

Random Iterated Averages

Alan Yan

Project Advisor: Professor Allan Sly

April 16, 2022

Contents

1	Introduction	2
2	Deterministic Generalized Averages	3
3	Random Bounded Averages	5
3.1	Logarithmic Length of the Interval Trajectory	7
3.2	Trying Random Stuff	9
3.3	New Process	9
4	Starting Anew	10

1 Introduction

In [1], Chatterjee, Diaconis, Sly, and Zhang are interested in the following Markov chain.

Definition 1 (Markov Chain I). *We have a Markov chain x_0, x_1, \dots where x_0 is some initial vector in \mathbb{R}^n . Notation-wise, let $x_k = (x_{k,1}, \dots, x_{k,n})$. The Markov chain progresses as follows: given x_k pick two distinct coordinates I and J uniformly at random, and replace $x_{k,I}$ and $x_{k,J}$ by their average $\frac{x_{k,I} + x_{k,J}}{2}$.*

We want to study the rate of convergence of this Markov chain to its average $(\bar{x}_0, \dots, \bar{x}_0)$ where

$$\bar{x}_0 = \frac{x_{0,1} + \dots + x_{0,n}}{n}.$$

Without loss of generality, we may assume $\bar{x}_0 = 0$ unless otherwise specified. We can measure the distance of our Markov chain to the average in many ways. The functions $T(k)$ and $S(k)$ defined below are the L^1 and L^2 measurements.

$$T(k) := \sum_{i=1}^n |x_{k,i} - \bar{x}_0|$$

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

For the L^2 norm measurement, an exact value is given in [1].

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

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

This formula gives us an exact formula for the decay of the L^2 distance of our Markov chain to the average. The main result in the paper was the behavior of the L^1 decay.

Theorem 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*

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

where the convergence is in probability.

Clearly, Theorem 1 implies the following result.

Corollary 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 $\theta > 1/(2 \log 2)$.*

In the proof of Theorem 1, to avoid issues of dependence in the discrete time setting, Markov chain I is approximated with a Poissonized continuous-time version of the same Markov chain.

Definition 2 (Markov Chain II). *Denote our Markov chain by $\{x_t\}_{t \geq 0}$. Give each pair of coordinates an independent Poisson clock with rate $\binom{n}{2}^{-1}$. When a clock rings, replace the corresponding pair of coordinates with their average.*

By giving each pair of coordinates a clock of rate $\binom{n}{2}^{-1}$, we make it so that on average, one averaging happens every unit of time. Thus, this Poissonized version is a suitable approximation of our original process. With this new Markov chain, we introduce the same L^1 and L^2 measurements from the average:

$$T'(k) := \sum_{i=1}^n |x_{t,i} - \bar{x}_0|$$

$$S'(k) := \sum_{i=1}^n |x_{t,i} - \bar{x}_0|^2.$$

Similar to the discrete case, we are still able to get an exact formula for the L^2 measurement.

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

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

This suffices for the proof. □

The same result L^2 decay result holds for the continuous-time case.

Theorem 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 $n \rightarrow \infty$.*

Proposition 3. *Theorem 1 and Theorem 2 are equivalent.*

Proof. TODO □

We now consider some extensions of the problem.

2 Deterministic Generalized Averages

The first generalization involves taking larger averages at each unit of time instead of simply averaging two positions. This variant is very similar to the original version, thus we won't go into all of the details. We will only do the initial computations to illustrate the similarities.

Definition 3. Let \mathbf{M}_m^n be the markov chain $\{x_k\}_{k \geq 0} \subset \mathbb{R}^n$ 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 $\mathbb{P}\mathbf{M}_m^n = \{x_t\}_{t \in \mathbb{R}_{\geq 0}}$ be the Poissonized version of the process where we assign independent clocks of rate $\binom{n}{m}^{-1}$ to every m -subset of coordinates. Whenever a clock rings, we replace the corresponding coordinates by their average.

