Chapter 8

Finite Markov Chains

A ***discrete system*** is characterized by a set V of "states" and ***transitions*** between the states. V is referred to as the **state space**. We think of the transitions as occurring at each time beat, so the state of the system at time t is a value $X_t \in V$ $(t = 0, 1, 2, \dots)$. The adjective "discrete" refers to discrete time beats.

A ***discrete stochastic process*** is a discrete system in which transitions occur randomly according to some probability distribution. The process is ***memoryless*** if the probability of an $i \to j$ transition does not depend on the history of the process (the sequence of previous states): $(\forall i, j, u_0, \dots, u_{t-1} \in V)(P(X_{t+1} = j \mid X_t = i, X_{t-1} = u_{t-1}, \dots, X_0 = u_0) = P(X_{t+1} = j \mid X_t = i))$. (Here the universal quantifier is limited to feasible sequences of states $u_0, u_1, \dots, u_{t-1}, i$, i. e., to sequences which occur with positive probability; otherwise the conditional probability stated would be undefined.) If in addition the transtion probability $p_{ij} = P(X_{t+1} = j \mid X_t = i\}$ does not depend on the time t, we call the process ***homogeneous***.

Definition 8.0.7. A **finite Markov chain** is a memoryless homogeneous discrete stochastic process with a finite number of states.

Let $\mathcal{M}$ be a finite Markov chain with n states, $V = [n] = \{1, 2, \dots, n\}$. Let p_{ij} denote the probability of transition from state i to state j, i. e., $p_{ij} = P(X_{t+1} = j \mid X_t = i)$. (Note that this is a conditional probability: the question of $i \to j$ transition only arises if the system is in state i, i. e., $X_t = i$.)

The finite Markov chain $\mathcal{M}$ is characterized by the $n \times n$ ***transition matrix*** $T = (p_{ij})$ $(i, j \in [n])$ and an ***initial distribution*** $q = (q_1, \dots, q_n)$ where $q_i = P(X_0 = i)$.

Definition 8.0.8. An $n \times n$ matrix $T = (p_{ij})$ is ***stochastic*** if its entries are nonnegative real numbers and the sum of each row is 1:

$(\forall i, j)(p_{ij} \geq 0)$ and $(\forall i)(\sum_{j=1}^{n} p_{ij} = 1)$.

Exercise 8.0.9. The transition matrix of a finite Markov chain is a stochastic matrix. Conversely, every stochastic matrix can be viewed as the transition matrix of a finite Markov chain.

Exercise 8.0.10. Prove: if T is a stochastic matrix then T^k is a stochastic matrix for every k.

Random walks on digraphs are important examples of finite Markov chains. They are defined by hopping from vertex to neighboring vertex, giving equal chance to each out-neighbor. The state space will be V, the set of vertices. The formal definition follows.

Let $G = (V, E)$ be a finite digraph; let $V = [n]$. Assume $(\forall i \in V)(\deg^+(i) \geq 1)$. Set $p_{ij} = 1/\deg^+(i)$ if $(i, j) \in E$; $p_{ij} = 0$ otherwise.

Exercise 8.0.11. Prove that the matrix (p_{ij}) defined in the preceding paragraph is stochastic.

Conversely, all finite Markov chains can be viewed as *weighted* random walks on a digraph, the weights being the transition probabilities. The formal definition follows.

Let $T = (p_{ij})$ be an arbitrary (not necessarily stochastic) $n \times n$ matrix. We associate with T a digraph $G = (V, E)$ as follows. Let $V = [n]$ and $E = \{(i, j) : p_{ij} \neq 0\}$. We label the edge $i \to j$ with the number $p_{ij} \neq 0$ (the "weight" of the edge).

This definition makes sense for any matrix T; edges indicate nonzero entries. If T is the transition matrix of a finite Markov chain $\mathcal{M}$ then we call the associated digraph the **transition digraph** of $\mathcal{M}$. The **vertices** of the transition digraph represent the **states** of $\mathcal{M}$ and the **edges** the **feasible transitions** (transitions that occur with positive probability).

Exercise 8.0.12. Prove that in the transition digraph of a finite Markov chain, $(\forall i)(\deg^+(i) \geq 1)$.

