Lecture Notes on Stochastic Processes with Applications in Biology

David F. Anderson

March 1, 2017

Preface

This book has two main purposes. The first is to provide a basic introduction to some of the most common stochastic models, such as discrete and continuous time Markov chains, renewal processes and point processes. The second purpose is to provide an introduction to the basic stochastic models utilized in cellular and molecular biology. This introduction can be found in chapter 5. While there are a number of very good textbooks that provide an introduction to Markov processes, renewal processes and point processes, there are few that introduce the stochastic models used in cellular and molecular biology, and it was this fact that motivated the writing of these notes.

Simulation methods are a core component of this book. It is my intention that each homework assignment should have at least one exercise that requires numerical simulation, and that these exercises should increase in complexity as the course progresses. These exercises begin in a gentle enough manner that no previous experience with simulation should be required.

A student who has taken courses in calculus, linear algebra, and introductory probability should find these notes accessible. Having taken a class in real analysis, or any other proof based course, is also desirable. No knowledge of measure theory will be assumed.

It is my hope that students with a mathematics background will gain an appreciation for some of the mathematical challenges in biology, and that students with a biological background will gain the basic mathematical tools necessary to study the models that are arising in their field. It is also one of my strong desires that both groups of students gain an appreciation for the other's field.

Various portions of these notes were, of course, greatly influenced by a number of other books on stochastic processes. The main such texts are Rick Durrett's *Essentials of Stochastic Processes* [10], Gregory Lawler's *Introduction to Stochastic Processes* [28], James Norris's *Markov chains* [29], Sidney Resnick's *Adventures in Stochastic Processes* [31], Darren Wilkinson's *Stochastic Modelling for Systems Biology* [36], and *Stochastic Analysis for Biochemical Systems* by myself and Thomas Kurtz [3].

There are undoubtedly many typos throughout these notes. If you find any, please email me at anderson@math.wisc.edu.

David Anderson
August 2015

Contents

Preface i

1 Discrete Time Markov Chains 1
- 1.1 Stochastic processes . . . 1
- 1.2 Discrete time Markov chains . . . 2
  - 1.2.1 Examples of discrete time Markov chains . . . 6
  - 1.2.2 Stopping times and the strong Markov property . . . 9
- 1.3 Simulation of discrete time Markov chains . . . 10
- 1.4 Higher Order Transition Probabilities . . . 12
- 1.5 Reducibility, periodicity, and recurrence . . . 16
  - 1.5.1 Reducibility . . . 16
  - 1.5.2 Periodicity . . . 20
  - 1.5.3 Recurrence and transience . . . 22
- 1.6 Absorption probabilities . . . 27
- 1.7 Stationary distributions and the long time behavior of Markov chains 34
  - 1.7.1 Markov chains with a finite state space . . . 36
  - 1.7.2 Markov chains with a countably infinite state space . . . 45
- 1.8 Exercises . . . 52

2 Discrete Time Markov Chains in the Life Sciences 61
- 2.1 Genetic Models . . . 61
  - 2.1.1 Mendelian inheritance . . . 61
  - 2.1.2 The Wright-Fisher Model . . . 64
  - 2.1.3 Phylogenetic Distance and the Jukes-Cantor Model . . . 66
- 2.2 Discrete Time Birth and Death models . . . 68
- 2.3 Branching Processes . . . 75
  - 2.3.1 Terminology and notation . . . 75
  - 2.3.2 Behavior of the mean . . . 76
  - 2.3.3 Probability of extinction . . . 77
- 2.4 Exercises . . . 84

3 Renewal and Point processes 89
- 3.1 Renewal Processes . . . 89
  - 3.1.1 The behavior of $N(t)$, as $t \to \infty$ . . . 91
  - 3.1.2 The renewal reward process . . . 95

3.2 Point Processes . . . 97
3.2.1 Preliminaries . . . 98
3.2.2 The Poisson process . . . 100
3.2.3 Transformations of Poisson processes . . . 105
3.3 Exercises . . . 118

4 Continuous Time Markov Chains 123
4.1 The basics . . . 123
4.2 Explosions . . . 130
4.3 Forward Equation, Backward Equation, and the Generator Matrix . . . 133
4.4 Stationary Distributions . . . 140
4.4.1 Classification of states . . . 141
4.4.2 Invariant measures . . . 141
4.4.3 Limiting distributions and convergence . . . 147
4.5 The Generator, Revisited . . . 148
4.6 Continuous Time Birth and Death Processes . . . 151
4.7 Exercises . . . 155

5 Continuous Time Markov Chains in biochemistry 158
5.1 Models of biochemical reaction systems . . . 158
5.1.1 The basic model . . . 158
5.1.2 Intensity functions: mass-action kinetics . . . 160
5.1.3 Example: Gene transcription and translation . . . 161
5.1.4 Example: Virus kinetics . . . 165
5.1.5 Example: Enzyme kinetics . . . 166
5.1.6 Generator of the process and the forward equations . . . 167
5.1.7 Population processes . . . 169
5.2 Simulation . . . 171
5.2.1 The stochastic simulation algorithm . . . 172
5.2.2 The next reaction method . . . 173
5.2.3 Euler's method . . . 174
5.3 First order reaction networks . . . 175
5.4 Relationship with Deterministic Model: the Law of Large Numbers . . . 178
5.4.1 Conversion of rate constants . . . 179
5.4.2 The classical scaling . . . 180
5.5 Exercises . . . 182

6 Brownian motion and stochastic integration 184
6.1 Brownian Motions . . . 184
6.1.1 Markov property and generator of a Brownian motion . . . 185
6.2 Integration with Respect to Brownian Motion . . . 187
6.3 Diffusion and Linear Noise Approximations . . . 192
6.3.1 Diffusion approximation . . . 192
6.3.2 Linear noise approximation . . . 193
6.4 Numerical methods for stochastic differential equations . . . 195

A Solving linear difference equations **197**
A.1 Linear Difference Equations 197
A.2 Exercises 200

B Some probability basics **201**
B.1 Exponential random variables 201
B.2 Transformations of random variables 202
B.3 Estimation via simulation: Monte Carlo 204
B.4 Exercises 207

Chapter 1

Discrete Time Markov Chains

This chapter presents an introduction to discrete time Markov chains with a discrete state space. Before proceeding to Markov chains, however, we begin with a quick discussion of the overall subject of this book, stochastic processes.

1.1 Stochastic processes

A *stochastic process* is a collection of random variables indexed by time. They are used to model the evolution of random processes in time. We begin with two examples, one in which it is natural to model time discretely and one in which it is natural to model time continuously.

Example 1.1. Repeatedly role a fair six-sided die and for $k \in \{1, 2, \dots\}$ let Z_k be the outcome of the kth roll. Let

$$X_n = Z_1 + \cdots + Z_n$$

be the accumulated total of the first n rolls. If $Z_1 = 3$, then $X_1 = 3$ and $X_2 \in \{4, 5, 6, 7, 8, 9\}$, each with equal probability. Note that time, indexed here by n, is discrete since we only update the process X_n after each roll of the die. $\triangle$

Example 1.2. Consider a frog who lives in a pond with k lily pads, which we have labeled 1 through k. The frog starts the day by sitting on a randomly chosen pad. For example, each lily pad may have an equal probability of being the first one chosen, in which case

$$P(\text{frog starts on lily pad } i) = \tfrac{1}{k}, \quad i \in \{1, \dots, k\}.$$

When the frog is sitting on a lily pad, it will jump to another lily pad after a random amount of time. For the sake of concreteness, we assume that this random time is modeled by a unit exponential random variable. When the frog jumps it chooses the pad it goes to according to some probability distribution on the remaining pads. Of course, this distribution may depend upon which lily pad the frog is currently enjoying. We assume that the frog is very fast and the jump happens instantaneously. Letting $t = 0$ denote the start of the day, we let $X(t) \in \{1, \dots, k\}$ denote the lily pad occupied by the frog at time t. In this example, time is naturally continuous. $\triangle$

As is customary in the field of probability, we will denote stochastic processes with upper case letters with the time index either in parenthesis, as in $X(t)$ in Example 1.2, or as a subscript, as in X_n in Example 1.1. Most often, the time parameter will be a subset of the nonnegative integers $\{0, 1, 2, \dots\}$, in which case it will most often be denoted by n, or a subset of $[0, \infty)$, the nonnegative real numbers, in which case it will most often be denote by t. When time is indexed by the nonnegative integers, we say it is a *discrete time* process, whereas when time is indexed by the nonnegative reals, we say it is a *continuous time* process. Each process will take values in a *state space*, which can be discrete (finite or countably infinite) or continuous (for example, the real line or $\mathbb{R}^d$). The natural state space for Example 1.1 is $\{1, 2, \dots\}$, whereas the state space for Example 1.2 is $\{1, \dots, k\}$.

When analyzing a stochastic process, we may be interested in computing probabilities, such as

$$P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n), \tag{1.1}$$

where $n \in \{0, 1, 2, \dots\}$ and $\{i_0, \dots, i_n\}$ is some finite sequence of states in the state space. We may also be interested in generating some realizations via simulation in order to visualize the process and see some possible behaviors. In this text, we will focus on both issues. That is, we will learn how to analytically compute probabilities of the form (1.1), together with other interesting objects such as expectations, exit probabilities, etc., and will also learn to simulate the processes in order to generate realizations.

1.2 Discrete time Markov chains

We begin our study of stochastic processes with discrete time Markov chains. We will focus on models with a discrete state space. Such models, while relatively straightforward to understand, have proven useful in many arenas, from biology, to ecology, to engineering, to finance, etc.

Definition 1.3. A discrete time stochastic process $X_n, n \geq 0$, is said to satisfy the *Markov property*, and is called a *discrete time Markov chain*, if conditioned on X_n (the present), the random variables $\{X_{n+1}, X_{n+2}, \dots\}$ (the future), are independent of the random variables $\{X_0, \dots, X_{n-1}\}$ (the past).

Definition 1.3 says that the probabilities associated with future states only depend upon the current state, and not on the history of the process. For example, if X_n is a discrete time Markov chain, then for any $n \geq 0$ and any set of states $i_0, \dots, i_n, i_{n+1}$ in the state space,

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n). \tag{1.2}$$

We state without proof the following fact, which says that the seemingly weaker condition given in (1.2) is actually equivalent to the Markov property given in Definition (1.3).

Proposition 1.4. *Let $X_n, n \geq 0$, be a discrete time stochastic process with a discrete state space that satisfies (1.2) for any choice of states $i_0, i_1, \ldots, i_{n+1}$. Then X_n satisfies the Markov property given in Definition 1.3.*

It is often the case that it is easier to show that a process satisfies the property (1.2) than Definition 1.3.

Example 1.5. Suppose that during each round of a game, a fair coin is tossed. If the coin lands on heads, a six-sided die is rolled. If the coin lands on tails, a twenty-sided die is rolled. We let Z_n be the value of the roll during the nth iteration of this game, and let $X_n = Z_1 + \cdots + Z_n$. Then $X_n, n \geq 1$, is a Markov chain.

Showing the general condition given in Definition 1.3 would be difficult. However, by the independence of the rolls and coin flips, we have that (1.2) holds since X_n is a function of X_{n-1} plus a random variable, Z_n, that is independent of all previous random variables. △

Example 1.6 (Example 1.1 continued)**.** Recall Example 1.1, where we let Z_k be the outcome of the kth roll of a fair die and let $X_n = Z_1 + \cdots + Z_n$. Assuming the rolls are independent, $X_n, n \geq 1$, is a Markov chain since X_n is a function of X_{n-1} and a random variable, Z_n, independent from all previous random variables.

We can calculate any probability of the form (1.1) for X_n. For example,

$$\begin{aligned} P(X_1 = 2, X_2 = 4, X_3 = 6) &= P(X_3 = 6 | X_1 = 2, X_2 = 4) P(X_1 = 2, X_2 = 4) \\ &= P(X_3 = 6 | X_2 = 4) P(X_2 = 4 \mid X_1 = 2) P(X_1 = 2) \\ &= P(Z_3 = 2) P(Z_2 = 2) P(Z_1 = 2) \\ &= \left(\tfrac{1}{6}\right)^3. \end{aligned}$$

where the Markov property was used in the second equality. △

Example 1.7 (Example 1.2 continued)**.** Suppose a frog can jump between three lily pads, labeled 1, 2, and 3. Suppose that if the frog is on lily pad number 1, it will jump next to lily pad number 2 with a probability of one. Similarly, if the frog is on lily pad number 3, it will next jump to lily pad number 2 with a probability of one. However, when the frog is on lily pad number 2, it will next jump to lily pad number 1 with probability 1/4, and to lily pad number 3 with probability 3/4. We can depict the state space and associated probabilities graphically via

$$\textcircled{1} \underset{1}{\overset{1/4}{\leftrightarrows}} \textcircled{2} \underset{3/4}{\overset{1}{\leftrightarrows}} \textcircled{3}.$$

Let X_n denote the position of the frog after the nth jump and assume that the initial distribution is

$$P(X_0 = 1) = \tfrac{1}{3}, \quad P(X_0 = 2) = \tfrac{2}{3}, \quad \text{and} \quad P(X_0 = 3) = 0.$$

Assuming that $X_n, n \geq 0$, is a Markov chain we have

$$\begin{aligned} P(X_0 = 1, X_1 = 2, X_2 = 3) &= P(X_2 = 3 | X_0 = 1, X_1 = 2) P(X_0 = 1, X_1 = 2) \\ &= P(X_2 = 3 | X_1 = 2) P(X_1 = 2 | X_0 = 1) P(X_0 = 1) \\ &= \tfrac{3}{4} \times 1 \times \tfrac{1}{3} = \tfrac{1}{4}, \end{aligned}$$

whereas $P(X_0 = 1, X_1 = 3, X_2 = 2) = 0$. △

It is useful to see at least one process that does not satisfy the Markov property.

Example 1.8. Suppose that $Z_n, n \geq 1$, are i.i.d. random variables for which $P(Z_n = 1) = P(Z_n = -1) = \frac{1}{2}$. For each $n \geq 1$, let $W_n = Z_1 + \cdots + Z_n$ and let

$$M_n = \max\{W_k : 1 \leq k \leq n\}.$$

Then the sequence $M_n, n \geq 1$, does not satisfy the Markov property.

