Chapter 4

Markov Chains and Applications to Population Genetics

4.1 Introduction

The population genetics models of Chapter 3 all impose the infinite population assumption: that is, the frequency $f_G$ of a genotype, $G$, in an offspring population is equal to the probability $p_G$ that an individual is born with that genotype. In any actual finite population, $f_G$ will deviate randomly from $f_G$. But the law of large numbers says that $f_G$ will converge to $p_G$ as the population size tends to infinity, assuming there is enough independence between different matings, and so $p_G$ may be thought of as the frequency of $G$ in an infinite population. The infinite population assumption posits a population so large that replacing any genotype frequency $f_G$ by its infinite population value $p_G$ is an accurate approximation for modeling.

The case of small or moderate sized populations is dramatically different. Chance effects instead dominate its genetic evolution. In particular, genotype frequencies can, and will often, differ from their expected values by a significant amount. For example, consider a population of 10 individuals created by independent random matings, each of which results in an offspring of genotype $G$ with probability 0.2. Then the probability that there are no $G$'s among the 10 offspring is $(1 - 0.2)^{10} = 0.8^{10} = 0.107$, or about one in ten. Thus there is a small, but not insignificant probability that the genotype $G$ disappears completely at a finite time, not because it is less fit, but simply by chance. Likewise, an allele can disappear from a population even in the absence of selection or mutation. This does not occur in the basic infinite population models, even when selection is present.

Therefore, to accurately model finite populations, one must treat the
4. Markov Chains and Genetics

Genotype frequencies \( \{f_G(0), f_G(1), f_G(2), \ldots \} \) in successive generations as a sequence of random variables evolving in time, that is, as a random or stochastic process. A special class of stochastic processes called Markov chains turns out to be especially well-suited to population genetics. This is not a surprise. Markov chains are ubiquitous in applied probability, being employed as models in all major engineering and scientific disciplines, and biology is no exception.

This chapter is a mixture of new mathematical theory and population genetics. Our mathematical goal is an introduction to discrete-time Markov chains. Our applied goal is their application to finite population models, and we will study the Moran and Wright-Fisher models in some detail. Markov chains find many other applications in biology and some of these will appear in subsequent chapters. Also, in chapters 7 and 9, we will extend the basic theory to hidden Markov models and to Markov chains in continuous time and explore their biological applications.

4.2 Markov Chains: Introduction

4.2.1 Discrete-time stochastic process models and path probabilities.

The purpose of this section is to define what a stochastic processes is, to establish basic notation, and to discuss what is meant by a stochastic process model.

The word “stochastic” is roughly synonymous with “random.” The term stochastic process can be used to refer to any family of random variables, but is especially applied when the family describes the random evolution of a system over space or through time. In this chapter we study stochastic processes in the form of sequences \( \{X(0), X(1), X(2), \ldots \} \), which we often abbreviate by \( X = \{X(t); t = 0, 1, \ldots \} \). Usually, \( t \) is thought of as a time index and, whether it actually represents time or not, \( X = \{X(t); t = 0, 1, \ldots \} \) is called a discrete-time stochastic process. Suppose that each \( X(t) \) can only take values in some set \( \mathcal{E} \). Then \( \mathcal{E} \) is called the state space of the stochastic process. We say \( \mathcal{E} \) is said to be discrete if it is finite or countably infinite. We deal mostly with discrete state spaces, usually subsets of the integers, in this chapter.

For a simple example of a discrete-time, discrete state space stochastic process, think of the number \( Y(t) \) of \( A \) alleles at time \( t \), \( t = 0, 1, \ldots \), in an evolving population. If the organism is diploid, if the population remains at constant size \( N \), and if it evolves by random mating, then, for each \( t \), \( Y(t) \) is a random variable taking values in \( \{0, 1, \ldots, 2N\} \). Hence \( Y = \{Y(t); t = 0, 1, \ldots \} \) is a stochastic process with state space \( \{0, 1, \ldots, 2N\} \).
What does it mean to *model* a stochastic processes observed in the real world? By *model*, we mean a set of assumptions that determines, at least in principle, the probabilities of all different possible outcomes of the process. Thus, if the process takes the form $X = \{X(t); t = 0, 1, \ldots\}$, a model must at least prescribe the joint distribution of $(X(0), X(1), \ldots, X(t))$ for every $t \geq 0$. (These joint distributions are also called the finite-dimensional distributions of the process.) For example, when the state space $\mathcal{E}$ is discrete, a model should specify

$$P_t(x_0, x_1, \ldots, x_t) \triangleq \mathbb{P}(X(0) = x_0, X(1) = x_1, \ldots, X(t) = x_t),$$

for all $x_0, x_1, \ldots, x_t \in \mathcal{E}$, for every $t \geq 0$. (4.1)

In fact, this is enough; by limit procedures, these finite dimensional distributions also determine the probabilities of events concerning the entire history of the process.

It is convenient and intuitively appealing to call a (finite) sequence of successive values realized by $X = \{X(t)\}_{t \geq 0}$ a *path*. Thus, $P_t(x_0, x_1, \ldots, x_t)$ in (4.1) is the probability that the process $X$ follows the path $(x_0, x_1, \ldots, x_t)$ up to time $t$, and so we shall often refer to $P_t(x_0, \ldots, x_t)$ as a path probability.

A model may simply consist of explicit formulae for the finite dimensional distributions. This is the case in the following simple and important class of examples.

**Example 4.1. I.i.d. sequences** Suppose we toss a coin repeatedly and let $X(t) = 1$ if toss $t$ comes up heads and $X(t) = 0$ if it comes up tails. The state space is $\mathcal{E} = \{0, 1\}$. Let the model be that the coin is fair, i.e. $\mathbb{P}(X(t) = 1) = \mathbb{P}(X(t) = 0) = 1/2$ for every $t$, and that the tosses are independent. These assumptions determine all joint probabilities: if $(x_1, \ldots, x_n)$ is any sequence of 0’s and 1’s:

$$P_t(x_1, \ldots, x_n) = \mathbb{P}(X(1) = x_1) \cdots \mathbb{P}(X(n) = x_n) = \frac{1}{2^n}.$$  

The first equality is due to the independence of the tosses, the second from the assumption that the coin is fair.

More generally, let $\{p(x); x \in \mathcal{E}\}$ be a probability mass function on a discrete set $\mathcal{E}$. We say that $(X(1), X(2), \ldots)$ are independent and identically distributed (abbreviated *i.i.d.*) if they are independent and if they all have the same probability mass function $p$. In this case, the joint probabilities are given by

$$P_t(x_1, \ldots, x_n) = \mathbb{P}(X(1) = x_1) \cdots \mathbb{P}(X(n) = x_n) = p(x_1)p(x_2)\cdots p(x_n)$$
4. Markov Chains and Genetics

for any path \((x_1, \ldots, x_n)\) through states in \(E\).

Simple i.i.d. models are useful and common—think for example, of random sampling or coin tossing. But in general, the random variables of a stochastic process will exhibit dependencies. Rarely do the physical mechanisms governing the process directly suggest explicit joint distributions in this case. Rather, they usually translate into rules describing how the process evolves from one moment to the next. For a (discrete-time, discrete state space) stochastic process \(\{X(0), X(1), X(2), \ldots\}\), they are encapsulated in the family of conditional probabilities

\[
P_t(x_{t+1} \mid (X(0), \ldots, X(t)) = (x_0, \ldots, x_t)), \quad t \geq 1, \quad x_0, x_1, \ldots, x_{t+1} \in E.
\]

If we interpret \(t\) as present time, the event \(\{X(0), \ldots, X(t) = (x_0, \ldots, x_t)\}\) is the entire past history of the process, these are conditional probabilities for where the process moves next, given its past. For these reason, we shall call them one-step ahead conditional probabilities. Markov chains, introduced in the next section, are processes for which the one-step ahead conditional probability depends on the past history only through the present state.

It turns out that modeling can be reduced (almost) to specifying the one-step ahead conditional probabilities. This is true because of the following simple formula: at any time \(t \geq 0\),

\[
P_{t+1}(x_0, \ldots, x_t, x_{t+1}) = P(X(t+1) = x_{t+1} \mid (X(0), \ldots, X(t)) = (x_0, \ldots, x_t))P_t(x_0, \ldots, x_t) \quad \text{(4.2)}
\]

It is easy to understand and remember this when it is expressed in words: the probability of following the path \((x_0, \ldots, x_t, x_{t+1})\) up to time \(t+1\) is the probability of following \((x_0, \ldots, x_t)\) \emph{times} the conditional probability given this past history that the process moves to \(x_{t+1}\) at time \(t+1\). We shall see for Markov chains, that (4.2) leads by iteration from one-step ahead conditional probabilities to path probabilities.

The proof of (4.2) is just an application of the formula

\[
P(A \cap B) = P(A \mid B)P(B), \quad \text{(4.3)}
\]

which is a direct consequence of the definition, \(P(A \mid B) = P(A \cap B)/P(B)\), of conditional probability. Indeed, (4.2) is just (4.3) with \(B = \{(X(0), \ldots, X(t)) = (x_0, \ldots, x_t)\}\) (the past history) and \(A = \{X(t+1) = x_{t+1}\}\) (the next step), because \(A \cap B = \{(X(0), \ldots, X(t), X(t+1)) = (x_0, \ldots, x_t, x_{t+1})\}\), and so

\[
P_{t+1}(x_0, \ldots, x_t, x_{t+1}) = P(A \cap B).
\]
4. Markov Chains and Genetics

4.2.2 Definition of a Markov chain

The classic example of a Markov chain is random walk on the integers. A drunk walks out of a bar. (I know this sounds like the beginning of a joke, but the punch line will be the definition of a Markov chain!) He begins to walk along the street, taking steps of unit size. To model the effects of inebriation, let us assume that each step is randomly directed and independent of all previous steps. If this is the case, the probability of where the drunk lands on the next step, conditioned on the entire past history of his meanderings, depends only on his current position: because each step is independent of all others, once we know he is at, say $x$, how he got there will not affect where he moves on the next step. In more technical language, his one-step ahead conditional probability depends only on his current position. This is called the Markov property, in honor of the Russian mathematician A. Markov(1856-1922), who first explored this type of conditional dependence. Discrete state space processes with the Markov property are called Markov chains. The following definition states the Markov property mathematically, defines Markov chains, and introduces basic terminology of Markov chain theory.