Exercise 8.0.13. Draw the transition digraph corresponding to the stochastic matrix

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

Label the edges with the transition probabilities.

The principal subject of study in the theory of Markov chains is the **evolution** of the system.

The initial distribution $q = (q_1, \dots, q_n)$ describes the probability that the system is in a particular state at time $t = 0$. So $q_i \geq 0$ and $\sum_{i=1}^n q_i = 1$.

Set $q(0) = q$ and let $q(t) = (q_{1t}, \dots, q_{nt})$ be the distribution of the states at time t, i. e., the distribution of the random variable X_t:

$$q_{it} = P(X_t = i).$$

The following simple equation describes the evolution of a finite Markov chain.

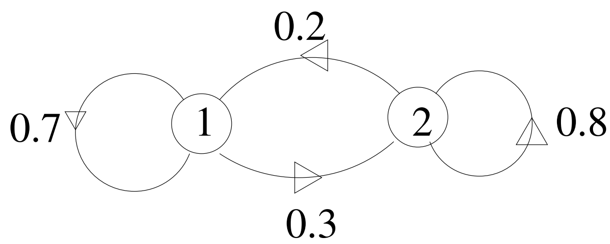

Figure 8.1: The solution to Exercise 8.0.13

Exercise 8.0.14. (Evolution of Markov chains) Prove: $q(t) = q(0)T^t$.

So the study of the *evolution of a finite Markov chain* amounts to studying the *powers of the transition matrix.*

Exercise 8.0.15. Experiment: study the powers of the matrix A defined in Exercise 8.0.13. Observe that the sequence $I, A, A^2, A^3, \ldots$ appears to converge. What is the limit?

Exercise$^+$ 8.0.16. Prove the convergence observed in the preceding exercise.

The study of the powers rests on the study of *eigenvalues* and *eigenvectors*.

Definition 8.0.17. A ***left eigenvector*** of an $n \times n$ matrix A is a $1 \times n$ vector $x \neq 0$ such that $xA = \lambda x$ for some (complex) number λ called the *eigenvalue* corresponding to x. A ***right eigenvector*** of A is an $n \times 1$ matrix $y \neq 0$ such that $Ay = \mu y$ for some (complex) number μ called the ***eigenvalue*** corresponding to y.

Remember that the zero vector is never an eigenvector. Note that if $x = (x_1, \ldots, x_n)$ is a $1 \times n$ vector, $A = (a_{ij})$ is an $n \times n$ matrix, and $z = (z_1, \ldots, z_n) = xA$ then

$$z_j = \sum_{i=1}^{n} x_i a_{ij}. \tag{8.1}$$

Note that if G is the digraph associated with the matrix A then the summation can be reduced to

$$z_j = \sum_{i:i\to j}^{n} x_i a_{ij}. \tag{8.2}$$

So the **left eigenvectors** to the eigenvalue λ is defined by the equation

$$\lambda x_j = \sum_{i:i\to j}^{n} x_i a_{ij}. \tag{8.3}$$

Exercise 8.0.18. State the equations for the left action and the right eigenvectors of the matrix A.

Theorem. The left and the right eigenvalues of a matrix are the same (but not the eigenvectors!).

Proof. Both the right and the left eigenvalues are the roots of the **characteristic polynomial** $f_A(x) = \det(xI - A)$ where I is the $n \times n$ identity matrix.

Exercise 8.0.19. Find the eigenvalues and the corresponding left and right eigenvectors of the matrix A from Exercise 8.0.13.

Hint. The characteristic polynomial is

$$f_A(x) = \begin{vmatrix} x - 0.7 & -0.3 \\ -0.2 & x - 0.8 \end{vmatrix} = x^2 - 1.5x + 0.5 = (x-1)(x-1/2).$$

So the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1/2$. Each eigenvalue gives rise to a system of linear equations for the coordinates of the corresponding (left/right) eigenvectors.

Exercise$^+$ 8.0.20. Prove: if λ is a (complex) eigenvalue of a stochastic matrix then $|\lambda| \leq 1$. *Hint.* Consider a right eigenvector to eigenvalue λ.