We will show

$$P(M_4 = 2 | M_1 = 1, M_2 = 1, M_3 = 1) \neq P(M_4 = 2 | M_1 = -1, M_2 = 0, M_3 = 1), \quad (1.3)$$

demonstrating that (1.2) does not hold and the history of the process does influence future probabilities. Starting with the right hand side of (1.3),

$$P(M_4 = 2 | M_1 = -1, M_2 = 0, M_3 = 1) = P(Z_3 = 1 | Z_1 = -1, Z_2 = 1, Z_3 = 1) = \tfrac{1}{2}.$$

Next, using that $\{M_1 = 1, M_2 = 1, M_3 = 1\}$ is equal to

$$\{Z_1 = 1, Z_2 = -1, Z_3 = -1\} \cup \{Z_1 = 1, Z_2 = -1, Z_3 = 1\},$$

it is an exercise in the use of conditional probabilities to show

$$P(M_4 = 2 | M_1 = 1, M_2 = 1, M_3 = 1) = \tfrac{1}{4}.$$

Hence, we have (1.3) and $M_n, n \geq 1$, does not satisfy the Markov property. △

We leave examples and return to the task of developing our mathematical machinery.

Definition 1.9. The *one-step transition probability* of a Markov chain from state i to state j, denoted by $p_{ij}(n)$, is

$$p_{ij}(n) \stackrel{\text{def}}{=} P(X_{n+1} = j \mid X_n = i).$$

If the transition probabilities do not depend upon n, in which case $p_{ij}(n) = P(X_1 = j | X_0 = i)$, then the processes is said to be *time homogeneous*, or simply *homogeneous*, and we will use the notation p_{ij} as opposed to $p_{ij}(n)$.

Unless explicitly stated otherwise, all discrete time Markov chain models considered in these notes will be time homogeneous and the one step transition probabilities will be denoted by p_{ij} or $p_{i,j}$.

Throughout, we will denote initial probability distributions by α. That is, for a discrete time Markov chain X_n we let

$$\alpha_j = P(X_0 = j), \quad j \in S.$$

Throughout, we will think of α as a row vector whose ith component is α_i.

Returning to (1.1), for a discrete time Markov chain we have

$$\begin{aligned} &P(X_0 = i_0, \cdots, X_n = i_n) \\ &\quad = P(X_n = i_n \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}) P(X_0 = i_0, \cdots, X_{n-1} = i_{n-1}) \\ &\quad = p_{i_{n-1}, i_n} P(X_0 = i_0, \cdots, X_{n-1} = i_{n-1}) \\ &\quad \vdots \\ &\quad = \alpha_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i_n}, \end{aligned} \tag{1.4}$$

and under the Markov assumption the problem of computing probabilities has been converted to one of multiplication. For example, returning to Example 1.7, we have

$$P(X_0 = 1, X_1 = 2, X_2 = 3) = \alpha_1 p_{12} p_{23} = \tfrac{1}{3} \times 1 \times \tfrac{3}{4} = \tfrac{1}{4}.$$

For processes with a finite state space, the one-step transition probabilities can be conveniently expressed in matrix form.

Definition 1.10. The *transition matrix* P for a Markov chain with finte state space $S = \{1, 2, \ldots, N\}$ and one-step transition probabilities p_{ij} is the $N \times N$ matrix

$$P \stackrel{\text{def}}{=} \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}.$$

Note that the matrix P satisfies

$$0 \leq P_{ij} \leq 1, \quad \text{for all } 1 \leq i, j, \leq N, \tag{1.5}$$

$$\sum_{j=1}^{N} P_{ij} = 1, \quad \text{for each } 1 \leq i \leq N. \tag{1.6}$$

Any square matrix satisfying the two conditions (1.5) and (1.6) is called a Markov or *stochastic* matrix, and can be the transition matrix for a Markov chain. If P also satisfies the condition

$$\sum_{i=1}^{N} P_{ij} = 1, \quad \text{for each } 1 \leq j \leq N, \tag{1.7}$$

so that the column sums are also equal to 1, then P is termed *doubly stochastic*.

1.2.1 Examples of discrete time Markov chains

Having some basic definitions in hand, we turn to some examples that will be returned to throughout these notes.

Example 1.11. This example, sometimes termed the *deterministically monotone* Markov chain, is simple but will serve as a building block for an important model in the continuous time setting (the Poisson process).

Consider a discrete time process X_n with state space $\{1, 2, \dots, \}$ and transition probabilities $p_{i,i+1} = 1$ for $i \geq 1$:

$$\textcircled{1} \overset{1}{\to} \textcircled{2} \overset{1}{\to} \textcircled{3} \overset{1}{\to} \cdots .$$

If α is the initial distribution and $\alpha_1 = 1$, then the process simply starts at state 1 and proceeds deterministically up the integers towards positive infinity. △

Example 1.12. Suppose that $X_n, n \geq 0$, are independent and identically distributed with

$$P(X_0 = k) = p_k, \quad k = 1, \dots, N,$$

where $p_k \geq 0$ and $\sum_{k=1}^{N} p_k = 1$. Then,

$$P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_{n+1}},$$

since X_{n+1} is independent of all previous random variables and takes the value i_{n+1} with probability $p_{i_{n+1}}$. Hence, the process satisfies the Markov property. For this example, the transition matrix takes a particularly simple form,

$$P = \begin{pmatrix} p_1 & \cdots & p_N \\ \vdots & \ddots & \vdots \\ p_1 & \cdots & p_N \end{pmatrix}.$$

△

Example 1.13 (2 state Markov chain)**.** Consider a gene that can be repressed by a protein. By $X_n = 1$, we mean the gene is free, or not repressed, at time n, and by $X_n = 2$ we mean that the gene is repressed, or bound. We make the following assumptions:

1. If the gene is free at time n, there is a probability of $p \geq 0$ that it is repressed at time $n + 1$.

2. If the gene is repressed at time n, there is a probability of $q \geq 0$ that it is free at time $n + 1$.

The process X_n can be modeled as a discrete time Markov chain with finite state space $S = \{1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}, \tag{1.8}$$

where the first row and first column are associated with state 1.

Note that *any* two state discrete time Markov chain has a transition matrix of the form (1.8). We can graphically represent the state space and transition probabilities for this chain via

$$(1-p) \circlearrowleft \textcircled{1} \underset{p}{\overset{q}{\leftrightarrows}} \textcircled{2} \circlearrowright (1-q)$$

△.

Example 1.14 (Random walk on $\{0, 1, \ldots, N\}$)**.** For an integer $N \geq 1$, let the state space S of a discrete time Markov chain be $\{0, \ldots, N\}$. We consider a walker who can only move left (decrement by one) or right (increment by one) per time interval. We may think of the walker as flipping a biased coin to decide whether or not to move to the right or left during each time increment. That is, for some $p > 0$, at each time-step the walker moves one step to the right with probability p (she flipped a heads) and to the left with probability $1 - p$ (she flipped a tails). If $p = \frac{1}{2}$, the walk is termed *symmetric* or *unbiased*, whereas if $p \neq \frac{1}{2}$, the walk is *biased*. The one step transition probabilities for $i \in \{1, \ldots, N-1\}$ are,

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p, \quad 0 < i < N,$$

though we must still give the transition probabilities at the boundaries.

One choice for the boundary conditions would be to assume that with probability one, the walker transitions away from the boundary during the next time step. That is, we could have

$$p_{0,1} = 1, \qquad p_{N,N-1} = 1.$$

We say such a process has *reflecting boundaries*. Note that Example 1.7 was a model of a random walk on $\{1, 2, 3\}$ with reflecting boundaries.

Another option for the boundary conditions is to assume there is *absorption*, yielding the boundary conditions

$$p_{00} = 1, \quad p_{NN} = 1,$$

in which case the chain is often called the *Gambler's ruin* (a terminology that can be understood by assuming $p < 1/2$ and thinking of the state of the process as the wealth of a gambler).

We are not limited to reflection and absorption and could have a partial reflection such as

$$p_{00} = 1 - p, \quad p_{01} = p, \quad p_{N,N-1} = 1 - p, \quad p_{NN} = p.$$

Of course, we could also have any combination of the above conditions at the different boundary points.

We can generalize the model to allow for the possibility of the walker choosing to stay at a given site $i \in \{1, \ldots, N-1\}$ during a time interval and by allowing the

transition probabilities to be state dependent. In this general case, we let q_i, p_i and r_i be the probabilities that the walker moves to the left, right, and stays, respectively, given that she is in state i. Assuming absorbing boundary conditions, the transition matrix for this model is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 & 0 \\ \vdots & \ddots & & \ddots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

where it is understood that $q_i, p_i, r_i \geq 0$ and $q_i + p_i + r_i = 1$ for $i \in \{1, \ldots, N-1\}$. △

Example 1.15 (Axonal transport)**.** One method of transport used in living cells is axonal transport in which certain (motor) proteins carry cargo such as mitochondria, other proteins, and other cell parts, on long microtubules. These microtubule can be thought of as the "tracks" of the transportation mechanism, with the motor protein as the transporter. One natural, and simple, mathematical model for such transport would begin by breaking the microtubule into N equally sized intervals, and then letting X_n be the position of the motor protein on the state space $\{1, \ldots, N\}$. We could then let the transition probabilities satisfy

$$p_{i,i+1} = p_i, \quad p_{i,i-1} = q_i, \quad p_{i,i} = r_i, \qquad i \in \{2, \ldots, N-1\},$$

where $p_i + q_i + r_i = 1$ with $p_i, q_i, r_i \geq 0$, and with boundary conditions

$$p_{1,1} = r_1, \quad p_{1,2} = p_1, \qquad p_{N,N} = 1,$$

with $r_1 + p_1 = 1$. Taking the state N to be the end of the microtubule, it would be natural to model $p_i > q_i$ for each i. △

Example 1.16 (Random walk on the integers)**.** This Markov chain is like that of Example 1.14, except now we assume that the state space is all of the integers $S = \mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$. Specifically, X_n is the position of a walker at time n, where for some $0 < p < 1$ the transition probabilities are given by

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p,$$

for all $i \in \mathbb{Z}$. This model is one of the most studied stochastic processes and will be returned to frequently as a canonical example. △

Example 1.17 (Random walk on $\mathbb{Z}^d$)**.** We let $\mathbb{Z}^d$ be the d-dimensional integer lattice:

$$\mathbb{Z}^d = \{(x_1, \ldots, x_d) \ : \ x_i \in \mathbb{Z}\}.$$

Note that for each $x \in \mathbb{Z}^d$ there are exactly $2d$ values y with $|x - y| = 1$ (as there are precisely d components that can be changed by a value of ± 1). We may let

$$p_{xy} = \begin{cases} \frac{1}{2d} & \text{if } |x - y| = 1 \\ 0 & \text{else} \end{cases}.$$

△

1.2.2 Stopping times and the strong Markov property

We sometimes need a stronger property than the Markov property. Towards that end, we introduce the concept of a stopping time.

Definition 1.18. A *stopping time*, τ, with respect to a discrete time stochastic process X_n, $n \geq 0$, is a random variable taking values in $\{0, 1, \dots\}$ for which the occurrence or non-occurrence of $\{\tau = n\}$ can be determined by observation of the random variables $\{X_0, \dots, X_n\}$.

For our purposes, the most important examples of stopping times are the return times and hitting times of a process to a state or set of states.

Example 1.19. Consider a stochastic process $X_n, n \geq 0$, with discrete state space S. For $i \in S$ let

$$\tau_i \stackrel{\text{def}}{=} \inf\{n \geq 1 \ : \ X_n = i\} \tag{1.9}$$

$$H_i \stackrel{\text{def}}{=} \inf\{n \geq 0 \ : \ X_n = i\}, \tag{1.10}$$

where we take $\tau_i = \infty$ and $H_i = \infty$ if the chain never hits state i for $n \geq 1$ and $n \geq 0$, respectively. Similarly, if $A \subset S$, then we define $\tau_A = \inf\{n \geq 1 \ : \ X_n \in A\}$ and $H_A = \inf\{n \geq 0 \ : \ X_n \in A\}$.

Both τ_i and H_i are stopping times (as are τ_A and H_A for $A \subset S$). This follows since if you have observed the random variables $\{X_0, \dots, X_n\}$, then you can answer whether or not n was the first time the chain entered state i. The stopping times τ_i and H_i are often called the return time and hitting time of state i, respectively. $\triangle$

The term stopping time is best understood in the following manner. Think of X_n as the wealth of a gambler after n plays of a game. Then any rule devised by the gambler to determine when to stop playing leads to a stopping time. For example, the rule could be: stop playing when I have either doubled my money, or lost everything. The important point is that τ must be determinable from only the plays of the game through the present. No knowledge of the future is permitted![1]

The following is a generalization of the Markov property.

Definition 1.20. A discrete time stochastic process X_n is said to satisfy the *strong Markov property* if for any stopping time τ the following holds: conditioned on both X_n (the present) and $\tau = n$, the random variables $\{X_{n+1}, X_{n+2}, \dots\}$ (the future), are independent of the random variables $\{X_0, \dots, X_{n-1}\}$ (the past).

Suppose that τ is a stopping time for a Markov chain X_n. Then, assuming $\tau < \infty$,

[1] Imagine how good a gambler would be if they could use the following rule: stop gambling the round *before* my first loss.

the one step transition probabilities still hold at τ:

$$\begin{aligned} P(X_{\tau+1} = j \mid X_\tau = i, \tau < \infty) &= \sum_{n=0}^{\infty} P(X_{n+1} = j, \tau = n \mid X_\tau = i, \tau < \infty) \\ &= \sum_{n=0}^{\infty} P(X_{n+1} = j \mid \tau = n, X_n = i) P(\tau = n \mid X_\tau = i, \tau < \infty) \\ &= \sum_{n=0}^{\infty} P(X_{n+1} = j \mid X_n = i) P(\tau = n \mid X_\tau = i, \tau < \infty) \\ &= p_{ij}, \end{aligned}$$

where the second to last equality used the Markov property in that the distribution of the random variable X_{n+1} only depends upon X_n and not on any information pertaining to the random variables $\{X_0, \ldots, X_{n-1}\}$.

The following result, which we do not prove, will be useful to us.

Theorem 1.21. *A discrete time Markov chain satisfies the strong Markov property.*

1.3 Simulation of discrete time Markov chains

We will present a straightforward method for the simulation of a single realization of a discrete time Markov chain, $X_n, n \geq 0$. Without loss of generality, we take the state space of the Markov chain to be $S = \{1, 2, 3, \ldots\}$, the positive integers. We assume that we are provided with:

(i) an initial distribution α, satisfying $\alpha_k = P(X_0 = k)$, for $k \in \{1, 2, \ldots\}$, and

(ii) the one step transition probabilities $p_{ij} = P(X_1 = j \mid X_0 = i)$, for each pair $i, j \in \{1, 2, \ldots\}$.

The basic idea is the following. We will utilize a uniform$[0, 1]$ random variable, U_0, and the transformation method of Theorem B.7, to generate the initial state of the chain. We will then use a sequence of independent uniform$[0, 1]$ random variables, $\{U_n\}_{n=1}^{\infty}$, which are also independent of the initial random variable U_0, together with the transformation method of Theorem B.7, to generate each sequential state of the chain. Specifically, we will use functions g and f so that

$$X_0 = g(U_0), \quad \text{and} \quad X_n = f(X_{n-1}, U_n), \text{ for } n \geq 1,$$

where for $u \in [0, 1]$ and $i \in \{1, 2, \ldots\}$,

$$g(u) = \min\left\{ j \geq 1 \mid \sum_{m=1}^{j} \alpha_m \geq u \right\} \quad \text{and} \quad f(i, u) = \min\left\{ j \geq 1 \mid \sum_{m=1}^{j} p_{i,m} \geq u \right\}.$$

Here is the algorithm.

