# Junior Seminar Notes

Alan Yan
Project Advisor: Professor Allan Sly

February 22, 2022

## Contents

# 1 The Initial Conditions

We first describe the initial 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 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

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

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

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

We have an exact formula for the expected $L^2$ norm. Without loss of generality, we can assume $\bar{x_0} = 0$.

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

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

*Proof.* The calculation is done in the paper. □

To avoid issues of dependence in the discrete time setting, we approximate Markov chain I with the following continuous time Markov chain.

**Definition 2** (Markov Chain II). *We have a Markov chain $\{x_t\}_{t\geq 0}$. 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.*

With this new Markov chain, we introduce the same measurements from the average.

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

We get a similar formula for the expected $L^2$ 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

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

This suffices for the proof. □

From here on out, we study the process starting from $x_0 = (1 - 1/n, -1/n, \ldots, -1/n)$. The main theorem that the paper proves is the following.

**Theorem 1** (Theorem 2.3 in the paper)**.** *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 $n \to \infty$.*

In the discrete case, we have the following theorem.

**Theorem 2** (Theorem 1.2 [1])**.** *For any $a \in \mathbb{R}$, we have that*

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

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

*Proof.* TODO □

## 1.1 Generalized Averages

**Definition 3.** *Let $\mathsf{M}_m^n$ 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 1.** *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

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

□

# References

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