**Definition 1**

(i) Let $X = \{X(t)\}_{t \geq 0}$ be a stochastic process taking values in the discrete set $\mathcal{E}$. $X$ is called a Markov chain, if for every $t \geq 0$,

$$P(X(t+1) = j \mid X(0) = i_0, \ldots, X(t) = i_t) = P(X(t+1) = j \mid X(t) = i_t) \quad (4.4)$$

for all possible values of $i_0, \ldots, i_t$ and $j$ in $\mathcal{E}$.

(ii) The identity (4.4) is called the Markov property.

(iii) The conditional probability

$$p_{ij}(t) \triangleq P(X(t+1) = j \mid X(t) = i)$$

is called the transition probability from state $i$ to state $j$ at time $t$.

(iv) If all the transition probabilities are independent of time, the Markov chain is said to be time-homogeneous and its transition probabilities are denoted by $p_{ij}$ rather than $p_{ij}(t)$. The collection of transition probabilities,

$$A = [p_{ij}]_{i,j \in \mathcal{E}},$$

considered as a matrix whose rows and columns are indexed by the states in $\mathcal{E}$, is called the transition probability matrix.

(v) The probability distribution of $X(0)$ is denoted by $\rho(0) = \{\rho_i(0) = P(X(0) = i), i \in \mathcal{E}\}$; it is called the initial distribution of the Markov chain.

We shall work with time-homogeneous Markov chains almost exclusively.
Example 4.2. The two-state Markov chain. This is the simplest, non-trivial Markov chain, and is a good example with which to illustrate the concepts of Definition 1. Imagine a system that can be in one of two states, labeled 0 and 1—hence the state space is $\mathcal{E} = \{0, 1\}$—and let $X(t)$ denote the state of the system at time $t$. Suppose that when its in state 0, it chooses either to move to state 1 at the next time step with some probability $\lambda$, independent of $t$, or stay put with probability $1-\lambda$. Mathematically this means that

$$p_{00} = \mathbb{P}(X(t+1)=1 \mid X(t)=0) = \lambda,$$

$$p_{01} = \mathbb{P}(X(t+1)=0 \mid X(t)=0) = 1-\lambda.$$

Likewise suppose that when the process is in state 1, it decides to move to state zero with a probability $\mu$ or stay put at 1 with probability $1-\mu$, so that $p_{11} = \mu$ and $p_{10} = 1-\mu$. Suppose in addition that the random choice of where to move next is independent of the history of the process previous to time $t$. Then, the conditional probabilities of where the process moves next will depend only on its present state, even if the past history is given, and hence the Markov property will be satisfied. Thus $\{X(t); t = 0, 1, \ldots\}$ will be a Markov chain with transition probability matrix,

$$A = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1-\lambda & \lambda \\ \mu & 1-\mu \end{pmatrix}.$$

A popular method for representing Markov chain models is the state transition diagram. The state transition diagram for the two-state process just described is:

Here, the labeled boxes represent the states of the Markov chain, the arrows indicate the possible transitions between states, and each arrow is labeled by the probability of its associated transition. (When drawing state transition diagrams, arrows corresponding to transitions that occur with probability known to be zero are suppressed. For example, had we set $\lambda = 1$ we would not draw the arrow from state 0 back to itself.) When the Markov chain is simple enough, state transition diagrams convey the probabilistic structure of the chain at a glance. For example, one can immediately read off the
state transition matrix of the two-state Markov chain from state transition diagram above.

To imagine a physically real, random experiment that produces the two-state chain, suppose that states 0 and 1 represent two stations. Each station is provided with a loaded coin such that the coin at station 0 comes up heads with probability \( \lambda \), that at station 1 comes up heads with probability \( \mu \), and all flips are independent of each other. Move between stations at unit time intervals according to the following rule. Flip the coin at the station you are currently at and if it’s heads, move to the other station; otherwise stay put. Because the coin tosses are independent one from another, the Markov property holds. When at station 0, you move to station 1 with probability \( \lambda \) and when at 1 you move to 0 with probability \( \mu \). Thus, your transition probabilities are exactly those of the state transition diagram above.

This coin flipping model can be used as the basis for simulating the two state chain. Simply generate two sequences of i.i.d. Bernoulli random variables, one with probability \( \lambda \) for heads, the other with probability \( \mu \), choose a state to start the process in, and use the coin flip sequences to move it from state to state. We will illustrate in the specific case that \( \lambda = 2/3 \) and \( \mu = 1/2 \). Using Maple’s random number generator, we generated a sequence of 14 independent Bernoulli random variables for which the probability of heads is \( \lambda = 2/3 \) and an independent sequence for which it is \( \mu = 1/2 \). The sequence of tosses corresponding to \( \lambda = 2/3 \) was \( H_1H_2T_3H_4H_5T_6H_7H_8H_9H_{10}H_{11}H_{12}H_{13}T_{14} \); we have used \( H \)'s and \( T \)'s to denote the outcomes so as not to confuse them with states 0 and 1, and we have included a subscript to record the sequence in which the tosses occurred. The simulation of the coin tosses corresponding to \( \mu = 1/2 \) yielded the sequence, \( t_1t_2h_3h_4t_5h_6t_7t_8h_9t_{10}t_{11}t_{12}h_{13}h_{14} \), where we have used lower case letters to distinguish this sequence from the previous one. Suppose that \( X(0) = 1 \). Remembering that the first sequence is used to determine transitions from state 0 and the second from state 1, and that heads means “move” and tails “stay put,” these coin tosses sequences cause the process to visit the succession of states shown below; the arrows indicate each successive transition and the letter above each arrow shows which toss from which sequence determined that transition:

\[
0 \xrightarrow{H_1} 1 \xrightarrow{t_1} 1 \xrightarrow{t_2} 1 \xrightarrow{h_3} 0 \xrightarrow{H_2} 1 \xrightarrow{h_4} 0 \xrightarrow{T_3} 0 \xrightarrow{H_3} 1 \xrightarrow{t_5} 1 \xrightarrow{...} \]

Remark.\(^1\) In applications, it is usually clear intuitively whether the Markov property holds or not. When we want to verify it rigorously, we must check (4.4) explicitly. \textit{A priori} this would seem to require separately

\(^1\)Suggestion to students: you may safely skip this remark, Lemma 1, and the paragraph after it.
computing both \( P(X(t+1) = j \mid X(0) = i_0, \ldots, X(t) = i_t) \) and \( P(X(t+1) = j \mid X(t) = i_t) \) and checking they are equal. However, we need only to compute \( P(X(t+1) = j \mid X(0) = i_0, \ldots, X(t) = i_t) \). Think of this as a function of the variables \((i_0, \ldots, i_t)\) and \(j\). If the Markov property holds then this function actually depends only on the values of \(i_t\) and \(j\). The converse is also true.

**Lemma 1** If, as a function of \((i_0, \ldots, i_t)\) and \(j\), \( P(X(t+1) = j \mid X(0) = i_0, \ldots, X(t) = i_t) \) depends only on \(i_t\) and \(j\), then \(\{X(t); t = 0, 1, \ldots\}\) satisfies the Markov property.

This lemma certainly makes good sense on the intuitive level. Its proof, with hints, is left to Exercise 4. of this section. Using Lemma 1, one can prove the Markov property and compute transition probabilities at the same time, as in the following example.

**Example 4.3. Simple random walk on the integers.** This is a formal model of the drunkard’s walk defined at the beginning of this section. Let \(\xi_1, \xi_2, \ldots\) be i.i.d. random variables with \(P(\xi_i = 1) = p\) and \(P(\xi_i = -1) = q = 1 - p\); these represent the drunkard’s random steps. Let \(X(0)\), his initial position, be an integer valued random variable independent of \(\xi_1, \xi_2, \ldots\). Then \(X(t) := X(0) + \sum_{s=1}^{t} \xi_s\) represents his position after \(t\) steps. It is a stochastic process, called simple random walk, whose state space is the set \(\mathcal{E} = \{-2, -1, 0, 1, 2, \ldots\}\) of all integers. If \(X(0) = m\) with probability one, then \(X(t) = m + \sum_{s=1}^{t} \xi_s\) is the simple random walk starting at \(m\).