Algorithm 1 (Simulation of discrete time Markov chains with state space $\{1, 2, \dots\}$)**.** Take as input an initial distribution α and transition rates $\{p_{ij}\}$. Let $n = 1$.

1. Generate a uniform$[0, 1]$ random variable, U_0. Find the smallest $j \in \{1, 2, \dots\}$ satisfying
$$\sum_{m=1}^{j} \alpha_m \geq U_0,$$
and set $X_0 = j$.

2. Generate a uniform$[0, 1]$ random variable, U_n, that is independent of all previous random variables generated.

3. Find the smallest $j \in \{1, 2, \dots\}$ satisfying
$$\sum_{m=1}^{j} p_{X_{n-1},m} \geq U_n,$$
and set $X_n = j$.

4. Set $n \leftarrow n + 1$, and return to step 2 or quit.

It is straightforward to verify that the process generated in Algorithm 1 is a discrete time Markov chain with the desired transition probabilities.

First, by Theorem B.7 we have that $X_0 \sim \alpha$. Next, we verify that the constructed process satisfies the Markov property with the correct one-step transition probabilities. The important thing to note is that $\{X_0, X_1, \dots, X_n\}$ is a function of $\{U_0, \dots, U_n\}$. We have,

$$\begin{aligned} P(X_{n+1} = j \,|\, X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) & \\ &= P(f(i, U_{n+1}) = j \,|\, X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) \\ &= P(f(i, U_{n+1}) = j) \\ &= p_{ij}, \end{aligned}$$

where the first equality follows by the construction in Algorithm 1, the second follows from the independence of U_{n+1} and $\{U_0, \dots, U_n\}$, and the final step follows from Theorem B.7.

Example 1.22. Consider a process with state space $S = \{1, 2, 3\}$, initial distribution $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

If the sequence of independent uniform$[0, 1]$ random variables

$$(U_0, U_1, U_2, U_3, \dots) = (0.814,\ 0.906,\ 0.127,\ 0.632, \dots)$$

is utilized with Algorithm 1, then the following realization of the process is produced,

$$X_0 = 3, \quad X_1 = 3, \quad X_2 = 1, \quad X_3 = 3, \dots .$$

On the other hand, if the sequence of independent uniform$[0, 1]$ random variables

$$(U_0, U_1, U_2, U_3, \dots) = (0.096,\ 0.278,\ 0.547,\ 0.958, \dots)$$

is utilized with Algorithm 1, then the realization

$$X_0 = 1, \quad X_1 = 2, \quad X_2 = 2, \quad X_3 = 2, \dots .$$

is produced. △

1.4 Higher Order Transition Probabilities

Consider again the frog in Example 1.7. Perhaps we would like to have some idea as to where the frog will be at time $n = 2$, or even $n = 200$. We could simulate the behavior of the frog multiple times using the method of the previous section, but this seems uninspired and would not allow us to conclude anything with surety. Something more is desired.

Leaving the example of the frog, we ask one of the most basic questions possible of a discrete time Markov chain: given an initial distribution α, and a transition matrix P, what is the probability that the Markov chain will be in state $i \in S$ at time $n \geq 1$? This question leads naturally to the following definition.

Definition 1.23. The *n-step transition probability*, denoted $p_{ij}^{(n)}$, is the probability of moving from state i to state j in n steps,

$$p_{ij}^{(n)} \stackrel{\text{def}}{=} P(X_n = j \mid X_0 = i).$$

Note that by time-homogeneity we also have $p_{ij}^{(n)} = P(X_{n+k} = j \mid X_k = i)$ for any $k \geq 0$.

The n-step transition probabilities are intimately linked with powers of the transition matrix, P, of Definition 1.10. In the case of a finite state space we let P_{ij}^n denote the i, jth entry of the matrix P^n. Likewise, if the state space is infinite, then we define

$$P_{ij}^2 = \sum_{k \in S} p_{ik} p_{kj},$$

which converges since $\sum_k p_{ik} p_{kj} \leq \sum_k p_{ik} = 1$. Similarly, for $n \geq 3$ we define

$$P_{ij}^n = \sum_{k \in S} P_{ik}^{n-1} p_{kj}.$$

Proposition 1.24. *Let $X_n, n \geq 0$, be a discrete time Markov chain with one-step transition probabilities p_{ij}. For all integers $n \geq 1$ and all $i, j \in S$,*

$$p_{ij}^{(n)} = P_{ij}^n.$$

Proof. We will show the result by induction on n. First, note that the case $n = 1$ follows by definition. Now assuming the result is true for a given $n \geq 1$, we have

$$\begin{aligned} p_{ij}^{(n+1)} &= P(X_{n+1} = j \mid X_0 = i) = \sum_{k \in S} P(X_{n+1} = j, X_n = k \mid X_0 = i) \\ &= \sum_{k \in S} P(X_{n+1} = j \mid X_n = k, X_0 = i) P(X_n = k \mid X_0 = i) \\ &= \sum_{k \in S} P(X_{n+1} = j \mid X_n = k) P(X_n = k \mid X_0 = i) \\ &= \sum_{k \in S} p_{ik}^{(n)} p_{kj} = \sum_{k \in S} P_{ik}^n P_{kj}, \end{aligned}$$

where the third equality uses the Markov property and the final equality follows from our inductive hypothesis. □

A slight generalization of the above computation yields a very useful set of equations, called the *Chapman-Kolmogorov* equation. For all $i, j \in S$ and all non-negative integers $m, n \geq 0$,

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}. \tag{1.11}$$

The argument goes as follows,

$$\begin{aligned} p_{ij}^{(m+n)} &= P(X_{m+n} = j \mid X_0 = i) = \sum_{k \in S} P(X_{m+n} = j, X_m = k \mid X_0 = i) \\ &= \sum_{k \in S} P(X_{m+n} = j \mid X_m = k) P(X_m = k \mid X_0 = i) = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}. \end{aligned}$$

The Chapman-Kolmogorov equation (1.11) has an intuitive interpretation: the chain *must* be somewhere after m steps, and we are simply summing over the associated probabilities. Note that in the case of a finite state space, Proposition 1.24 shows that the Chapman-Kolmogorov equations are simply a reproduction of the matrix identity

$$P^{m+n} = P^m P^n.$$

We may now answer our original question pertaining to the probability that the Markov chain X_n with initial distribution α and transition matrix P will be in state $i \in S$ at time $n \geq 0$. Writing P_α for the probabilities associated with a chain with initial distribution α, we have

$$\begin{aligned} P_\alpha(X_n = i) &= \sum_{k \in S} P_\alpha(X_n = i, X_0 = k) = \sum_{k \in S} P_\alpha(X_n = i \mid X_0 = k) P_\alpha(X_0 = k) \\ &= \sum_{k \in S} \alpha_k P_{ki}^n = (\alpha P^n)_i, \end{aligned} \tag{1.12}$$

where α is interpreted as a row vector in the last equality. Thus, the probability vector αP^n gives the probabilities associated at time n, and calculating higher order transition probabilities is computationally equivalent to computing powers of the transition matrix.

Example 1.25. Consider again Example 1.13 pertaining to the gene that can be repressed. Suppose that $p = \frac{1}{3}$ and $q = \frac{1}{8}$ and that the gene is unbound at time 0, in which case

$$\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Suppose we want to know the probability that the gene is unbound at time $n = 4$, $P(X_4 = 1|X_0 = 1) = P^4_{11}$. Since

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{8} & \frac{7}{8} \end{pmatrix}, \quad \text{we have} \quad P^4 = \begin{pmatrix} \frac{13907}{41472} & \frac{27565}{41472} \\ \frac{27565}{110592} & \frac{83027}{110592} \end{pmatrix} \tag{1.13}$$

and

$$\alpha P^4 = \left(\tfrac{13907}{41472}, \tfrac{27565}{41472}\right).$$

Thus, the desired probability is $\frac{13907}{41472} \approx 0.33533$. △

Two state Markov chains can be handled in a general fashion.

Example 1.26 (Two-state Markov chain)**.** Consider a Markov chain with state space $S = \{1, 2\}$ and transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}. \tag{1.14}$$

Since $P^{n+1} = P^n P$ we have

$$p^{(n+1)}_{11} = p^{(n)}_{11}(1-p) + (1-p^{(n)}_{11})q = q + p^{(n)}_{11}(1-q-p),$$

where we utilized that $p^{(n)}_{12} = 1 - p^{(n)}_{11}$. Together with the initial condition $p^{(0)}_{11} = 1$, this recurrence relation can be solved (see Exercise 1.1 and Appendix A.1) to yield

$$p^{(n)}_{11} = \begin{cases} \frac{q}{q+p} + \frac{p}{q+p}(1-q-p)^n & \text{if } p+q > 0 \\ 1 & \text{if } p+q = 0, \end{cases}$$

which also yields $p^{(n)}_{12} = \frac{p}{q+p} - \frac{p}{q+p}(1-q-p)^n$ so long as $p > 0$. We may similarly solve for the second row of $P^{(n)}$ yielding

$$P^{(n)} = \begin{pmatrix} \frac{q}{q+p} + \frac{p}{q+p}(1-q-p)^n & \frac{p}{q+p} - \frac{p}{q+p}(1-q-p)^n \\ \frac{q}{q+p} - \frac{q}{q+p}(1-q-p)^n & \frac{p}{q+p} + \frac{q}{q+p}(1-q-p)^n \end{pmatrix}, \tag{1.15}$$

so long as $p + q > 0$. △

Example 1.27. Consider a Markov chain with state space $S = \{1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{4} & \frac{5}{12} \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{2}{7} & \frac{1}{7} & \frac{4}{7} \end{pmatrix}. \tag{1.16}$$

Then, utilizing a computer,

$$P^5 \approx \begin{pmatrix} 0.285 & 0.417 & 0.298 \\ 0.281 & 0.481 & 0.238 \\ 0.286 & 0.402 & 0.312 \end{pmatrix},$$

and, for example,

$$P(X_5 = 3 | X_0 = 1) = P^5_{13} \approx 0.298.$$

If $\alpha = [\frac{9}{10}, \frac{1}{10}, 0]$, then

$$\alpha P^5 \approx (0.285, \quad 0.423, \quad 0.292),$$

and

$$P_\alpha(X_5 = 3) = 0.292.$$

△

A natural question, and the focus of Section 1.7, is the following: for large n, what are the values $P_\alpha(X_n = i)$, where $i \in S$. By Proposition 1.24, we see that this question, at least in the case of a finite state space, can be understood simply through matrix multiplication.

For example, suppose that X_n is a two-state Markov chain with transition matrix given in (1.13) above,

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{8} & \frac{7}{8} \end{pmatrix}.$$

It is easy to check with a computer, or linear algebra, that for very large n,

$$P^n \approx \begin{pmatrix} \frac{3}{11} & \frac{8}{11} \\ \frac{3}{11} & \frac{8}{11} \end{pmatrix} \stackrel{\text{def}}{=} \Pi.$$

Note that the rows of Π are identical and equal to $\pi = \left(\frac{3}{11}, \frac{8}{11}\right)$. Therefore, if α is an initial distribution, we see

$$\lim_{n\to\infty} \alpha P^n = \alpha \Pi = \pi,$$

and for this example we may conclude that

$$\lim_{n\to\infty} P(X_n = 1) = \tfrac{3}{11}, \quad \text{and} \quad \lim_{n\to\infty} P(X_n = 2) = \tfrac{8}{11},$$

no matter the initial distribution.

More generally, we see from (1.15) that for a two-state Markov chain with transition matrix (1.14)

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{q}{q+p} & \frac{p}{q+p} \\ \frac{q}{q+p} & \frac{p}{q+p} \end{pmatrix},$$

so long as $q + p > 0$. Thus, for any initial distribution α, the probability distribution of the chain limits to $\pi = \left(\frac{q}{q+p}, \frac{p}{q+p}\right)$.

Such a vector π will eventually be termed a stationary, or invariant, distribution of the process, and is often of great interest to anyone wishing to understand the underlying model. Natural questions now include: does every discrete time Markov chain have such a stationary distribution? If so, is it unique? Can we quantify how long it takes to converge to a stationary distribution? To answer these questions[2] we need more terminology and mathematical machinery that will be developed in the next section. We will return to these questions in Section 1.7.

1.5 Reducibility, periodicity, and recurrence

We introduce the concepts of reducibility, periodicity, and recurrence. These notions play an important role throughout our study of Markov chains, both in discrete and continuous time.

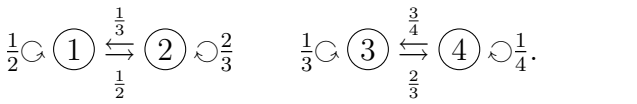
1.5.1 Reducibility

While Definition 1.30 below gives the formal definition of when a chain is reducible or irreducible, the concept is quite intuitive and can be stated as follows: a Markov chain is irreducible if for any states i and j, there is a sequence of possible transitions (which depend upon i and j) that take you from i to j. Otherwise, the chain is said to be reducible.

For example, suppose that X_n is a Markov chain with state space $S = \{1, 2, 3, 4\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}. \tag{1.17}$$

The state space and transitions probabilities can be represented via



If the chain starts in either state 1 or 2, then it will remain in $\{1, 2\}$ for all time, whereas if the chain starts in state 3 or 4, it will remain in $\{3, 4\}$ for all time. It seems natural to study this chain by analyzing separately the reduced chains consisting of state spaces $S_1 = \{1, 2\}$ and $S_2 = \{3, 4\}$, and transition matrices

$$P_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \qquad P_2 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix},$$

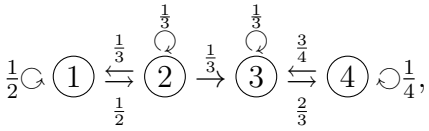
respectfully. This is an example of a reducible Markov chain.

[2]The answers are: no, sometimes, yes.

If instead of (1.17) the transition matrix for the chain is

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \tag{1.18}$$

which can be visualized via

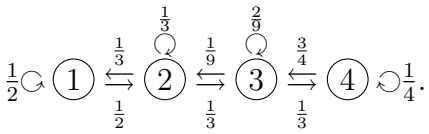


then it should be intuitively clear that even if $X_0 \in \{1, 2\}$, the chain will eventually transition into the set of states $\{3, 4\}$. However, there is no way for the chain to transition from the set of states $\{3, 4\}$ to the set of state $\{1, 2\}$. Thus, this is also an example of a reducible Markov chain.

