

Cutoff for Random Iterated Averages

Alan Yan

Junior Paper Advisor: Professor Allan Sly

April 25, 2022

Abstract

In this paper, we consider a generalization of the iterated averages Markov chain defined in [1]. Instead of restricting to binary averages, we allow for random, but bounded, averages at each step. In particular, we prove Theorem 2.1 which gives cutoff for the L^1 distance.

Contents

1	Introduction	2
2	Random Bounded Averages	4
2.1	Explicit Formulae for the Squared Distance to Mean	4
2.2	Coupling with Fragmentation and Branching Processes	6
2.3	Logarithmic Length of Interval Trajectories	8
2.4	Proof of Theorem 2.1	10
3	Future Work and Open Questions	16
3.1	Unbounded Random Averages	17
3.2	Maximum Decay over All Starting Points	17
3.3	Second Order Behavior of Random Averages	17
3.4	Stationary Distribution of Normalized Markov Chain	17
	Acknowledgements	17
	Honor Code	17
	References	17

1 Introduction

In [1], Chatterjee, Diaconis, Sly, and Zhang analyze the Markov chain defined in Definition 1.1 which evolves through binary averages.

Definition 1.1 (Binary Markov Chain). *Let $\mathbf{B} = \{x_t\}_{t \in \mathbb{Z}_{\geq 0}}$ where $x_0 \in \mathbb{R}$ is some initial vector in \mathbb{R}^n . The vector x_{k+1} is determined picking two distinct coordinates of x_k uniformly at random and replacing both of them by their average.*

The binary Markov chain in Definition 1.1 traces a discrete trajectory on the hyperplane $\{x \in \mathbb{R}^n : \langle x - x_0, \mathbf{1} \rangle = 0\}$ in \mathbb{R}^n where $\mathbf{1}$ is the vector of all ones. As $t \rightarrow \infty$, the Markov chain will converge to $x^\mu \cdot \mathbf{1}$ almost surely where $x^\mu = \frac{1}{n} \langle x_0, \mathbf{1} \rangle$ is the average of the coordinates of x_0 . In this paper, we will be interested in the rate of convergence to the average for generalizations of this Markov chain. Then, we consider two metrics to measure the distance from the average. These two metrics correspond to the L^1 and L^2 distances. Following [1], we define the functions $T_{\mathbf{B}}$ and $S_{\mathbf{B}}$ as

$$T_{\mathbf{B}}(k) := \sum_{i=1}^n |x_{k,i} - x^\mu|, \quad S_{\mathbf{B}}(k) := \sum_{i=1}^n |x_{k,i} - x^\mu|^2.$$

Note that $T_{\mathbf{B}}$ is defined using the L^1 norm exactly while $S_{\mathbf{B}}$ is defined using the L^2 norm squared. We are now interested in the rate of convergence and the existence of a cutoff time. In [1], an explicit formula for the expected L^2 distance is found and is given in Proposition 1.1.

Proposition 1.1 (Proposition 2.1 in [1]). *For any $k \geq 0$, we have*

$$\mathbb{E}[S_{\mathbf{B}}(k+1) | \mathcal{F}_k] = \left(1 - \frac{1}{n-1}\right) S_{\mathbf{B}}(k).$$

In Proposition 1.1, we let \mathcal{F}_k be the sigma algebra which consists of all of the information up to time k . We keep this convention throughout the paper. From the formula in Proposition 1.1, the behavior of the L^2 distance S is well-understood. It also provides a way to prove for fixed n that the Markov chain converges to the average almost surely. Before proving this fact, we first prove Lemma 1.1 which is an elementary fact from probability theory.

Lemma 1.1. *Let $\{Z_n\}_{n \in \mathbb{N}_0}$ be a sequence of decreasing non-negative random variables. If $\mathbb{E}[Z_n] \rightarrow 0$, then $Z_n \rightarrow 0$ almost surely.*

Proof. If $\mathbb{E}[Z_n] \rightarrow 0$, then $Z_n \rightarrow 0$ in probability by Markov's inequality. From a standard argument, convergence in probability implies that there is a subsequence which converges almost surely. This limit point must evidently be 0. Moreover, since the sequence is decreasing, this implies that the whole sequence must converge to 0 almost surely. \square

Proposition 1.2. *For fixed n , we have that $x_k \rightarrow x^\mu \cdot \mathbf{1}$ almost surely as $k \rightarrow \infty$ where $\mathbf{1}$ is the vector of all ones.*

Proof. From the formula in Proposition 1.1, we have that

$$\mathbb{E}[\|x_k - x^\mu \cdot \mathbf{1}\|_2^2] = c \left(1 - \frac{1}{n-1}\right)^k$$

where $c > 0$ is some constant depending on x_0 . Taking $k \rightarrow \infty$, we have $\mathbb{E}[\|x_k - x^\mu \cdot \mathbf{1}\|_2^2] \rightarrow 0$. From Lemma 1.1 and that $S_{\mathbf{B}}$ is monotonically decreasing we have proven the result. \square

The main result [1] was concerning cutoff of the L^1 norm. In particular, the main theorem in [1] given in Theorem 1.1 finds the second order behavior of the L^1 distance.

Theorem 1.1 (Theorem 1.2 in [1]). *Let $\Phi : \mathbb{R} \rightarrow [0, 1]$ be the cumulative distribution function of the standard normal distribution $N(0, 1)$. For $x_0 = (1, 0, \dots, 0)$, and any $a \in \mathbb{R}$, as $n \rightarrow \infty$ we have*

$$T(\lfloor n(\log_2(n) + a\sqrt{\log_2(n)}) \rfloor) \xrightarrow{\mathbb{P}} 2\Phi(-a).$$

By weakening the result in Theorem 1.1, Corollary 1.1 follows easily and has a more succinct statement.

Corollary 1.1 (Theorem 1.1 in [1]). *For $x_0 = (1, 0, \dots, 0)$ as $n \rightarrow \infty$ we have $T(\theta n \log n) \xrightarrow{\mathbb{P}} 2$ for $\theta < 1/(2 \log 2)$ and $T(\theta n \log n) \xrightarrow{\mathbb{P}} 0$ for $\theta > 1/(2 \log 2)$.*

One of the main techniques in the proof of Theorem 1.1 was to poissonize the process in order to avoid issues of dependence in the discrete time setting. This poissonized version of the Markov chain is described in Definition 1.2.

Definition 1.2 (Poissonized Binary Markov Chain). *Let $\mathbf{B}' = \{x_t\}_{t \in \mathbb{R}_{\geq 0}}$ where $x_0 \in \mathbb{R}$ is some initial vector in \mathbb{R}^n . We can associate an independent Poisson clock of rate $\binom{n}{2}^{-1}$ to each pair of coordinates. When one of the clocks ring, average the corresponding coordinates.*

By giving each pair of coordinates an independent clock of rate $\binom{n}{2}^{-1}$, we equivalently have a single clock of rate 1 which, upon ringing, picks two distinct coordinates uniformly at random and replaces them by their average. Thus, \mathbf{B}' seems to be suitable approximation of \mathbf{B} . Let $T'_\mathbf{B}$ and $S'_\mathbf{B}$ be the corresponding L^1 and L^2 distances from the average of \mathbf{B}' . Similar to the discrete case, there is still an explicit formula for the L^2 distance.