Exercise 8.0.21. Let A be an $n \times n$ matrix. Prove: if x is a left eigenvector to eigenvalue λ and y is a right eigenvector to eigenvalue μ and $\lambda \neq \mu$ then x and y are **orthogonal,** i.e., $xy = 0$. *Hint.* Consider the product xAy.

Definition 8.0.22. A ***stationary distribution*** (also called ***equilibrium distribution***) for the Markov chain is a probability distribution $q = (q_1, \ldots, q_n)$ ($q_i \geq 0$, $\sum_{i=1}^n q_i = 1$) which is a left eigenvector to the eigenvalue 1: $\quad qA = q$.

Exercise 8.0.23. If at time t, the distribution $q(t)$ is stationary then it will remain the same forever: $q(t) = q(t+1) = q(t+2) = \ldots$.

Exercise 8.0.24. Prove: if T is a stochastic matrix then $\lambda = 1$ is a right eigenvalue. *Hint.* Guess the (very simple) eigenvector.

Observe the consequence that $\lambda = 1$ is also a *left* eigenvalue. This is significant because it raises the possibility of having stationary distributions.

Exercise 8.0.25. Find a *left* eigenvector $x = (x_1, x_2)$ to the eigenvalue 1 for the stochastic matrix A defined in Exercise 8.0.13. Normalize your eigenvector such that $|x_1| + |x_2| = 1$. Observe that x is a stationary distribution for A.

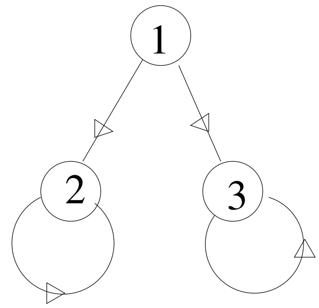

Figure 8.2: A graph with transition probabilities. FIX THIS!

Exercise 8.0.26. Let T be a stochastic matrix. Prove: **if** the limit $T^{\infty} = \lim_{t\to\infty} T^t$ **exists** then every row of T^{∞} is a stationary distribution.

Exercise 8.0.27. Consider the stochastic matrix

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Prove that the sequence $I, B, B^2, B^3, \ldots$ does **not** converge, yet B does have a stationary distribution.

Exercise 8.0.28. Let $\vec{C}_n$ denote the directed cycle of length n. Prove that the powers of the transition matrix of the random walk on $\vec{C}_n$ do not converge; but a stationary distribution exists.

Exercise 8.0.29. Consider the following digraph: $V = [3]$, $E = \{1 \to 2, 1 \to 3, 2 \to 2, 3 \to 3\}$.

Write down the transition matrix of the random walk on the graph shown in Figure 8. Prove that the random walk on this graph has 2 stationary distributions.

Definition 8.0.30. A stochastic matrix $T = (p_{ij})$ is called ***doubly stochastic*** if its column sums are equal to 1: $(\forall j \in [n])(\sum_{i=1}^{n} p_{ij} = 1)$.

In other words, T is doubly stochastic if both T and its transpose are stochastic.

Exercise 8.0.31. Let T be the transition matrix for a finite Markov chain M. Prove that the uniform distribution is stationary if and only if T is doubly stochastic.

A matrix is called **non-negative** if all entries of the matrix are non-negative. The *Perron–Frobenius theory of non-negative matrices* provides the following fundamental result.

Theorem (Perron–Frobenius, abridged) If A is a non-negative $n \times n$ matrix then A has a non-negative left eigenvector.

Exercise 8.0.32. Prove that a non-negative matrix has a non-negative right eigenvector. (Use the Perron–Frobenius Theorem.)

Exercise 8.0.33. Let T be a stochastic matrix and x a non-negative left eigenvector to eigenvalue λ. Prove: $\lambda = 1$. *Hint.* Use Exercise 8.0.21.

Exercise 8.0.34. Prove: **every finite Markov chain has a stationary distribution.**

Exercise$^+$ 8.0.35. Let A be a non-negative matrix, x a non-negative left eigenvector of A, and G the digraph associated with A. Prove: if G is strongly connected then all entries of x are positive. *Hint.* Use equation (8.3).