Finally, suppose the transition matrix for the chain is

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \tag{1.19}$$

which can be visualized via



There is way to get from every state to every other state (possibly through a sequence of steps, and not just one), and so the chain is irreducible.

The following definitions will allow us to be precise.

Definition 1.28. The state $j \in S$ is *accessible* from the state $i \in S$, and we write $i \to j$, if there is an $n \geq 0$ for which

$$p_{ij}^{(n)} > 0.$$

Thus, state j is accessible from state i if there is a positive probability of the chain ever hitting state j if it starts in state i. Note that we allow $n = 0$ in the definition above, so that state i is always accessible from itself.

For example, for the chain with transition matrix (1.17) we have the relations $1 \to 2$, $2 \to 1$, $3 \to 4$, and $4 \to 3$, together with all the relations $i \to i$. However, for the chain with transition matrix (1.18), we have all the relations $i \to i$ and

- $1 \to 2$, $1 \to 3$, $1 \to 4$,
- $2 \to 1$, $2 \to 3$, $2 \to 4$,
- $3 \to 4$,
- $4 \to 3$,

which can be seen from

$$P^4 = \begin{pmatrix} \frac{19}{72} & \frac{5}{18} & \frac{5}{18} & \frac{13}{72} \\ \frac{10}{27} & \frac{97}{216} & \frac{1}{8} & \frac{1}{18} \\ 0 & 0 & \frac{107}{216} & \frac{109}{216} \\ 0 & 0 & \frac{109}{192} & \frac{83}{192} \end{pmatrix},$$

combined with the fact that the bottom left 2×2 sub-matrix of P^n will always consist entirely of zeros. For the chain with transition matrix (1.19), we have $i \to j$ for all pairs of states i, j.

Definition 1.29. States $i, j \in S$ of a Markov chain *communicate* with each other, and we write $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

It is straightforward to verify that the relation $\leftrightarrow$ is

1. Reflexive: $i \leftrightarrow i$.
2. Symmetric: $i \leftrightarrow j$ implies $j \leftrightarrow i$.
3. Transitive: $i \leftrightarrow j$ and $j \leftrightarrow k$ implies $i \leftrightarrow k$.

The first two conditions follow directly from the definitions above. In order to verify the third condition, we utilize the Chapman-Kolmogorov equations (1.11).

Verification of transitivity of $\leftrightarrow$:
We will show that if $i \to j$ and $j \to k$, then $i \to k$.
Since $i \to j$, there is an $n \geq 0$ for which $p_{ij}^{(n)} > 0$. Since $j \to k$, there is an $m \geq 0$ for which $p_{jk}^{(m)} > 0$. Therefore, by the Chapman-Klomogorov equations (1.11)

$$p_{ik}^{(n+m)} = \sum_{\ell \in S} p_{i\ell}^{(n)} p_{\ell k}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0,$$

and $i \to k$. □

Because $\leftrightarrow$ is reflexive, symmetric, and transitive, it is an equivalence relation and we may decompose the state space using the relation $\leftrightarrow$ into disjoint equivalence classes called *communication classes*. For example, the Markov chain with transition matrix (1.17) has two communication classes: $\{1, 2\}$ and $\{3, 4\}$. The Markov chain with transition matrix (1.18) has the same communication classes: $\{1, 2\}$ and $\{3, 4\}$.

The Markov chain with transition matrix (1.19) has the single communication class $\{1, 2, 3, 4\}$.

For the deterministically monotone chain of Example 1.11, each singleton $\{i\}$, $i \geq 0$, is its own communication class (remember that $n = 0$ is allowed in Definition 1.28).

For the symmetric random walk of Example 1.14 with absorbing boundaries (the Gambler's Ruin problem) the communication classes are $\{0\}$, $\{N\}$, and $\{1, \ldots, N-1\}$, whereas for the symmetric random walk with reflecting boundaries the only communication class is the entire state space $\{0, \ldots, N\}$.

For the random walk on the integer lattice $\mathbb{Z}^d$ described in Example 1.17, the only communication class is all of $\mathbb{Z}^d$.

Definition 1.30. A discrete time Markov chain is *irreducible* if there is only one communication class. That is, if $i \leftrightarrow j$ for all $i, j \in S$. Otherwise, the chain is called *reducible*.

Having the notion of irreducibility in hand, we turn to a related concept, that of a *closed* subset of the state space. Consider again the Markov chains with transition matrices (1.17) and (1.18), each of which have communication classes $\{1, 2\}$ and $\{3, 4\}$. For both chains, the set of states $\{1, 2\}$ is a communication class. However, it should be clear that the behavior of the chains on $\{1, 2\}$ are quite different as the chain with transition matrix (1.18) will eventually leave those states (assuming it starts there), never to return.

Definition 1.31. A subset of the state space $C \subset S$, is said to be *closed* if $p_{ij} = 0$ when $i \in C$ and $j \notin C$. We say that the state j is *absorbing* if $\{j\}$ is closed.

Thus $C \subset S$ is closed if it is impossible to reach any state outside of C from any state inside C via one-step transitions. Note that j is an absorbing state if and only if $p_{jj} = 1$.

The set $\{1, 2\}$ is closed for the chain with transition matrix (1.17), whereas it is not for the chain with transition matrix (1.18). The set $\{3, 4\}$ is closed for both chains.

For the deterministically monotone system of Example 1.11, the subset $\{n, n + 1, n + 2, \ldots\}$ is closed for any $n \geq 0$.

For the Gambler's ruin problem in Example 1.14, which is a random walk on $\{0, \ldots, N\}$ with absorbing boundary conditions, only $\{0\}$ and $\{N\}$ are closed, and hence the states 0 and N are absorbing.

It is important to note that if $C \subset S$ is closed, then the matrix with elements p_{ij} for $i, j \in C$ (i.e. the transition matrix restricted to C) is also a stochastic matrix. This fact follows because for any $i \in C$,

$$\sum_{j \in C} p_{ij} = 1, \quad \text{and} \quad \sum_{j \in C^c} p_{ij} = 0.$$

Therefore, if we restrict our attention to any closed subset of the state space, we can treat the resulting model as a discrete time Markov chain itself.

1.5.2 Periodicity

As a canonical example of a periodic Markov chain, consider a random walker with state space $S = \{0, 1, 2, 3, 4\}$ and reflecting boundary conditions: $p_{01} = p_{43} = 1$. Suppose that the chain starts in an even state. Then the chain must be in an even state for every even n and an odd state for every odd n. We will see below that this is a chain with a period of two.

For another example, consider the Markov chain on $\{0, 1, 2\}$ with

$$p_{01} = p_{12} = p_{20} = 1.$$

Thus, the chain deterministically moves from state 0 to state 1, then to state 2, then back to 0, etc. Here, if the chain starts in state i, it can (and will) only return to state i at times that are multiples of 3. This chain will be seen to have a period of three.

On the other hand, consider the random walk on $S = \{0, 1, 2, 3, 4\}$ with boundary conditions

$$p_{00} = \frac{1}{2}, \quad p_{01} = \frac{1}{2}, \quad \text{and} \quad p_{43} = 1.$$

In this case, if the chain starts at state 0, there is no condition similar to those above on the times that the chain can return to state 0. This chain will be said to be aperiodic.

With the intuition of these examples in hand, we turn to formal definitions in order to make these concepts precise.

Definition 1.32. The *period* of state $i \in S$ is

$$d(i) = \gcd\{n \geq 1 \ : \ p_{ii}^{(n)} > 0\},$$

where gcd stands for greatest common divisor. If $\{n \geq 1 \ : \ p_{ii}^{(n)} > 0\} = \emptyset$,[3] we take $d(i) = 1$. If $d(i) = 1$, we say that i is *aperiodic*, and if $d(i) > 1$, we say that i is *periodic* with a period of $d(i)$.

The following is often useful.

Proposition 1.33. *If $p_{ii} > 0$, then the state $i \in S$ is aperiodic.*

Proof. If $p_{ii} > 0$, then for all $n \geq 1$ we have $p_{ii}^{(n)} \geq (p_{ii})^n > 0$, and the greatest common divisor is 1. □

The proof of the following theorem can be found in either [28, Chapter 1] or [31, Chapter 2].

Theorem 1.34. *Let $X_n, n \geq 0$, be a Markov chain with state space S. If $i, j \in S$ are in the same communication class, then $d(i) = d(j)$. That is, they have the same period.*

[3] This happens, for example, for the deterministically monotone chain of Example 1.11.

Therefore, we may speak of the period of a communication class, and if the chain is irreducible, we may speak of the period of the Markov chain itself. Any property which necessarily holds for all states in a communication class is called a *class property*. Periodicity is, therefore, the first class property we have seen, though recurrence and transience, which are discussed in the next section, are also important class properties.

Periodicity is often obvious when powers of the transition matrix are taken.

Example 1.35. Consider a random walk on $\{0,1,2,3\}$ with reflecting boundary conditions. This chain is periodic with a period of two. Further, we have

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

and for any $n \geq 1$,

$$P^{2n} = \begin{pmatrix} * & 0 & * & 0 \\ 0 & * & 0 & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \end{pmatrix}, \quad \text{and} \quad P^{2n+1} = \begin{pmatrix} 0 & * & 0 & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \\ * & 0 & * & 0 \end{pmatrix},$$

where $*$ is a generic placeholder for a positive number. △

Example 1.36. Consider the random walk on $S = \{0,1,2,3,4\}$ with boundary conditions

$$p_{00} = \frac{1}{2}, \quad p_{01} = \frac{1}{2}, \quad \text{and} \quad p_{43} = 1.$$

The transition matrix is

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and

$$P^8 = \begin{pmatrix} \frac{71}{256} & \frac{57}{256} & \frac{1}{4} & \frac{9}{64} & \frac{7}{64} \\ \frac{57}{256} & \frac{39}{128} & \frac{29}{256} & \frac{21}{64} & \frac{1}{32} \\ \frac{1}{4} & \frac{29}{256} & \frac{49}{128} & \frac{9}{256} & \frac{7}{32} \\ \frac{9}{64} & \frac{21}{64} & \frac{9}{256} & \frac{63}{128} & \frac{1}{256} \\ \frac{7}{32} & \frac{1}{16} & \frac{7}{16} & \frac{1}{128} & \frac{35}{128} \end{pmatrix},$$

showing that $d(i) = 1$ for each $i \in S$. △

In the previous example, we used the basic fact that if each element of P^n is positive for some $n \geq 1$, then P^{n+k} has strictly positive elements for all $k \geq 0$. This follows because (i) each element of P is nonnegative, (ii) the rows of P sum to one, and (iii) $P^{n+k} = PP^{n+k-1}$.

1.5.3 Recurrence and transience

A state $i \in S$ of a Markov chain is called recurrent if, with a probability of one, after every visit to state i the chain will eventually return for another visit to state i. Otherwise, we will call the state transient.

More formally, we begin by recalling the definition for the first return time to state i (Definition 1.9):

$$\tau_i \stackrel{\text{def}}{=} \min\{n \geq 1 \ : \ X_n = i\}.$$

Also, for $i \in S$, we define the probability measure P_i by

$$P_i(A) \stackrel{\text{def}}{=} P(A|X_0 = i),$$

where A is any event of interest. For example, we would write

$$P_i(X_2 = i_2, X_1 = i_1) = P(X_2 = i_2, X_1 = i_1 \mid X_0 = i).$$

Note that we have changed notation slightly in that we should write P_{e_i} for the above probability measure, where e_i is the vector with a one in the ith component and zeros elsewhere. We let $\mathbb{E}_i$ be the expected value associated with the probability measure P_i.

Definition 1.37. The state $i \in S$ is *recurrent* if

$$P_i(\tau_i < \infty) = 1,$$

and *transient* if $P_i(\tau_i < \infty) < 1$. Note that an equivalent formulation for transience is that $P_i(\tau_i = \infty) > 0$.

To begin our study of the difference between a recurrent and transient state we let

$$R = \sum_{n=0}^{\infty} 1_{\{X_n = i\}}$$

denote the random variable giving the total number of times the chain visits state i. Computing the expectation of R under the assumption that $X_0 = i$, we see that

$$\mathbb{E}_i[R] = \sum_{n=0}^{\infty} P_i(X_n = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)},$$

where we made use of the fact that the expectation of an indicator function is the probability of the event being indicated upon.

We begin by supposing that the chain is transient. We let

$$p \stackrel{\text{def}}{=} P_i(\tau_i < \infty) < 1.$$

Note that p is simply the probability that, given the chain starts in state i, the chain returns to state i at least once. Since R simply counts the number of times the chain visits state i (including time $n = 0$), the strong Markov property implies that the random variable R is geometric with parameter $1 - p > 0$. That is, for $k \geq 1$

$$P_i(R = k) = p^{k-1}(1 - p).$$

Therefore,

$$\mathbb{E}_i[R] = \frac{1}{1-p} < \infty. \tag{1.20}$$

Note that equation (1.20) also allows us to conclude that for a transient chain we have

$$P_i(R = \infty) = 0,$$

for otherwise we would have $\mathbb{E}_i[R] = \infty$. Hence, we may conclude that there is, with a probability of one, a *last* time the chain visits the site i. Similarly, if state i is recurrent, then $P_i(R = \infty) = 1$ and $\mathbb{E}_i[R] = \infty$. Combining the above yields the following.

Theorem 1.38. *A state $i \in S$ is transient if and only if the expected number of returns to that state is finite, which occurs if and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$. Further, if i is recurrent, then with a probability of one, X_n returns to i infinitely often, whereas if i is transient, there is a last time a visit occurs.*

The following theorem shows that recurrence, and hence transience, is a class property. Thus, when the chain is irreducible and each state is recurrent, we may say that the chain itself is recurrent.

Theorem 1.39. *Suppose that $i \leftrightarrow j$. Then state i is recurrent if and only if state j is recurrent.*

Proof. The following argument is the intuition needed to understand the result (which is also the basis of the proof): because state i is recurrent, we return to it an infinite number of times with a probability of one. We also know that there is an $n > 0$ for which $p_{ij}^{(n)} > 0$. Thus, every time we are in state i, which happens an infinite number of times, there is a positive probability that we get to state j in n steps. Thus, we must enter state j an infinite number of times. The formal proof, based upon this intuition, is below.