Proposition 1.3. *For any $t \geq s \geq 0$, we have that*

$$\mathbb{E}[S'_\mathbf{B}(t) | \mathcal{F}_s] = \exp\left(-\frac{t-s}{n-1}\right) S'_\mathbf{B}(s).$$

Proof. Let E_k be the event that there are k rings between times s and t . Then

$$\begin{aligned} \mathbb{E}[S'_\mathbf{B}(t) | \mathcal{F}_s] &= \sum_{k=0}^{\infty} \mathbb{E}[S'_\mathbf{B}(t) | \mathcal{F}_s \text{ and } E_k] \cdot \Pr[E_k] \\ &= \sum_{k=0}^{\infty} \left(1 - \frac{1}{n-1}\right)^k S'_\mathbf{B}(s) \cdot e^{-(t-s)} \cdot \frac{(t-s)^k}{k!} \\ &= \exp\left((t-s) \left(1 - \frac{1}{n-1}\right)\right) \cdot e^{-(t-s)} S'_\mathbf{B}(s) \\ &= \exp\left(-\frac{t-s}{n-1}\right) \cdot S'_\mathbf{B}(s). \end{aligned}$$

This suffices for the proof. □

[1] proves Theorem 1.2 which is a continuous version of Theorem 2.2.

Theorem 1.2 (Theorem 2.3 in [1]). *Take $\Phi : \mathbb{R} \rightarrow [0, 1]$ as the cdf of the standard normal distribution. For any $a \in \mathbb{R}$, we have $T'(n(\log_2(n) + a\sqrt{\log_2(n)})/2) \xrightarrow{\mathbb{P}} 2\Phi(-a)$ in probability as $n \rightarrow \infty$.*

It turns out that the approximation is robust enough to have Theorem 1.2 imply Theorem 1.1. We will prove this in the more general setting of Theorem 2.2. This completes the discussion about the motivating problem. In the sequel, we generalize the Markov chain to allow our averages to be of random (but bounded) size.

2 Random Bounded Averages

In this section, we will consider the same Markov chain as in Definition 1.2 except when we average, we will average some random number of coordinates. The random number of coordinates will be sampled from a fixed bounded distribution.

Definition 2.1. Let Q be a random variable supported in a finite set $\{2, 3, \dots, N\}$ such that $\Pr[Q = k] = q_k$ for all k , $2 \leq k \leq N$. Let $P_Q := \{x_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a Markov chain with $x_0 = (1, 0, \dots, 0)$. The Markov chain R_Q is equipped with a Poisson clock of rate 1. Whenever the clock rings, we independently sample a number m from the law of Q . We then uniformly pick m distinct coordinates and replace them by their average. Let S and T be the L^2 and L^1 distances to the mean.

Note that in Definition 2.1, we set the initial value as $x_0 = (1, 0, \dots, 0)$. But, according to an argument in [1], we do not lose much generality in this choice because the Markov chain takes the longest to converge at this choice. Moreover, even if we did choose a different starting position, the main ideas of the argument do not change.

2.1 Explicit Formulae for the Squared Distance to Mean

It will be useful to consider a less general version of Definition 2.1 but more general than Definition 1.1. Instead of averaging two elements every time, we average some fixed number of coordinates m .

Definition 2.2. Let B_m be the Markov chain $\{x_t\}_{t \in \mathbb{Z}_{\geq 0}} \subset \mathbb{R}^n$ where x_0 is some initial vector in \mathbb{R}^n and x_{k+1} is obtained from x_k by uniformly picking m distinct coordinates at random and replacing them by their average. Let B'_m be the poissonized version of B_m .

Let S_m and S'_m be the corresponding L^2 distances from the mean.

Proposition 2.1. For any $k \geq 0$, we have

$$\mathbb{E}[S_m(k+1)|\mathcal{F}_k] = \left(1 - \frac{m-1}{n-1}\right) S_m(k).$$

Proof. Witout loss of generality, we can assume that the average of the coordinates is 0. Then, we have

$$\begin{aligned} \mathbb{E}[S_m(k+1)|\mathcal{F}_k] &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \left\{ S_m(k) - \sum_{j=1}^m x_{k,i_j}^2 + m \left(\frac{x_{k,i_1} + \dots + x_{k,i_m}}{m} \right)^2 \right\} \\ &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \left\{ S_m(k) - \sum_{j=1}^m x_{k,i_j}^2 + \frac{1}{m} \sum_{j=1}^m x_{k,i_j}^2 + \frac{1}{m} \sum_{1 \leq u < v \leq m} x_{k,i_u} x_{k,i_v} \right\} \\ &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \left\{ S_m(k) + \frac{1-m}{m} \sum_{j=1}^m x_{k,i_j}^2 + \frac{1}{m} \sum_{1 \leq u < v \leq m} x_{k,i_u} x_{k,i_v} \right\} \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3 \end{aligned}$$

where

$$\begin{aligned}\Sigma_1 &:= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} S_m(k) \\ \Sigma_2 &:= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \frac{1-m}{m} \sum_{j=1}^m x_{k,i_j}^2 \\ \Sigma_3 &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \frac{1}{m} \sum_{1 \leq u < v \leq m} 2x_{k,i_u} x_{k,i_v}.\end{aligned}$$

For the first sum, we immediately get $\Sigma_1 = S_m(k)$. For the second sum, we get

$$\Sigma_2 = \binom{n}{m}^{-1} \cdot \frac{1-m}{m} \cdot \binom{n-1}{m-1} S_m(k) = \frac{1-m}{n} S_m(k).$$

For the last sum, we get

$$\Sigma_3 = \binom{n}{m}^{-1} \cdot \frac{1}{m} \cdot \binom{n-2}{m-2} \sum_{1 \leq u < v \leq n} 2x_{k,u} x_{k,v} = \frac{1-m}{n(n-1)} S_m(k)$$

where we used the fact that

$$S_m(k) + \sum_{1 \leq u < v \leq n} 2x_{k,u} x_{k,v} = (x_{k,1} + \dots + x_{k,n})^2 = 0.$$

Then

$$\mathbb{E}[S_m(k+1)|\mathcal{F}_k] = \left(1 + \frac{1-m}{n} + \frac{1-m}{n(n-1)}\right) S_m(k) = \left(1 - \frac{m-1}{n-1}\right) S_m(k).$$

□

Corollary 2.1. *For $0 \leq s < t$, we have*

$$\mathbb{E}[S'_m(t)|\mathcal{F}_s] = \exp\left(-\frac{(t-s)(m-1)}{n-1}\right) S'_m(s).$$

Proof. We have

$$\begin{aligned}\mathbb{E}[S'_m(t)|\mathcal{F}_s] &= \sum_{k=0}^{\infty} \Pr[k \text{ rings}] \cdot \mathbb{E}[S'_m(t)|\mathcal{F}_s \text{ and } k \text{ rings}] \\ &= \sum_{k=0}^{\infty} \frac{(t-s)^k}{k!} \cdot e^{-(t-s)} \cdot \left(1 - \frac{m-1}{n-1}\right)^k S'_m(s) \\ &= \exp\left(-\frac{(t-s)(m-1)}{n-1}\right) S'_m(s).\end{aligned}$$

□

Proposition 2.1 and Corollary 2.1 will allow us to easily compute the L^2 distance for the random bounded averages case.

