Random Iterated Averages

Alan Yan

Project Advisor: Professor Allan Sly

April 20, 2022

Contents

1 Introduction

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

Definition 1 (Binary Markov Chain). *Let $\mathsf{B} = \{x_t\}_{t \in \mathbb{Z}_{\geq 0}}$ 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 traces a discrete trajectory on a $n-1$-dimensional hyperplane in $\mathbb{R}^n$. As $t \to \infty$, the Markov chain will converge to the uniform vector of the average of the coordinates almost surely. In this paper, we will be interested in the rate of convergence to the average for generalizations of this Markov chain. Let the x^μ be the average value of the coordinates of x_0. Then, we consider two measurements of the distance from the average which correspond to the L^1 and L^2 distances. Following [1], we define the T and S functions explicitly as

$$T(k) := \sum_{i=1}^{n} |x_{k,i} - \overline{x}_0|, \quad S(k) := \sum_{i=1}^{n} |x_{k,i} - \overline{x}_0|^2.$$

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 was found and is given in Proposition 1.

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

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

Propostion 1 implies that the L^2 norm decays exponentially. The main result [1] was concerning the decay and cutoff of the L^1 norm. In particular, the main theorem in [1] given in Theorem 1 finds the second order behavior of the L^1 distance.

Theorem 1 (Theorem 1.2 in [1]). *Let $\Phi : \mathbb{R} \to [0,1]$ be the cumulative distribution function of the standard normal distribution $N(0,1)$. For $x_0 = (1,0,\ldots,0)$, and any $a \in \mathbb{R}$, as $n \to \infty$ 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, Corollary 1 follows easily and has a more succinct statement.

Corollary 1 (Theorem 1.1 in [1]). *For $x_0 = (1,0,\ldots,0)$ as $n \to \infty$ 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 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 2.

Definition 2 (Poissonized Binary Markov Chain). *Let $\mathsf{B}' = \{x_t\}_{t \in \mathbb{R}_{\geq 0}}$ 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 are equivalently having a single clock of rate 1 which then, upon ringing, picks two distinct coordinates uniformly at random and replaces them by their average. Thus, B' seems to be suitable approximation of B. Let T' and S' be the corresponding L^1 and L^2 distances from the average of B'. Similar to the discrete case, there is still an explicit formula for the L^2 distance.

Proposition 2. *For any $t \geq s \geq 0$, we have that*

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

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

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

This suffices for the proof. $\qquad \square$

Then, [1] focuses on the proof of Theorem 2 as this a good enough approximation of Theorem 1 to imply it.

Theorem 2 (Theorem 2.3 in [1]). *Take $\Phi : \mathbb{R} \to [0,1]$ 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 $n \to \infty$.*

Proposition 3. *Theorem 3 and Theorem 2 are equivalent.*

Proof. TODO $\qquad \square$

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 Definition 2, but at each averaging we average some random number of coordinates where the random number is sampled from a fixed distribution.

Definition 3. *Let Q be a random variable supported in a finite set $\{2, 3, \ldots, N\}$ such that $\Pr[Q = k] = q_k$ for all k, $2 \leq k \leq N$. Let $\mathsf{P}_Q := \{x_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a Markov chain with $x_0 = (1, 0, \ldots, 0)$. 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 3, we set the initial value as $x_0 = (1, 0, \ldots, 0)$. 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 3 but more general than Definition 1. Instead of averaging two elements every time, we average some fixed number of coordinates m.

Definition 4. *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 4. *For any* $k \geq 0$, *we have*

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

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

$$\mathbb{E}\left[S_m(k+1)|\mathcal{F}_k\right] = \binom{n}{m}^{-1} \sum_{i_1 < \ldots < i_m} \left\{ S_m(k) - \sum_{j=1}^{m} x_{k,i_j}^2 + m\left(\frac{x_{k,i_1} + \ldots + x_{k,i_m}}{m}\right)^2 \right\}$$

$$= \binom{n}{m}^{-1} \sum_{i_1 < \ldots < i_m} \left\{ S_m(k) - \sum_{j=1}^{m} x_{k,i_j}^2 + \frac{1}{m}\sum_{j=1}^{m} x_{k,i_j}^2 + \frac{1}{m}\sum_{1 \leq u < v \leq m} x_{k,i_u}x_{k,i_v} \right\}$$