We want to study the rate of convergence of this Markov chain to its average $(\bar{x}_0, \dots, \bar{x}_0)$ where

$$\bar{x}_0 := \frac{\langle x_0, \mathbf{1} \rangle}{n}.$$

Remark. Without loss of generality, $\bar{x}_0 = 0$ and in general the slowest rates of convergence seem to take place when all of the weight is concentrated on a single coordinate vector. Thus, we could consider initial vectors

$$x_0 = \left(1 - \frac{1}{n}, -\frac{1}{n}, \dots, -\frac{1}{n}\right).$$

We have two ways to measure the distance from the average:

$$\begin{aligned} T(k) &:= \|x_k - x_0\|_{L^1} \\ S(k) &:= \|x_k - x_0\|_{L^2}. \end{aligned}$$

Proposition 4. Let $\{x_k\}_{k \geq 0} = \mathbf{M}_m^n$. For any $k \geq 0$, we have

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

Proof. We have

$$\begin{aligned} \mathbb{E}[S(k+1)|\mathcal{F}_k] &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \left\{ S(k) - \sum_{j=1}^m x_{k,i_j}^2 + m \left(\frac{x_{k,i_1} + \dots + x_{k,i_m}}{m} \right)^2 \right\} \\ &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \left\{ S(k) - \sum_{j=1}^m x_{k,i_j}^2 + \frac{1}{m} \sum_{j=1}^m x_{k,i_j}^2 + \frac{1}{m} \sum_{1 \leq u < v \leq m} x_{k,i_u} x_{k,i_v} \right\} \\ &= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \left\{ S(k) + \frac{1-m}{m} \sum_{j=1}^m x_{k,i_j}^2 + \frac{1}{m} \sum_{1 \leq u < v \leq m} x_{k,i_u} x_{k,i_v} \right\} \\ &= S_1 + S_2 + S_3 \end{aligned}$$

where

$$\begin{aligned} S_1 &:= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} S(k) \\ S_2 &:= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \frac{1-m}{m} \sum_{j=1}^m x_{k,i_j}^2 \\ S_3 &:= \binom{n}{m}^{-1} \sum_{i_1 < \dots < i_m} \frac{1}{m} \sum_{1 \leq u < v \leq m} 2x_{k,i_u} x_{k,i_v}. \end{aligned}$$

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

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

For the last sum, we get

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

where we used the fact that

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

Then

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

□

Corollary 2. *Let $\{x_t\}_{t \in \mathbb{R}_{\geq 0}} = \mathbb{PM}_m^n$. For $0 \leq s < t$, we have*

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

Proof. We have

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

□

As the previous computations show, the results are very similar and a similar result as Theorem 1 should hold (Maybe find the specific constant)?.

3 Random Bounded Averages

Now, we consider the variant where at each unit time we average a random number of elements. Let Q be the random variable with distribution (q_2, \dots, q_N) where $N \geq 1$ is a fixed positive integer and $\Pr[Q = k] = q_k$ for all k , $2 \leq k \leq N$. We will fix this distribution throughout the section.

Definition 4. *Let $\mathbf{RM} := \{x_k\}_{k \geq 0} \subset \mathbb{R}^n$ be the Markov chain where x_0 is some initial vector and x_{k+1} is obtained from x_k by uniformly picking Q_k distinct coordinates at random and replacing them by their average where the random variables Q_1, Q_2, \dots are independent, identically distributed with the same law as Q . Let $\mathbf{PRM} := \{x_k\}_{k \geq 0} \subset \mathbb{R}^n$ be the Poissonized version of \mathbf{RM} where we have a Poisson clock of rate 1 and every time it rings we perform an averaging in the same way as in the discrete-time setting.*

Before we do further analysis, we first show that we can get explicit formulas for the expectations of the L^2 bound. The first formula is for the base version of the process RM.