Proposition 2.2. *Let $S_Q(k)$ be the L^2 distance from mean squared at time k of the discrete version of P_Q . Then*

$$\mathbb{E}[S_Q(k+1)|\mathcal{F}_k] = \left(1 - \frac{\mu-1}{n-1}\right) S_Q(k).$$

Proof. We have

$$\begin{aligned} \mathbb{E}[S_Q(k+1)|\mathcal{F}_k] &= \sum_{m=2}^N \Pr[m \text{ averaging}] \cdot \mathbb{E}[S_Q(k+1)|\mathcal{F}_k \text{ and } m \text{ averaging}] \\ &= \sum_{m=2}^N q_m \cdot \left(1 - \frac{m-1}{n-1}\right) S_Q(k) \\ &= \left(1 + \frac{1}{n-1} - \frac{1}{n-1} \sum_{m=2}^N m q_m\right) S_Q(k) \\ &= \left(1 + \frac{1}{n-1} - \frac{\mu}{n-1}\right) S_Q(k) \\ &= \left(1 - \frac{\mu-1}{n-1}\right) S_Q(k). \end{aligned}$$

□

Corollary 2.2. *Let $S(t)$ be the L^2 distance at time t of P_Q . Then*

$$\mathbb{E}[S(t)|\mathcal{F}_s] = \exp\left(-\frac{(t-s)(\mu-1)}{n-1}\right) S(s).$$

Proof. The proof is very similar to the proof of Corollary 2.1 now that we have Proposition 2.2. We can compute

$$\begin{aligned} \mathbb{E}[S(t)|\mathcal{F}_s] &= \sum_{k=0}^{\infty} \Pr[k \text{ rings}] \cdot \mathbb{E}[S(t)|\mathcal{F}_s \text{ and } k \text{ rings}] \\ &= \sum_{k=0}^{\infty} \frac{(t-s)^k}{k!} \exp(-(t-s)) \cdot \left(1 - \frac{\mu-1}{n-1}\right)^k S(s) \\ &= \exp\left(-\frac{(t-s)(\mu-1)}{n-1}\right) S(s). \end{aligned}$$

□

2.2 Coupling with Fragmentation and Branching Processes

We recall the definition of our original process for the random bounded averages case and we define two new processes, the first of which is a type of fragmentation process and the second of which is a type of branching process. From now on, let $t := \gamma n(\log n - (\log n)^{3/4})$, $\gamma := (\mathbb{E}[Q \log Q])^{-1}$, and $\mu := \mathbb{E}[Q]$. Define the constants

$$a_n := n^{-1+\alpha(n)}, b_n := n^{-1+\beta(n)}, \alpha(n) := \frac{1}{2}(\log n)^{-1/4}, \text{ and } \beta(n) := 2(\log n)^{-1/4}.$$

In particular, the constants are chosen so that $n^{-1} \ll a_n < b_n = n^{-1+o(1)}$.

Definition 2.3 (Original Process). Let $\mathbf{P} := \{x_t\}$ be the Markov chain where $x_0 = (1, 0, \dots, 0) \in \mathbb{R}^n$. We have a Poisson clock of rate 1. Everytime the clock rings, we pick some random coordinates to average where the random number is sampled randomly from Q .

Definition 2.4 (Fragmentation Process). Let $\mathbf{P}_{\text{FRAG}} := \{\bar{x}_t\}$ be the Markov chain where $\bar{x}_0 = (1, 0, \dots, 0) \in \mathbb{R}^n$. It evolves in the same way as $\{x_t\}$ except when we average at least two positive coordinates, we throw away all of the coordinates. Moreover, when intervals become smaller than a_n , we also throw them away. There are also other moments when we throw away coordinates according to the coupling with Definition 2.5 which we define below when describing the coupling. This is a fragmentation process version of the original process where we throw out piles under certain rules.

Definition 2.5 (Branching Process). Let $\mathbf{P}_{\text{BRANCH}} := \{I_t\}$ be a Markov chain where $I_0 = \{[0, 1]\}$. Each I_t will consist of a collection of intervals which partition $[0, 1]$. To each interval of I_t , we attach a clock of rate μ/n and when it rings we split that interval up into a random number of equal parts according to the law of Q' with distribution $(2q_2/\mu, \dots, Nq_N/\mu)$. In particular, when an interval's clock rings, we choose to split it into k equal parts for $2 \leq k \leq N$ with probability q_k/μ .

We now describe how we couple the processes \mathbf{P} , \mathbf{P}_{FRAG} , and $\mathbf{P}_{\text{BRANCH}}$. The process \mathbf{P}_{FRAG} will act as an approximation of \mathbf{P} , and $\mathbf{P}_{\text{BRANCH}}$ an approximation of \mathbf{P}_{FRAG} . Our approximations will allow us to transfer information about the L^1 convergence. We now describe the coupling between \mathbf{P} , \mathbf{P}_{FRAG} , and $\mathbf{P}_{\text{BRANCH}}$.

(The Coupling) We couple \mathbf{P} and \mathbf{P}_{FRAG} in the obvious way: whenever we average some piles on \mathbf{P} , we average the same exact piles in \mathbf{P}_{FRAG} while keeping the removal rules consistent. The coupling between \mathbf{P}_{FRAG} and $\mathbf{P}_{\text{BRANCH}}$ is a bit more complicated. In order to couple the two processes, we attach to I_t a labelling function $\pi_t : I_t \rightarrow [n] \cup \{0\}$ which assigns labels to intervals. The labelling function will always be injective on the intervals which are not given the label 0. If $I \in I_t$ satisfies $\pi_t(I) = k$ for some $k \in [n]$, the interval I “represents” the pile k and will satisfy $|I| = \bar{x}_{t,k}$. The coupling between \mathbf{P}_{FRAG} and $\mathbf{P}_{\text{BRANCH}}$ will also describe how the labelling function π_t evolves. For each interval $I \in I_t$ and subset $S \subseteq [n]$ of $k-1$ elements, where k ranges from 2 to N , we can define an independent Poisson clock $N_{I,S}$ of rate $q_k \binom{n}{k}^{-1}$ which signals for I to split into k equal parts. Then to the interval I , we can attach the Poisson clock