$$= \binom{n}{m}^{-1} \sum_{i_1 < \ldots < i_m} \left\{ S_m(k) + \frac{1-m}{m}\sum_{j=1}^{m} x_{k,i_j}^2 + \frac{1}{m}\sum_{1 \leq u < v \leq m} x_{k,i_u}x_{k,i_v} \right\}$$

$$= \Sigma_1 + \Sigma_2 + \Sigma_3$$

where

$$\Sigma_1 := \binom{n}{m}^{-1} \sum_{i_1 < \ldots < i_m} S_m(k)$$

$$\Sigma_2 := \binom{n}{m}^{-1} \sum_{i_1 < \ldots < i_m} \frac{1-m}{m}\sum_{j=1}^{m} x_{k,i_j}^2$$

$$\Sigma_3 = \binom{n}{m}^{-1} \sum_{i_1 < \ldots < i_m} \frac{1}{m}\sum_{1 \leq u < v \leq m} 2x_{k,i_u}x_{k,i_v}.$$

For the first sum, we immediately get $S_1 = S_m(k)$. For the second sum, we get

$$S_2 = \binom{n}{m}^{-1} \cdot \frac{1-m}{m} \cdot \binom{n-1}{m-1} S_m(k) = \frac{1-m}{n} S_m(k).$$

For the last sum, we get

$$S_3 = \binom{n}{m}^{-1} \cdot \frac{1}{m} \cdot \binom{n-2}{m-2} \sum_{1 \leq u < v \leq n} 2x_{k,u}x_{k,v} = \frac{1-m}{n(n-1)} S_m(k)$$

where we used the fact that

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

Then

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

$\square$

Corollary 2. *For $0 \le s < t$, 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

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

$\square$

Proposition 4 and Corollary 2 will allow us to easily compute the L^2 distance for the random bounded averages case.

Proposition 5. *Let $S(k)$ be the L^2 distance at time k of the discrete version of P_Q. Then*

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

Proof. We have

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

$\square$

5

Corollary 3. *Let $S(t)$ be the L^2 distance at time t of P_Q. Then*

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

Proof. The proof is very similar to the proof of Corollary 2 now that we have Proposition 5. We can compute

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

$\square$

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.

Definition 5 (Original Process). *Let $\mathsf{P} := \{x_t\}$ be the Markov chain where $x_0 = (1,0,\ldots,0) \in \mathbb{R}^n$. We have a Poisson clock of rate 1. Everytime this clock rings, we pick some random coordinates to average where the random number is sampled randomly from Q.*

Definition 6 (Fragmentation Process). *Let $\mathsf{P}_{\mathsf{FRAG}} := \{\overline{x}_t\}$ be the Markov chain where $\overline{x}_0 = (1,0\ldots,0) \in \mathbb{R}^n$. 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. This is a fragmentation process version of the original process where we throw out piles which become too small.*

Definition 7 (Branching Process). *Let $\mathsf{P}_{\mathsf{BRANCH}} := \{I_t\}$ 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 distibution $(2q_2/\mu,\ldots,Nq_N/\mu)$. That is, for any $2 \le k \le N$, we have $\Pr[Q' = k] = kq_k/\mu$.*

We now describe how we couple the processes P, $\mathsf{P}_{\mathsf{FRAG}}$, and $\mathsf{P}_{\mathsf{BRANCH}}$ so that $\mathsf{P}_{\mathsf{FRAG}}$ acts as an approximation of P, and $\mathsf{P}_{\mathsf{BRANCH}}$ is an approximation of $\mathsf{P}_{\mathsf{FRAG}}$ where our approximations are fine enough to conclude information about the L^1 convergence.

> We couple P and $\mathsf{P}_{\mathsf{FRAG}}$ in the obvious way: whenever we average some piles on P, we average the same exact piles in $\mathsf{P}_{\mathsf{FRAG}}$ while keeping the averaging rules consistent in each process. The coupling between $\mathsf{P}_{\mathsf{FRAG}}$ and $\mathsf{P}_{\mathsf{BRANCH}}$ is a bit more complicated. In order to couple the two processes, we attach to each interval set I_t a labelling function $\pi_t : I_t \to [n] \cup \{0\}$ which assigns labels to intervals. The labelling function will always be injective on the intervals which are not given the label 0. Morally, if $I \in I_t$ satisfies

