

Chapter 1

Ratios and Mixing Times

1.1 Introduction

In this chapter, we'll talk about consecutive ratio sequence simulations.

1.2 Markov Chains

Definition 1.2.1 (Markov Chain). *Say a set of random variables X_i each take a value in a set, called the state space, $S_M = \{1, 2, \dots, M\}$. Then, a sequence of such random variables X_0, X_1, \dots, X_n is called a Markov Chain if the following conditions are satisfied:*

- $\forall X_i : X_i$ has support and range $S_M = \{1, 2, \dots, M\}$.
- (*Markov Property*) The transition probability from state $i \rightarrow j$, $P(X_{n+1} = j \mid X_n = i)$ is conditionally independent from all past events in the sequence $X_{n-1} = i', X_{n-2} = i'', \dots, X_0 = i^{(n-1)}$, excluding the present/last event in the sequence. In other words, given the present, the past and the future are conditionally independent.

$$\forall i, j \in S_M : P(X_{n+1} = j \mid X_n = i) = P(X_{n+1} = j \mid X_n = i, X_{n-1} = i', \dots, X_0 = i^{(n-1)})$$

Definition 1.2.2 (Transition Matrix). *Let X_0, X_1, \dots, X_M be a Markov Chain with state space S_M . Letting $q_{ij} = P(X_{n+1} = j \mid X_n = i)$ be the transition probability from $i \rightarrow j$, then the matrix $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M] : Q = (q_{ij})$ is the transition matrix of the chain. Q must satisfy the following conditions to be a valid transition matrix:*

Definition 1.2.3 (Transition Matrix). *Take a Markov Chain with states $1, \dots, M$. Letting $q_{ij} = P(X_{n+1} = j \mid X_n = i)$ be the transition probability from $i \rightarrow j$, then the matrix $Q = (q_{ij})$ is the transition matrix of the chain. For this transition matrix to be valid, its rows have to be stochastic, meaning their entries sum to 1; $\forall i \in 1, \dots, M : \sum_{j \in 1, \dots, M} q_{ij} = 1$.*

- Q is a non-negative matrix. That is, note that $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M]$ so every $q_{ij} \in \mathbb{R}^+$. This follows because probabilities are necessarily non-negative values.
- The entries of every row i of Q must sum up to 1. This may be understood as applying the law of total probability to the event of transitioning from any given state $\forall i \in S_M$. In other words, the chain has to go somewhere with probability 1.

$$\forall i \in S_M : \sum_{j \in S_M} q_{ij} = 1$$

- Note, it is NOT necessary that the converse holds. The columns of our transition matrix need not sum to 1 for it to be a valid transition matrix.

Definition 1.2.4 (n-Step Transition Probability). *The n -step transition probability of $i \rightarrow j$ is the probability of being at j exactly n steps after being at i . We denote this value $q_{ij}^{(n)}$:*

$$q_{ij}^{(n)} : P(X_n = j \mid X_0 = i)$$

Realize:

$$q_{ij}^{(2)} = \sum_{k \in S_M} q_{ik} \cdot q_{kj}$$

Because by definition, a Markov Chain is closed under a support/range of S_M so the event $i \rightarrow j$ may have taken any intermediate step $k \in S_M$. Realize by notational equivalence, $Q^2 = (q_{ij}^{(2)})$. Inducting over n , we then obtain that:

$$q_{ij}^{(n)} \text{ is the } (i, j) \text{ entry of } Q^n$$

Definition 1.2.5 (Marginal Distribution of X_n). *Let $\mathbf{t} = (t_1, t_2, \dots, t_M)$ such that $\forall i \in S_M : t_i = P(X_0 = i)$ So, $\mathbf{t} \in \mathcal{M}_{\mathbb{R}}[1, M]$. Then, the marginal distribution of X_n is given by the product of the vector $\mathbf{t}Q^n \in \mathcal{M}_{\mathbb{R}}[1, M]$. That is, the j^{th} component of that vector is $P(X_n = j)$ for any $j \in S_M$. We may call \mathbf{t} an initial state distribution.*

1.2.1 Classification of states

- A state $i \in S_M$ is said to be **recurrent** if starting from i , the probability is 1 that the chain will *eventually* return to i . If the chain is not recurrent, it is **transient**, meaning that if it starts at i , there is a non-zero probability that it never returns to i .
- Caveat: As we let $n \rightarrow \infty$, our Markov chain will guarantee that all transient states will be left forever, no matter how small the probability is. This can be proven by letting the probability be some ε , then realizing that by the support of $\text{Geom}(\varepsilon)$ is always some finite value, then the equivalence between the Markov property and independent Geometric trials guarantees the existence of some finite value such that there is a success of never returning to i .

Definition 1.2.6 (Reducibility). A Markov chain is said to be **irreducible** if for any $i, j \in S_M$, it is possible to go from $i \rightarrow j$ in a finite number of steps with positive probability. In other words:

$$\forall i, j \in S_M : \exists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

- From our quantifier formulation of irreducible Markov chains, note that we can equivalently say that a chain is irreducible if there is integer $n \in \mathbb{N}$ such that the (i, j) entry of Q^n is positive for any i, j .
- A Markov chain is **reducible** if it is not **irreducible**. Using our quantifier formulation, it means that it suffices to find transient states so that:

$$\exists i, j \in S_M : \nexists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

1.3 Erdos-Renyi Graphs

Definition 1.3.1 (Erdos-Renyi Graph). An Erdos-Renyi graph is a graph $G = (V, E)$ with a set of vertices $V = 1, \dots, nM$ and edges $E = \mathbf{1}_{i,j \in V} \sim \text{Bern}(p_{ij})$. It is homogenous if $p_{ij} = p$ is fixed for all i, j .