$$N_I = \begin{cases} \sum_{k=2}^N \sum_{S \in \binom{[n]}{k-1} : \pi_t(I) \cap S = \emptyset} N_{I,S} & \text{when } \pi_t(I) \neq 0 \\ \sum_{k=2}^N \sum_{S \in \binom{[n-1]}{k-1}} N_{I,S} & \text{when } \pi_t(I) = 0. \end{cases}$$

where the size of the splitting is given by the specific clock that rings. Here, $[n]$ refers to the set $\{1, \dots, n\}$ and the notation $\binom{S}{k}$ for a set S refers to the collection of subsets of S of size k . We want I to evolve as described in Definition 2.5 but we want to phrase it in a way to facilitate a coupling with \mathbf{P}_{FRAG} . Thus, in the case $\pi_t(I) \neq 0$, we need one clock for every possible combination of averages which include the pile corresponding to I . In the case of $\pi_t(I) = 0$, it doesn't matter which clocks we associated (as long as they have the correct weights) since the interval is not associated to any pile. Attached to \bar{x}_t , for every $S \subseteq [n]$ with $|S| = k \in \{2, \dots, N\}$ we have a clock \bar{N}_S of rate $q_k \binom{n}{k}^{-1}$. When the clock \bar{N}_S rings, we perform an averaging on the coordinates S on \bar{x}_t . This is also the same clock used in \mathbf{P} according to the coupling between \mathbf{P} and \mathbf{P}_{FRAG} . We now construct the following coupling: if $S' = x \sqcup S$ where $|S| = k-1$ and $\bar{x}_{t,x} \neq 0$ while $\bar{x}_{t,s} = 0$ for all $s \in S$, then we couple $\bar{N}_{S'}$ with $N_{\pi_t^{-1}(x), S}$ where $\pi_t^{-1}(x)$ is the interval which is associated to the pile x . The only problem is that there may be times where an interval I is associated some label $\pi_t(I) = x$

and $S \subseteq [n] \setminus x$ has some non-zero piles and $N_{I,S}$ rings but the corresponding averging $\overline{N}_{x \cup S}$ does not happen. In this case, we set all coordinates of \overline{x}_t in positions $x \cup S$ to zero. This is the extra removal mentioned in Definition 2.4.

We now describe how the labelling function evolves. Whenever we perform a removal in \overline{x}_t , the corresponding labels are removed. Whenever we perform a valid averaging, we redistribute the labels in any suitable way. For example, we can redistribute them in increasing order from left to right. In this way, the interval and pile pairs always have the same length.

To prove that this coupling is valid, we need to prove that the marginal distributions are correct. By construction, in Definition 2.4, we defined the marginal distribution to be the one dictated by the averging in \mathbf{P} alongside the addition removal in $\mathbf{P}_{\text{BRANCH}}$. Thus, by definition, the marginal distribution is correct. Now, we need to prove that the martingale distribution in $\mathbf{P}_{\text{BRANCH}}$ is correct. But this is clear by looking at an individual interval. Suppose the interval is labelled by some pile. For any $k \in \{2, \dots, N\}$, we have the following clocks for I which dictate I to split into k equal parts:

$$\left\{ N_{I,S} : S \subseteq \binom{[n] \setminus \pi_t(I)}{k-1} \right\}.$$

There are $\binom{n-1}{k-1}$ such clocks each of rate $q_k \binom{n}{k}^{-1}$. This is equivalent to saying that I has one clock of rate

$$\frac{\binom{n-1}{k-1}}{\binom{n}{k}} q_k = \frac{k q_k}{n}$$

which dictates when the interval splits into k equal parts. The analysis for the case where I is not labelled is exactly the same, if not even simpler. Thus, the marginal distributions are correct and the coupling is valid.

2.3 Logarithmic Length of Interval Trajectories

In this section, we analyze the distribution of $|I_t^x|$ for any fixed $x \in [0, 1]$ where I_t^x is defined to be the (leftmost) interval in I_t which contains x . We first prove Lemma 2.1, which is a well-known result about Poisson concentration.

Lemma 2.1. *Suppose $X_t \sim \text{Poisson}(t)$ is a Poisson random variable of rate t . Then*

$$\frac{X_t - t}{\sqrt{t}} \longrightarrow N(0, 1)$$

in distribution as $t \rightarrow \infty$.

Proof. Let Y_n for $n \geq 1$ be iid Poisson random variables of rate 1 and let Z_t be a Poisson random variable of rate $\{t\}$. Then X_t has the same distribution as

$$X_t \sim \sum_{i=1}^{\lfloor t \rfloor} Y_i + Z_t.$$

Thus, we have that

$$\frac{X_t - t}{\sqrt{t}} = \frac{\sum_{i=1}^{\lfloor t \rfloor} Y_i + Z_t - t}{\sqrt{t}} = \frac{\sqrt{\lfloor t \rfloor}}{\sqrt{t}} \cdot \frac{\sum_{i=1}^{\lfloor t \rfloor} Y_i - \lfloor t \rfloor}{\sqrt{\lfloor t \rfloor}} + \frac{Z_t - \{t\}}{\sqrt{t}}.$$

Since $(Z_t - \{t\})/\sqrt{t} \rightarrow 0$ in probability, the Lemma follows from the Central Limit Theorem. \square

Lemma 2.1 will allow us to show that with high probability, $|I_t^x|$ lies in the interval (a_n, b_n) .

Lemma 2.2. *Let $x \in [0, 1]$ be a fixed real number. Then*

$$\lim_{n \rightarrow \infty} \Pr[|I_t^x| \in (a_n, b_n)] = 1.$$

Proof. Consider the process $\{I_s^x\}_{0 \leq s \leq t}$. In $\mathbf{P}_{\text{BRANCH}}$, this is the interval trajectory traced by the point x . To these intervals, we have a clock of rate μ/n . When the clock rings, we split into k equal segments with probability kq_k/μ . Equivalently, for each k , we can give the interval a clock of rate $\frac{\mu}{n} \cdot \frac{kq_k}{\mu} = \frac{kq_k}{n}$. When this clock rings, then we split the interval into k equal parts. When the k th clock rings, we multiply the length by $\frac{1}{k}$. Taking logarithms, we see that

$$\log |I_t^x| = \sum_{k=2}^N \log \left(\frac{1}{k} \right) \cdot \text{Poisson} \left(\frac{q_k k}{n} \cdot t \right)$$

in distribution. To reduce the clutter of the notation, let P_k be the independent Poisson clock of rate $\frac{kq_k}{n} \cdot t$. Then, we can rewrite

$$\begin{aligned} \log |I_t^x| &= \sum_{k=2}^N \log \left(\frac{1}{k} \right) \cdot \left(P_k - \frac{q_k k}{n} \cdot t \right) + \sum_{k=2}^N \log \left(\frac{1}{k} \right) \cdot \frac{q_k k}{n} \cdot t \\ &= \sqrt{\frac{t}{n}} \cdot \sum_{k=2}^N \log \left(\frac{1}{k} \right) \cdot \sqrt{q_k k} \cdot \left(\frac{P_k - \frac{q_k k}{n} \cdot t}{\sqrt{\frac{q_k k}{n} t}} \right) - \frac{t}{n} \mathbb{E}[Q \log Q]. \end{aligned}$$