$\pi_t(I) = k$ for some $k \in [n]$, the interval I "represents" the interval k and will satisfy $|I| = \overline{x}_{t,k}$. To see how this labelling function evolves, we first initialize $\pi_0([0,1]) = 1$. This reflects the initial vector $(1, 0, \ldots, 0)$ since the whole interval represents the first pile and the other piles are assigned. Consider the sequence of averagings in $\overline{x}_t$ which we can say occur at times $t_1 < t_2 < \ldots$ since the times occur at a discrete set of points almost surely. We initially remove all of the clocks from all of the intervals. Now, suppose we are at time t_i awaiting the evolution of our process I_t. If at time t_{i+1}, a positive pile is averaged with only zero piles, we take the interval corresponding to this positive pile and split it into equal pieces while assigning the piles we averaged to the equal pieces in a one-to-one manner. If at time t_{i+1}, we average at least two positive piles together, then we split each of the intervals corresponding to the positive piles into the correct number of equal pieces, remove the labels of the intervals corresponding to these piles, and attach Poisson clocks of rate μ/n to each of them such that when it rings we split it into equal pieces based on the law of Q'. We also let the other intervals currently with clocks evolve naturally. To prove that this forms a valid coupling, it suffices to prove that the marginal law of $\mathsf{P_{BRANCH}}$ is as described in Definition 7. To see this, we require to go back to our original process $\{x_t\}$. We can split our Poisson clock of rate 1 into separate independent Poisson clocks $C_2, \ldots, C_N$ where C_k is a Poisson clock of rate q_k. When C_k rings, we pick k distinct coordinates uniformly and random and replace them by their average. If we fix a pile in $\{\overline{x}_{t,i}\}$ the probability that it gets picked to get averaged when C_k rings is $\frac{k}{n}$. In particular, if C_k rings, the probability that the corresponding interval in I_t gets split into k parts is also $\frac{k}{n}$. This is also true of intervals which are not associated with any pile in $\overline{x}_t$ by construction. Hence, every interval has a clock of rate $\frac{kq_k}{n}$ for all k which dictates when that interval should shatter into k equal pieces. By amalgmating all these clocks, we can equivalently view each interval having a clock of rate $\sum_{k=2}^{N} q_k \frac{k}{n} = \mu/n$ where whenever it rings, the probability we split the interval into k parts is

$$\frac{q_k \cdot (k/n)}{\mu/n} = \frac{kq_k}{\mu}.$$

This proves the coupling is a valid coupling.

From now on, let $t := \gamma n(\log n - (\log n)^{3/4})$ and $\gamma = (\mathbb{E}[Q \log Q])^{-1}$. 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 where chosen that $\frac{1}{n} \ll a_n < b_n = n^{-1+o(1)}$.

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 the following well-known result in Lemma 1 about Poisson concentration.

Lemma 1. *Suppose $X_t \sim \mathsf{Poisson}(t)$ is a Poisson random variable of rate t. Then*

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

where the convergence is in distribution.

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

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

Thus, we have that

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

Since $(Z_t - \{t\})/\sqrt{t} \to 0$ almost surely, the Lemma follows from the central limit theorem. $\square$

Lemma 1 will allow us to show that with high probability, $|I_t^x|$ lies in the interval (a_n, b_n).

Lemma 2. *Let $x \in [0, 1]$ be a fixed real number. Then*

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

Proof. Consider the process $\{I_s^x\}_{0 \leq s \leq t}$. In $\mathsf{P}_{\mathsf{BRANCH}}$, 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 kth clock rings, we multiply the length by $\frac{1}{k}$. Taking logarithms, we see that

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

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

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

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

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 1 then implies the following convergence in distribution:

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

8

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

$$A_n = \sqrt{\frac{n}{t}} \cdot \left(\log a_n + \frac{t}{n}\mathbb{E}[Q \log Q]\right)$$

$$B_n = \sqrt{\frac{n}{t}} \cdot \left(\log b_n + \frac{t}{n}\mathbb{E}[Q \log Q]\right).$$

It is a standard calculation to check that $A_n \to -\infty$ and $B_n \to \infty$ as $n \to \infty$. Hence,

$$\lim_{n\to\infty} \Pr(|I_t^x| \in (a_n, b_n)) = \lim_{n\to\infty} \Pr\left(\frac{\log |I_t^x| + \frac{t}{n}\mathbb{E}[Q \log Q]}{\sqrt{t/n}} \in (A_n, B_n)\right) = 1$$

since the random variable converges to a Gaussian. This proves the lemma.