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

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

Proof. We have

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

□

Corollary 3. *Let $S^*(t)$ be the L^2 distance at time t for the Poissonized process. 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 ???. We can compute

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

□

We can equivalently formulate PRM by splitting our clock of rate 1 to $N-1$ independent Poisson clocks C_2, \dots, C_N where C_k is a Poisson clock of rate p_k . Then our original clock is then equivalent to the amalgamation of our clocks C_2, \dots, C_N where when the k th clock rings we pick k random positions and replace them by their average. In the original version of the problem, one key approximation step was to condition on the event that no average was performed on two non-zero elements. Let's assume in this case that the probability that we average at least two non-zero elements is negligible. (PROVE LATER?) Then, we can view our Markov chain as a type of branching process on an interval $[0, 1]$. The interval $[0, 1]$ in its initial state represents all of

the weight being in one position. Averagings then are represented by an interval being split into multiple equal length intervals. The clock C_k rings with rate q_k . When it rings, there is a probability of k/n that any given interval will be split into k equal parts. Equivalently, to every interval, we can attach 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

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

Let Q' be the random variable with distribution $(2q_2/\mu, \dots, Nq_N/\mu)$ where $\Pr[Q' = k] = kq_k/\mu$. So to each interval we attach a clock of rate μ/n and when it rings we split into a random number of equal parts according to the law of Q' . Let I_t be the set of intervals at time t . In the proof of the original problem, we found a time where the total content of the intervals were “close” to $1/n$. At this time, we can then apply the L^2 bound which will give the correct cutoff time. In other words, we want to find a time $t(n)$ and constants a_n, b_n where $1/n \ll a_n < b_n \sim n^{-1+o(1)}$ such that

$$\Pr \left(\sum_{I \in I_{t(n)}} \mu(I) \cdot \mathbb{1}_{a_n < \mu(I) < b_n} \geq 1 - o(1) \right) \geq 1 - o(1).$$

For $x \in [0, 1]$, let I_t^x be the interval of the point x at time t . Consider $\log |I_t^x|$. To this interval we have clocks D_2, \dots, D_N where D_k has rate $q_k \cdot k/n$. When clock D_k rings, then $\log |I_t^x|$ increases by $\log(1/k)$. This we can represent the process of $\log |I_t^x|$ as

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

Thus, the distribution of $\log |I_t^x|$ for fixed x is well understood. We are interested in the random variable

$$X = \sum_{I \in I_{t(n)}} \mu(I) \cdot \mathbb{1}_{a_n < \mu(I) < b_n}.$$

The expected value of X can be computed by exploiting the symmetry of the interval:

$$\mathbb{E}[X] = \mathbb{E} \int_0^1 \mathbb{1}_{|I_{t(n)}^x| \in (a_n, b_n)} dx = \int_0^1 \Pr[|I_{t(n)}^x| \in (a_n, b_n)] dx = \Pr[|I_{t(n)}^x| \in (a_n, b_n)]$$

from Fubini's Theorem. To have any hope in producing the bound (??), we need to understand the probability $\Pr[|I_{t(n)}^x| \in (a_n, b_n)]$. Note that

$$\Pr[|I_{t(n)}^x| \in (a_n, b_n)] = \Pr[\log |I_{t(n)}^x| \in (\log a_n, \log b_n)].$$

3.1 Logarithmic Length of the Interval Trajectory

Recall that for any fixed $t > 0$, we have that

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

Lemma 1. *Suppose $X_t \sim \text{Poisson}(t)$ for all t . Then*

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

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

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

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

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

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

From the Lemma, this gives

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

in distribution. Define the constants

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

Suppose $a_n = n^{-1+a(n)}$ and $b_n = n^{-1+b(n)}$.