Rearranging, we get that

$$\frac{\log |I_t^x| + \frac{t}{n} \mathbb{E}[Q \log Q]}{\sqrt{t/n}} = \sum_{k=2}^N \log \left(\frac{1}{k} \right) \cdot \sqrt{q_k k} \cdot \left(\frac{P_k - \frac{q_k k}{n} \cdot t}{\sqrt{\frac{q_k k}{n} t}} \right).$$

Lemma 2.1 then implies the following convergence in distribution:

$$\frac{\log |I_t^x| + \frac{t}{n} \mathbb{E}[Q \log Q]}{\sqrt{t/n}} \rightarrow N \left(0, \sum_{k=2}^N \log(k)^2 k q_k \right) = N(0, \mathbb{E}[Q(\log Q)^2]).$$

Then, we can compute

$$\Pr(|I_t^x| \in (a_n, b_n)) = \Pr \left(\frac{\log |I_t^x| + \frac{t}{n} \mathbb{E}[Q \log Q]}{\sqrt{t/n}} \in (A_n, B_n) \right)$$

where

$$\begin{aligned}
A_n &= \sqrt{\frac{n}{t}} \cdot \left(\log a_n + \frac{t}{n} \mathbb{E}[Q \log Q] \right) \\
&= \sqrt{\frac{n}{t}} (\log n)^{3/4} [(\log n)^{1/4} \alpha(n) - 1] \\
&= -\frac{1}{2} \sqrt{\frac{n}{t}} (\log n)^{3/4} \\
B_n &= \sqrt{\frac{n}{t}} \cdot \left(\log b_n + \frac{t}{n} \mathbb{E}[Q \log Q] \right) \\
&= \sqrt{\frac{n}{t}} (\log n)^{3/4} [(\log n)^{1/4} \beta(n) - 1] \\
&= \sqrt{\frac{n}{t}} (\log n)^{3/4}.
\end{aligned}$$

It is in this calculation where the importance of the constant γ appears. In particular, it allows us to get the correct asymptotics such that $A_n \rightarrow -\infty$ and $B_n \rightarrow \infty$. Indeed, we have that

$$\sqrt{\frac{n}{t}} (\log n)^{3/4} = \Omega \left(\frac{(\log n)^{3/4}}{(\log n)^{1/2}} \right) = \Omega((\log n)^{1/4})$$

which diverges to infinity as $n \rightarrow \infty$. Hence,

$$\begin{aligned}
\lim_{n \rightarrow \infty} \Pr(|I_t^x| \in (a_n, b_n)) &= \lim_{n \rightarrow \infty} \Pr \left(\frac{\log |I_t^x| + \frac{t}{n} \mathbb{E}[Q \log Q]}{\sqrt{t/n}} \in (A_n, B_n) \right) \\
&= \Pr [N(0, \mathbb{E}[Q(\log Q)^2]) \in (-\infty, \infty)] = 1.
\end{aligned}$$

This completes the proof of the lemma. □

We now have all the tools needed to prove Theorem 2.1.

2.4 Proof of Theorem 2.1

Proposition 2.3. *As $n \rightarrow \infty$,*

$$\sum_{I \in I_t} |I| \cdot \mathbf{1}_{|I| \in (a_n, b_n)} \rightarrow 1$$

in probability.

Proof. Note that $\sum_{I \in I_t} |I| \cdot \mathbf{1}_{|I| \in (a_n, b_n)} \leq 1$. Hence, it suffices to prove that its expected value converges to 1. Then, we have

$$\mathbb{E} \left[\sum_{I \in I_t} |I| \cdot \mathbf{1}_{|I| \in (a_n, b_n)} \right] = \mathbb{E} \left[\int_0^1 \mathbf{1}_{|I_t^x| \in (a_n, b_n)} dx \right] = \int_0^1 \mathbb{E}[\mathbf{1}_{|I_t^x| \in (a_n, b_n)}] dx$$

where the last equality follows from Fubini's Theorem. Note that

$$\mathbb{E}[\mathbf{1}_{|I_t^x| \in (a_n, b_n)}] = \Pr[|I_t^x| \in (a_n, b_n)].$$

Miraculously, the probability on the right is independent of the point x we pick! Thus, the expected value is

$$\mathbb{E} \left[\sum_{I \in I_t} |I| \cdot \mathbb{1}_{|I| \in (a_n, b_n)} \right] = \Pr[|I_t^x| \in (a_n, b_n)]$$

where $x \in [0, 1]$ is a fixed element. By taking the limit as $n \rightarrow \infty$, Lemma 2.2 finishes the proof of the proposition. \square

Our next step is to prove Lemma 2.3, which states that the processes \mathbf{P} and \mathbf{P}_{FRAG} are close in L^1 at time t .

Lemma 2.3. $\|x_t - \bar{x}_t\|_1 \rightarrow 0$ as $n \rightarrow \infty$ in probability.

Proof. Since \mathbf{P} and \mathbf{P}_{FRAG} are coupled so that they only differ in that \mathbf{P}_{FRAG} loses mass occasionally, we must have that $x_{t,i} \geq \bar{x}_{t,i}$ for all coordinates i , $1 \leq i \leq n$. The value $\|x_t - \bar{x}_t\|_1$ is simply the amount of mass we remove by time t . There are three ways we can throw out mass:

- (R1) We average a positive pile with only zero piles, but the pieces become too small.
- (R2) We average multiple positive piles together.
- (R3) We remove mass via the extra condition in the coupling with $\mathbf{P}_{\text{BRANCH}}$.

First we consider mass removed because of (R1). Due to the coupling between \mathbf{P}_{FRAG} and $\mathbf{P}_{\text{BRANCH}}$, the mass thrown out due to pieces being too small will be dominated by the sum of lengths of the intervals in I_t which have length less than a_n . Thus, the total mass thrown out in this way will be $o(1)$ in probability from Lemma 2.3. Now consider the mass removed because of (R2). Let D_t be the total length removed by averaging multiple positive piles by time t . Linearity of expectation will give us

$$\mathbb{E}[D_{t+\Delta} - D_t | \mathcal{F}_t \text{ and one ring in } (t, t + \Delta)] = \mathbb{E} \left(\sum_i \bar{x}_{t,i} \cdot p_i | \mathcal{F}_t \right)$$

where p_i is the probability we pick i and some other j such that $\bar{x}_{t,j} \neq 0$. Now consider the process where we average exactly N elements in each ring. The probability in this case that we pick i and some other j such that $\bar{x}_{t,j} \neq 0$ will be at least p_i since larger averages will only increase this probability. In this sense, the N -averaging-restricted process dominates the original one. The probability that we pick $\bar{x}_{t,i}$ in an averaging is then N/n . The probability that we pick $\bar{x}_{t,i}$ and no other positive pile is at least $\frac{\binom{n-1/a_n}{N-1}}{\binom{n}{N}}$ since there are at most $1/a_n$ positive piles. Thus, we have that

$$p_i \leq \frac{N}{n} - \frac{\binom{n-1/a_n}{N-1}}{\binom{n}{N}} \leq \frac{N}{n} - N \cdot \frac{(n-1/a_n)^{N-1}}{(n-N+1)^N} \leq N \left(\frac{(n-N+1)^N - (n-1/a_n)^N}{n(n-N+1)^N} \right).$$