$\square$

We now have all the tools needed to prove Theorem 3.

2.4   Proof of Theorem 3

Proposition 6. *As $n \to \infty$,*

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

almost surely.

Proof. Note that $\sum_{I \in I_t} |I| \cdot \mathbb{1}_{|I| \in (a_n, b_n)} \leq 1$. Hence, it suffices to prove that its expected value converges to 1. For $x \in [0, 1]$, define I_t^x to be the interval in I_t which contains x. If there are two such intervals, take the leftmost interval. Then, we have

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

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

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

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 $n \to \infty$, Lemma 2 finishes the proof of the proposition.

$\square$

Moreover, by time t, the piles which satisfy $\overline{x}_{t,i} \in (a_n, b_n)$ correspond to intervals whose lengths are in the same range. For the rest of the paper, our goal will be to prove Theorem 3. The first step in this journey will be to prove Lemma 3 which says that at time t the processes P and $\mathsf{P}_{\mathsf{FRAG}}$ are close in L^1.

Lemma 3. $\|x_t - \overline{x}_t\|_1 \to 0$ *as* $n \to \infty$ *almost surely.*

Proof. Since P and $\mathsf{P}_{\mathsf{FRAG}}$ are coupled so that the exact same averagings happen and $\mathsf{P}_{\mathsf{FRAG}}$ has the potential to lose mass in each averaging, we have that $x_{t,i} > \overline{x}_{t,i}$ for all coordinates i, $1 \leq i \leq n$. The value $\|x_t - \overline{x}_t\|_1$ is simply the amount of mass we throw out up to time t. The only ways we can throw out mass is if we average multiple positive piles together or we split up a pile into piles which are too small. In the latter case, due to the coupling between $\mathsf{P}_{\mathsf{FRAG}}$ and $\mathsf{P}_{\mathsf{BRANCH}}$, the mass thrown out in this way will be dominated by the lengths of the intervals not in (a_n, b_n). Thus, the total mass thrown out in this way will be $o(1)$ almost surely from Lemma 3. Thus, it suffices to prove that the total length removed by averaging multiple positive piles is $o(1)$ almost surely. Let D_t be the total length removed by averaging multiple positive piles by time t. Linearity of expectation will give us the equation

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

where p_{ij} is the probability we pick i and some other j such that $\overline{x}_{t,j} \neq 0$. 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 $\overline{x}_{t,j} \neq 0$ will be at least p_{ij} since bigger averages will only increase this probability. In this sense, the N-averaging-restricted process dominates the original one. The probability that we pick $\overline{x}_{t,i}$ in an averaging is then N/n. The probability that we pick $\overline{x}_{t,i}$ 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 $t \in (n-1/a_n, n-N+1)$. 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

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

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 $D_t \to 0$ almost surely and $\|x_t - \overline{x}_t\| = D_t + o(1) \to 0$ almost surely. This proves the lemma. $\quad\square$

We now prove Lemma 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 4. *We have that*

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

almost surely.

Proof. From our coupling, it is clear that all piles satisfying $\overline{x}_{t,i} \neq 0$ 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

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

The only reason why the right sum is larger than the left are 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 $\mathsf{P}_{\mathsf{FRAG}}$. But this mass is $o(1)$ almost surely from Lemma 3. This proves the other bound

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

and suffices for the proof of the lemma. $\qquad\square$

We currently have enough to prove one side of Theorem 3.

Proposition 7. $T(t) \to 2$ *almost surely as $n \to \infty$.*

Proof. We can bound

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

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

$$2 \sum_{i=1}^{n} \left(\overline{x}_{t,i} - \frac{1}{n} \right)^{+} \geq 2 \sum_{i=1}^{n} \overline{x}_{t,i} \cdot \mathbb{1}_{x_{t,i} \in (a_n, b_n)} - \frac{2}{n} |\{i \in [n] : \overline{x}_{t,i} > 0\}$$

$$\geq 2 \sum_{I \in I_t} |I| \cdot \mathbb{1}_{|I| \in (a_n, b_n)} - \frac{2}{n a_n} + o(1)$$

$$\geq 2 - o(1).$$

This proves the proposition. $\qquad\square$

To prove the other side of Theorem 3, we want to run our original process starting from a "smoothed" version of the point $\overline{x}_t$. The L^2 bound on this new process will be more effective than if we were to use it on the original process.