We will show that the simple random walk is a Markov chain and compute its transition probabilities. The first step uses the simple fact that \(X(t+1) = X(t) + \xi_{t+1}\). When conditioning on \(X(t) = i_t\) we can thus replace \(X(t+1)\) by \(i_t + \xi_{t+1}\):

\[
P(X(t+1) = j \mid X(t) = i_t, X(t-1) = i_{t-1}, \ldots, X(0) = i_0) = P(i_t + \xi_{t+1} = j \mid X(t) = i_t, X(t-1) = i_{t-1}, \ldots, X(0) = i_0)
\]

Next observe that since \(X(r) = X(0) + \sum_{s=1}^{r} \xi_s\) and \(\xi_{t+1}\) is independent of \(X(0), \xi_1, \ldots, \xi + t\), it is independent of \(X(0), X(1), \ldots, X(t)\). Therefore, from the previous calculation,

\[
P(\xi(t+1) = j - i_t \mid X(t) = i_t, X(t-1) = i_{t-1}, \ldots, X(0) = i_0) = P(\xi(t+1) = j - i_t)
\]

\(^{2}\)Suggestion to students: you may safely skip the details here, and only focus on the formula for the transition probabilities.
This depends only on \( i_t \) and \( j \) and so, by Lemma 1, a simple random walk is a Markov chain whose transition probabilities are (replacing \( i_t \) by \( i \))

\[
p_{ij} = \begin{cases} 
p, & \text{if } j = i + 1; 
q, & \text{if } j = i - 1; 
0, & \text{otherwise.} 
\end{cases}
\]

The state transition diagram for simple random walk is

\[
\cdots \quad p \quad -1 \quad p \quad 0 \quad p \quad 1 \quad p \quad 2 \quad p \quad \cdots
\]

(The diagram extends by repetition to all integers.)

Since there are an infinite number of states, the state transition matrix is infinite dimensional. It is not particularly useful to write down.

\[\diamond\]

**Example 4.4.** Let \( Z_1, Z_2, \ldots \) be any sequence of i.i.d. integer valued random variables. Let \( X(0) \) be integer valued and independent of \( Z_1, Z_2, \ldots \). By the same reasoning as for simple random walk, \( X(t) = \sum_t^t Z_i \) defines a Markov chain with transition probabilities, \( p_{ij} = \mathbb{P}(Z_1 = i - j) \).

### 4.2.3 Path probabilities

Because of the Markov property, the transition probabilities and the initial distribution of a Markov chain are enough to calculate all its path probabilities. This is the content of the next result.

**Theorem 1** Let \( \{X_t\}_{t \geq 1} \) be Markov chain with transition probabilities \( \{p_{ij} \}; i,j \in \mathcal{E} \) and initial distribution \( \rho(0) = \{\rho_i(0); i \in \mathcal{E} \} \). Then for any \( t \geq 0 \) and any path \( \{i_0, \ldots, i_t\} \), the path probability is

\[
p_t(i_0, \ldots, i_t) = \rho_{i_0}(0)p_{i_0i_1}p_{i_1i_2}p_{i_2i_3} \cdots p_{i_{t-1}i_t}. \tag{4.5}
\]

This formula is easy to remember. Think of a path \( (i_0, i_1, \ldots, i_t) \) as a starting value \( i_0 \) and a series of transitions \( i_0 \to i_1 \to i_2 \to \cdots \to i_t \); the probability that a Markov chain follows this path is the probability that it starts at \( i_0 \) times the product of the probabilities all the transitions.

The proof of (4.5) is by recursive application of the identity (4.2). Since \( X \) is a time-homogeneous chain, the one-step ahead conditional probability on the right-hand side of (4.2) equals \( p_{i_t, i_{t+1}} \). Hence,

\[
p_{t+1}(i_0, \ldots, i_t, i_{t+1}) = p_t(i_0, \ldots, i_t)p_{i_t, i_{t+1}} \tag{4.6}
\]
For $t = 0$, this says
\[ p_1(i_0, i_1) = p_0(i_0)p_{i_0i_1} = \rho_{i_0}(0)p_{i_0i_1}, \]
which is (4.5) for $t = 1$. Now use this result in (4.6) when $t = 2$; the result is
\[ p_2(i_0, i_1, i_2) = p_1(i_0, i_1)p_{i_1i_2} = \rho_{i_0}(0)p_{i_0i_1}p_{i_1i_2}. \]
This is (4.5) for $t = 2$. Repeated iteration of this argument, or equivalently, mathematical induction, proves (4.5) for all $t$.

Example 4.4. Consider the two state Markov chain defined in Example 4.3. Suppose that $X(0) = 0$ with probability $1/4$ and otherwise $X(0) = 1$. Find the probability that $(X(0), X(1), \ldots, X(5)) = (0, 1, 1, 0, 0, 1)$. According to (4.6) this probability is $(1/4) \cdot \lambda \cdot (1 - \mu) \cdot \mu \cdot (1 - \lambda) \cdot \lambda$. 

4.2.4 Stochastic matrices and Markov chain models.

Let $A = [a_{ij}]_{i,j \in \mathcal{E}}$ be the transition probability matrix of a Markov chain. For any state $i$, $\mathbb{P}(X(t+1) \in \mathcal{E} \mid X(t) = i) = 1$, since by definition, $\mathcal{E}$ contains the set of all possible states to which $X$ can move at time $t+1$. But
\[ \mathbb{P}(X(t+1) \in \mathcal{E} \mid X(t) = i) = \sum_{j \in \mathcal{E}} \mathbb{P}(X(t+1) = j \mid X(t) = i) = \sum_{j \in \mathcal{E}} a_{ij}. \]
Therefore, for any $i$
\[ \sum_{j \in \mathcal{E}} a_{ij} = 1, \quad (4.7) \]
That is, every row of $A$ sums to one. Any square matrix whose entries are non-negative numbers and whose rows sum to one is called a stochastic matrix. Thus, the transition probability matrix of any Markov chain is a stochastic matrix. Conversely, any stochastic matrix defines the transition probability matrix of a Markov chain.

By Markov chain model for a randomly evolving system, we shall mean a state space and a stochastic matrix defining transitions between its states. In Section 4.2.1 we said that a stochastic process model should completely determine all path probabilities. But we know from Theorem 1 that calculating path probabilities of a Markov chain requires both the transition probabilities and an initial distribution. Hence a transition probability matrix alone rather defines a family of models, one for each possible initial distribution. But usually the initial distribution is incidental and varies from situation to situation. The transition probabilities incorporate the more important information, the rules governing how the process evolves from any starting value. Thus, we shall consider a Markov chain model to be specified once the transition probability matrix is given.
4.2.5 The Moran and Wright-Fisher models

In this section we define two basic models that are studied in population genetics, the Moran model and the Wright-Fisher model. Actually, in terms of genetics, the Moran model is really what we call a “toy” model. The assumptions underlying it are overly simplified and do not apply to sexual reproduction. However, it is has the advantage of being very simple to define and serves to introduce the concept of a birth-and-death chain. The Wright-Fisher model is more important for population genetics. It is the simplest stochastic version of what in Chapter 3 we called the basic model, defined there by assumptions (A.1)-(A.5).

The Moran model. This is named in honor of the Australian applied probabilist P.A.P. Moran (1917-1988). It is a finite population model with overlapping generations, but without selection, mutation or migration. However, it assumes that reproduction is asexual, by simple duplication. The model treats a population divided into two types, which shall be denoted by the letters $B$ and $b$. One might think of these letters as denoting a genotype, but that is not necessary. The population changes over time according to the following mechanism. At each time $t$, two individuals are selected from the population by independent random sampling with replacement. The first individual gives birth to a copy of itself, which joins the population. The second individual is removed from the population (it dies). The random samplings at different times are all independent. Since one individual is added and one removed at each stage, the population remains constant in size. Also, since the the choice of who reproduces and who dies is made with replacement, the individual chosen to reproduce may be the same as the one chosen to die, and the end result is that it is just replaced with a copy of itself. There is no selection in the model because each individual has the same chance, irrespective of genotype, to be chosen for both reproduction and death.

Let $X(t)$ denote the number of individuals of type $B$ at stage $t$, and let $N$ denote the total size of the population. We will argue that $X = \{X(t)\}_{t \geq 0}$ is a Markov chain and calculate its transition probabilities.

Suppose $X(t) = i$ where $0 < i < N$, and perform the random selection, copy and replacement experiment described above. There are three possibilities. The first is that a type $B$ is chosen to reproduce and a type $b$ to die. In this case the number of $B$’s is increased by one and hence $X(t+1) = i + 1$. The second possibility is that a type $b$ is chosen to reproduce and a type $B$ to die, in which case $X(t+1) = i - 1$. Finally, it can happen that both individuals selected are the same type; then either a type $b$ replaces a type $b$ or a type $B$ replaces a type $B$ and $X(t) = i$.

