# Random <br> Compositions 

Evangelos Kranakis<br>School of Computer Science<br>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, $[n]=\{1,2, \ldots, n\}$.
- $\mathcal{F}(m, n)$ is the space of functions $f:[m] \rightarrow[n]$.
- $\mathcal{F}(n):=\mathcal{F}(n, n)$.
- 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?

## Values of Random Functions

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

$$
\begin{aligned}
\operatorname{Pr}[f(x)=y] & =\frac{n^{m-1}}{n^{m}} \\
& =\frac{1}{n}
\end{aligned}
$$

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

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

for $k=0,1, \ldots, n$.

## Graph of a Random Function

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

$$
O_{f}(x):=\left\{x, f(x), f(f(x)), \ldots, f^{n-1}(x)\right\}
$$

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

$$
\begin{aligned}
\operatorname{Pr}\left[\mathcal{L}_{n}=j\right] & =\sum_{k=j}^{n} \frac{(n)_{k}}{n^{k+1}}, j=1, \ldots, n \\
E\left[\mathcal{L}_{n}\right] & =\sum_{j=1}^{n} j \sum_{k=j}^{n} \frac{(n)_{k}}{n^{k+1}}
\end{aligned}
$$

## 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

$$
\begin{aligned}
\operatorname{Pr}\left[\mathcal{K}_{n}=j\right] & =\sum_{k=j}^{n}\binom{n-1}{k-1} \frac{S(k, j)}{n^{k}}, j=1,2, \ldots, n \\
E\left[\mathcal{K}_{n}\right] & =\frac{1}{2} \log n(1+o(1)) \\
\operatorname{Var}\left[\mathcal{K}_{n}\right] & =\frac{1}{2} \log n(1+o(1))
\end{aligned}
$$

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

$$
\begin{aligned}
\operatorname{Pr}\left[\mathcal{Z}_{n}=k\right] & =\frac{k(n)_{k}}{n^{k+1}}, k=0,1, \ldots, n \\
E\left[\mathcal{Z}_{n}\right] & =\sqrt{\frac{\pi n}{2}}(1+o(1)) \\
\operatorname{Var}\left[\mathcal{Z}_{n}\right] & =\left(2-\frac{\pi}{2}\right) n(1+o(1))
\end{aligned}
$$

where $(n)_{k}=n(n-1) \cdots(n-k+1)$.

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

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

## Random Functions Cause Shrinkage!

Given a random function $f \in \mathcal{F}(n)$,

$$
E[|\operatorname{Range}(f)|]=n-n\left(1-\frac{1}{n}\right)^{n} \approx n(1-1 / e) .
$$



## What Causes Shrinkage?

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

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

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 $|\operatorname{Range}(f)|$ for a Random Function $f:[m] \rightarrow[n]$
Let $X$ be the r.v. that counts the size of Range $(f)$ and $U=n-X$. Consider the indicator function $I_{i}^{\prime}: I_{i}^{\prime}=1$ if $i \notin \operatorname{Range}(f)$, and $I_{i}^{\prime}=1$, otherwise. Observe that

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

## 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 $F: S \rightarrow S$.
- There is a probability distribution $\mu$ 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 $X_{t+1}=F\left(X_{t}\right)$ where $F$ is a random function $F \in \mathcal{F}$.


## Example I: Linear Affine Functions

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

$$
\begin{aligned}
& F_{+}: \quad \mathbf{R} \rightarrow \mathbf{R}: x \rightarrow F_{+}(x)=a x+1 \\
& F_{-}:
\end{aligned} \quad \mathbf{R} \rightarrow \mathbf{R}: x \rightarrow F_{-}(x)=a x-1, ~ l
$$

where $0<a<1$.

- The probability distribution $\mu$ on $\mathcal{F}$ is $\mu\left(F_{+}\right)=\mu\left(F_{-}\right)=1 / 2$. Let $\xi_{i}= \pm 1$ with probability $1 / 2$, respectively.
- The process starts with $\xi_{0}$ and inductively defines $X_{t+1}=F\left(X_{t}\right)$ where $F$ is a random function $F \in \mathcal{F}$.
- Clearly, $X_{t+1}=a X_{t}+\xi_{t}$ 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

$$
F: \mathbf{R}^{d} \rightarrow \mathbf{R}^{d}: x \rightarrow F(x)=A x+B
$$

$A$ 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 $\mu$ 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 $\left(A_{t}, B_{t}\right)$ 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 $\left\{o_{f}: f \in \mathcal{F}(n)\right\}$ maps the state space into itself as follows: Given a function $g \in \mathcal{F}(n)$ already in state $s$,