Exercise 8.0.36. Let A be a non-negative matrix, x and x' two non-negative eigenvectors of A, and G the digraph associated with A. Prove: if G is strongly connected then x and x' belong to the same eigenvalue. *Hint.* Use the preceding exercise and Exercise 8.0.21.

Exercise$^+$ 8.0.37. Let A be a non-negative matrix; let x be a non-negative left eigenvector to the eigenvalue λ and let x' be another left eigenvector with real coordinates to the same eigenvalue. Prove: if G is **strongly connected** then $(\exists \alpha \in \mathbb{R})(x' = \alpha x)$. *Hint.* WLOG (without loss of generality we may assume that) all entries of x are positive (why?). Moreover, WLOG $(\forall i \in V)(x'_i \leq x_i)$ and $(\exists j \in V)(x'_j = x_j)$ (why?). Now prove: if $x_j = x'_j$ and $i \to j$ then $x_i = x'_i$. Use equation (8.3).

Finite Markov chains with a **strongly connected** transition digraph (every state is accessible from every state) are of particular importance. Such Markov chains are called **irreducible.** To emphasize the underlying graph theoretic concept (and reduce the terminology overload), we shall deviate from the accepted usage and use the term **strongly connected Markov chains** instead of the classical and commonly used term "irreducible Markov chains."

Our results are summed up in the following exercise, an immediate consequence of the preceding three exercises.

Exercise 8.0.38. Prove: **A strongly connected finite Markov chain (a) has exactly one stationary distribution; and (b) all probabilities in the stationary distribution are positive.**

As we have seen (which exercise?), strong connectivity is not sufficient for the powers of the transition matrix to converge. One more condition is needed.

Definition 8.0.39. The ***period*** of a vertex v in the digraph G is the g.c.d. of the lengths of all closed directed walks in G passing through v. If G has no closed directed walks through v, the period of v is said to be 0. If the period of v is 1 then v is said to be ***aperiodic.***

Exercise 8.0.40. (a) Show that it is not possible for every state of a finite Markov chain to have period 0 (in the transition digraph). (b) Construct a Markov chain with n states, such that all but one state has period 0.

Note that a **loop** is a closed walk of length 1, so if G has a loop at v then v is automatically aperiodic. A **lazy random walk** on a digraph stops at each vertex with probability 1/2 and divides the remianing 1/2 evenly between the out-neighbors ($p_{ii} = 1/2$, and if $i \to j$ then $p_{ij} = 1/2 \deg +(i)$). So the lazy random walks are aperiodic at each vertex.

Exercise 8.0.41. Let $G = (V, E)$ be a digraph and $x, y \in V$ two vertices of G. Prove: if x and y belong to the same strong component of G (i. e., x and y are mutually accessible from one another) then the periods of x and y are equal.

It follows that **all states of a strongly connected finite Markov chain have the same period.** We call this common value the **period** of the strongly connected Markov chain. A Markov chain is **aperiodic** if every node has period 1.

Exercise 8.0.42. Recall that (undirected) graphs can be viewed as digraphs with each pair of adjacent vertices being connected in both directions. Let G be an undirected graph viewed as a digraph. Prove: every vertex of G has period 1 or 2. The period of a vertex v is 2 if and only the connected component of G containing v is bipartite.

Exercise 8.0.43. Suppose a finite Markov chain $\mathcal{M}$ is strongly connected and NOT aperiodic. (It follows that the period ≥ 2 (why?).)
Prove: the powers of the transition matrix do not converge.
Hint. If the period is d, prove that the transition graph is a "blown-up directed cycle of length d" in the following sense: the vertices of the transition graph can be divided into d disjoint subsets $V_0, V_1, \ldots, V_{d-1}$ such that $(\forall k)$ all edges starting at V_k end in V_{k+1}, where the subscript is read modulo d (wraps around). – Once you have this structure, observe that any t-step transition would take a state in V_k to a state in V_{k+t} (the subscript again modulo d).

Now we state the Perron–Frobenius Theorem in full.