Consider the one-step ahead conditional probability of the first case,
supposing that we know the entire past history: \(X(0) = i_0, \ldots, X(t-1) = i_{t-1}\) and \(X(t) = i\). Then \(X(t+1) = i + 1\) only if a \(B\) is selected to reproduce and a \(b\) to die. But these selections are by independent random sampling from a population of size \(N\) and \(X(t) = i\) type \(B\)'s and \(N - i\) type \(b\)'s. The probability of selecting a \(B\) is thus \(i/N\) and the probability of selecting a \(b\) is \((N-i)/N\). Hence

\[
\mathbb{P}(X(t+1) = i+1 \mid (X(0), \ldots, X(t)) = (i_0, \ldots, i_{t-1}, i)) = \frac{i(N-i)}{N^2}.
\]

Notice that this depends only on \(X(t) = i\), because the only factor relevant to computing the probabilities of selecting \(B\) or \(b\) is the number \(i\) of type \(B\)'s at time \(t\). It follows immediately from Lemma 1 that

\[
\mathbb{P}(X(t+1) = i+1 \mid X(t) = i) = \mathbb{P}(X(t+1) = i+1 \mid (X(0), \ldots, X(t)) = (i_0, \ldots, i_{t-1}, i)) = \frac{i(N-i)}{N^2}.
\]

The second and third case are argued similarly.

\[
\mathbb{P}(X(t+1) = i-1 \mid (X(0), \ldots, X(t)) = (i_0, \ldots, i_{t-1}, i)) = \frac{i(N-i)}{N^2},
\]

because to get \(X(t+1) = i-1\) requires selecting \(b\) (with probability \((N-i)/N\)) to reproduce and \((B\) with probability \(i/N\)) to die. Again by Lemma 1,

\[
\mathbb{P}(X(t+1) = i-1 \mid X(t) = i) = \mathbb{P}(X(t+1) = i+1 \mid (X(0), \ldots, X(t)) = (i_0, \ldots, i_{t-1}, i)) = \frac{i(N-i)}{N^2}.
\]

By the same type of reasoning,

\[
\mathbb{P}(X(t+1) = i \mid X(t) = i) = \mathbb{P}(X(t+1) = i+1 \mid (X(0), \ldots, X(t)) = (i_0, \ldots, i_{t-1}, i)) = \frac{i^2}{N^2} + \frac{(N-i)^2}{N^2}.
\]

These calculations verify that the Moran process is a Markov chain with transition probabilities,

\[
p_{ij} = \begin{cases} 
\frac{i(N-i)}{N^2}, & \text{if } 0 < i \leq N \text{ and } j = i - 1; \\
\frac{i^2}{N^2} + \frac{(N-i)^2}{N^2}, & \text{if } 0 < i < N \text{ and } j = i; \\
\frac{i(N-i)}{N^2}, & \text{if } 0 \leq i < N \text{ and } j = i + 1; \\
0, & \text{if } |j - i| > 1. 
\end{cases}
\]

Observe that the Moran Markov chain can move at most by one unit in every time step. Markov chains whose state space is the integers, or a
subset thereof, and that move by at most unit steps, are called *birth and death chains*. Thus the transition probabilities of a birth and death chain take the general form

\[ p_{i,i−1} = \mu_i, \quad p_{ii} = 1 − (\lambda_i + \mu_i), \quad p_{i,i+1} = \lambda_i, \quad P_{ij} = 0 \text{ if } |j − i| > 1. \]

where \( \lambda_i \) is the probability the chain increases by 1 from \( i \) (birth), and \( \mu_i \) it decreases by 1 from \( i \) (death).

Simple random walk on the integers, as defined in Example 4.2, is a birth and death chain in which \( \lambda_i = p \) and \( \mu_i = 1−p \) for all \( i \). Conversely, a birth and death chain may be thought of as a random walk on the integers in which the probability of moving left, right, or staying put depends on the current state.

**The Wright-Fisher Model**

This model is named after R.A. Fisher (1890-1962), a British statistician, and Sewall Wright (1889-1988), an American geneticist. They both played major roles in creating the field of theoretical population genetics.

The Wright-Fisher model is the simplest stochastic version of the basic population model introduced in Section 3.3.1. There we introduced five assumptions: random mating, monecious species, non-overlapping generations, no selection or mutation, and an infinite population. The Wright-Fisher model is the result of keeping all these assumptions except an infinite population assumption, which is replaced by the following: the population maintains a constant size \( N \) from generation to generation and each generation produces the next by \( N \) independent random matings. (4.8)

It is also assumed that the random matings of any one generation are independent of the random matings of other generations.

Consider a locus \( \ell \) and an allele \( A \) that can occur at \( \ell \). (There can be one, two, or many other alleles at \( \ell \); how many there are is immaterial to the model.) Consider a population of size \( N \) which evolves according to the Wright-Fisher assumptions and let \( X(t) \) denote the number of alleles \( A \) in the allele pool for locus \( \ell \). We are really interested in the frequency of \( A \), \( f_A(t) = X(t)/2N \), but the notation is simpler if we study \( X \) instead. \( \{X(t)\}_{t \geq 0} \) is the Wright-Fisher process. We will show it is a Markov chain and calculate its transition probabilities.

Suppose that \( X(t) = i \) and that the past values \( X(0) = i_0, \ldots, X(t−1) = i_{t−1} \) are also given. Random mating, as defined in Section 3.2.4, means producing a new individual by choosing the parents by independent random sampling and choosing a gamete from each parents’ gamete pool by random
sampling. Hence, as we saw in Lemma 3.2.4, the allele each parent donates is really a random sample from the allele pool of the parent population, and the genotype of offspring is created two independent random samples, one for each parent. In the Wright-Fisher model, generation $t + 1$ is created by $N$ independent random matings. Hence its allele pool is, in effect, created by $2N$ independent random samples of the allele pool of generation $t$. If $X(t) = i$, the frequency of $A$ in generation $t$ is $X(t) = i/2N$, and hence the probability of drawing $A$ in a single random sample is $i/2N$, no matter what the past history of $X$ is. If we think of drawing $A$ as a success, the number $X(t+1)$ of $A$ in the allele pool of generation $t+1$ is therefore the number of successes in $2N$ independent trials. This means that, conditional on $X(t) = i$ and $X(0) = i_0, \ldots, X(t-1) = i_{t-1}, X(t+1)$ is a binomial random variable with parameters $n = 2N$ and $p = i/2N$. Since this conditional distribution depends only on the value $X(t) = i$, and not on any values of $X(s)$ for $s = 0, \ldots, t-1$, $X$ has the Markov property and its transition probabilities are given by the binomial formulae (see Section 2.2.2)

$$p_{ij} = \binom{2N}{j} \left( \frac{i}{2N} \right)^j \left( 1 - \frac{i}{2N} \right)^{2N-j}, \quad 0 \leq j \leq 2N. \quad (4.9)$$

The Wright-Fisher model is more complicated than the Moran model because the probability of transition from any state other than 0 or $2N$ to any other state is positive.

### 4.3 Analysis of Markov chains and applications

This section is devoted to basic principles of Markov chain theory with a focus on finite state space chains. More theory is presented than is strictly needed to discuss the Moran and Wright-Fisher models, but this theory will be useful later.

#### 4.3.1 Multi-step transition probabilities and evolution of distributions

When the infinite population assumption is dropped, genotype and allele frequencies are random and it no longer makes sense to try to predict them exactly for all future times, as we did with infinite population models. Instead, we want to use models to calculate the probability distributions of gene frequencies in future generations and to predict how these probability distributions evolve. For Markov chain models, this entails calculating the probability distribution of $X(t)$ at any time $t$ from knowledge of the initial distribution and the state transition matrix. This section addresses this
Let $\{X(t)\}$ be a Markov chain with a discrete state space $\mathcal{E}$, with transition probability matrix $A$, and with initial distribution $\rho(0) = \{\rho(j) = \mathbb{P}(X(0) = j); j \in \mathcal{E}\}$. Our goal is to compute the probability mass function of $X(t)$ at future times $t$; denote this mass function by $\rho_t = \{\rho_t(j) = \mathbb{P}(X(t) = j); j \in \mathcal{E}\}$. We will discover that this is simple to do, at least in theory.

Our approach will be to derive a difference equation for $\rho_t$ by using the rule of total probabilities (see (2.12) in Chapter 2). The events $\{X(t-1) = i\}$, $i \in \mathcal{E}$, are disjoint and cover all possibilities. Therefore, by (2.12),

$$
\rho_t(j) = \mathbb{P}(X(t) = j) = \sum_{i \in \mathcal{E}} \mathbb{P}(X(t) = j | X(t-1) = i) \mathbb{P}(X(t-1) = i) = \sum_{i \in \mathcal{E}} \rho_{t-1}(i)p_{ij}.
$$

This equation tells us how to compute $\rho_t$ in terms of $\rho_{t-1}$. It is more complicated-looking than the ordinary difference equation because it must be solved at each $t$ for all the values $\rho_t(j)$, $j \in \mathcal{E}$. However, there is an interpretation of the equation that makes its solution almost trivial. The trick is to use vector and matrix algebra. It is easiest to understand this in the case that $\mathcal{E}$ is finite. It does not really matter how these states are labeled, so, for convenience, because we usually deal with this case, suppose

$$
\mathcal{E} = \{0, 1, 2, \ldots, N\}.
$$

Then the state transition matrix is

$$
A = \begin{pmatrix}
p_{00} & p_{01} & \cdots & p_{0N} 
p_{10} & p_{11} & \cdots & p_{1N} 
\vdots & \vdots & \ddots & \vdots 
p_{N0} & p_{N1} & \cdots & p_{NN}
\end{pmatrix}.
$$

The crucial idea is to interpret the probability mass function at each time $t$ as the row vector,

$$
\rho_t = (\rho_t(0), \rho_t(1), \cdots, \rho_t(N)).
$$

Now consider multiplying the row vector $\rho_{t-1}$ times $A$. The rules of matrix multiplication say that the result is a row vector whose $j^{th}$ term is the product of $\rho_{t-1}$ with the $j^{th}$ column of $A$:

$$
(\rho_{t-1} \cdot A)_j = \sum_{i=0}^{N} \rho_{t-1}(i)p_{ij}.
$$
Compare this to (4.10). The right-hand side is precisely $\rho_t(j)$, and hence $\rho_t \cdot A$ is equals $\rho_t$. Clearly this reasoning works no matter how we label the state of $E$, so long as the order of states in $\rho$ matches the order in which they label the rows of $A$. Also the reasoning works for infinite state spaces and infinite row vectors as long as we interpret $(\rho \cdot A)_j$ as the infinite sum $\sum_{i \in E} \rho(i)p_{ij}$. We enshrine this important interpretation of (4.10) and some useful corollaries in the following theorem.