$$
s \rightarrow \circ_{f}(s):=\text { size of range of } f \circ g .
$$

- The probability distribution on $\left\{o_{f}: f \in \mathcal{F}(n)\right\}$ 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 $X_{t+1}=\circ_{f}\left(X_{t}\right)$ 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 $f:[n] \rightarrow[n / 2]$.
- A family of functions $\left\{o_{h}: h \in \mathcal{H}(n)\right\}$ maps the state space into itself as follows: Given a function $g \in \mathcal{H}(n)$ already in state $s$,

$$
s \rightarrow \circ_{h}(s):=\text { size of range of } h \circ g
$$

- The probability distribution on $\left\{{ }_{o}: h \in \mathcal{H}(n)\right\}$ 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 $X_{t+1}=o_{h}\left(X_{t}\right)$ 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 $\left|\operatorname{Range}\left(f^{(t)}\right)\right|=r$.
- $\tau_{r}=\left|\left\{t:\left|\operatorname{Range}\left(f^{(t)}\right)\right|=r\right\}\right|$ 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.,

$$
T=\sum_{r=1}^{n} \tau_{r}
$$


E. Kranakis, Fall 2004

## 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: $E[T]=2 n(1+o(1))$, as $n \rightarrow \infty$.


## 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 $\binom{n}{j}$ 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 2 nd kind, and
- $n^{n-i}$ ways to map them into $[n]$.

It follows that

$$
p(i, j):=\operatorname{Pr}\left[f^{(t)} \in s_{j} \mid f^{(t-1)} \in s_{i}\right]=\binom{n}{j} \frac{S(i, j) j!}{n^{i}}
$$

## 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 $S(i, j) \leq(2 j)^{i}$.
- For $i=1: S(1, j) \leq 2 j$
- We have that

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

## Eigenvalues of the Transition Matrix

The transition matrix $P:=\left(p(i, j)_{i, j}\right.$ is lower diagonal.
Eigenvalues are the diagonal elements of the matrix, i.e.,

$$
\begin{aligned}
\lambda_{r} & =p(r, r) \\
& =\binom{n}{r} \frac{S(r, r) r!}{n^{r}} \\
& =\prod_{i=1}^{r-1}\left(1-\frac{i}{n}\right) \\
& =1-\frac{\binom{r}{2}}{n}+O\left(\frac{r^{4}}{n^{2}}\right) .
\end{aligned}
$$

Note: $1>1-\frac{1}{n}=\lambda_{2} \geq \cdots \geq \lambda_{n}>0$ and $\frac{1}{1-\lambda_{r+d}} \leq n-1$

## 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 $\operatorname{rank}\left(A^{(t)}\right)=r$.
- $\tau_{r}=\left|\left\{t: \operatorname{rank}\left(A^{(t)}\right)=r\right\}\right|$ is the amount of time in state $s_{r}$.
- Open Question: For $j \leq i$, compute the transition probabilities

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

This is equivalent to computing the number of matrices $B$ such that $\operatorname{rank}(B A)=j$, given that $\operatorname{rank}(A)=i$.

## Lower Bound on $E[T]$

We have the identity

$$
\begin{aligned}
E[T] & =E\left[\sum_{r=2}^{n} \tau_{r}\right] \\
& =\sum_{r=2}^{n} E\left[\tau_{r}\right] \\
& =\sum_{r=2}^{n} E\left[\tau_{r} \mid \tau_{r}>0\right] \cdot \operatorname{Pr}\left[\tau_{r}>0\right]
\end{aligned}
$$

It remains

- to compute $E\left[\tau_{r} \mid \tau_{r}>0\right]$, and
- give a lower bound on $\operatorname{Pr}\left[\tau_{r}>0\right]$.


## Computing $E\left[\tau_{r} \mid \tau_{r}>0\right]$

This is the expected amount of time you stay in state $\tau_{r}$, given that you visit it?

Given $\tau_{r}>0, \tau_{r}$ follows the geometric distribution, with probability of success $p(r, r)=\lambda_{r}$.

$$
\begin{aligned}
E\left[\tau_{r} \mid \tau_{r}>0\right] & =\sum_{t=1}^{\infty} t \lambda_{r}^{t-1}\left(1-\lambda_{r}\right) \\
& =\frac{1}{1-\lambda_{r}} \\
& =\frac{n}{\binom{r}{2}}\left(1+O\left(\frac{r^{2}}{n}\right)\right)
\end{aligned}
$$

## Estimating $\operatorname{Pr}\left[\tau_{r}=0\right]$

We give an upper bound on $\operatorname{Pr}\left[\tau_{r}=0\right]$.