3.2 Trying Random Stuff

Let's say $t := t_n = \gamma n(\log n - h(n))$ where $\gamma = \mathbb{E}[Q \log Q]^{-1}$ and $h(n) = o(\log n)$. Then we have

$$\begin{aligned} A_n &:= \sqrt{\frac{1}{\gamma(\log n - h(n))}} \cdot ((-1 + a(n)) \log n + n(\log n - h(n))) \\ &= \frac{\log n}{\sqrt{\gamma(\log n - h(n))}} \cdot \left(a(n) - \frac{h(n)}{\log n} \right) \\ B_n &:= \frac{\log n}{\sqrt{\gamma(\log n - h(n))}} \cdot \left(b(n) - \frac{h(n)}{\log n} \right) \end{aligned}$$

Then we can let,

$$\begin{aligned} a(n) &= \frac{h(n)}{2 \log n} \\ b(n) &= \frac{2h(n)}{\log n} \end{aligned}$$

for example to ensure $A_n \rightarrow -\infty$ and $B_n \rightarrow \infty$. This gives us $\Pr[|I_t^x| \in (a_n, b_n)] \rightarrow 1$.

The main questions are now:

- (a) Approximating the process with a new one, where we delete intervals if they get too small. Use this to prove that deleting intervals shouldn't be too bad.
- (b)
- (c)

3.3 New Process

Now that we have a_n , do a new process: the same but when an interval gets smaller than a_n we throw it out. Moreover, if an interval gets averaged with a non-zero one, then we also throw it out.

Question 1. *Let $x \in [0, 1]$. Then $\Pr[I_t^x \text{ gets averaged with non-zero pile}] \rightarrow 0?$ uniformly*

At any given point, there are at most $1/a_n$ intervals.

4 Starting Anew

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

Definition 6 (Fragmentation Process). *Let $\{\bar{x}_t\}$ 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. This is a fragmentation process version of the original process where we throw out piles which become too small.*

Definition 7 (Branching Process). *Let 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 $(2q_2/\mu, \dots, Nq_N/\mu)$.*

(describe the coupling?) there are floating labels on the intervals, and when an averaging happens, the labels change. This is the valid coupling. We can couple I_t and \bar{x}_t in a way such that at time t , when $\bar{x}_{t,i} \neq 0$, there exists some $\pi_t(i)$ such that $\bar{x}_{t,i} = |I_{\pi_t(i)}|$ where $I_{\pi_t(i)} \in I_t$ and when $i \neq j$ we have $\pi_t(i) \neq \pi_t(j)$. We couple the processes $\{x_t\}$ and $\{\bar{x}_t\}$ such that the same averagings happen to both of them. To the fragmentation process, there is an interval interpretation. For $x \in [0, 1]$, consider I_t^x , the interval containing x . From now on, let $t := \gamma n(\log n - (\log n)^{3/4})$. Define

$$\begin{aligned} a_n &= n^{-1+a(n)} \\ b_n &= n^{-1+b(n)} \\ a(n) &= \frac{1}{2}(\log n)^{-1/4} \\ b(n) &= 2(\log n)^{-1/4}. \end{aligned}$$

Let $t := \gamma n(\log n - (\log n)^{3/4})$.

Lemma 2. $\|x_t - \bar{x}_t\|_1 \rightarrow 0$ as $n \rightarrow \infty$ almost surely.

Proof. Clearly $x_{t,i} \geq \bar{x}_{t,i}$ for all coordinates $1 \leq i \leq n$ because \bar{x}_t is the same process as x_t except we throw out piles when they become too small or throw them out when we average at least two positive piles. The statement $\|x_t - \bar{x}_t\|_1 \rightarrow 0$ then states that the amount we throw out in the process \bar{x}_t is $o(1)$. The only way we throw out things is if we average at least two positive piles or we split up a pile into piles which are too small. In the latter case, the process I_t is the process \bar{x}_t if we condition on the event that we never average two positive piles and we don't throw away the small piles. Hence the total length of the intervals in I_t that are smaller than a_n will be greater than the amount of intervals we throw away in \bar{x}_t due to being too small. By Lemma (???), this length is $o(1)$ almost surely. Thus, it suffices to prove that the total length removed by averaging at least two positive piles is $o(1)$. Let D_t be the total length removed by averaging at least two positive intervals at time t . By linearity of expectation, we have that

$$\mathbb{E}[D_{t+\Delta} - D_t | \text{one ring}] = \mathbb{E} \sum_i \bar{x}_{t,i} \cdot p_i.$$