Suppose that state i is recurrent. We must show that state j is recurrent. Because $i \leftrightarrow j$, there are nonnegative integers n and m that satisfy $p_{ij}^{(n)}, p_{ji}^{(m)} > 0$. Let k be a non-negative integer. By the Chapman-Kolmogorov equations

$$p_{jj}^{(m+n+k)} \geq p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)},$$

which says that one way to get from state j to state j in precisely $m+n+k$ steps is to first go to i in m steps, then return to i in k steps, then transition to j in n steps. Using this observation, we have

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} \geq \sum_{k=0}^{\infty} p_{jj}^{(m+n+k)} \geq \sum_{k=0}^{\infty} p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)}$$
$$= p_{ji}^{(m)} p_{ij}^{(n)} \sum_{k=0}^{\infty} p_{ii}^{(k)}.$$

Because i is recurrent, Theorem 1.38 says that the sum $\sum_{k=0}^{\infty} p_{ii}^{(k)}$ is infinite. Thus, $\sum_{k=0}^{\infty} p_{jj}^{(k)}$ is infinite and, again by Theorem 1.38, state j is recurrent. □

Corollary 1.40. *Let i and j be transient states in the same communication class. Then, $\lim_{n\to\infty} p_{ij}^{(n)} = 0$.*

Proof. Suppose, in order to find a contradiction, that $\lim_{n\to\infty} p_{ij}^{(n)} \neq 0$. Then there is a sequence of integers $\{n_k\}_{k=1}^{\infty}$ and a $c > 0$ for which $p_{ij}^{(n_k)} \geq c$ for all k. By irreducibility, there is an m for which $p_{ji}^{(m)} > 0$. Since i is transient, we may apply Theorem 1.38 to conclude that $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$. Hence, we conclude

$$\infty > \sum_{n=0}^{\infty} p_{ii}^{(n)} \geq \sum_{k=1}^{\infty} p_{ii}^{(n_k+m)} \geq \sum_{k=1}^{\infty} p_{ij}^{(n_k)} p_{ji}^{(m)} \geq c \cdot p_{ji}^{(m)} \sum_{k=1}^{\infty} 1 = \infty,$$

a contradiction. Thus, we may conclude that $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. □

Note that Theorems 1.38 and 1.39 together guarantee the following:

Fact: All states of an irreducible, finite state space Markov chain are recurrent.

The above fact holds by the following logic: if the states were not recurrent, they are each transient. Hence, there is a last time, call it T_i, that a particular realization of the chain visits state i. Therefore, $\max_i\{T_i\}$ is the last time the realization visits *any* state, which can not be. Things are significantly less clear in the infinite state space setting as the next few examples demonstrate.

Example 1.41 (Example 1.16 continued)**.** Consider a one dimensional random walk on the integer lattice $S = \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ where for some $0 < p < 1$ we have

$$p_{i,i+1} = p, \quad p_{i,i-1} = q, \quad \text{with} \quad q \stackrel{\text{def}}{=} 1-p, \quad \text{for } i \in \mathbb{Z}.$$

This chain is irreducible and has a period of 2. We will show that this chain is recurrent if $p = \frac{1}{2}$ and transient if $p \neq \frac{1}{2}$. To do so, we will use Theorem 1.38 to determine wether state zero is recurrent or transient. We will then use Theorem 1.39 to extend the result to the entire state space.

Begin by noticing that the periodicity of the system implies $p_{00}^{(2n+1)} = 0$, for all $n \geq 0$. Therefore,

$$\sum_{n=0}^{\infty} p_{00}^{(n)} = \sum_{n=0}^{\infty} p_{00}^{(2n)}.$$

Thus we restrict our attention to the chain at even times.

Next, notice that if for some $n \geq 1$ we have that $X_0 = X_{2n} = 0$, then within the first $2n$ steps of the chain, it must have moved to the right exactly n times and to the left exactly n times. Each such sequence of steps has a probability of $p^n q^n$. Because there are exactly $\binom{2n}{n}$ such paths, we see

$$p_{00}^{(2n)} = \binom{2n}{n}(pq)^n = \frac{(2n)!}{n!n!}(pq)^n.$$

Thus, we may conclude that

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} = \sum_{n=0}^{\infty} \frac{(2n)!}{n!n!}(pq)^n.$$

Recall that Stirling's formula states that for $m \gg 1$,

$$m! \sim m^m e^{-m}\sqrt{2\pi m},$$

where by $f(m) \sim g(m)$ we mean

$$\lim_{m\to\infty} \frac{f(m)}{g(m)} = 1.$$

Verification of Stirling's formula can be found in a number of places, for example in [13]. Stirling's formula yields

$$p_{00}^{(2n)} = \frac{(2n)!}{n!n!}(pq)^n \sim \frac{\sqrt{4\pi n}(2n)^{2n}e^{-2n}}{2\,\pi\, n\, n^{2n}\, e^{-2n}}(pq)^n = \frac{1}{\sqrt{\pi n}}(4pq)^n.$$

Therefore, we may conclude that there is an $N > 0$ such that $n \geq N$ implies

$$\frac{1}{2\sqrt{\pi n}}(4pq)^n < p_{00}^{(2n)} < \frac{2}{\sqrt{\pi n}}(4pq)^n.$$

The function $4pq = 4p(1-p)$ is strictly less than one for all $p \in [0,1]$ with $p \neq 1/2$. However, when $p = \frac{1}{2}$, we have that $4p(1-p) = 1$. Thus, in the case $p = \frac{1}{2}$ we have

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} > \sum_{n=N}^{\infty} p_{00}^{(2n)} > \sum_{n=N}^{\infty} \frac{1}{2\sqrt{\pi n}} = \infty,$$

and the chain is recurrent by Theorem 1.38. When $p \neq 1/2$, let $\rho = 4pq < 1$. We have

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} < N + \sum_{n=N}^{\infty} \frac{2}{\sqrt{\pi n}}\rho^n < \infty,$$

and the chain is transient by Theorem 1.38. △

Example 1.42 (Example 1.17 continued)**.** Consider symmetric random walk on the integer lattice $\mathbb{Z}^d$, which has transition probabilities

$$p_{ij} = \begin{cases} \frac{1}{2d} & \text{if } |i-j| = 1 \\ 0 & \text{else} \end{cases}.$$

We again consider starting the walk at the origin $\vec{0} = (0, 0, \ldots, 0)$. The chain has a period of 2, and so as in the one dimensional setting, $p_{\vec{0},\vec{0}}^{(2n+1)} = 0$ for all $n \geq 0$. Thus, to apply Theorem 1.38 we only need an expression for $p_{\vec{0},\vec{0}}^{(2n)}$. The derivation of an asymptotic formula for $p_{\vec{0},\vec{0}}^{(2n)}$ is beyond the scope of this book. However, it can be shown that

$$p_{\vec{0},\vec{0}}^{(2n)} \sim C_d \left(\frac{1}{n}\right)^{d/2},$$

for some $C_d > 0$, which depends upon the dimension d. Recalling that $\sum_{n=1}^{\infty} n^{-a} < \infty$ if and only if $a > 1$, we see that $\sum_{n=0}^{\infty} p_{\vec{0},\vec{0}}^{(2n)} = \infty$ if d is 1 or 2, and $\sum_{n=0}^{\infty} p_{\vec{0},\vec{0}}^{(2n)} < \infty$ if $d \geq 3$.

Thus, simple random walk in $\mathbb{Z}^d$ is recurrent if $d = 1$ or 2 and is transient if $d \geq 3$. This points out the general phenomenon that dynamics, in general, are quite different in dimensions greater than or equal to three than in dimensions one and two. Essentially, a path restricted to a line or a plane is much more restricted than one in space.[4] △

The following result should, at this point, be intuitive.

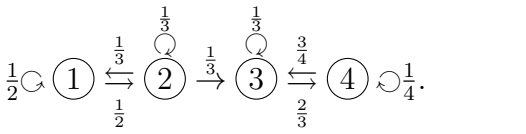
Theorem 1.43. *Every recurrent class of a Markov chain is a closed set.*

Proof. Suppose, in order to find a contradiction, that C is a recurrent class that is not closed. Then, there exists $i \in C$ and $j \notin C$ such that $p_{ij} > 0$, but it is impossible to return to state i (otherwise, $i \leftrightarrow j$). Therefore, the probability of starting in i and never returning is at least $p_{ij} > 0$, a contradiction with the class, and hence state i, being recurrent. □

Example 1.44. Consider the Markov chain with state space $\{1, 2, 3, 4\}$ and transition matrix (1.18)

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix},$$

which can be visualized via



[4] The video game "Tron" points this out well. Imagine how the game would play in three dimensions.

The communication class $\{1, 2\}$ is not closed and so by Theorem 1.43 the states 1 and 2 are transient. △

Note that the converse of Theorem 1.43 is, in general, false. For example, for the deterministic monotone chain, each set $\{n, n+1, \dots\}$ is closed, though no state is recurrent.

1.6 Absorption probabilities

Let $X_n, n \geq 0$, be a Markov chain with a finite state space and let t be a transient state. Let $\{C_k\}$ be the recurrent communication classes of the chain. Conditioned on $X_0 = t$, what is the probability the chain is absorbed into the communication class C_k?

We begin with an example, and then find the general answer to this, and related, questions. The following example should serve as a cautionary tale to anyone who steps foot in a casino.

Example 1.45 (Gambler's ruin, Example 1.14, continued)**.** Consider a gambler who repeatedly plays a game. Suppose he bets \$1 per game, and that his probability of winning any round is $0 < p < 1$. Of course, we assume that the outcomes of successive rounds of the game are independent. Denote the wealth of the gambler at time $n \geq 0$ by X_n. Then X_n is a discrete time Markov chain on the integers with transition probabilities $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p$.

Let A and B be nonnegative integers and let $\tau = \inf\{n \geq 0 \ : \ X_n \in \{-B, A\}\}$. We are interested in computing $P_0(X_\tau = A) = P(X_\tau = A \mid X_0 = 0)$. Hence, we are computing the probability that the gambler goes home a winner.

The random variable τ is a stopping time (recall Definition 1.18) and X_τ denotes the value of the process at time $\tau \geq 0$. Note that by even writing X_τ we are implicitly assuming that $\tau < \infty$ with a probability of one. We can prove that $P_0(\tau < \infty)$ in the following manner. Consider the chain with the same transition probabilities except states A and $-B$ are absorbing. Then the states $\{-B+1, \dots, A-1\}$ are transient, and by Theorem 1.38, with a probability of one there is a last time they are visited. Hence, τ is finite with a probability of one.

Returning to the problem of computing $P_0(X_\tau = A)$, the method of argument employed to solve the problem, first step analysis, is quite useful and will be used repeatedly in these notes. For each $k \in \{-B, \dots, A\}$ define

$$g_k = P_k(X_\tau = A) = P(X_\tau = A | X_0 = k).$$

Note that $g_{-B} = 0$ and $g_A = 1$. Next, for $k \in \{-B+1, \dots, A-1\}$, the Markov

p	0.49	0.48	0.47	0.45	0.40
$P_0(X_\tau = 50)$	0.119	0.018	0.0025	4.4×10^{-5}	1.6×10^{-9}

Table 1.1: Probabilities associated with winning \$50 before losing \$50 for various values of $p < \frac{1}{2}$.

property as given in Definition 1.3, implies

$$\begin{aligned} g_k &= P(X_\tau = A, X_1 = k+1 | X_0 = k) + P(X_\tau = A, X_1 = k-1 | X_0 = k) \\ &= P(X_\tau = A | X_1 = k+1, X_0 = k) \cdot p_{k,k+1} \\ &\qquad + P(X_\tau = A | X_1 = k-1, X_0 = k) \cdot p_{k,k-1} \\ &= P(X_\tau = A | X_1 = k+1) \cdot p + P(X_\tau = A | X_1 = k-1) \cdot (1-p) \\ &= p \cdot g_{k+1} + (1-p) \cdot g_{k-1}. \end{aligned} \tag{1.21}$$

If $p = \frac{1}{2}$, these equations have a simple solution satisfying (see Section A.1),

$$g_k = c_1 + c_2 k, \quad g_{-B} = 0, \quad g_A = 1,$$

which yields $g_k = \frac{B}{A+B} + \frac{1}{A+B}k$. Thus, when $p = \frac{1}{2}$,

$$P_0(X_\tau = A) = g_0 = \tfrac{B}{A+B}.$$

When $p \neq \frac{1}{2}$, the general solution to the difference equation (1.21) is (see Section A.1),

$$g_k = c_1 + c_2 \left(\frac{1-p}{p} \right)^k,$$

where c_1 and c_2 need to be determined from the boundary conditions $g_{-B} = 0$ and $g_A = 1$. Solving yields

$$g_k = \frac{(\frac{1-p}{p})^{k+B} - 1}{(\frac{1-p}{p})^{A+B} - 1},$$

and, in particular,

$$P_0(X_\tau = A) = g_0 = \frac{(\frac{1-p}{p})^{B} - 1}{(\frac{1-p}{p})^{A+B} - 1}. \tag{1.22}$$

It is hard to gain intuition from (1.22), so we choose some concrete numbers. First suppose that $A = B = 50$. Since it is the usual case that $p < \frac{1}{2}$, Table 1.1 gives the associated probabilities for $p = 0.49, 0.48, 0.47, 0.45$, and 0.40. When $A = B = 100$, the numbers are even worse. See Table 1.2. △

We now turn to a more general setting and ask the following.

Question 1. If $X_0 = t_i$ is a transient state of a Markov chain with finite state space, and the recurrent classes are denoted $C_1, C_2, \ldots, C_\ell$, what is the probability that the chain eventually ends up in recurrent class C_k, for $k \in \{1, \ldots, \ell\}$?

p	0.49	0.48	0.47	0.45	0.40
$P_0(X_\tau = 100)$	0.018	0.00033	6.1×10^{-6}	1.9×10^{-9}	2.4×10^{-18}

Table 1.2: Probabilities associated with winning \$100 before losing \$100 for various values of $p < \frac{1}{2}$.

To begin, note that we can assume each recurrent class consists of a single point as we can just group the states of each recurrent class together. Therefore, we denote the recurrent classes as $r_1, r_2, \ldots, r_k$, with $p_{r_i,r_i} = 1$. We let $t_1, t_2, \ldots$ denote the transient states.

We now follow Example 1.45 and use a first step analysis. Let $\tau = \min\{n \geq 0 \ : X_n = r_k \text{ for some } k\}$ and note that by Theorem 1.38 the stopping time τ is finite with a probability of one. For any transient state t_i and recurrent class k, we define

$$g_k(t_i) \stackrel{\text{def}}{=} P(X_\tau = r_k \mid X_0 = t_i).$$

For recurrent states r_k, r_i we define

$$g_k(r_i) = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}.$$

For transient state t_i, a first step analysis yields

$$\begin{aligned} g_k(t_i) &= \sum_{j \in S} p_{t_i,j} \cdot P(X_\tau = r_k \mid X_1 = j) = \sum_{j \in S} p_{t_i,j} g_k(j) \\ &= \sum_{r_j} p_{t_i,r_j} g_k(r_j) + \sum_{t_j} p_{t_i,t_j} g_k(t_j) \\ &= p_{t_i,r_k} + \sum_{t_j} p_{t_i,t_j} g_k(t_j), \end{aligned} \tag{1.23}$$

where the first sum of (1.23) is over the recurrent states and the second sum is over the transient states.