**Theorem 2** If $\{X(t)\}$ is a time-homogeneous Markov chain with state transition matrix $A$, then

$$\rho_t = \rho_{t-1} \cdot A \quad \text{for } t \geq 1 \tag{4.11}$$

As a consequence,

$$\rho_t = \rho_0 \cdot A^t \quad \text{for } t \geq 1. \tag{4.12}$$

Finally, for any states $i$ and $j$, the $s$-step ahead transition probability of the chain is given by

$$\mathbb{P}(X(t+s) = j \mid X(t) = i) = [A^s]_{ij}. \tag{4.13}$$

In this theorem (4.11) is the matrix algebra version of (4.10). It is called the forward equation for $\{\rho_t\}$, because it is derived by conditioning on the value of the chain at the time $t-1$ immediately preceding $t$. It is the simplest possible linear difference equation, and its solution, $\rho_t = \rho_0 \cdot A$, presented in (4.12), is derived simply by iteration:

$$\rho_t = \rho_{t-1} \cdot A = \rho_{t-2} \cdot A \cdot A = \rho_{t-3} \cdot A \cdot A \cdot A = \cdots = \rho_0 \cdot A^t.$$  

To prove the last claim of the theorem, consider the special case in which $\mathbb{P}(X(0) = i) = 1$. If this is true

$$\mathbb{P}(X(s) = j) = \mathbb{P}(X(t) = j, X(0) = i) = \mathbb{P}(X(s) = j \mid X(0) = i)\mathbb{P}(X(0) = i)$$

$$= \mathbb{P}(X(s) = j \mid X(0) = i).$$

On the other hand, $\rho_0 = (0, \ldots, 0, 1, \ldots, 0)$, where 1 is in the spot corresponding to state $i$. Thus, by (4.12) and matrix multiplication

$$\mathbb{P}(X(s) = j) = [\rho_0 \cdot A^s]_j = [A^s]_{ij}.$$  

Comparing the two results gives (4.13) for $t = 0$. Because the chain is time-homogeneous, the $s$-step ahead transition probabilities do not depend on $t$ and so (4.13) is true for all $t \geq 0$.

It will be convenient to use the notation $p_{ij}^s$ for the $s$-step ahead transition probability in (4.13). Thus, from (4.13) $p_{ij}^s$ is the entry in row $i$ and column $j$ of $A^s$.  

$$\diamond$$
Example 4.5. Consider the Moran model with \( N = 4 \) and \( X(0) = 1 \) with probability one. Find the probability mass functions of \( X(1) \) and \( X(2) \).

The state transition for the Moran model with \( N = 4 \) and states listed in the order \( \{0, 1, 2, 3, 4\} \) is

\[
A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
3 & 16 & \frac{3}{16} & \frac{3}{16} & 0 \\
0 & \frac{1}{4} & \frac{7}{16} & \frac{1}{16} & 0 \\
0 & 0 & \frac{3}{16} & \frac{10}{16} & \frac{3}{16} \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

Thus

\[
A^2 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
39 & 128 & 56 & 128 & 0 \\
6 & 128 & 36 & 128 & 6 \\
0 & 6 & 128 & 27 & 128 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

Since \( \rho_0 = (0, 1, 0, 0, 0) \), the product \( \rho \cdot B \), where \( B \) is a \( 5 \times 5 \) matrix, picks out the second row of \( B \). Therefore the probability distributions of \( X(1) \) and \( X(2) \) are

\[
\rho_1 = \rho_0 \cdot A = \left( \frac{3}{16}, \frac{10}{16}, \frac{3}{16}, 0, 0 \right)
\]

\[
\rho_1 = \rho_0 \cdot A^2 = \left( \frac{39}{128}, \frac{56}{128}, \frac{27}{128}, \frac{6}{128}, 0 \right)
\]

Example 4.6. Consider a population of \( N = 10 \) individuals at time \( t = 0 \): \( AA, AA, AA, Aa, Aa, Aa, Aa, Aa, Aa, Aa \). Assume it evolves according to the Wright-Fisher model, and let \( X(t) \) denote the number of alleles of type \( A \) in generation \( t \). Find the probability mass function of \( X(2) \).

Notice that \( X(0) = 13 \), \( \mathbb{P}(X(0) = 13) = 1 \). In this case, rather than write out the transition probability matrix, which is large and ugly, we will just work directly with the coordinate by coordinate definition of the matrix product.

\[
\mathbb{P}(X(2) = j) = \sum_{j=0}^{20} \rho_0(i)[A^2]_{i,k} = [A^2]_{13,k} = \sum_{k=0}^{20} p_{13,k}p_{k,j}
\]

\[
= \sum_{k=0}^{20} \binom{20}{k} \binom{13}{20}^k \binom{7}{20}^{20-k} \binom{20}{j} \binom{k}{20} \binom{20-k}{20}^{20-j}.
\]
The fact that we have a simple and explicit formula for computing $\rho_t$ for every $t$ is gratifying. But we see from the last example that it does not necessarily yield tractable expressions. Nor does matrix multiplication alone reveal what we really would like to know: how does the probability mass function $\rho_t$ behave in the limit as $t \to \infty$? Markov chain theory provides a very complete and elegant solution to this question. Under general conditions it is possible to prove that

$$A^\infty \overset{\Delta}{=} [p^\infty_{ij}] \overset{\Delta}{=} \lim_{t \to \infty} A^t$$

in the sense that $p^\infty_{ij} = \lim_{k \to \infty} P(X(t+k) \mid X(t) = i)$ exists for any pair of states $i$ and $j$. When this is the case

$$\lim_{t \to \infty} \rho_t = \lim_{t \to \infty} \rho_0 \cdot A^t = \rho_0 \cdot A^\infty.$$

We take take this issue up, with applications to the Moran and Wright-Fisher models, in the following sections, after a brief digression on computing expected values in the Markov chain setting.

### 4.3.2 Expectation and Markov chains

For this section, the reader may wish to review conditional expectations—see Chapter 2, Section 2.3.7.

Let $\{X(t)\}$ be a Markov chain with state space $E$ and let $g$ be a function whose domain is $E$. The conditional expectation of $g(X(t+1))$ given $\{X(t) = i\}$ is, by definition of conditional expectation,

$$E[ g(X(t+1)) \mid X(t) = i] = \sum_{j \in E} g(j) P(X(t+1) = j \mid X(t) = i) = \sum_{j \in E} p_{ij} g(j)$$

(4.15)

**Example 4.7.** Consider the Wright-Fisher model with state space $E = \{0, 1, \ldots, 2N\}$. For the identity function $h(x) = x$, we have

$$E[X(t+1) \mid X(t) = i] = \sum_{j=0}^{2N} \binom{2N}{j} \left( \frac{i}{2N} \right)^j \left( 1 - \frac{i}{2N} \right)^{2N-j} \cdot j$$

This looks complicated until we remember that, given $\{X(t) = i\}$, $X(t+1)$ is a binomial random variable with $n = 2N$ and $p = i/2N$. Hence its conditional expectation given $\{X(t) = i\}$ is $np = 2N(i/2N) = i$. Therefore

$$E[X(t+1) \mid X(t) = i] = i, \quad \text{for each state } i, \ 0 \leq i \leq 2N. \quad \diamond \quad (4.16)$$

**Exercise 4.1.** Show that (4.16) is true for the Moran model as well.
Equation (4.15) has a matrix algebra interpretation similar to equation (4.10) for the probability mass distribution vector \( \rho_t \). This requires that we interpret a function \( g \) on \( \mathcal{E} \) as a column vector indexed by the states in \( \mathcal{E} \). For emphasis, we shall denote write \( g \) to denote \( g \) interpreted in this way. For example, the identity function \( h(x) = x \) used in Example 4.6 should be considered the same thing as the vector,

\[
h = \begin{pmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ 2N \end{pmatrix}.
\]

With this notation, we see that the right-hand side of (4.15) is the same thing as the \( i \)th component of \( A \cdot g \): thus,

\[
E [g(X(t+1)) \mid X(t) = 1] = [A \cdot g]_i, \quad (4.17)
\]

**Example 4.7, continued.** The calculation in the first part of Example 4.6 shows that for the Wright-Fisher model, and \( h(x) = x \), \( i = h(i) = E[h(X(t+1)) \mid X(t) = i] = [A \cdot h]_i \), for each state \( i \), where \( A \) is the Wright-Fisher transition probability matrix. This is the same as saying,

\[
h = A \cdot h \quad (4.18)
\]

This identity can be used to show the expected value of any Wright-Fisher chain is constant in \( t \). That is if \( X(t) \) is a Wright-Fisher chain, then, whatever its initial distribution,

\[
E[X(t)] = E[X(0)] \quad \text{for all } t \geq 0. \quad (4.19)
\]

To show why this is true, we apply the formula of Theorem 11 in Chapter 2 to compute \( E[X(t+1)] \) in terms of conditional expectations with respect to \( X(t) \). This formula implies,

\[
E[X(t+1)] = \sum_{i=0}^{2N} E[X(t+1) \mid X(t) = i] P(X(t) = i).
\]

But from (4.16), or equivalently (4.18), the conditional expectation in the sum is simply \( i \). Thus

\[
E[X(t+1)] = \sum_{i=0}^{2N} i P(X(t) = i) = E[X(t)].
\]
This is true for any $t \geq 1$, and, as a result, $E[X(t)] = E[X(t-1)] = E[X(t-2)] = \cdots = E[X(0)]$.

The fact that the expectation of a Wright-Fisher chain is constant is the stochastic counterpart to the fact that allele frequencies are constant in the infinite population model under assumptions (A.1)—(A.5) of Chapter 3. ♦

Notice that, using the result of Exercise 4.1, the same arguments show that expectation of $X(t)$ is constant for the Moran model also.