We then apply the Mean Value Theorem in the form

$$\frac{(n-N+1)^N - (n-1/a_n)^N}{(n-N+1) - (n-1/a_n)} = N \cdot t_*^{N-1} \leq N(n-N+1)^{N-1}$$

where $t_* \in (n-1/a_n, n-N+1)$ is some element. Thus, we have

$$p_i \leq \frac{N^2}{(n-N+1)^{N+1}} \cdot (1/a_n - N + 1) \cdot (n-N+1)^{N-1} \leq \frac{N^2}{a_n(n-N+1)^2} \leq \frac{C}{a_n n^2}.$$

for some universal constant C depending on N . This gives us

$$\mathbb{E}[D_{t+\Delta} - D_t | \mathcal{F}_t \text{ and one ring in } (t, t + \Delta)] \leq \frac{C}{a_n n^2} \cdot \sum_i \bar{x}_{t,i} \leq \frac{C}{a_n n^2}.$$

Between 0 and t , we are expected to have t rings. Hence, we have

$$\mathbb{E}[D_t] \leq \frac{C \gamma n \log n}{a_n n^2} = O\left(\frac{\log n}{a_n n}\right) = O\left(\log n \cdot n^{-\frac{1}{2}(\log n)^{-1/4}}\right) = o(1).$$

Thus $D_t \rightarrow 0$ in probability and $\|x_t - \bar{x}_t\| = D_t + o(1) \rightarrow 0$ in probability. This proves that the mass thrown out because of (R2) will be $o(1)$ in probability. Finally, we consider the mass removed because of (R3). Note that this removal happens because there is some $I \in I_t$ with $s_1 = \pi_t(I)$ and $s_2, \dots, s_k \in [n]$ such that at least one of the piles associated to s_2, \dots, s_k is positive and N_{I, s_2, \dots, s_k} rings but the $\bar{N}_{s_1, \dots, s_k}$ does not. Note that the rate that this happens is the exact same as the rate that $\bar{N}_{s_1, \dots, s_k}$ rings. There are at most N clocks associated to a single $\bar{N}_{s_1, \dots, s_k}$, so in the worst case if we want to bound the mass we remove, we can think of this amount being dominated by the amount we remove in (R2) if averagings were to happen $N + 1$ times faster. This amount of mass removed is still $o(1)$, which suffices for the proof. \square

We now prove Lemma 2.4 which states the total mass in piles of size in (a_n, b_n) is well-approximated by the total length of intervals of similar length.

Lemma 2.4. *We have that*

$$\left| \sum_{i=1}^n \bar{x}_{t,i} \cdot \mathbf{1}_{\bar{x}_{t,i} \in (a_n, b_n)} - \sum_{I \in I_t} |I| \cdot \mathbf{1}_{|I| \in (a_n, b_n)} \right| \rightarrow 0.$$

in probability.

Proof. From our coupling, it is clear that all piles satisfying $\bar{x}_{t,i} \neq 0$ is associated with a unique interval in I_t at time t . Moreover, this associated interval has length equal to the mass in the pile. This proves that

$$\sum_{i=1}^n \bar{x}_{t,i} \cdot \mathbf{1}_{\bar{x}_{t,i} \in (a_n, b_n)} \leq \sum_{I \in I_t} |I| \cdot \mathbf{1}_{|I| \in (a_n, b_n)}.$$

The only reason why the right hand side is possibly larger than the left hand side is because of the intervals $I \in I_t$ whose length lies in $|I| \in (a_n, b_n)$ but there is no pile associated to it. Since $|I| > a_n$, this implies that it must have lost its associated label via multiple positive piles being averaged together. Thus the length in $|I|$ is included in the mass we throw out in \mathbf{P}_{FRAG} . But this mass is $o(1)$ in probability from Lemma 2.3. This proves the other bound

$$\sum_{I \in I_t} |I| \cdot \mathbf{1}_{|I| \in (a_n, b_n)} \leq \sum_{i=1}^n \bar{x}_{t,i} \cdot \mathbf{1}_{\bar{x}_{t,i} \in (a_n, b_n)} + o(1)$$

and suffices for the proof of the lemma. \square

We currently have enough to prove one side of Theorem 2.1.

Proposition 2.4. $T(t) \rightarrow 2$ in probability as $n \rightarrow \infty$.

Proof. We can bound

$$T(t) = 2 \sum_{i=1}^n \left(x_{t,i} - \frac{1}{n} \right)^+ \geq 2 \sum_{i=1}^n \left(\bar{x}_{t,i} - \frac{1}{n} \right)^+$$

since x_t is greater than \bar{x}_t coordinate-wise. We can simplify the expression on the right hand side as

$$\begin{aligned} 2 \sum_{i=1}^n \left(\bar{x}_{t,i} - \frac{1}{n} \right)^+ &\geq 2 \sum_{i=1}^n \bar{x}_{t,i} \cdot \mathbb{1}_{x_{t,i} \in (a_n, b_n)} - \frac{2}{n} |\{i \in [n] : \bar{x}_{t,i} > 0\}| \\ &\geq 2 \sum_{I \in I_t} |I| \cdot \mathbb{1}_{|I| \in (a_n, b_n)} - \frac{2}{na_n} + o(1) \\ &\geq 2 \sum_{I \in I_t} |I| \cdot \mathbb{1}_{|I| \in (a_n, b_n)} - o(1). \end{aligned}$$

The first term converges to 2 in probability, which completes the proof of the proposition. \square

To prove the other side of Theorem 2.1, we want to run our original process starting from a “smoothed” version of the point \bar{x}_t . The L^2 bound on this new process will be more effective than if we were to use it on the original process.

Definition 2.6. Consider the process \hat{x} starting at time t with initial condition

$$\hat{x}_{t,i} = \bar{x}_{t,i} \cdot \mathbb{1}_{\bar{x}_{t,i} \in (a_n, b_n)} + \frac{1}{n} \left(1 - \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \bar{x}_{t,i} \right).$$

In particular, we have that $\|\hat{x}_t\|_1 = 1$. Let \hat{x} evolve according to the same averages as x .

Lemma 2.5. $\|\hat{x}_t - \bar{x}_t\|_1 \rightarrow 0$ as $n \rightarrow \infty$ in probability.