where p_{ij} is the probability we pick i and some j such that $\bar{x}_{t,j} \neq 0$. We can bound p_i by above by coupling with the process where at each averaging we pick N positions to average since this will

increase the probability that we average $\bar{x}_{t,i}$ with a positive entry. There are at most $1/a_n$ positive piles. The probability that we pick $\bar{x}_{t,i}$ in an avering is then N/n . The probability that we pick $\bar{x}_{t,i}$ and no other positive pile is at least $\frac{\binom{n-1/a_n}{N-1}}{\binom{n}{N}}$. Thus, we have that

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

where the last inequality follows from the mean value theorem. Thus, we have

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

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

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

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

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

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

Lemma 3. *We have that*

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

almost surely.

Proof. For the piles $\bar{x}_{t,i} \neq 0$, there is an injection into I_t into intervals of equal length. The other intervals correspond to $\bar{x}_{t,i} = 0$ intervals, but these intervals have been averaged with positive piles. But the total length of these intervals is dominated by $\|x_t - \bar{x}_t\|_1$ which we proved to be $o(1)$. Hence, we have that

$$\left| \sum_{i=1}^n \bar{x}_{t,i} \cdot \mathbb{1}_{\bar{x}_{t,i} \in (a_n, b_n)} - \sum_{t \in I_t} |I| \cdot \mathbb{1}_{I \in (a_n, b_n)} \right| \leq \|x_t - \bar{x}_t\|_1 = o(1).$$

This proves the lemma. \square

Proposition 6. $T(t) \rightarrow 2$ almost surely as $n \rightarrow \infty$.

Proof. We can write

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

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

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

This proves the proposition. □

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

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

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

Lemma 4. $\|\hat{x}_t - \bar{x}_t\|_1 \rightarrow 0$ as $n \rightarrow \infty$ almost surely.

Proof. When $\bar{x}_{t,i} \in (a_n, b_n)$, then we have that

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

where $z := \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \bar{x}_{t,i} \rightarrow 1$ almost surely as $n \rightarrow \infty$. When $\bar{x}_{t,i} \notin (a_n, b_n)$, the

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

Thus, we have that

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

As $n \rightarrow \infty$, we have $z \rightarrow 1$ and hence $\|\hat{x}_t - \bar{x}_t\|_1 \rightarrow 0$ as desired. □

Lemma 5. $\left\| \hat{x}_t - \frac{1}{n} \cdot \mathbb{1} \right\|_2^2 \leq \frac{2b_n^2}{a_n}$ for sufficiently large n .

Proof. We have that

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

where z is the quantity

$$z = \sum_{i: \bar{x}_{t,i} \in (a_n, b_n)} \bar{x}_{t,i} \in [0, 1].$$

Thus, we can bound

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

It is easily checked that $b_n^2/a_n \gg 1/n$, so for sufficiently large n we have the desired bound. \square

Lemma 6. *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 $\|x_0\|_1 = \|y_0\|_1 = 1$ and x_1, y_1 the vectors after performing the same averaging on x_0, y_0 respectively, then $\|x_1 - y_1\|_1 \leq \|x_0 - y_0\|_1$. By considering $z_0 = x_0 - y_0$ and $z_1 = x_1 - y_1$, it suffices to prove that if the average coordinates of z_0 is 0 and z_1 is the result of averaging some of the coordinates of z_0 , then $\|z_1\|_1 \leq \|z_0\|_1$. Suppose we averaged the values u_1, \dots, u_k . Then, we have that

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

This proves the lemma. \square

Now let $\Omega > 0$ be a constant which is larger than both 2γ and $7/2(\mu - 1)$. Let $s = \Omega(n - 1)(\log n)^{3/4}$. Then, we have the following lemma.

Lemma 7. *$T(s + t) \rightarrow 0$ as $n \rightarrow \infty$ almost surely.*

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

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

From the Cauchy-Schwartz inequality, we have that

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

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

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

This proves the lemma. \square

References

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