < m - \beta$$

and

$$|\text{Range}(f^{(t)})| = |f^{(t)}([n])| \\ = |f_t(f^{(t-1)}([n]))| \\ = |h(\text{Range}(f^{(t-1)}))| \\ = R.$$

Claim 2:
$$E[B] \ge \xi_2 - n + n(1 - \frac{1}{n})^{\xi_2} >> \frac{n}{\log^4 n}$$
.
Observe that

$$E[B] = E[m - R]$$

$$= m - E[R]$$

$$= m - n + n\left(1 - \frac{1}{n}\right)^m$$

$$> \xi_2 - n + n(1 - \frac{1}{n})^{\xi_2}$$

$$= 2\beta$$

$$>> \frac{n}{\log^4 n}.$$
Also recall, $Var(B) =$

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

Claim 4: $\Pr[\forall 1 \le \delta \le \beta(\tau_{m-\delta} = 0) | \tau_m > 0]$ By Chebeshev's Inequality and since $m > \xi_2$ we have $\Pr[|B| \le \beta] \le \Pr\left[|B| \le \frac{1}{2}E[B]\right]$ $\le \frac{4Var(B)}{(E[B])^2}$ $= O\left(\frac{m\log^8 n}{n^2}\right)$ = o(1).

Hence, $\Pr[|B| > \beta] = 1 - o(1)$. It follows, by Claim 1,

$$\Pr[\forall 1 \le \delta \le \beta(\tau_{m-\delta} = 0) | \tau_m > 0] = 1 - o(1).$$

Back to the 3rd Sum: $= \sum_{m=\xi_2+1}^n E[\tau_m | \tau_m > 0] \cdot \Pr[\tau_m > 0]$ Define $\chi_m = 1$ if $\tau_m > 0$ and $\chi_m = 0$, otherwise. Recall that 3rd Sum $\leq 2 \sum \Pr[\tau_m > 0]$ $m = \xi_2 + 1$ $= 2 \sum^{n} E[\chi_m]$ $m = \xi_2 + 1$ $= 2E\left[\sum_{m=\xi_2+1}^n \chi_m\right]$ = 2E[V],

where $V := \sum_{m=\xi_2+1}^{n} \chi_m$. Define $W := \sum_{m=\xi_2+1}^{n} (1-\chi_m)$ and observe that $V + W = n - \xi_2$.

Back to the 3rd Sum:
$$= \sum_{m=\xi_2+1}^{n} E[\tau_m | \tau_m > 0] \cdot \Pr[\tau_m > 0]$$
Note that if $\tau_m > 0$ and $\forall 1 \le \delta \le \beta(\tau_{m-\delta} = 0)$ then these β missed states contribute exactly β to W . Hence, if we define $J_m = \chi_m \cdot \prod_{\delta=1}^{\beta} (1 - \chi_{m-\delta})$ then $W \ge \beta \sum_{m>\xi_2} J_m$. Hence,
$$E[W] \ge \beta \sum_{m>\xi_2} E[J_m]$$
$$= \beta \sum_{m>\xi_2} \Pr[J_m = 1]$$
$$= \beta \sum_{m>\xi_2} \Pr[\tau_m > 0] \Pr[\forall 1 \le \delta \le \beta(\tau_{m-\delta} = 0) | \tau_m > 0]$$
$$\ge \beta(1 + o(1)) \sum_{m>\xi_2} \Pr[\tau_m > 0]$$
$$= \beta(1 + o(1)) E[V].$$

Back to the 3rd Sum: $= \sum_{m=\xi_2+1}^{n} E[\tau_m | \tau_m > 0] \cdot \Pr[\tau_m > 0]$ It follows that

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

Hence,

3rd Sum
$$\leq 2E[V]$$

 $\leq \frac{2(n-\xi_2)}{1+\beta(1+o(1))}$
 $= O(\log^4 n)$
 $= o(n).$

This completes the proof of the theorem.

Another Idea: Hitting Times

For r = 1, 2, ..., n, let t_r be the (hitting) time that the chain spends in transient state r until absorbtion by state 1.

- Let the $(n-1) \times (n-1)$ matrix Q be obtained from the transition matrix P by removing the first row and column.
- It is easy to see that $t_1 = 1$ and $t_r = 1_{r=1} + \sum_{r'=2}^r p(r, r') t_{r'}$
- If I is the unit matrix and t is the vector of hitting times then t = I + Qt, which is equivalent to t(I Q) = I. So to compute the hitting times it is enough to compute the inverse of the matrix I Q.
- Easy to prove: if lim_{k→∞} Q^k = 0 then (I Q)⁻¹ = ∑_{k=0} Q^k.
 Speed of convergence depends on 2nd largest eigenvalue!

Open Questions

- Consider the space $\mathcal{H}(n) := \mathcal{F}(n, n/2)$ of hash functions.
- Consider different probability distributions on $\mathcal{F}(n)$. The reason is that in practice one has preference over certain types of random functions.
- Consider a space {(A, B)} of pairs of d × d (random) matrices and the functions x → Ax + B. States are determined by the rank of a random matrix. This problem is equivalent to estimating the time T until the product of T random matrices is 0.