Equation (1.23) looks daunting, but there is actually a nice way to solve it using linear algebra when the state space is finite. The key is to note that we may write the transition matrix as

$$P = \left(\begin{array}{c|c} I & 0 \\ \hline U & Q \end{array}\right),$$

where we put the recurrent (absorbing) states first, Q is the matrix giving the transition probabilities from transient states to transient states, and U is the matrix giving the transition probabilities from transient states to recurrent states. Note that by raising P to the power $n \geq 1$ shows that for transient states i, j, we have $p_{ij}^{(n)} = Q_{ij}^n$.

If G is the matrix whose i, k^{th} entry is $g_k(t_i)$, then (1.23) can be written in matrix form,

$$G = U + QG.$$

Assuming $(I-Q)^{-1}$ exists, the solution is

$$G = (I-Q)^{-1}U. \tag{1.24}$$

The following proposition shows that the eigenvalues of Q all have absolute value strictly less than one. Because of this, the equation $(I-Q)v = 0$ has no solutions and $I-Q$ is invertible. Further, we have

$$(I-Q)^{-1} = I + Q + Q^2 + \cdots, \tag{1.25}$$

where the second equality follows from the identity

$$(I + Q + Q^2 + \cdots)(I-Q) = I.$$

Proposition 1.46. *The eigenvalues of Q all have absolute values strictly less than one.*

Proof. Let i and j be transient states. By Corollary 1.40

$$Q^n_{ij} = p^{(n)}_{ij} \to 0, \text{ as } n \to \infty.$$

This implies the result: if there were a λ and v with $|\lambda| \geq 1$ and $vQ = \lambda v$, then $|vQ^n| = |\lambda|^n|v|$ would not converge to zero. □

Example 1.47 (Example 1.45 continued)**.** We look at a special case of the Gambler's ruin problem. Consider unbiased random walk on $\{0,1,2,3,4\}$ with absorbing boundaries. We order the states $S = \{0,4,1,2,3,\}$ and have

$$P = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{array}\right).$$

Then,

$$U = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad (I-Q)^{-1} = \begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} \\ 1 & 2 & 1 \\ \frac{1}{2} & 1 & \frac{3}{2} \end{pmatrix}, \quad \text{and} \quad (I-Q)^{-1}U = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

For example, starting at state 1, the probability that the walk is eventually absorbed at state 0 is $\frac{3}{4}$. Note that this result agrees with that of Example 1.45 by taking $B = 0$, $A = 4$, and $p = \frac{1}{2}$. △

The next logical question is the following.

Question 2. Given that $X_0 = i$ is a transient state, how many steps do we expect the chain to make before being absorbed by a recurrent class?

To answer this question, we let j be a transient state (which may or may not be state i) and define R_j to be the total number of visits to state j,

$$R_j = \sum_{n=0}^{\infty} 1_{\{X_n=j\}},$$

where we explicitly note that if the chain starts in state j, then we count that as one visit. Note $R_j < \infty$ with a probability of one by Theorem 1.38.

We have

$$\mathbb{E}_i[R_j] = \sum_{n=0}^{\infty} P(X_n = j \mid X_0 = i) = \sum_{n=0}^{\infty} p_{ij}^{(n)}.$$

Therefore, $\mathbb{E}_i[R_j]$ is the i, j^{th} entry of

$$I + P + P^2 + \cdots,$$

which, because both i and j are transient, is the same as the i, j^{th} entry of

$$I + Q + Q^2 + \cdots = (I - Q)^{-1},$$

where, as above, Q is the matrix containing the transition probabilities from transient states to transient states. Therefore, we conclude that the expected number of visits to the transient state j, given that the chain starts in state i, is $(I-Q)^{-1}_{ij}$. Moreover, the expected time until absorption into some recurrent class, τ, satisfies

$$\mathbb{E}_i[\tau] = \mathbb{E}_i\left[\sum_j R_j\right] = \sum_j (I - Q)^{-1}_{ij},$$

where the sum is over all the transient states.

We have shown the following.

Proposition 1.48. *Consider a Markov chain with finite state space. Let Q be the matrix containing the transition probabilities from transient states to transient states. Let i and j be transient states and let R_j the total number of times the chain visits state j. Then, $\mathbb{E}_i[R_j] = (I - Q)^{-1}_{ij}$.*

Moreover, if τ is the first time the chain enters a recurrent class, then $\mathbb{E}_i[\tau] = \sum_{t_j} (I - Q)^{-1}_{i,t_j}$, where the sum is over the transient states.

Example 1.49. Consider the the Markov chain with state space $\{1, 2, 3, 4\}$ and transition matrix given by (1.18),

$$P = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

Note that states 1 and 2 are transient. After reordering the elements of the state space as $\{3, 4, 1, 2\}$ the new transition matrix is

$$\begin{array}{c} 3 \\ 4 \\ 1 \\ 2 \end{array} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \tag{1.26}$$

and for this example

$$Q = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Thus,

$$(I - Q)^{-1} = \begin{pmatrix} \frac{2}{3} & -\frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix}^{-1} = \begin{pmatrix} 3 & 3 \\ \frac{3}{2} & 3 \end{pmatrix}.$$

We see that starting in state 1 the expected number of visits to state 2 before being absorbed to the recurrent states is equal to $(I - Q)^{-1}_{12} = 3$. Starting in state 2, the expected number of visits to state 1 is $(I - Q)^{-1}_{21} = \frac{3}{2}$.

Let $\tau = \min\{n \geq 0 \ : \ X_n \in \{3, 4\}\}$. Then, for example,

$$\mathbb{E}_2[\tau] = \mathbb{E}_2[R_1] + \mathbb{E}_2[R_2] = (I - Q)^{-1}_{21} + (I - Q)^{-1}_{22} = 4.5.$$

△

We tackle one final question, which is now quite simple to answer.

Question 3. For given states $i, j \in S$ of an irreducible chain with a finite state space, what is the expected number of steps needed to go from state i to state j?

We begin by reordering the state space so that j is the first element. Hence, the transition matrix is

$$P = \left(\begin{array}{c|c} p_{jj} & K \\ \hline U & Q \end{array}\right),$$

where the row vector K has the transition probabilities $p_{jk}, k \neq j$. Now simply note that the answer to the question is unchanged if we make j an absorbing state. Thus, we can consider the problem on the system with transition matrix

$$\tilde{P} = \left(\begin{array}{c|c} 1 & 0 \\ \hline U & Q \end{array}\right),$$

However, this is now exactly the same problem solved above and we see the answer is the ith component of the vector $(I - Q)^{-1}\vec{1}$, where $\vec{1}$ is the vector of all ones.

Example 1.50. Suppose that P is the transition matrix for a random walk on $\{0,\dots,5\}$ with reflecting boundary:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \end{array} \\ \left(\begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \\ \hline \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right) \end{array}.$$

What is the expected number of steps to reach state $j = 0$ if the chain starts at state $i \geq 1$?

In this case, we have

$$Q = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \text{and} \quad (I-Q)^{-1} = \begin{pmatrix} 2 & 2 & 2 & 2 & 1 \\ 2 & 4 & 4 & 4 & 2 \\ 2 & 4 & 6 & 6 & 3 \\ 2 & 4 & 6 & 8 & 4 \\ 2 & 4 & 6 & 8 & 5 \end{pmatrix}.$$

Thus,

$$(I-Q)^{-1}\vec{1} = (9,\ 16,\ 21,\ 24,\ 25)^T.$$

For example, the expected number of steps needed to get from state 3 to state 0 is 21. △

Example 1.51. Consider the Jukes-Cantor model of DNA mutation, which tracks the state of a nucleotide. This model has state space $S = \{1,2,3,4\}$ (or {A,G,C,T}), and will be studied more extensively in Section 2.1.3.

For now, it is sufficient to know that the transition matrix for this model is

$$P = \begin{pmatrix} 1-\rho & \rho/3 & \rho/3 & \rho/3 \\ \rho/3 & 1-\rho & \rho/3 & \rho/3 \\ \rho/3 & \rho/3 & 1-\rho & \rho/3 \\ \rho/3 & \rho/3 & \rho/3 & 1-\rho \end{pmatrix},$$

where $\rho < 1$ is some parameter. If at time zero the nucleotide is in state 1, how many steps do we expect to take place before it enters states 3 or 4?

Recalling that the different states are A, G, C, and T, we note that A (adenine) and G (guanine) are *purines* and that C (cytosine) and T (thymine) are *pyrimidines*. Thus, this question is asking for the expected time until a given purine converts to a pyrimidine.

We make $\{3,4\}$ absorbing states, reorder the state space as $\{3,4,1,2\}$, and find the new transition matrix

$$\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline \rho/3 & \rho/3 & 1-\rho & \rho/3 \\ \rho/3 & \rho/3 & \rho/3 & 1-\rho \end{array}\right),$$

with Q and $(I-Q)^{-1}$ given via

$$Q = \begin{pmatrix} 1-\rho & \rho/3 \\ \rho/3 & 1-\rho \end{pmatrix}, \quad \text{and} \quad (I-Q)^{-1} = \begin{pmatrix} \frac{9}{8\rho} & \frac{3}{8\rho} \\ \frac{3}{8\rho} & \frac{9}{8\rho} \end{pmatrix}.$$

Therefore, the expected number of transitions needed to go from state 1 (A) to states 3 or 4 (C or T) is

$$(I-Q)^{-1}_{11} + (I-Q)^{-1}_{12} = \frac{9}{8\rho} + \frac{3}{8\rho} = \frac{3}{2\rho}.$$

Note that this value goes to ∞ as $\rho \to 0$, which is reasonable. △

1.7 Stationary distributions and the long time behavior of Markov chains

In the previous section, we showed how to compute limiting probabilities for systems with absorbing states. As we will see in this section, it is often possible to find limiting probabilities when the state space is irreducible. These limiting probabilities are given by a stationary distribution.

Definition 1.52. Consider a Markov chain with transition matrix P. A non-negative vector π is said to be an *invariant measure* if for all $i \in S$

$$\pi_i = \sum_{j \in S} \pi_j p_{ji}. \tag{1.27}$$

Written in vector notation, the condition is

$$\pi P = \pi. \tag{1.28}$$

If π also satisfies $\sum_k \pi_k = 1$, then π is called a *stationary distribution* for the chain. *Equilibrium distribution* and *steady state distribution* are other commonly used terms for π.

Thus, a stationary distribution is a probability vector that is also a left eigenvector of the transition matrix with associated eigenvalue equal to one.

A stationary distribution can be interpreted as a fixed point for the distribution of the Markov chain: if the initial distribution of the chain is π, then the distribution at all times $n \geq 1$ is also π,

$$\pi P^n = \pi P P^{n-1} = \pi P^{n-1} = \cdots = \pi,$$

where we are using equation (1.12). Some questions immediately come to mind:

1. Under what conditions on a Markov chain will a stationary distribution exist?

2. When is a stationary distribution unique?

3. Under what conditions can we guarantee convergence to a unique stationary distribution? That is, when is there a stationary distribution π for which $\lim_{n\to\infty} \alpha P^n = \pi$ for all initial distributions α?

We note that we have already seen some examples in which all of the above questions where answered. Recall that in Section 1.4, we showed that if a two-state Markov chain has transition matrix

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{8} & \frac{7}{8} \end{pmatrix}, \tag{1.29}$$

then,

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{3}{11} & \frac{8}{11} \\ \frac{3}{11} & \frac{8}{11} \end{pmatrix} = \Pi.$$

The important point was that the rows of Π are identical and equal to $\pi = (\frac{3}{11}, \frac{8}{11})$. Thus, if α is an arbitrary initial distribution

$$\lim_{n\to\infty} \alpha P^n = \alpha\Pi = \pi,$$

and

$$\lim_{n\to\infty} P_\alpha(X_n = 1) = \tfrac{3}{11}, \quad \text{and} \quad \lim_{n\to\infty} P_\alpha(X_n = 2) = \tfrac{8}{11}.$$

It is straightforward to check that $(\frac{3}{11}, \frac{8}{11})$ is the unique left eigenvector of P with an eigenvalue of 1,

$$\left(\tfrac{3}{11}, \tfrac{8}{11}\right) \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{8} & \frac{7}{8} \end{pmatrix} = \left(\tfrac{3}{11}, \tfrac{8}{11}\right).$$

We have also studied the general two-state Markov chain.

Example 1.53 (Example 1.26 revisited)**.** If a two state Markov chain has transition matrix given by (1.14),

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

then $P^{(n)}$ is given by (1.15),

$$P^n = \begin{pmatrix} \frac{q}{q+p} + \frac{p}{q+p}(1-q-p)^n & \frac{p}{q+p} - \frac{p}{q+p}(1-q-p)^n \\ \frac{q}{q+p} - \frac{q}{q+p}(1-q-p)^n & \frac{p}{q+p} + \frac{q}{q+p}(1-q-p)^n \end{pmatrix},$$

and the limiting distribution is $\pi = (\frac{q}{p+q}, \frac{p}{p+q})$. It is straightforward to check that $\pi P = \pi$. △

Let us consider at least one more example before proceeding with theory.

Example 1.54. Suppose that X_n is a three state Markov chain with transition matrix

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{12} & \frac{5}{8} & \frac{7}{24} \\ 0 & \frac{1}{8} & \frac{7}{8} \end{pmatrix}. \tag{1.30}$$

Then, for large n

$$P^n \approx \begin{pmatrix} \frac{3}{43} & \frac{12}{43} & \frac{28}{43} \\ \frac{3}{43} & \frac{12}{43} & \frac{28}{43} \\ \frac{3}{43} & \frac{12}{43} & \frac{28}{43} \end{pmatrix} = \Pi,$$

where we again note that each row of Π is identical. Therefore, regardless of the initial distribution α we have

$$\lim_{n\to\infty} P_\alpha(X_n = 1) = \tfrac{3}{43}, \quad \lim_{n\to\infty} P_\alpha(X_n = 2) = \tfrac{12}{43}, \quad \text{and} \quad \lim_{n\to\infty} P_\alpha(X_n = 3) = \tfrac{28}{43}.$$

We again note that it is straightforward to check that $(\frac{3}{43}, \frac{12}{43}, \frac{28}{43})$ is the unique left eigenvalue of P with an eigenvalue of 1. △

We will answer the three questions posed above first in the finite state space setting, where many of the technical details can be reduced to linear algebra. We then extend all the results to the infinite state space setting.

1.7.1 Markov chains with a finite state space

Consider a discrete time Markov chain with finite state space S and transition matrix P. We wish to understand P^n for large n and, relatedly, to find conditions that guarantee a unique limiting stationary distribution exists. However, and in order to gain intuition on what can go wrong, we first provide a few examples showing when such a unique limiting stationary distribution *does not* exist.