- If $\tau_{r}=0$ then the state $s_{r}$ is never visited.
- Therefore there must exist a transition

$$
s_{r+d} \longrightarrow s_{r-j}
$$

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 $\operatorname{Pr}\left[\tau_{r}=0\right]$

We can show that

$$
\begin{aligned}
\operatorname{Pr}\left[\tau_{r}=0\right] & =\sum_{d=1}^{n-r} \sum_{j=1}^{r-1} \operatorname{Pr}\left[\tau_{r+d}>0\right] \sum_{t=0}^{\infty} p(r+d, r-j) p(r+d, r+d)^{t} \\
& =\sum_{d=1}^{n-r} \sum_{j=1}^{r-1} \operatorname{Pr}\left[\tau_{r+d}>0\right] \frac{p(r+d, r-j)}{1-\lambda_{r+d}} \\
& \leq \sum_{d=1}^{n-r} \sum_{j=1}^{r-1}\binom{n}{r-j} \frac{S(r+d, r-j)(r-j)!}{n^{r+d}\left(1-\lambda_{r+d}\right)} \\
& \leq(n-1) \sum_{d=1}^{n-r} \frac{1}{n^{d}} \sum_{j=1}^{r-1} \frac{S(r+d, r-j)}{n^{j}}
\end{aligned}
$$

Recall: 1) $\left.\operatorname{Pr}\left[\tau_{r+d}>0\right] \leq 1,2\right) 1>1-\frac{1}{n}=\lambda_{2} \geq \cdots \geq \lambda_{n}>0$ and $\frac{1}{1-\lambda_{r+d}} \leq n-1$

Lower Bound on $\operatorname{Pr}\left[\tau_{r}>0\right]$
Hence we obtain for $r \leq\lfloor\log \log n\rfloor$

$$
\begin{aligned}
\operatorname{Pr}\left[\tau_{r}>0\right] & \geq 1-(n-1) \sum_{d=1}^{n-r} \frac{1}{n^{d}} \sum_{j=1}^{r-1} \frac{S(r+d, r-j)}{n^{j}} \\
& \geq 1-(n-1) \sum_{d=1}^{n-r} \frac{1}{n^{d}} \sum_{j=1}^{r-1} \frac{(2(r-j))^{r+d}}{n^{j}} \\
& \geq 1-(n-1) \sum_{d=1}^{n-r} \frac{1}{n^{d}} \frac{r(2 r)^{r+d}}{n} \\
& \geq 1-O\left(\frac{(2 \ell)^{\ell+2}}{n}\right) \\
& =1-o(1) .
\end{aligned}
$$

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

$$
\begin{aligned}
E[T] & \geq \sum_{r=1}^{\ell} E\left[\tau_{r} \mid \tau_{r}>0\right] \cdot \operatorname{Pr}\left[\tau_{r}>0\right] \\
& \geq \sum_{r=2}^{\ell} \frac{n}{\binom{r}{2}} \\
& =2 n \sum_{r=2}^{\ell} \frac{1}{r(r-1)} \\
& =2 n \sum_{r=2}^{\ell}\left(\frac{1}{r-1}-\frac{1}{r}\right) \\
& =2 n\left(1-\frac{1}{\ell}\right)
\end{aligned}
$$

This completes the proof of the lower bound.

## Idea of Proof of Upper Bound

Recall that $T=\sum_{m=2}^{n} \tau_{m}$. We split $E[T]$ into three sums

$$
\begin{aligned}
E[T] & =\sum_{m=2}^{n} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right] \\
& =\sum_{m=2}^{\xi_{1}} \cdots+\sum_{m=\xi_{1}+1}^{\xi_{2}} \cdots+\sum_{m=\xi_{2}+1}^{n} \cdots
\end{aligned}
$$

whereby

$$
\begin{aligned}
& \xi_{1}=\left\lfloor\sqrt{\frac{n}{\log n}}\right\rfloor \\
& \xi_{2}=\left\lfloor\frac{n}{\log ^{2} n}\right\rfloor
\end{aligned}
$$

and make upper bound estimates on each of them.

$$
\text { 1st Sum: }=\sum_{m=2}^{\xi_{1}} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right]
$$