4.3.3 Behavior of chains with absorbing states

In this section we show how to compute $A^\infty = \lim_{t \to \infty} A^t$ in a special case that applies to the Moran and Wright-Fisher models.

Consider a Markov chain with state space $E$. If $i$ and $j$ are states, we say that $i \rightarrow j$ if, starting from $i$, the probability that the chain eventually visits $j$ is positive. If $i, i_1, i_2, \ldots, i_n, j$ is a sequence of states starting from $i$ and ending at $j$, we know from Theorem 1 that

$$
\mathbb{P}\left(X(t+1)=i_1, X(t+2)=i_1, \ldots, X(t+n+1)=j \mid X(t)=i\right) = p_{i_1}p_{i_1i_2}\cdots p_{i_nj}.
$$

Therefore, $i \rightarrow j$ if and only if there is a path from $i$ to $j$ along which the transition probability between each pair of successive states is positive. As a convention, $i \rightarrow i$ for all states $i$. To illustrate, consider the Moran model for a population of size $N$. It is clear that if $0 < i < N$ and $j$ is any state ($0 \leq j \leq N$), then $i \rightarrow j$, because the Moran model moves up and down by at most unit steps and for each $i$, $0 < i < N$, $p_{i,i+1} = p_{i,i-1} > 0$. However $p_{00} = p_{NN} = 1$ and hence once the chain enters these states it can reach no others. For this reason we say that $0$ and $N$ are absorbing. In general, a state $s$ in a Markov chain is absorbing if $p_{ss} = 1$.

In the Wright-Fisher model for a population of size $N$, if $0 < i < 2N$, $p_{ij} > 0$ for all $j$ and hence again $i \rightarrow j$ for all states $j$. Again, also, states $0$ and $2N$ are absorbing.

If $i \rightarrow j$ and $j \rightarrow i$, we write $i \leftrightarrow j$ and we say that $i$ and $j$ communicate. In the Wright-Fisher model model $i \leftrightarrow j$ as long as $0 < i,j < 2N$, that is, as long as neither $i$ nor $j$ is one of the absorbing states. The same is also true for the Moran model.

Both the Moran and Wright-Fisher chains are finite state space chains with the following property: for every state $i$ there is at least one absorbing state $s$ such that $i \rightarrow s$. It is interesting to know whether or not such a chain must eventually enter an absorbing state with probability one. For the Moran and Wright-Fisher models this is asking whether or not one of the alleles must eventually die out, just as a result of random fluctuation. The next result says that absorption must occur and derives from this fact the limit of $A^t$ as $t \to \infty$. 
Theorem 3 Let $A$ be the state transition function for a Markov chain \{X(t)\} with the following property: for every state $i$ there is an absorbing state $s$ such that $i \rightarrow s$. Then for each state $i$

$$
\mathbb{P}\left(\{X(t) \text{ eventually enters an absorbing state}\} \mid X(0) = i\right) = 1 \quad (4.20)
$$

In particular, $X(t)$ enters an absorbing state eventually no matter what its initial distribution is.

As a consequence of (4.13), if $i$ is any state and $j$ is any non-absorbing state,

$$
p_{ij}^\infty = \lim_{t \rightarrow \infty}[A^t]_{ij} = 0 \quad (4.21)
$$

If $i$ is any state,

$$
p_{is}^\infty = \lim_{t \rightarrow \infty}[A^t]_{is} \text{ exists for every absorbing } s;
$$

$$
p_{is}^\infty = \mathbb{P}\left(\{X(t) \text{ eventually enters state } s \mid X(0) = i\}\right) \text{ if } s \text{ is absorbing};
$$

$$
1 = \sum_{s \text{ is absorbing}} p_{is}^\infty.
$$

This theorem gives a complete description of $A^\infty$ if we can compute the absorption probabilities $p_{is}^\infty$ for any state $i$ and absorbing state $s$. To understand intuitively why it is true, suppose $X(t) = i$. Then in order that $X(t+1) = i, X(t+2) = i, \ldots, X(t+k) = i$, the chain must make $k$ successive transitions from $i$ to $i$, and the probability this happens is $p_{ii}^k$. If $i$ is non-absorbing, $p_{ii} < 1$ and so $p_{ii}^k \rightarrow 0$ as $k \rightarrow \infty$. Therefore the probability that the chain stays in state $i$ forever is 0, which means as long as $X(t)$ does not enter an absorbing state it moves around among the finite number of non-absorbing states. As it does so, it will have repeated opportunities to enter, by chance, an absorbing state, and eventually it must do so.

After discussing its application to genetics models, we will present a detailed mathematical demonstration of Theorem 3, because it involves many important ideas and methods for dealing with Markov chains.

It is an immediate consequence of Theorem 3 that both the Moran and Wright-Fisher models must eventually enter an absorbing state. Thus in both models, one of the alleles takes over (“fixes”), as a consequence of random genetic drift. This contrasts sharply with the infinite population models, in which allele frequency is constant, and, hence, if two alleles are present in the population initially, they remain in the population for all time.

Using Theorem 3 more can be said. It is possible to compute the probabilities of absorption into the different absorbing states.
Example 4.8. The two absorbing states of the Wright-Fisher model are 0 and \(2N\). We will show that for any other state,
\[
p_{i,0}^\infty = \frac{2N - i}{2N}, \quad p_{i,2N}^\infty = \frac{i}{2N},
\]
(4.22)
Since \(p_{ij}^\infty = 0\) if \(j \neq 0\) or \(j \neq 2N\),
\[
A^\infty = \lim_{t \to \infty} A^t = \begin{pmatrix}
1 & 0 & \cdots & \cdots & 0 & 0 \\
\frac{2N-1}{2N} & 0 & \cdots & \cdots & 0 & \frac{1}{2N} \\
\frac{2N-2}{2N} & 0 & \cdots & \cdots & 0 & \frac{2}{2N} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{1}{2N} & 0 & \cdots & \cdots & 0 & \frac{2N-1}{2N} \\
0 & 0 & \cdots & \cdots & 0 & 1
\end{pmatrix}.
\]
Let \(U\) be the event that the chain is absorbed by state \(2N\) and define
\[
g(i) = \mathbb{P}(U \mid X(0) = i).
\]
This defines a function \(g\) on the states of the chain; as in the previous section, \(g\) represents the column vector, indexed by the states, defined by \(g\). We will derive a linear equation with “boundary condition” for \(g\).

By the Markov property,
\[
\mathbb{P}(U \mid X(1) = j, X(0) = i) = \mathbb{P}(U \mid X(1) = j) = g(j)
\]
because we can tell if \(U\) occurs by looking at \(X(t)\) for times \(t \geq 1\) and once the \(X(1) = j\) is known, the value of \(X(0)\) does not affect the conditional probabilities of future events. The rule of total probabilities (see (2.12)), generalized to conditional probabilities, implies that for every \(i\),
\[
g(i) = \sum_{j=0}^{2N} \mathbb{P}(U \mid X(1) = j, X(0) = i) \mathbb{P}(X(1) = j \mid X(0) = i)
\]
(4.23)
(A proof is given below of this identity.) Thus,
\[
g(i) = \sum_{j=0}^{2N} p_{ij} g(j) \quad \text{for all } i, \text{ or, equivalently } \quad \mathbf{g} = A \cdot \mathbf{g}
\]
(4.24)
On the other hand, it is clear from the definition of \(g\) that
\[
g(0) = 0 \quad \text{and} \quad g(2N) = 1
\]
(4.25)
Therefore we will have a candidate for the function $g$, if we solve (4.24)–(4.25). But this is easily done. We know from Example 4.7 that the identity function $h(i) = i$, for all $i$ solves $h = A \cdot h$, and $h(0) = 0$ and $h(2N) = 2N$. Therefore $\tilde{g}(i) = h(i)/2N = i/2N$ is a solution satisfying $\tilde{g}(0) = 0$ and $\tilde{g}(2N) = 1$. To finish we need to show that in fact $\tilde{g} = g$. To do this, observe that by iterating the equation $\tilde{g} = A \cdot \tilde{g}$, one derives that $\tilde{g} = A^{t} \cdot \tilde{g}$ by iterating again, $\tilde{g} = A^{3} \cdot \tilde{g}$, and, continuing in the manner, $\tilde{g} = A^{t} \cdot \tilde{g}$ for all $t \geq 1$. Equivalently, for each $i$,

$$
\tilde{g}(i) = \sum_{j=0}^{2N} [A^{t}]_{ij} \tilde{g}(j).
$$

Now take $t \to \infty$ in this equation. Since all states but 0 and 2N are absorbing, Theorem 3 implies

$$
\frac{i}{2N} = \tilde{g}(i) = p_{00}^{\infty} \tilde{g}(0) + p_{i,2N}^{\infty} \tilde{g}(2N) = p_{i,2N}^{\infty}.
$$

This is exactly the expression for absorption at 2N that was stated in (4.22). Finally, Theorem 3 says that absorption must occur with probability one. Since 0 is the only other absorbing state, it follows that $p_{00}^{\infty} = 1 - p_{i,2N}^{\infty} = (2N - i)/2N$, as claimed.

**Exercise 4.2.** Using the result of the previous example and of Exercise 4.1, show that for the Moran model, the probability that the chain is absorbed by state $N$ if it starts at $i$ is $i/N$.