Example 1.55. Consider simple random walk on $\{0, 1, 2\}$ with reflecting boundaries. In this case we have

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}.$$

Note that for all $n \geq 1$,

$$P^{2n} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \quad \text{and} \quad P^{2n-1} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}.$$

Therefore, P^n does not converge as $n \to \infty$ and there is no hope of finding a limiting distribution. On the other hand, it is easy to verify that the probability vector $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ is a unique stationary distribution for the chain. Thus, we see that stationary distributions *do not* have to be limiting distributions of the form $\lim_{n\to\infty} \alpha P^n$. △

Example 1.56. Consider simple random walk on $\{0, 1, 2, 3\}$ with absorbing boundaries. That is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This model has communication classes $\{0\}$, $\{3\}$, and $\{1, 2\}$, and is therefore reducible. Note that each of the vectors $\alpha_1 = (1, 0, 0, 0)$ and $\alpha_2 = (0, 0, 0, 1)$ is a stationary distribution for the process, and so is any vector of the form $\rho\alpha_1 + (1 - \rho)\alpha_2$, for $0 \leq \rho \leq 1$. Thus, there is not a unique stationary distribution. Further, for n large we have

$$P^n \approx \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and we also see that there is no unique limiting distribution. For example, we have $\lim_{n\to\infty} \alpha_1 P^n = (1, 0, 0, 0)$, $\lim_{n\to\infty} \alpha_2 P^n = (0, 0, 0, 1)$, and $\lim_{n\to\infty}(0, 0, 1, 0)P^n = (\frac{1}{3}, 0, 0, \frac{2}{3})$. △

The following theorem shows that so long as the Markov chain is irreducible and aperiodic, there is a stationary distribution and, moreover, the stationary distribution is the unique limiting distribution no matter the initial distribution.

Theorem 1.57. *Consider a Markov chain with finite state space S, and transition matrix P. Suppose that the Markov chain is (i) irreducible and (ii) aperiodic. Then, there is a unique stationary distribution π,*

$$\pi P = \pi,$$

for which $\pi_i > 0$ for each i. Further, $\lim_{n\to\infty} P^n$ is the matrix with rows identically equal to π. In particular, if α is any probability vector, then

$$\lim_{n\to\infty} \alpha P^n = \pi.$$

The remainder of this sub-section consists of verifying Theorem 1.57. We begin with the following proposition, which connects stationary distributions with limiting distributions in the finite state space setting.

Proposition 1.58. *Let P be the transition matrix for a Markov chain with finite state space S. Suppose that for each $i \in S$,*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j^{(i)}, \qquad \text{for all } j \in S,$$

Then $\pi^{(i)}$ is a stationary distribution for the chain. If $\pi^{(i)} = \pi$ for each $i \in S$, then π is the unique stationary distribution of the chain.

Note that proposition 1.58 shows that in order to prove Theorem 1.57, we must only demonstrate that an irreducible, aperiodic Markov chain with a finite state space has a unique limiting distribution.

Proof of Proposition 1.58. We begin by showing that $\pi^{(i)}$ is necessarily a probability vector:

$$\sum_{j\in S}\pi_j^{(i)} = \sum_{j\in S}\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty}\sum_{j\in S} p_{ij}^{(n)} = 1, \tag{1.31}$$

where the finiteness of the state space was used to change the limit with the summation.

Next, for each $j \in S$, we apply the Chapman-Kolmogorov equations (1.11) to conclude

$$\pi_j^{(i)} = \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty}\sum_{k\in S} p_{ik}^{(n-1)}p_{kj} = \sum_{k\in S}\left[\lim_{n\to\infty} p_{ik}^{(n-1)}\right]p_{kj} = \sum_{k\in S}\pi_k^{(i)}p_{kj},$$

where the finiteness of the state space was again used to interchange the limit with the summation. Thus, π is a stationary distribution for the process.

Now suppose that $\pi^{(i)} = \pi$ for all $i \in S$ and suppose that ν is a stationary distribution of the chain. We must show that $\pi = \nu$. This is now straightforward,

$$\nu = \lim_{n\to\infty} \nu P^n = \pi,$$

where the first equality follows from the fact that $\nu P^n = \nu$ for all n, and the second follows because π is assumed to be a unique limiting distribution. □

It remains to be shown that all irreducible, aperiodic Markov chains with a finite state space admit a unique limiting distribution. Before proving this fact, we attempt to gain intuition by reexamining some examples.

Example 1.59. Consider the Markov chain with transition matrix P given in (1.29):

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{8} & \frac{7}{8} \end{pmatrix}.$$

It is simple to compute the eigenvalues of P,

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = \tfrac{13}{24} < 1.$$

Let $\pi^{(1)}$ and $\pi^{(2)}$ denote the left eigenvectors associated with λ_1 and λ_2, respectively, and let α be an arbitrary probability vector. Because $\pi^{(1)}$ and $\pi^{(2)}$ are necessarily linearly independent, there are constants c_1 and c_2 for which $\alpha = c_1\pi^{(1)} + c_2\pi^{(2)}$. Hence,

$$\alpha P^n = (c_1\pi^{(1)}P^n + c_2\pi^{(2)}P^n) = c_1\pi_1 + c_2(\tfrac{13}{24})^n\pi_2 \to c_1\pi^{(1)}, \quad \text{as } n\to\infty.$$

Note that since αP^n is necessarily a probability vector, $c_1\pi^{(1)}$ is also a probability vector (as in the argument (1.31)). Taking $\pi = c_1\pi^{(1)}$ shows that π is the unique

limiting distribution for the chain. By Proposition 1.58, it must also be the unique stationary distribution.

Note that we have concluded that the properly normalized eigenvector associated with the eigenvalue of 1 is the unique limiting distribution of the chain without even calculating the vector, and without resorting to raising P to higher and higher powers. Of course calculating π as the left eigenvector with an eigenvalue of 1 is simple and yields $\pi = (\frac{3}{11}, \frac{8}{11})$, as already demonstrated. △

We will consider one more example before proceeding with the theory.

Example 1.60 (Example 1.54 continued)**.** Let

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{12} & \frac{5}{8} & \frac{7}{24} \\ 0 & \frac{1}{8} & \frac{7}{8} \end{pmatrix}$$

be the transition matrix for a three-state Markov chain. The eigenvalues of P are $\lambda_1 = 1$ and $\lambda_2, \lambda_3 = \frac{14 \pm \sqrt{14}}{24}$. Thus, $|\lambda_i| < 1$ for $i \in \{2, 3\}$, and $\lambda_1 = 1$ is again the dominant eigenvalue. Let $\pi^{(i)}$ be the eigenvector associated with λ_i for each of $i \in \{1, 2, 3\}$. Let α be a probability vector and let c_1, c_2, and c_3 be chosen so that $\alpha = c_1\pi^{(1)} + c_2\pi^{(2)} + c_3\pi^{(3)}$. Then, as in Example 1.59,

$$\begin{aligned} \alpha P^n &= (c_1\pi^{(1)} + c_2\pi^{(2)} + c_3\pi^{(3)})P^n \\ &= (c_1\pi^{(1)} + c_2(\tfrac{14+\sqrt{14}}{24})^n\pi^{(2)} + c_3(\tfrac{14-\sqrt{14}}{24})^n\pi^{(3)}) \to c_1\pi^{(1)}, \quad \text{as } n \to \infty. \end{aligned}$$

Thus, $\pi = c_1\pi^{(1)} = (\frac{3}{43}, \frac{12}{43}, \frac{28}{43})$ is shown to be the unique limiting distribution of the chain. △

The above examples shed light on a method of proof for Theorem 1.57. We will demonstrate that the transition matrix of an irreducible, aperiodic discrete time Markov chain has an eigenvalue of 1, and that all other eigenvalues satisfy $|\lambda_i| < 1$. The following proposition, which loosely says that there is a time N at which the chain can be found in any state, is instrumental in showing this.

Proposition 1.61. *Suppose that P is the transition matrix for an aperiodic, irreducible Markov chain with finite state space S. Then, there is an $N \geq 1$ for which P^N has strictly positive entries.*

Proof. We take the following fact for granted, which follows from a result in number theory: if the chain is aperiodic, then for each state $i \in S$, there is an $N(i)$ for which $p_{ii}^{(n)} > 0$ for all $n \geq N(i)$ (see Exercise 1.9).

To prove the proposition, we must demonstrate that there is an $N > 0$ so that if $n \geq N$, then P^n has strictly positive entries. Let $i, j \in S$. By the irreducibility of the chain, there is an $n(i, j)$ for which

$$p_{ij}^{(n(i,j))} > 0.$$

Thus, for all $n \geq N(i)$, the Chapman-Kolmogorov equations (1.11) allow us to conclude

$$p_{ij}^{(n+n(i,j))} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(n(i,j))} \geq p_{ii}^{(n)} p_{ij}^{(n(i,j))} > 0.$$

The proposition is therefore shown by letting N be the maximum over $N(i) + n(i,j)$, which exists since the state space is finite. □

We have all the building blocks necessary for the proof of Theorem 1.57.

Proof of Theorem 1.57. In light of Proposition 1.61, the Perron-Frobenius theorem applies to P^N, where N is given in the statement of the proposition. We may conclude the following:

1. 1 is a simple eigenvalue of P^N.
2. There are no other eigenvalues of P^N with a modulus of 1 (since P^N is aperiodic).
3. The left eigenvector of P^N with an eigenvalue of 1 can be chosen to have strictly positive components.
4. All other eigenvalues of P^N have a modulus less than 1. That is, $|\lambda_i| < 1$.

We may now draw some conclusions about the transition matrix P.

Since $P\vec{1} = \vec{1}$, where $\vec{1}$ is the vector with all ones, P has an eigenvalue of 1. Let π be the left eigenvector of P associated with this eigenvalue. Since $\pi P^N = \pi P P^{N-1} = \pi P^{N-1} = \cdots = \pi$, we see that π is the eigenvector of P^N with an eigenvalue of one, and therefore can be chosen to have strictly positive components. We choose π so that it is a probability vector.

If P has a full set of linearly independent eigenvectors, the proof now follows similarly to the examples above. We let $\pi^{(i)}$ and λ_i, for $i \in \{2, \ldots, d\}$, where d is the size of the state space, be the other eigenvalues and eigenvectors. Then, for any probability vector α, there are $c, c_2, \ldots, c_d \in \mathbb{R}$ for which

$$\alpha = c\pi + \sum_{i=2}^{d} c_i \pi^{(i)},$$

and, since $|\lambda_i| < 1$ for all $i \in \{2, \ldots, d\}$,

$$\lim_{n\to\infty} \alpha P^n = \lim_{n\to\infty} \left(c\pi + \sum_{i=2}^{d} c_i \pi^{(i)} \right) P^n = \lim_{n\to\infty} \left(c\pi + \sum_{i=2}^{d} c_i \lambda_i^n \pi^{(i)} \right) = c\pi.$$

Since π is a probability vector, we see that $c = 1$, and the proof is complete.

If P does not have a full set of linearly independent eigenvectors then slightly more care is required. In this case, we use a Jordan decomposition to conclude that

$$J = Q^{-1} P Q,$$

where J is in Jordan canonical form, Q^{-1} has a first row consisting of the probability vector π, and Q has a first column that contains all ones. Further, because all eigenvalues except 1 have modulus less than 1, we know that J is of the form

$$J = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{pmatrix},$$

where $A^n \to 0$, as $n \to \infty$. Thus,

$$\lim_{n\to\infty} \alpha P^n = \lim_{n\to\infty} \alpha Q J^n Q^{-1} = \alpha Q \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{0} & \\ 0 & & & \end{pmatrix} Q^{-1} = \alpha \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix} = \pi.$$

Thus, Theorem 1.57 is proven in general. □

We see that if we wish to understand the large time probabilities associated with an irreducible, aperiodic Markov chain with a finite state space, then it is sufficient to calculate the unique left eigenvector of the transition matrix with eigenvalue equal to one. Such computations can be carried out by hand for small examples, but are usually performed with software for larger systems.

Example 1.62. Consider a Markov chain with state space $\{0, 1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

Find $\lim_{n\to\infty} P_\alpha(X_n = 2)$, where α is an arbitrary initial distribution.

The graph associated with the Markov chain is strongly connected, and so the chain is irreducible. Moreover, $p_{11} = \frac{1}{4}$, and so the chain is aperiodic. Thus, Theorem 1.57 applies. The eigenvector of P (normalized to be a probability distribution) associated with the eigenvalue 1 is

$$\pi = \left(\tfrac{25}{67}, \tfrac{12}{67}, \tfrac{22}{67}, \tfrac{8}{67}\right).$$

Thus, regardless of α, $\lim_{n\to\infty} P_\alpha(X_n = 2) = \frac{12}{67}$. △

In the next subsection we will consider what changes to the theory when we drop the irreducibility, but not the aperiodicity, assumption. We will consider the periodic case when we turn to infinite state space Markov chains in Section 1.7.2.

Reducible chains with aperiodic, recurrent communication classes

In order to gain intuition, we begin with a number of examples.

Example 1.63 (Examples 1.14 and 1.45 continued)**.** Consider a random walk on $\{0, 1, \ldots, N\}$ with transition probabilities

$$p_{i,i+1} = p_{i,i-1} = p_{i,i} = \tfrac{1}{3}, \qquad \text{for } i \in \{1, \ldots, N-1\},$$

and absorbing boundary conditions: $p_{0,0} = p_{N,N} = 1$. For some $0 \leq \rho \leq 1$, let

$$\pi_\rho = (\rho, 0, \ldots, 0, 1-\rho).$$

It is straightforward to show that $\pi_\rho P = \pi_\rho$, where P is the transition matrix for the model, and that all stationary distributions are of the form π_ρ. Thus, there are uncountably many stationary distributions for this example. However, note that each such distribution is a linear combination of the two distributions $\pi^{(0)} = (1, 0, \ldots, 0)$ and $\pi^{(N)} = (0, \ldots, 0, 1)$. Observe that Example 1.56 is a special case of this example, with $N = 3$.