Observe that

$$
\begin{aligned}
1 \text { st Sum } & \leq \sum_{m=2}^{\xi_{1}} \frac{1}{1-\lambda_{m}} \\
& =\sum_{m=2}^{\xi_{1}} \frac{1}{\frac{\binom{m}{2}}{n}+O\left(m^{4} / n^{2}\right)} \\
& =\sum_{m=2}^{\xi_{1}} \frac{n}{\binom{m}{2}+O\left(m^{4} / n\right)} \\
& \leq 2 n \sum_{m=2}^{\xi_{1}} \frac{1}{m(m-1)} \\
& =2 n\left(1-1 / \xi_{1}\right) \\
& =2 n(1+o(1))
\end{aligned}
$$

$$
\text { 2nd Sum: }=\sum_{m=\xi_{1}+1}^{\xi_{2}} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right]
$$

Observe that $\lambda_{\xi_{1}}=1-\frac{1}{2 \log n}+O\left(1 / \log ^{2} n\right)$.

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

$$
\text { 3rd Sum: }=\sum_{m=\xi_{2}+1}^{n} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right]
$$

Observe that $\max _{m>\xi_{2}} \frac{1}{1-\lambda_{m}}=\frac{1}{1-\lambda_{\xi_{2}+1}}$. Hence,

$$
\begin{aligned}
3 \text { rd Sum } & \leq \sum_{m=\xi_{2}+1}^{n} \frac{1}{1-\lambda_{m}} \operatorname{Pr}\left[\tau_{m}>0\right] \\
& \leq \frac{1}{1-\lambda_{\xi_{2}+1}}\left(\sum_{m=\xi_{2}+1}^{n} \operatorname{Pr}\left[\tau_{m}>0\right]\right) \\
& \leq \frac{1}{1-\exp \left(-\binom{\xi_{2}}{2} / n\right)}\left(\sum_{m=\xi_{2}+1}^{n} \operatorname{Pr}\left[\tau_{m}>0\right]\right) \\
& \leq 2 \sum_{m=\xi_{2}+1}^{n} \operatorname{Pr}\left[\tau_{m}>0\right]
\end{aligned}
$$

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

## On Skipping Large States

To deal with $\sum_{m=\xi_{2}+1}^{n} \operatorname{Pr}\left[\tau_{m}>0\right]$ we will show that "every hit (i.e., $\tau_{m}>0$ )" is followed by "many (i.e., $\beta$ ) misses (i.e., $\tau_{m-\delta}=0$, forall $\delta$ 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

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

Suppose we are in state $s_{m}$ at time $t-1$, i.e., $\left|\operatorname{Range}\left(f^{(t-1)}\right)\right|=m$, and select the next function $f_{t}$ at random.

Claim 1: If $B>\beta$ then $\tau_{m-\delta}=0$, for $1 \leq \delta \leq \beta$.
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

$$
\begin{aligned}
B>\beta & \Rightarrow m-R>\beta \\
& \Rightarrow R<m-\beta
\end{aligned}
$$

and

$$
\begin{aligned}
\left|\operatorname{Range}\left(f^{(t)}\right)\right| & =\left|f^{(t)}([n])\right| \\
& =\left|f_{t}\left(f^{(t-1)}([n])\right)\right| \\
& =\left|h\left(\operatorname{Range}\left(f^{(t-1)}\right)\right)\right| \\
& =R .
\end{aligned}
$$

$$
\text { Claim 2: } E[B] \geq \xi_{2}-n+n\left(1-\frac{1}{n}\right)^{\xi_{2}} \gg \frac{n}{\log ^{4} n}
$$

Observe that

$$
\begin{aligned}
E[B] & =E[m-R] \\
& =m-E[R] \\
& =m-n+n\left(1-\frac{1}{n}\right)^{m} \\
& >\xi_{2}-n+n\left(1-\frac{1}{n}\right)^{\xi_{2}} \\
& =2 \beta \\
& >\frac{n}{\log ^{4} n}
\end{aligned}
$$

Also recall, $\operatorname{Var}(B)=$
$n^{2}\left(\left(1-\frac{2}{n}\right)^{m}-\left(1-\frac{1}{n}\right)^{2 m}\right)+n\left(\left(1-\frac{1}{n}\right)^{m}-\left(1-\frac{2}{n}\right)^{m}\right)=O(m)$.