Here is a statement and derivation of the general rule behind equation (4.23). Suppose $B_1, \ldots, B_n$ are disjoint events which partition the set of all possible outcomes (so that $\mathbb{P}(B_1 \cup \cdots \cup B_n) = 1$). Then

$$
\mathbb{P}(A \mid C) = \sum_{i=1}^{n} \mathbb{P}(A \mid C \cap B_i) \mathbb{P}(B_i \mid C) \tag{4.26}
$$

The derivation of this rule is very much like the derivation of the total probability rule (2.12) in Chapter 2. Because the events $B_1, \ldots, B_n$ partition the set of all outcomes,

$$
\mathbb{P}(A \cap C) = \sum_{i=1}^{n} \mathbb{P}(A \cap C \cap B_i) = \mathbb{P}(A \mid C \cap B_i) \mathbb{P}(C \cap B_i).
$$

Thus, since $\mathbb{P}(A \mid C) = \mathbb{P}(A \cap C)/\mathbb{P}(C)$,

$$
\mathbb{P}(A \mid C) = \sum_{i=1}^{n} \frac{\mathbb{P}(A \mid C \cap B_i) \mathbb{P}(C \cap B_i)}{\mathbb{P}(C)} = \sum_{i=1}^{n} \frac{\mathbb{P}(A \mid C \cap B_i) \mathbb{P}(B_i \mid C)}{\mathbb{P}(C)}
$$
Proof of the statements in Theorem 3. The technique we want to emphasize in this proof was already used in Example 4.8. It is possible to calculate many quantities of interest related to Markov chains by deriving systems of linear equations using conditioning and the Markov property. We did this in Example 4.8 to calculate the probabilities of absorption into different states. We will use the same technique to show that under the hypotheses of Theorem 3, the chain must eventually enter an absorbing state.

(A) We will first prove that \( \lim_{t \to \infty} p_{ij}^t = 0 \) whenever \( j \) is not an absorbing state, in the special case in which there is exactly one absorbing state, labeled 0, and \( i \to 0 \) for all other state \( i \).

(i) \( p_{i0}^t \) is, by definition, the probability that the chain has been absorbed by time \( t \), given that \( X(0) = i \). This probability cannot decrease as \( t \) increases, because if the chain has entered 0 by time \( t \), it has certainly entered by time \( t+1 \). Also,

\[
\mathbb{P}(\text{chain is absorbed eventually} \mid X(0) = i) = \lim_{t \to \infty} \mathbb{P}(\text{chain is absorbed by} t \mid X(0) = i) = \lim_{t \to \infty} p_{i0}^t. \tag{4.27}
\]

(ii) For each state \( i \), \( i \to 0 \), and so there is a smallest time \( t_i \) such that \( p_{i0}^{t_i} > 0 \). Since there are only a finite number of states, the maximum value of \( t_i \) over all states is some finite time \( k \). Thus, because of (i), \( p_{i0}^k > 0 \) for all states \( i \).

(iii) Let \( U \) be the event that \( X(t) \) never enters the absorbing state 0, and define, \( h(i) = \mathbb{P}(U \mid X(0) = i) \). We will show that

\[
h = A^t \cdot h, \text{ for all } t \geq 1, \text{ and } h(0) = 0. \tag{4.28}
\]

The fact that \( h(0) = 0 \) follows directly from the definition of \( h \). The same type of argument that was applied to \( g \) in Example 4.8 can be used to show \( h = A \cdot h \). Indeed, by conditioning and using the Markov property,

\[
h(i) = \sum_{j \in \mathcal{E}} \mathbb{P}(U \mid X(1) = j, X(0) = i) \mathbb{P}(X(1) = j \mid X(0) = i) = \sum_{j \in \mathcal{E}} p_{ij} h(j),
\]

for each \( i \), which is the same as \( h = A \cdot h \). By iterating this equation \( h = A^t h \) for any \( t \) as claimed.

(iv) Our object is to show that the only solution to (4.28) must be the zero vector—\( h(i) = 0 \) for all \( i \). To this end, let \( r = \max h(j); j \in \mathcal{E} \). Then using (4.28) for \( t = k \), where \( k \) was defined in (ii), and the fact that \( h(0) = 0 \)

\[
h(i) = \sum_{j \in \mathcal{E}} p_{ij}^k h(j) \leq r \sum_{j \in \mathcal{E}, j \neq 0} p_{ij}^k = r(1 - p_{i0}^k).
\]
By taking a maximum over \( i \), it follows that
\[
0 \leq r \leq r \max_{i \in E} (1 - p_{i0}^k).
\] (4.29)

But by the choice of \( k \), \( p_{i0}^k > 0 \) for all \( k \) and so \( \max_{i \in E} (1 - p_{i0}^k) < 1 \). Therefore, the only solution \( r \) to (4.29) is \( r = 0 \). This proves \( h(i) = 0 \) for all \( i \).

(v) It follows from \( h(i) = 0 \) and (4.27) that
\[
\lim_{t \to \infty} p_{i0}^t = 1.
\]

Since \( 1 = \sum_{j \in E} p_{ij}^t = 1 \) for all \( t \), it follows that \( \lim_{t \to \infty} p_{ij}^t = 0 \) whenever \( j \neq 0 \).

(B) We next remove the restriction that there be just one absorbing state. Indeed suppose there are several. Define a new chain \( X' \) as follows. Replace the collection of absorbing states of \( X \) by a single absorbing state 0. Assume that \( X' \) is the same as \( X \) until \( X \) enters on of its absorbing states; at this point send \( X' \) to the state 0. The new chain \( X' \) will behave just like the old chain until the time of absorption. Thus if \( i \) and \( j \) are not absorbing, \( p_{ij}^t \) is the same for both chains. But we have proved above that \( X'(t) \) must enter its absorbing state and \( \lim_{t \to \infty} p_{ij}^t = 0 \) whenever \( j \) is non-absorbing. Therefore, \( X(t) \) must enter one of its absorbing states and \( \lim_{t \to \infty} p_{ij}^t = 0 \) whenever \( j \) is non-absorbing.

(C) Finally, we justify the remaining statements of Theorem 3. The argument we gave in (A) (i) shows that \( p_{is}^t \) is non-decreasing as \( t \) increases whenever \( s \) is absorbing. A non-decreasing, bounded sequence of numbers must have a limit and so \( p_{is}^\infty \) exists whenever \( s \) is absorbing. It is equal to the probability that \( X \) gets absorbed by state \( s \) when \( X(0) = i \) because the events \{ \( X \) is absorbed by \( s \) by time \( t \) \} increase with \( t \) and their union over all \( t \) is the event of eventual absorption by \( s \). Finally, since \( 1 = \sum_{j \in E} p_{ij}^t \) for all \( t \), it follows by taking limits as \( t \to \infty \) and using (4.21), that
\[
\sum_{s \text{ is absorbing}} p_{is}^\infty = 1. \quad \diamond
\]

4.4 Recurrent chains

The motivation for this section is analyzing modifications of the Wright-Fisher and Moran models that allow mutation. When mutation can occur, these Markov chains may no longer have absorbing states, and with no states to settle into permanently, the chains will behave very differently in the long run. What can be said about \( \lim_{t \to \infty} A^t \) and hence about the
limiting behavior of the distribution of $X(t)$? There is an elegant answer to this question for general Markov chains.

**Example 4.9.** Here is a simple modification of the Moran model to allow mutation—compare to the mutation model in Section 3.3.5. When a type $B$ reproduces, its offspring is a $b$ with probability $u$ and a $B$ with probability $1-u$; when a $b$ reproduces, its offspring is a $B$ with probability $v$ and a $b$ with probability $1-v$. The reproduction events are independent of each other. One can think of $u$ and $v$ as mutation, or copying error, probabilities. It is left as an exercise to show that

\[
\begin{align*}
    p_{i,i+1} &= \frac{i(N-i)}{N^2}(1-u) + \frac{(N-i)^2}{N^2} \cdot v \\
    p_{i,i-1} &= \frac{i(N-i)}{N^2}(1-v) + \frac{i^2}{N^2} \cdot u \\
    p_{ii} &= \frac{i(N-i)}{N^2}(u+v) + \frac{(N-i)^2}{N^2}(1-v) + \frac{i^2}{N^2}(1-u)
\end{align*}
\]  

(4.30)  

(4.31)

Observe that $p_{01} = v$ and $p_{N,N-1} = u$. Thus, if $u > 0$ and $v > 0$, neither 0 nor $N$ are absorbing. Since $p_{i,i+1} > 0$ and $p_{i,i-1} > 0$ for all other states $i$, $i \leftrightarrow j$ for all states $i$ and $j$ of the chain.

Consider now an arbitrary Markov chain, which, like the Moran model with mutation, evolves in a finite state space $E$ and has the property that any two states $i$ and $j$ in $E$ communicate. Thus no states are absorbing, the Markov chain cannot remain in any state forever, and hence one expects that in fact, with probability one, it will visit all states infinitely often. A state $i$ with the property that $P\left(X(t) \text{ returns to } i \text{ infinitely often } \bigg| X(0) = i \right) = 1$ is said to be **recurrent**. We expect therefore that in a finite state space chain whose states all communicate, all states are recurrent. That this is indeed the case is part of the following general theorem, which gives a sense of how chains with recurrent states behave. It should be noted that in parts (a) and (b), it is not assumed that the state space must be finite.
Theorem 4

(a) If \( P(X(t) \text{ returns to } i \text{ for at least one } t \geq 1 \mid X(0) = i) = 1 \), then \( i \) is recurrent.

(b) Let \( i \) be a recurrent state. If \( i \leftrightarrow j \), then \( j \) is also recurrent and \( P(X(t) \text{ visits } j \text{ infinitely often} \mid X(0) = i) = 1 \).

(c) If the state space of the chain is finite and all states communicate, then all states are recurrent.