Definition 8. *Consider the process $\widehat{x}$ starting at time t with initial condition*

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

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

Lemma 5. $\|\widehat{x}_t - \overline{x}_t\|_1 \to 0$ *as $n \to \infty$ almost surely.*

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

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

where $z := \sum_{i:\overline{x}_{t,i} \in (a_n, b_n)} \overline{x}_{t,i} \to 1$ almost surely as $n \to \infty$. When $\overline{x}_{t,i} \notin (a_n, b_n)$, then we can bound

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

Adding these estimates together, we have that

$$\|\widehat{x}_t - \overline{x}_t\|_1 \leq n \cdot \frac{1-z}{n} + \sum_{i:\overline{x}_{t,i} \notin (a_n, b_n)} \overline{x}_{t,i}$$
$$\leq (1-z) + (1-z)$$
$$= 2(1-z).$$

As $n \to \infty$, we have $z \to 1$ almost surely. Hence, $\|\widehat{x}_t - \overline{x}_t\|_1 \to 0$ as desired. $\qquad\square$

Lemma 6. $\left\|\widehat{x}_t - \frac{1}{n} \cdot \mathbb{1}\right\|_2^2 \leq \frac{2b_n^2}{a_n}$ *for sufficiently large n where $\mathbb{1}$ is the uniform vector of 1's.*

Proof. We have that

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

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

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

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

Lemma 7. *For any $s > 0$, we have that $\|x_{t+s} - \widehat{x}_{t+s}\|_1 \leq \|x_t - \widehat{x}_t\|_1$.*

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 $u_1, \ldots, u_k$. Then,

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

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

Now let $\Omega = \max\left\{\gamma, \frac{7}{2(\mu-1)}\right\} + 1$ be a universal constant. We define $s := \Omega(n-1)(\log n)^{3/4}$. Then, we have the following lemma.

Lemma 8. $T(s+t) \to 0$ as $n \to \infty$ almost surely.

Proof. Let T', S' be the L^1 and L^2 distances of the process $\widehat{x}$. From Lemma 3, we have that

$$S'(t+s) = \exp\left(-\frac{s(\mu-1)}{(n-1)}\right) S'(t)$$

$$\leq \exp(-\Omega(\mu-1)(\log n)^{3/4}) \cdot \frac{2b_n^2}{a_n}$$

$$= 2\exp\left(\left\{\frac{7}{2} - \Omega(\mu-1)\right\}(\log n)^{3/4}\right) \cdot \frac{1}{n}.$$

From the Cauchy-Schwartz inequality, we have that

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

This gives us $\left\|\widehat{x}_{t+s} - \frac{1}{n} \cdot \mathbb{1}\right\|_1 = o(1)$. From the triangle inequality, we have

$$\left\|x_{t+s} - \frac{1}{n} \cdot \mathbb{1}\right\|_1 \leq \|x_{t+s} - \widehat{x}_{t+s}\|_1 + \left\|\widehat{x}_{t+s} - \frac{1}{n} \cdot \mathbb{1}\right\|_1$$

$$\leq \|x_t - \widehat{x}_t\| + o(1)$$

$$\leq \|x_t - \overline{x}_t\|_1 + \|\overline{x}_t - \widehat{x}_t\|_1 + o(1)$$

$$= o(1).$$

This proves the lemma. □

Note that there is a $\varepsilon > 0$ sufficiently small such that $(\Omega - \varepsilon - \gamma) > 0$. In particular, for sufficiently large n, we have

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

Thus, we can slightly weaken our current results to get a succint formulation in Theorem 3.

Theorem 3. *Let $t = \theta n \log n$ for some $\theta > 0$. Then, the following are true:*

 (a) *If $\theta < \gamma$, then $T(t) \to 2$ almost surely.*

 (b) *If $\theta > \gamma$, then $T(t) \to 0$ almost surely.*

Proof. Both results follow from Theorem 3 and the monotonicity of T and S □

In particular, we have proven that $\gamma n \log n$ is a cutoff time for the L^1 distance where $\gamma = 1/\mathbb{E}[Q \log Q]$.

3 Future Work

Acknowledgements

References

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