Random Iterated Averages

Alan Yan
Project Advisor: Professor Allan Sly

April 5, 2022

Contents

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 $x_0, x_1, \ldots$ where x_0 is some initial vector in $\mathbb{R}^n$. Notation-wise, let $x_k = (x_{k,1}, \ldots, x_{k,n})$. 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 $(\overline{x}_0, \ldots, \overline{x}_0)$ where

$$\overline{x_0} = \frac{x_{0,1} + \ldots + x_{0,n}}{n}.$$

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.

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

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

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 formla 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 $\Phi : \mathbb{R} \to [0,1]$ 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*

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

where the convergence is in probability.

Clearly, Theorem 1 implies the following result.

Corollary 1 (Theorem 1.1 in [1])**.** *For $x_0 = (1, 0, \ldots, 0)$ 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 $\theta > 1/(2\log 2)$.*

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:

$$
\begin{aligned}
T'(k) &:= \sum_{i=1}^{n} |x_{t,i} - \overline{x}_0| \\
S'(k) &:= \sum_{i=1}^{n} |x_{t,i} - \overline{x}_0|^2 .
\end{aligned}
$$

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

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

This suffices for the proof. □

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

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 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* 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* $\mathbb{P}\mathsf{M}_m^n = \{x_t\}_{t\in\mathbb{R}_{\geq 0}}$ *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 $(\overline{x}_0, \ldots, \overline{x}_0)$ where

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

Remark. Without loss of genearlity, $\overline{x}_0 = 0$ 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

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

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* $\{x_k\}_{k\geq 0} = \mathsf{M}_m^n$*. For any* $k \geq 0$*, we have*

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

Proof. We have

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

where

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

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

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

where we used the fact that

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

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 $\{x_t\}_{t \in \mathbb{R}_{\ge 0}} = \mathbb{P}\mathsf{M}_m^n$. For $0 \le s < t$, we have*

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

Proof. We have

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

□

As the prevous computatoins 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 $(q_2, \ldots, q_N)$ where $N \ge 1$ is a fixed positive integer and $\Pr[Q = k] = q_k$ for all k, $2 \le k \le N$. We will fix this distribution throughout the section.

Definition 4. *Let $\mathsf{RM} := \{x_k\}_{k \ge 0} \subset \mathbb{R}^n$ 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 $Q_1, Q_2, \ldots$ are independent, identically distributed with the same law as Q. Let $\mathsf{PRM} := \{x_k\}_{k \ge 0} \subset \mathbb{R}^n$ be the Poissonized version of 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.*

We can equivalently formulate PRM by splitting our clock of rate 1 to $N-1$ independent Poisson clocks $C_2, \ldots, C_N$ where C_k is a Poisson clock of rate p_k. Then our original clock is then equivalent to the amalgamation of our clocks $C_2, \ldots, C_N$ where when the kth 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 $(2q_2/\mu, \ldots, Nq_N/\mu)$ 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 $1/n << a_n < b_n \sim n^{-1+o(1)}$ 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 $D_2, \ldots, D_N$ 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

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

Thus, the distribution of $\log |I_t^x|$ for fixed x is well understood. Now note that

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

is the probability that if we pick a random point uniformly from $[0,1]$, the point will lie in an interval of length between a_n and b_n.

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

$$=$$

References

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