Theorem (Perron–Frobenius, unabridged) Let A be a non-negative $n \times n$ matrix and G the associated digraph. Let $f_A(x) = \prod_{i=1}^{n}(x - \lambda_i)$ be the characteristic polynomial of A factored over the complex numbers. (So the λ_i are the eigenvalues, listed with multiplicity.) Then

(a) There is an eigenvalue λ_1 such that

  (a1) λ_1 is real and non-negative;

(a2) $(\forall i)(\lambda_1 \geq |\lambda_i|)$;

(a3) there exists a non-negative eigenvector to eigenvalue λ_1.

(b) If G is strongly connected and **aperiodic** then $(\forall i)(\lambda_1 > |\lambda_i|)$.

Definition 8.0.44. A strongly connected aperiodic Markov chain is called ***ergodic.***

The significance of aperiodicity is illuminated by the following exercises.

Exercise 8.0.45. Prove that the eigenvalues of the random walk on the directed n-cycle are exactly the n-th roots of unity. (So all of them have unit absolute value.)

More generally, we have the following:

Exercise 8.0.46. Let A be a (not necessarily non-negative) $n \times n$ matrix and G the associated digraph. Suppose d is a common divisor of the periods of G. Let ω be a complex d-th root of unity (i. e., $\omega^d = 1$). Then, if λ is an eigenvalue of A then $\lambda\omega$ is also an eigenvalue of A. *Hint.* Equation (8.3).

The following consequence of the Perron–Frobenius Theorem is the fundamental result in the theory of finite Markov chains.

Exercise* 8.0.47. (Convergence of ergodic Markov chains.) Prove: if T is the transition matrix of an **ergodic Markov chain** then the powers of T **converge.** *Hint.* There exists an invertible complex matrix S such that $U = S^{-1}TS$ is an upper triangular matrix of which the first row is $[1, 0, 0, \dots, 0]$. (This follows, for example, from the Jordan normal form.) Now the diagonal entries of U are the eigenvalues, starting with $\lambda_1 = 1$; all other eigenvalues satisfy $|\lambda_i| < 1$. Prove that as a consequence, the sequence U^t $(t \to \infty)$ converges to the matrix N which has a 1 in the top left corner and 0 everywhere else. Now $T^k \to M := SNS^{-1}$ (why?).

Exercise 8.0.48. Prove: if T is the transition matrix of an ergodic Markov chain and $\lim_{t\to\infty} T^t = M$ then all rows of M are equal.

Exercise 8.0.49. Prove: if a finite Markov chain is ergodic then from any initial distributrion, the process will approach the unique stationary distribution. In other words, let T be the transition matrix, s the stationary distribution, and q an arbitrary initial distribution. Then

$$\lim_{t\to\infty} qT^t = s.$$

The following example illuminates the kind of Markov chains encountered in combinatorics, theoretical computer science, and statistical physics.

Random recoloring: a class of large Markov chains. Let $G = (V, E)$ be a graph with n vertices and maximum degree Δ; and let $Q \geq \Delta + 1$. Let S be the set of all legal colorings of G with Q colors, i. e., S is the set of functions $f : V \to [Q]$ such that if $v, w \in V$ are adjacent then $f(v) \neq f(w)$. This "random recoloring process" is a Markov chain which takes S as its set of states (the "state space"). The transitions from a legal coloring are defined as follows. We pick a vertex $v \in V$ at random, and recolor it by one of the available colors (colors not used by the neighbors of v), giving each available color an equal chance (including the current color of v).

Exercise 8.0.50. Prove: if $Q \geq \Delta + 2$ then the random recoloring process is an ergodic Markov chain.

Exercise 8.0.51. Prove that the number of states of the random recoloring process is between $(Q - \Delta - 1)^n$ and Q^n. So if $Q \geq \Delta + 2$ then the state space is exponentially large.

Exercise 8.0.52. Prove: if $Q \geq \Delta + 2$ then the stationary distribution for the random recoloring process is uniform.