The proof of this theorem is omitted, but it is easy to understand. Part (a) says that if the chain must return at least once to \( i \) if it starts out there, then it must return infinitely often. This is ultimately a consequence of the Markov property; the future of the process at the time of first return to \( i \) is just a new Markov chain starting at \( i \) with the same transition probabilities. Thus, since the probability to return to \( i \) is one, the chain will return to \( i \) a second time. Continuing the same line of reasoning, once it returns a second time, it must return a third, and so on. We have not justified the steps in this argument, but it can be made rigorous.

To understand part (b), one reasons there is the same positive chance to visit \( j \) during every excursion from \( i \) back to \( i \) and these chances are independent. It’s as if one flipped a coin to determine whether to visit \( j \) during an excursion. If the probability of heads is positive, heads will turn up infinitely often in an infinite sequence of flips. Likewise, a visit to \( j \) will occur in an infinite number of excursions from \( i \) back to \( i \).

Finally, if the state space is finite, there must be at least one state that the chain returns to with probability one. This state is then recurrent, and if the rest communicate with it, all states are recurrent by part (b).

Part (c) is not true for infinite state space chains. For example, consider the simple random walk with \( p > q > 0 \). Recall that this can be represented in the form \( X(t) = X(0) + \sum_{s=1}^{t} \xi_s \), where \( \xi_1, \xi_2, \ldots \) are independent Bernoulli random variables with \( P(\xi_s = 1) = p \) and \( P(\xi_s = -1) = q \), which means they have mean \( E[\xi_s] = p - q > 0 \). Given any two states \( i \) and \( j \), the path that moves from \( i \) to \( j \) in successive unit steps occurs with positive probability, so any two state communicate. But when \( p > q \), the random walker moves right more often than left and hence will drift right. In fact, by the law of large numbers \( \lim_{t \to \infty} (1/t) \sum_{s=0}^{t} \xi_s = p - q > 0 \), which implies that \( \lim_{t \to \infty} X(t) = \infty \) with probability one. This means that with probability one, \( X(t) \) can visit any state only a finite number of times, and hence no states are recurrent.

Observe that an absorbing state is certainly recurrent. A state that is not recurrent is said to be transient. For example in the Moran and Wright-Fisher models without mutation, all states except the absorbing states are transient, because starting from any state the chain eventually reaches an
4. Markov Chains and Genetics

absorbing state from which it cannot move. All the states in the simple random walk, when either \( p > q \) or \( q > p \), are transient.

To describe the limiting behavior of recurrent chains, it is useful to define the notion of an invariant or stationary distribution. A stationary distribution for a Markov chain with state space \( \mathcal{E} \) and transition probability matrix \( A \) is a probability mass function \( \{r(j); j \in \mathcal{E}\} \) on the states of \( \mathcal{E} \) such that

\[
r(j) = \sum_{i \in \mathcal{E}} r(i) p_{ij} \quad \text{for every } i \in \mathcal{E}
\]

(4.32)

In linear algebra language, if \( r \) is the row vector indexed by the states of \( \mathcal{E} \), then \( r \) is a solution of

\[
r = r \cdot A, \quad r(j) \geq 0 \quad \text{for every } j \in \mathcal{E} \quad \text{and} \quad \sum_{j \in \mathcal{E}} r(j) = 1.
\]

(4.33)

In practice, one finds stationary distributions by solving these equations.

Solutions to (4.33) are called stationary or invariant for the following reason. Suppose that the probability mass function of \( X(0) \) is given by \( r \). Let \( \rho_t \) denote the probability mass function of \( X(t) \), considered as a row vector, as in section 4.3. Then, by Theorem 2 and (4.33)

\[
\rho_t = \rho_0 \cdot A^t = r \cdot A^t = r \cdot A^{t-1} = \cdots = r \cdot A = r = \rho_0.
\]

Thus, \( \rho_t \) does not change; although the chain moves from state to state, \( \rho_t(j) = P(X(t) = j) = P(X(0) = j) \) for all \( t \). The mass distribution defined by \( r \) is thus the stochastic equivalent of a fixed point in the deterministic models.

Example 4.10. We show that there is a unique stationary probability distribution for the two state Markov chain defined in Example 4.3, if \( \lambda > 0 \) and \( \mu > 0 \). It is necessary to solve the equations

\[
(r(0)r(1)) = (r(0)r(1)) \cdot \begin{pmatrix} 1 - \lambda & \lambda \\ \mu & 1 - \mu \end{pmatrix} \quad \text{and} \quad r(0) + r(1) = 1.
\]

This is equivalent to the three equations \( r(0) = (1 - \lambda)r(0) + \mu r(1) \), \( r(1) = \lambda r(0) + (1 - \mu)r(1) \), and \( r(0) + r(1) = 1 \). There are three equations for two unknowns here, so it looks as if this system of equations is over-determined. However, it is always the case that at least one of the equations in \( r = r \cdot A \) is redundant. In the example, the first two equations give the same result: \( r(1) = (\lambda / \mu) r(0) \). Substitution of this result into \( r(0) + r(1) = 1 \) allows one to find \( r(0) \). The final result is \( (r(0), r(1)) = (\mu / (\lambda + \mu), \lambda / (\lambda + \mu)) \).

Stationary distributions arise as limits of Markov chain distributions. Suppose that \( \{X(t)\} \) is a Markov chain with transition probability matrix
A, and assume that \( \rho_\infty = \lim_{t \to \infty} \rho_t \) exists and defines a probability mass function; then, by using the identity \( \rho_t = \rho_{t-1} \cdots A \) from Theorem 2,

\[
\rho_\infty = \lim_{t \to \infty} \rho_t = \lim_{t \to \infty} \rho_{t-1} \cdots A = \rho_\infty \cdot A
\]

Hence if the mass function settles into a limit, that limit must be a stationary distribution. This argument is valid whatever the recurrence or transience properties of the states are.

The limiting behavior of recurrent chains is particularly nice. However, to describe it we have to take care of one technical point. Suppose that starting from \( i \), the chain can only return to \( i \) at times \( d, 2d, 3d, \ldots \). If \( d > 1 \), state \( i \) is said to have period \( d \). When \( d = 1 \), state \( i \) is said to be aperiodic. Any state \( i \) for which \( p_{ii} > 0 \) is aperiodic because then the chain can return to \( i \) after one time increment with positive probability. Thus all states in the Moran and Wright-Fisher models are aperiodic, with or without mutation—we say then that the chain is aperiodic. The simple random walk has period 2; because the chain moves up or down one unit every time step, it can only return to its starting state at even times. If a state has period \( d > 1 \), then \( \lim_{t \to \infty} p_{ii}^{(t)} \) may not exist, because \( \lim_{s \to \infty} p_{ii}^{(sd)} \) may be positive, but \( p_{ii}^{(t)} = 0 \) when \( t \) is not an integer multiple of \( d \).

It turns out that if the period of \( i \) is \( d \) and \( i \leftrightarrow j \), then the period of \( j \) is \( d \) also. It is not hard to check for aperiodicity. As we have seen, \( p_{ii} > 0 \) implies aperiodicity. A finite state space chain will be aperiodic if and only if all the entries \( A^k \) are strictly positive for some integer \( k \geq 1 \).

**Theorem 5**  
(a) Let \( \{X(t)\} \) be a finite state, aperiodic Markov chain all of whose states communicate. Then it admits a unique invariant distribution \( \pi \) and for every pair of states \( i \) and \( j \),

\[
\lim_{t \to \infty} p_{ij}^{(t)} = \pi(j)
\]

Furthermore, whatever the initial distribution \( \lim_{t \to \infty} \rho_t = \pi \).

(b) Let \( \{X(t)\} \) be a Markov chain (whose state space is possibly countably infinite). Assume that there exists a stationary distribution, and that all the states communicate. Then the states are recurrent and (4.34) is true.

**Example 4.11.** Consider the Moran model with mutation as defined in Example 4.6. Assume that \( N = 3 \) and \( u = v = 1/4 \). Then the probability transition matrix is

\[
A = \begin{pmatrix}
3/4 & 1/4 & 0 & 0 \\
7/36 & 19/36 & 10/36 & 0 \\
0 & 10/36 & 19/36 & 7/36 \\
0 & 0 & 1/4 & 3/4 \\
\end{pmatrix}
\]
4. Markov Chains and Genetics

First, we will find the invariant distribution \( \mathbf{r} \). The equation \( \mathbf{r} = \mathbf{r} \cdot \mathbf{A} \) is equivalent to

\[
\begin{align*}
    r(0) &= \frac{3}{4} r(0) + \frac{7}{36} r(1) \\
    r(1) &= \frac{1}{4} r(0) + \frac{19}{36} r(1) + \frac{10}{36} r(2) \\
    r(2) &= \frac{10}{36} r(1) + \frac{19}{36} r(2) + \frac{1}{4} r(3) \\
    r(3) &= \frac{7}{36} r(2) + \frac{3}{4} r(3)
\end{align*}
\]

Use these equations in succession to solve for \( r(1), r(2), r(3) \) in terms of \( r(0) \). The result is \( \mathbf{r} = r(0)(1, \frac{9}{7}, \frac{9}{7}, 1) \). Since the terms of \( \mathbf{r} \) must sum to one, \( 1 = r(0)[1 + (9/7) + (9/7) + 1] = (32/7)r(0) \). Thus \( \mathbf{r} = (7/32, 9/32, 9/32, 7/32) \).

By Theorem 5, this vector is the same as \( \rho(\infty) \). No matter what the initial distribution of the chain is, the long time limits of the probabilities to be in states 0, 1, 2, or 3 are, respectively, 7/32, 9/32, 9/32, and 7/32. If the chain is in, say, state 0, it will take on average 32/7 time steps to return to state 0. \( \diamond \)