Proof. Whenever $\bar{x}_{t,i} \in (a_n, b_n)$, we have the bound

$$|\bar{x}_{t,i} - \hat{x}_{t,i}| = \frac{1-z}{n}$$

where $z := \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \bar{x}_{t,i} \rightarrow 1$ in probability. When $\bar{x}_{t,i} \notin (a_n, b_n)$, then we can bound

$$|\bar{x}_{t,i} - \hat{x}_{t,i}| \leq \bar{x}_{t,i} + \frac{1-z}{n}.$$

Adding these estimates together, we have that

$$\begin{aligned} \|\hat{x}_t - \bar{x}_t\|_1 &\leq n \cdot \frac{1-z}{n} + \sum_{i: \bar{x}_{t,i} \notin (a_n, b_n)} \bar{x}_{t,i} \\ &\leq (1-z) + (1-z) = 2(1-z). \end{aligned}$$

Since $1-z \rightarrow 0$ in probability, we get the desired result. \square

Lemma 2.6. $\|\hat{x}_t - \frac{1}{n} \cdot \mathbf{1}\|_2^2 \leq \frac{2b_n^2}{a_n}$ for sufficiently large n .

Proof. We have that

$$\left\| \hat{x}_t - \frac{1}{n} \cdot \mathbf{1} \right\|_2^2 \leq \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \left(\bar{x}_{t,i} - \frac{z}{n} \right)^2 + \sum_{i: \bar{x}_{t,i} \notin (a_n, b_n)} \frac{z^2}{n^2} \leq \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \bar{x}_{t,i}^2 + \frac{1}{n}$$

where z is the quantity $z = \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \bar{x}_{t,i}$ and $z \in [0, 1]$. Thus, we can bound

$$\left\| \hat{x}_t - \frac{1}{n} \cdot \mathbf{1} \right\|_2^2 \leq \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} (\bar{x}_{t,i})^2 + \frac{1}{n} \leq \frac{b_n^2}{a_n} + \frac{1}{n}.$$

Our choices for a_n and b_n guarantee that $b_n^2/a_n \gg 1/n$. Thus, for sufficiently large n , we get the desired bound. \square

Lemma 2.7. *For any $s > 0$, we have that $\|x_{t+s} - \hat{x}_{t+s}\|_1 \leq \|x_t - \hat{x}_t\|_1$.*

Proof. It suffices to prove that given two initial vectors $x_0, y_0 \in [0, 1]^n$ with $\|x_0\|_1 = \|y_0\|_1 = 1$, then $\|x_1 - y_1\|_1 \leq \|x_0 - y_0\|_1$ where x_1, y_1 are the vectors after performing the same averaging on x_0, y_0 respectively. Now, consider the vectors $z_0 = x_0 - y_0$ and $z_1 = x_1 - y_1$. It suffices to prove that if the average of the coordinates of z_0 is 0 and z_1 is the result of averaging some of the coordinates of z_0 , then $\|z_1\|_1 \leq \|z_0\|_1$. To prove this, suppose that we averaged the values u_1, \dots, u_k . Then,

$$k \cdot \left| \frac{u_1 + \dots + u_k}{k} \right| = |u_1 + \dots + u_k| \leq |u_1| + \dots + |u_k|$$

which proves that the L^1 distance decreases. This suffices for the proof. \square

Now let $\Omega = \max \left\{ \gamma, \frac{7}{2(\mu-1)} \right\} + 1$ be a universal constant and $s := \Omega(n-1)(\log n)^{3/4}$. Then, Proposition 2.5 proves other side of Theorem 2.1.

Proposition 2.5. *$T(s+t) \rightarrow 0$ as $n \rightarrow \infty$ in probability.*

Proof. Let T', S' be the L^1 and L^2 distances of the process \hat{x} . From Lemma 2.2, we have that

$$\begin{aligned} \mathbb{E}[S'(t+s) | \mathcal{F}_t'] &= \exp \left(-\frac{s(\mu-1)}{(n-1)} \right) S'(t) \\ &\leq \exp(-\Omega(\mu-1)(\log n)^{3/4}) \cdot \frac{2b_n^2}{a_n} \\ &= 2 \exp \left(\left\{ \frac{7}{2} - \Omega(\mu-1) \right\} (\log n)^{3/4} \right) \cdot \frac{1}{n}. \end{aligned}$$

From the Cauchy-Schwartz inequality, we have that

$$\mathbb{E}[(T'(t+s))^2] \leq \mathbb{E}[nS'(t+s)] \leq 2 \exp \left(\left\{ \frac{7}{2} - \Omega(\mu-1) \right\} (\log n)^{3/4} \right) = o(1).$$

From Jensen's inequality, we have $\mathbb{E}[T'(t+s)] \leq \{\mathbb{E}[T'(t+s)^2]\}^{1/2}$. This gives us $\|\hat{x}_{t+s} - \frac{1}{n} \cdot \mathbf{1}\|_1 = o(1)$ in probability. From the triangle inequality, we have

$$\begin{aligned} \left\| x_{t+s} - \frac{1}{n} \cdot \mathbf{1} \right\|_1 &\leq \|x_{t+s} - \hat{x}_{t+s}\|_1 + \left\| \hat{x}_{t+s} - \frac{1}{n} \cdot \mathbf{1} \right\|_1 \\ &\leq \|x_t - \hat{x}_t\|_1 + \left\| \hat{x}_{t+s} - \frac{1}{n} \cdot \mathbf{1} \right\|_1 \\ &\leq \|x_t - \bar{x}_t\|_1 + \|\bar{x}_t - \hat{x}_t\|_1 + \left\| \hat{x}_{t+s} - \frac{1}{n} \cdot \mathbf{1} \right\|_1. \end{aligned}$$

But this is $o(1)$ in probability from Lemma 2.3 and Lemma 2.5. This suffices for the proof. \square

Note that for sufficiently large n , we have

$$s + t \leq \gamma n (\log n + (\Omega - \gamma)(\log n)^{3/4}) \leq \gamma n (\log n + (\log n)^{4/5}).$$

Thus, we can slightly weaken our current results to get a succinct formulation in Theorem 2.1.

Theorem 2.1. *Let $t = \theta n \log n$ for some $\theta > 0$. Then, the following are true:*

- (a) *If $\theta < \gamma$, then $T(t) \rightarrow 2$ in probability.*
- (b) *If $\theta > \gamma$, then $T(t) \rightarrow 0$ in probability.*

Proof. Both results follow from Proposition 2.4 and Proposition 2.5 and the monotonicity of T and S . \square

In particular, we have proven that $\gamma n \log n$ is a cutoff time for the L^1 distance where $\gamma = 1/\mathbb{E}[Q \log Q]$. We can extend these results easily to the discrete case. We will redefine the processes so that $\{x_t\}_{t \in \mathbb{N}_0}$ is the discrete time process where x_{k+1} is obtained from x_k by sampling a number of Q , then averaging a randomly chosen set of coordinates of that size. The process $\{x'_t\}$ will then be the Poissonized version where averages happening according to a Poisson clock of rate 1. Then Theorem 2.1 holds for the discrete process.