As a consequence, the random recoloring process will converge to a uniformly distributed random legal Q-coloring of G. Just how quickly the process approaches the uniform distribution is an open problem. While the state space is exponential, it is expected that the process distribution will be close to uniform within a polynomial (n^{const}) number of steps. This phenomenon is called **rapid mixing**. Marc Jerrum proved in 1995 that for $Q > 2\Delta$, the random recoloring process does indeed mix rapidly; Jerrum proved an $O(n \log n)$ bound on the mixing time. In a recent (2000) paper, published in the *Journal of Mathematical Physics*, Eric Vigoda showed that the 2Δ bound was not best possible; he proved that rapid mixing already occurs for $Q > (11/6)\Delta$; under this weaker condition Vigoda shows a somewhat less rapid, $O(n^2 \log n)$ mixing. The techniques leading to such improvements are expected to be widely applicable in combinatorics, theoretical computer science, and statistical physics.

Concluding remarks. Markov chains are widely used models in a variety of areas of theoretical and applied mathematics and science, including statistics, operations research, industrial engineering, linguistics, artificial intelligence, demographics, genomics. Markov chain models are used in performance evaluation for computer systems ("if the system goes down, what is the chance it will come back?"), in queuing theory (server queuing, intelligent transportation systems). Hidden Markov models (where the transition probabilities are not known) are a standard tool in the design of intelligent systems, including speech recognition, natural language modelling, pattern recognition, weather prediction.

In discrete mathematics, theoretical computer science, and statistical physics, we often have to consider finite Markov chains with an enormous number of states. Card shuffling is an example of a Markov chain with 52! states. The "random recoloring process," discussed above, is an example of a class of Markov chains which have exponentially many states compared

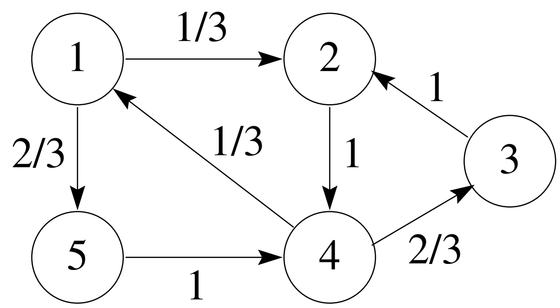

Figure 8.3: Transition graph for a Markov chain.

to the length of the description of the Markov chain. (The description of an instance of the random recoloring process consists of specifying the graph G and the parameter Q.) We remark that the random recoloring process is but one instance of a class of Markov chains referred to as "Glauber dynamics," originating in statistical physics.

An example from computer science: if the state of a memory unit on a computer chip can be described by a bit-string of length k then the number of states of the chip is 2^k. (Transitions can be defined by changing one bit at a time.)

This exponential behavior is typical of combinatorially defined Markov chains.

Because of the exponential growth in the number of states, it is not possible to store the transition matrices and to compute their powers; the size of the matrices becomes prohibitive even for moderate values of the description length of the states. (Think of a $52! \times 52!$ matrix to study card shuffling!)

The evolution of such "combinatorially defined" Markov chains is therefore the subject of intense theoretical study. It is of great importance to find conditions under which the distribution is guaranteed to get **close** to the stationary distribution very fast (in a polynomial number of steps). As noted above, this circumstance is called **rapid mixing**. Note that rapid mixing takes place much faster than it would take to visit each state! (Why is this not a paradox?)

8.1 Problems

Exercise 8.1.1. Let $\mathcal{M}$ be the Markov chain shown in Figure 8.3.

1. Is $\mathcal{M}$ strongly connected?
2. Write down the transition matrix T for $\mathcal{M}$.
3. What is the period of vertex 1?

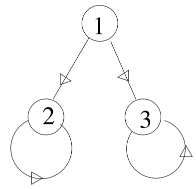

Figure 8.4: The transition graph for a Markov chain.

4. Find a stationary distribution for $\mathcal{M}$. You should describe this distribution as a 1×5 matrix.

5. Prove that $\lim_{t\to\infty} T^t$ does not exist. Prove this directly, do not refer to the Perron-Frobenius theorem.

Exercise 8.1.2. Consider the following digraph: $V = [3]$, $E = \{1 \to 2, 1 \to 3, 2 \to 2, 3 \to 3\}$. Write down the transition matrix of the random walk on this graph, with transition probabilities as shown in Figure 8.1. State two different stationary distributions for this Markov chain.