Essentially, an Erdos-Renyi graph is a graph whose 'connectedness' is parameterized by a probability p (assuming it's homogenous, which this document will unless otherwise noted). As $p \rightarrow 0$, we say that graph becomes more sparse; analogously, as $p \rightarrow 1$ the graph becomes more connected.

Recall from probability theory that a sum of i.i.d Bernoulli random variables is a Binomial variable. As such, we may alternatively say that the degree of each vertex v is distributed as $\deg(v) \sim \text{Bin}(M, p)$. This is helpful to know because the process of simulating graphs becomes much simpler.

1.4 Mixing Time Simulations

With the Erdos-Renyi graph defined, we may now motivate the simulation of random walks on them. First, however, we need to generate their corresponding transition matrices. An algorithm for this is outlined below.

Suppose we have simulated a transition matrix for an Erdos-Renyi graph called Q . Now, fixing some initial probability distribution $\vec{x} \in \mathbb{R}^M$, we may consider the evolution sequence of a random walk on this Erdos-Renyi graph by taking its evolution sequence $\mathcal{S}(Q, x)$.

Definition 1.4.1 (Random Batches). Let \mathbb{F} be a field, and fix some $M \in \mathbb{N}$. Let $\mathcal{B}_\lambda \subset \mathbb{F}^M$ be a uniformly random batch of points in the M -hypercube of length λ . That is,

$$\mathcal{B}_\lambda = \{\vec{x} \mid x_i \sim \text{Unif}(-\lambda, \lambda) \text{ for } i = 1, \dots, M\}$$

Note: If $\mathbb{F} = \mathbb{C}$, then take $\vec{x} \in \mathcal{B}_\lambda$ to mean $\vec{x} = a + bi$ where $a, b \sim \text{Unif}(-\lambda, \lambda)$.

Definition 1.4.2 (Evolution Sequence). *An evolution sequence of a vector $\vec{\pi}$ and a transition matrix Q is defined as the sequence $\mathcal{S}(Q, \pi) = (\pi'_n)_{n=1}^N$ where $\pi'_n = \pi Q^n$*

Definition 1.4.3 (Finite Evolution Sequences). *Suppose we sample a random point from \mathcal{B}_λ , emulating a random point $\vec{v} \in \mathbb{F}^M$. Additionally, let $Q \in \mathbb{F}^{M \times M}$ be a transition matrix over \mathbb{F} . Fixing a maximum power ('time') $T \in \mathbb{N}$, define the evolution sequence of \vec{v} as follows:*

$$\mathcal{S}(v, Q, T) = (\alpha_n)_{n=1}^T \text{ where } \alpha_k = vQ^k$$

If we do not impose a finiteness constraint on the sequence, we consider powers for $n \in \mathbb{N}$ or $t = \infty$

Definition 1.4.4 (Consecutive Ratio Sequences). *Accordingly, define the consecutive ratio sequence (CST) of \vec{v} as follows:*

$$\mathcal{R}(v, Q, T) = (r_n)_{n=2}^T \text{ where } (r_n)_j = \frac{(\alpha_n)_j}{(\alpha_{n-1})_j} \text{ for } j = 1, \dots, M$$

*In other words, the consecutive ratio sequence of v can be obtained by performing **component-wise division** on consecutive elements of the evolution sequence of v .*

Definition 1.4.5 (Near Convergence). *Because these sequences may never truly converge to eigenvectors of the matrix, we formalize a notion of "near convergence". As a preliminary, we first define ε -equivalence. Let \mathbb{F} be a field, and fix $\varepsilon \in \mathbb{R}^+$. Suppose we have vectors $v, v' \in \mathbb{F}^M$. Then, $v \sim_\varepsilon v'$ if $\|v - v'\| < \varepsilon$ where $\|\cdot\|$ is the norm on \mathbb{F} .*

Let $\varepsilon \in \mathbb{R}^+$, and suppose we have an evolution sequence $(a[\vec{v}])_n$. Then, a_n ε -converges at $N \in \mathbb{N}$ if:

$$\forall n \geq N \mid a_N \sim_\varepsilon a_n$$

1.5 Questions

1. How are the entries of the CRS distributed? Are they normal, and if so, what is its mean?
2. Are the entries of the CRS i.i.d as $t \rightarrow \infty$?
3. For an Erdos-Renyi matrix, is the mixing time t dependent on the parameter p ?
4. What impact does the running time parameter T have on σ (the variance of the distribution of the CRS entries)?

1.5.1 Questions

It seems to be the case that the **log-transformed** entries of the CRS are Cauchy distributed about $\log \lambda_1$ where $\lambda_1 = \max(\sigma(Q))$, the largest eigenvalue of Q . That is,

$$r_i \sim \text{Cauchy}(\ln \lambda_1) \text{ for } i = 1, \dots, M$$