Theorem 2.2. *Let $t = \theta n \log n$ for some $\theta > 0$. Let T be the L^1 distance of the discrete process. Then the following are true:*

- (a) *If $\theta < \gamma$, then $T(t) \rightarrow 2$ in probability.*
- (b) *If $\theta > \gamma$, then $T(t) \rightarrow 0$ in probability.*

Proof. The key idea is that we want to relate the discrete process with the Poissonized version. Luckily, the Poissonized version can be interpreted in terms of the discrete version in the following natural way: Let N be a Poisson random variable of rate t , then x_N has the same distribution as x'_t .

- (a) Let $t := \gamma n (\log n - (\log n)^{3/4})$ and $m := t - \sqrt{t} \log t$. We can lower bound m by

$$m \geq \gamma n (\log n - (\log n)^{3/4}) - 2n^{3/4} \log n \geq \gamma n (\log n - 2(\log n)^{3/4}).$$

In particular, if $\theta < \gamma$, we have $m \geq \theta n \log n$ for sufficiently large n . Let N be a Poisson random variable of rate t . We can then bound

$$\begin{aligned} \Pr \left[\left\| x_m - \frac{1}{n} \cdot \mathbf{1} \right\|_1 < 2 - \varepsilon \right] &\leq \Pr[N < m] + \Pr \left[\left\| x_m - \frac{1}{n} \cdot \mathbf{1} \right\|_1 < 2 - \varepsilon, N \geq m \right] \\ &\leq \Pr[N < m] + \Pr \left[\left\| x_N - \frac{1}{n} \cdot \mathbf{1} \right\|_1 < 2 - \varepsilon \right] \\ &\leq \Pr[N < m] + \Pr \left[\left\| x'_t - \frac{1}{n} \cdot \mathbf{1} \right\|_1 < 2 - \varepsilon \right] \end{aligned}$$

From the more exact versions of the thresholds in Theorem 2.1, we have that

$$\Pr \left[\left\| x'_t - \frac{1}{n} \cdot \mathbf{1} \right\|_1 < 2 - \varepsilon \right] \rightarrow 0$$

as $n \rightarrow \infty$. Moreover, we have that

$$\Pr[N < m] = \Pr\left[\frac{N-t}{\sqrt{t}} < \frac{m-t}{\sqrt{t}}\right] = \Pr\left[\frac{N-t}{\sqrt{t}} < -\log t\right].$$

From Lemma 2.1, we have that

$$\Pr[N < m] \rightarrow \Pr[N(0, 1) < -\infty] = 0.$$

This proves the desired result by monotonicity.

- (b) The proof is very similar to the previous part, but we include it for the sake of completeness. Let $t := \gamma n(\log n + (\log n)^{4/5})$ and $m := t + \sqrt{t} \log t$. One crude bound for m for sufficiently large n is

$$m \leq \gamma n(\log n + (\log n)^{4/5}) + 2n^{3/4} \log n \leq \gamma n(\log n + 2(\log n)^{4/5})$$

where to bound \sqrt{t} we used $t \ll n^{3/2}$ and to bound $\log t$ we used $t \ll n^2$. In particular, if $\theta > \gamma$, then $\theta n \log n \geq m$ for sufficiently large n . Let N be a Poisson random variable of rate t . We can then bound

$$\begin{aligned} \Pr\left[\left\|x_m - \frac{1}{n} \cdot \mathbf{1}\right\|_1 > \varepsilon\right] &\leq \Pr[N > m] + \Pr\left[\left\|x_m - \frac{1}{n} \cdot \mathbf{1}\right\|_1 > \varepsilon, N \leq m\right] \\ &\leq \Pr[N > m] + \Pr\left[\left\|x_N - \frac{1}{n} \cdot \mathbf{1}\right\|_1 > \varepsilon\right] \\ &\leq \Pr[N > m] + \Pr\left[\left\|x'_t - \frac{1}{n} \cdot \mathbf{1}\right\|_1 > \varepsilon\right]. \end{aligned}$$

where we used the monotonicity of the L^1 norm in the second inequality. From the more exact versions of the thresholds of Theorem 2.1, we have that

$$\Pr\left[\left\|x'_t - \frac{1}{n} \cdot \mathbf{1}\right\|_1 > \varepsilon\right] \rightarrow 0$$

as $n \rightarrow \infty$. Moreover, we have that

$$\Pr[N > m] = \Pr\left[\frac{N-t}{\sqrt{t}} > \frac{m-t}{\sqrt{t}}\right] = \Pr\left[\frac{N-t}{\sqrt{t}} > \log t\right].$$

From Lemma 2.1, we have that

$$\Pr[N > m] \rightarrow \Pr[N(0, 1) > \infty] = 0$$

as $n \rightarrow \infty$. This proves that $T(m) \rightarrow 0$ in probability. From monotonicity, the result follows.

This suffices for the proof. □

3 Future Work and Open Questions

There are still many open questions about the iterated averages Markov chain.

3.1 Unbounded Random Averages

In this paper, one of the important assumptions was that the law of Q was bounded. What are the weakest regularity conditions needed on the distribution so that some meaningful statement can be made about the cutoff?

3.2 Maximum Decay over All Starting Points

Beyond the cutoff time, there are other problems we can consider about this Markov chain. [1] first picks an initial vector and their analysis was focused on the L^1 trajectory of this specific process. From here, they prove that the maximum L^1 cutoff over all initial starting points occurs when all of the mass is concentrated at a single pile. Instead, we can consider the reverse: consider a different process for each initial starting vector where the processes are evolving simultaneously via the same averagings. At time t , we can consider the maximum L^1 distance among all starting points. What is the cutoff time of this process?

3.3 Second Order Behavior of Random Averages

In [1], the second order behavior of the binary iterated averages Markov chain was completely understood. With slight changes to the current techniques, it should be possible to also understand the second order behavior of the random averages case.

3.4 Stationary Distribution of Normalized Markov Chain

Alternatively, we may choose to abandon cutoff and instead consider a L^1 normalized version of our Markov chain. Suppose n is not a power of two, and let x_0 be a vector whose average is 0. Then any averaging or scalar multiple will preserve the zero average. We let the Markov chain evolve as in the averaging case, except after every averaging we normalize the vector in L^1 . Hence, the process is some trajectory in the intersection between a hyperplane and the L^1 ball. Does this Markov chain have a stationary distribution? If so, what is it?

Acknowledgements

I would like to thank Professor Allan Sly for all of his guidance throughout this junior paper. My knowledge of probability and mathematics has grown immensely because of our weekly meetings.

Honor Pledge

This paper represents my own work in accordance with University regulations. Alan Yan, April 25th, 2022.

References

- [1] Sourav Chatterjee, Persi Diaconis, Allan Sly, and Lingfu Zhang. A phase transition for repeated averages, 2021.