$$
\text { Claim 4: } \operatorname{Pr}\left[\forall 1 \leq \delta \leq \beta\left(\tau_{m-\delta}=0\right) \mid \tau_{m}>0\right]
$$

By Chebeshev's Inequality and since $m>\xi_{2}$ we have

$$
\begin{aligned}
\operatorname{Pr}[|B| \leq \beta] & \leq \operatorname{Pr}\left[|B| \leq \frac{1}{2} E[B]\right] \\
& \leq \frac{4 \operatorname{Var}(B)}{(E[B])^{2}} \\
& =O\left(\frac{m \log ^{8} n}{n^{2}}\right) \\
& =o(1)
\end{aligned}
$$

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

$$
\operatorname{Pr}\left[\forall 1 \leq \delta \leq \beta\left(\tau_{m-\delta}=0\right) \mid \tau_{m}>0\right]=1-o(1)
$$

Back to the 3rd Sum: $=\sum_{m=\xi_{2}+1}^{n} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right]$ Define $\chi_{m}=1$ if $\tau_{m}>0$ and $\chi_{m}=0$, otherwise. Recall that

$$
\begin{aligned}
3 \text { rd Sum } & \leq 2 \sum_{m=\xi_{2}+1}^{n} \operatorname{Pr}\left[\tau_{m}>0\right] \\
& =2 \sum_{m=\xi_{2}+1}^{n} E\left[\chi_{m}\right] \\
& =2 E\left[\sum_{m=\xi_{2}+1}^{n} \chi_{m}\right] \\
& =2 E[V]
\end{aligned}
$$

where $V:=\sum_{m=\xi_{2}+1}^{n} \chi_{m}$. Define $W:=\sum_{m=\xi_{2}+1}^{n}\left(1-\chi_{m}\right)$ and observe that $V+W=n-\xi_{2}$.

Back to the 3rd Sum: $=\sum_{m=\xi_{2}+1}^{n} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right]$ Note that if $\tau_{m}>0$ and $\forall 1 \leq \delta \leq \beta\left(\tau_{m-\delta}=0\right)$ then these $\beta$ missed states contribute exactly $\beta$ to $W$. Hence, if we define $J_{m}=\chi_{m} \cdot \prod_{\delta=1}^{\beta}\left(1-\chi_{m-\delta}\right)$ then $W \geq \beta \sum_{m>\xi_{2}} J_{m}$. Hence,

$$
\begin{aligned}
E[W] & \geq \beta \sum_{m>\xi_{2}} E\left[J_{m}\right] \\
& =\beta \sum_{m>\xi_{2}} \operatorname{Pr}\left[J_{m}=1\right] \\
& =\beta \sum_{m>\xi_{2}} \operatorname{Pr}\left[\tau_{m}>0\right] \operatorname{Pr}\left[\forall 1 \leq \delta \leq \beta\left(\tau_{m-\delta}=0\right) \mid \tau_{m}>0\right] \\
& \geq \beta(1+o(1)) \sum_{m>\xi_{2}} \operatorname{Pr}\left[\tau_{m}>0\right] \\
& =\beta(1+o(1)) E[V]
\end{aligned}
$$

Back to the 3rd Sum: $=\sum_{m=\xi_{2}+1}^{n} E\left[\tau_{m} \mid \tau_{m}>0\right] \cdot \operatorname{Pr}\left[\tau_{m}>0\right]$
It follows that

$$
\begin{aligned}
E[V] & \left.=n-\xi_{2}-E[W]\right) \\
& \left.\leq n-\xi_{2}-\beta(1+o(1)) E[V]\right)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
3 \text { rd Sum } & \leq 2 E[V] \\
& \leq \frac{2\left(n-\xi_{2}\right)}{1+\beta(1+o(1))} \\
& =O\left(\log ^{4} n\right) \\
& =o(n) .
\end{aligned}
$$

This completes the proof of the theorem.

## Another Idea: Hitting Times

For $r=1,2, \ldots, n$, 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 $t_{r}=1_{r=1}+\sum_{r^{\prime}=2}^{r} p\left(r, r^{\prime}\right) t_{r^{\prime}}$
- If $I$ is the unit matrix and $t$ is the vector of hitting times then $t=I+Q t$, 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 $\lim _{k \rightarrow \infty} Q^{k}=\mathbf{0}$ then $(I-Q)^{-1}=\sum_{k=0} Q^{k}$. 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 $d \times d$ (random) matrices and the functions $x \rightarrow A x+B$. 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 $\mathbf{0}$.