Let $\tau = \min\{n \geq 0 \ : \ X_n \in \{0, N\}\}$ and for each $i \in \{0, 1, \ldots, N\}$ let

$$g_0(i) = P(X_\tau = 0 \mid X_0 = i), \quad \text{and} \quad g_N(i) = P(X_\tau = N \mid X_0 = i).$$

We showed how to compute $g_0(i)$ and $g_N(i)$ in Section 1.6. We see that for $i \in \{0, 1, \ldots, N\}$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} g_0(i), & j = 0 \\ 1 - g_0(i), & j = N \\ 0, & j \in \{1, \ldots, N-1\} \end{cases}.$$

That is, the limiting probability vector is $\pi_{g_0(i)} = g_0(i)\pi^{(0)} + (1 - g_0(i))\pi^{(N)}$, which explicitly depends upon the initial state i. △

Example 1.64. Consider the Markov chain with transition matrix (1.17),

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix},$$

which can be visualized via

$$\tfrac{1}{2}\circlearrowleft \textcircled{1} \underset{\frac{1}{2}}{\overset{\frac{1}{3}}{\leftrightarrows}} \textcircled{2} \circlearrowright \tfrac{2}{3} \qquad\qquad \tfrac{1}{3}\circlearrowleft \textcircled{3} \underset{\frac{2}{3}}{\overset{\frac{3}{4}}{\leftrightarrows}} \textcircled{4} \circlearrowright \tfrac{1}{4}.$$

This chain is reducible, with communication classes $C_1 = \{1, 2\}$ and $C_2 = \{3, 4\}$. Viewed separately, these communication classes are irreducible, aperiodic Markov chains with transition matrices

$$P_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

The unique limiting stationary distribution for the Markov chain with state space C_1 and transition matrix P_1 is $\pi^{(1)} = (\frac{2}{5}, \frac{3}{5})$, whereas the unique limiting stationary distribution for the chain with state space C_2 and transition matrix P_2 is $\pi^{(2)} = (\frac{9}{17}, \frac{8}{17})$. Finally, for any $0 \leq \rho \leq 1$, a vector of the form

$$\pi_\rho = \rho\left(\tfrac{2}{5},\ \tfrac{3}{5},\ 0,\ 0\right) + (1-\rho)\left(0,\ 0,\ \tfrac{9}{17},\ \tfrac{8}{17}\right),$$

is a stationary distribution for the chain with transition matrix P. Moreover, these vectors account for all possible stationary distributions of the chain.

The limiting distributions again depend upon the initial state. For $i \in C_1$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} \pi_j^{(1)} & j \in C_1 \\ 0 & j \in C_2 \end{cases}$$

whereas for $i \in C_2$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} 0 & j \in C_1 \\ \pi_j^{(2)} & j \in C_2 \end{cases}.$$

In particular, for an initial distribution of α,

$$\lim_{n\to\infty} P_\alpha(X_n = j) = \lim_{n\to\infty} \sum_{i=1}^{4} P_\alpha(X_n = j, X_0 = i) = \sum_{i=1}^{4} \left[\lim_{n\to\infty} p_{ij}^{(n)}\right] \alpha_i$$

$$= \begin{cases} (\alpha_1+\alpha_2)\frac{2}{5} & j = 1 \\ (\alpha_1+\alpha_2)\frac{3}{5} & j = 2 \\ (\alpha_3+\alpha_4)\frac{9}{17} & j = 3 \\ (\alpha_3+\alpha_4)\frac{8}{17} & j = 4 \end{cases},$$

and the limiting distribution is

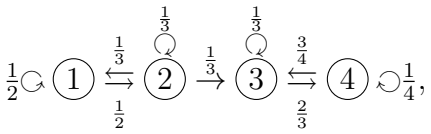
$$\pi_{\alpha_1+\alpha_2} = (\alpha_1+\alpha_2)[\tfrac{2}{5},\ \tfrac{3}{5},\ 0,\ 0] + (\alpha_3+\alpha_3)[0,\ 0,\ \tfrac{9}{17},\ \tfrac{8}{17}].$$

△

Example 1.65. Sometimes reducible Markov chains can have a unique, limiting stationary distribution. Consider the Markov chain with state space $S = \{1, 2, 3, 4\}$ and transition matrix (1.18),

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \tag{1.32}$$

which can be visualized via



The chain is reducible, with communication classes $T_1 = \{1, 2\}$ and $C_1 = \{3, 4\}$. By Theorem 1.43 the states 1 and 2 are transient. Viewed in isolation, the communication class C_1 has transition matrix

$$P_1 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix},$$

which has a unique stationary distribution of $\pi^{(1)} = (\frac{9}{17}, \frac{8}{17})$. The unique stationary distribution of the full chain with transition matrix P is

$$\pi = \left(0,\ 0,\ \tfrac{9}{17},\ \tfrac{8}{17}\right).$$

Finally, since the states 1 and 2 are transient and because there is a unique recurrent communication class, we see that

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} 0 & j = 1 \\ 0 & j = 2 \\ \frac{9}{17} & j = 3 \\ \frac{8}{17} & j = 4 \end{cases},$$

and π is the unique limiting stationary distribution. △

The examples above show how to proceed. We begin with some notation. We will consider a discrete time Markov chain, X_n, with finite state space S. We denote the recurrent communication classes by $C_1, \ldots, C_r$. We assume each recurrent communication class is aperiodic, in which case Theorem 1.57 guarantees that a unique limiting stationary distribution, $\pi^{(k)}$, exists for the kth recurrent class. Denote by $\bar{\pi}^{(k)}$ the vector of size $|S|$ for which

$$\bar{\pi}_j^{(k)} = \begin{cases} \pi_j^{(k)}, & \text{if } j \in C_k \\ 0, & \text{if } j \notin C_k \end{cases}.$$

Let $\tau = \min\{n \geq 0 \mid X_n \in C_k \text{ for some } k\}$. For each $i \in S$ we let

$$g_k(i) = P(X_\tau \in C_k \mid X_0 = i). \tag{1.33}$$

We saw how to calculate the $\{g_k(i)\}$ of (1.33) in Section 1.6.

We may conclude that for any $i, j \in S$,

$$\begin{aligned} \lim_{n\to\infty} p_{ij}^{(n)} &= \lim_{n\to\infty} P(X_n = j | X_0 = i) \\ &= \lim_{n\to\infty} \sum_{k=1}^{r} P(X_n = j, X_\tau \in C_k | X_0 = i) \\ &= \sum_{k=1}^{r} \left[\lim_{n\to\infty} P(X_n = j | X_\tau \in C_k, X_0 = i)\right] g_k(i) \\ &= \sum_{k=1}^{r} \left[\lim_{n\to\infty} P(X_n = j | X_0 \in C_k)\right] g_k(i) \\ &= \sum_{k=1}^{r} \bar{\pi}_j^{(k)} g_k(i), \end{aligned}$$

where we have applied Theorem 1.57 to each recurrent class. By Proposition 1.58 the vectors

$$\sum_{k=1}^{r} g_k(i)\bar{\pi}^{(k)}$$

are stationary distributions of the chain. See Exercise 1.10, where it is shown that all stationary distributions can be written in such a manner.

1.7.2 Markov chains with a countably infinite state space

We now extend the results of the previous section to discrete time Markov chains with countably infinite state spaces. We do not prove every result in this section, though we point to the relevant proofs in the literature. Finally, we note that every result stated in this section also holds in the finite state space case.

In order to gain some intuition, we begin with an example demonstrating a behavior that is not possible in the finite state space setting.

Example 1.66 (Example 1.16 continued)**.** Consider symmetric random walk on the integers. That is, suppose the state space of the chain is $S = \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ and the transition probabilities satisfy

$$p_{i,i+1} = p_{i,i-1} = \tfrac{1}{2}, \text{ for all integers } i \in \{\dots, -1, 0, 1, \dots\}.$$

From Example 1.41, we know that this chain is recurrent. Is there a stationary distribution?

Any invariant measure must satisfy two conditions: $\pi_j \geq 0$ and

$$\pi_j = \sum_{k=-\infty}^{\infty} \pi_k p_{kj} = \pi_{j-1}p_{j-1,j} + \pi_{j+1}p_{j+1,j} = \frac{1}{2}(\pi_{j-1} + \pi_{j+1}),$$

for all $j \in \mathbb{Z}$. These conditions can be satisfied by taking $\pi_j \equiv c$, where c is a nonnegative constant.

We will see in Theorem 1.72 that all irreducible, recurrent chains admit a unique (up to multiplication by a constant) invariant measure. Since the measure with $\pi_j \equiv c$ can not be normalized to give a probability vector, we conclude that no stationary distribution exists for this chain. △

The above example demonstrates that in the infinite state space setting, we need a stronger condition than recurrence to guarantee the existence of a stationary distribution. Recall the definition of τ_i in (1.9):

$$\tau_i = \min\{n \geq 1 \ : \ X_n = i\},$$

with $\tau_i = \infty$ if the chain never hits state i for $n \geq 1$. It is useful to subdivide the set of recurrent states further.

Definition 1.67. The value $\mu_i = \mathbb{E}_i[\tau_i]$ is called the *mean recurrence time* or *mean first return time* for state i. We say that a recurrent state i is *positive recurrent* if $\mathbb{E}_i[\tau_i] < \infty$, and otherwise say it is *null recurrent*.

Note that we trivially have $\mu_i = \infty$ if i is a transient state since in this case $P_i(\tau_i = \infty) > 0$.

The following is stated without proof. See [31, Chapter 3] for a proof utilizing renewal theory.

Theorem 1.68. *Consider a recurrent, irreducible, aperiodic Markov chain. For any $i, j \in S$*

$$\lim_{n\to\infty} p_{ji}^{(n)} = \frac{1}{\mu_i}.$$

If $\mu_i = \infty$, then we interpret the right hand side of the above equation as zero.

A similar theorem for periodic chains is the following.

Theorem 1.69. *Consider a recurrent, irreducible, d-periodic Markov chain. For any $i \in S$*

$$\lim_{n\to\infty} p_{ii}^{(nd)} = \frac{d}{\mu_i}.$$

If $\mu_i = \infty$, then we interpret the right hand side of the above equation as zero.

In Theorem 1.39, we showed that recurrence is a class property. The following shows that positive recurrence is also a class property.

Theorem 1.70. *Suppose that $i \leftrightarrow j$ and that state i is positive recurrent. Then state j is positive recurrent.*

Proof. We assume the chain is aperiodic. The proof in the periodic case is similar and is left as Exercise 1.15. Consider only the irreducible component of the state space containing states i and j. Theorem 1.68 implies

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j}, \tag{1.34}$$

By (1.34), we may conclude that $\mu_j < \infty$ if we demonstrate that $\lim_{n\to\infty} p_{ij}^{(n)} > 0$.

Since $i \leftrightarrow j$, there is an $m > 0$ for which $p_{ij}^{(m)} > 0$. Therefore,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} p_{ij}^{(n+m)} \geq \lim_{n\to\infty} p_{ii}^{(n)} p_{ij}^{(m)} = p_{ij}^{(m)} \lim_{n\to\infty} p_{ii}^{(n)} = p_{ij}^{(m)} \frac{1}{\mu_i} > 0,$$

where the final equality holds from Theorem 1.68 applied to state i. □

Therefore, we can speak of positive recurrent or null recurrent communication classes. Further, if the Markov chain is irreducible we may call the chain itself positive or null recurrent.

Example 1.71 (Example 1.16 continued)**.** Consider again the symmetric ($p = \frac{1}{2}$) random walk on the integer lattice. In Example 1.41, we showed that the chain is recurrent and that

$$p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}.$$

Therefore, $\lim_{n\to\infty} p_{00}^{(n)} = 0$ and by Theorem 1.69 we have that $\mu_0 = \infty$. Thus, the symmetric random walk on the integers is null recurrent. △

Following our results from the finite state space setting, it seems reasonable to suspect that for positive recurrent chains the limiting distribution provided in Theorem 1.68 is also a stationary distribution. We direct the reader to [31, Chapter 2.12] for a proof of the following theorem.

Theorem 1.72. *If a Markov chain is irreducible and recurrent, then there is an invariant measure π, unique up to multiplicative constants, that satisfies $0 < \pi_j < \infty$ for all $j \in S$. Further, if the Markov chain is positive recurrent then*

$$\pi_i = \frac{1}{\mu_i}, \tag{1.35}$$

where μ_i is the mean recurrence time of state i, $\sum_i \pi_i = 1$, and π is a stationary distribution of the Markov chain. If the Markov chain is also aperiodic, then $p_{ji}^{(n)} \to \pi_i$, as $n \to \infty$, for all $i, j \in S$.

Note that Theorem 1.72 guarantees the existence of a stationary distribution even if the chain is periodic. Note also that Theorem 1.72 provide an effective algorithm for computing mean return times: compute the invariant distribution using

$$\pi = \pi P,$$

and invert the component of interest.

Example 1.73. Consider reflecting random walk on $\{1, 2, 3, 4\}$. That is, the Markov chain with transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

This chain has period two, and for large n we have

$$P^{2n} \approx \begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{pmatrix}, \quad P^{2n+1} \approx \begin{pmatrix} 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \end{pmatrix}.$$

The unique stationary distribution of the chain guaranteed to exist by Theorem 1.72 is $\pi = (\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6})$. While π does not, in this case, give the long run probabilities of the associated chain, we will see in Theorem 1.78 a useful interpretation of π as giving the average amount of time spent in each state. △

A question still remains: can the invariant measure of a null recurrent chain ever be normalized to give a stationary distribution? The answer, given in the following theorem, is no. We refer to [29], Theorem 1.7.7, for a proof.

Theorem 1.74. *Suppose a Markov chain is irreducible and that a stationary distribution π exists:*

$$\pi = \pi P, \qquad \sum_{j\in S} \pi_j = 1, \quad \pi_j \geq 0.$$

Then, the Markov chain is positive recurrent.

Thus, a necessary and sufficient condition for determining positive recurrence is simply demonstrating the existence or non-existence of a stationary distribution.

Example 1.75 (Random walk with partially reflecting boundaries)**.** Consider a random walker on $S = \{0, 1, 2, \dots\}$ with transition probabilities

$$\begin{aligned} p_{j,j+1} &= p, \quad p_{j,j-1} = 1-p, \qquad \text{if } j \geq 1, \\ p_{01} &= p, \quad p_{00} = 1-p, \end{aligned}$$

where $0 < p < 1$. This Markov chain is irreducible and aperiodic (since $p_{00} > 0$). We want to determine when this model will have a limiting stationary distribution, and, hence, when it is positive recurrent.

A stationary distribution for this system must satisfy

$$\pi_j = \pi_{j+1} \cdot (1-p) + \pi_{j-1} \cdot p, \qquad \text{for } j > 0 \tag{1.36}$$

$$\pi_0 = \pi_1 \cdot (1-p) + \pi_0 \cdot (1-p), \tag{1.37}$$

with the condition that $\pi_j \geq 0$ and $\sum_{j=0}^{\infty} \pi_j = 1$. We may solve the difference equations (see Section A) and the general solution to equation (1.36) is

$$\pi_j = \begin{cases} c_1 + c_2 \left(\frac{p}{1-p}\right)^j, & p \neq \frac{1}{2} \\ c_1 + c_2 j, & p = \frac{1}{2} \end{cases}, \tag{1.38}$$

where c_1 and c_2 are unknown. Rearranging equation (1.37) yields

$$\pi_0 = \frac{1-p}{p} \cdot \pi_1. \tag{1.39}$$

Plugging (1.38) with $j = 0$ and $j = 1$ into (1.39) shows $c_1 = 0$ in the $p \neq \frac{1}{2}$ case, and that $c_2 = 0$ in the $p = \frac{1}{2}$ case. Therefore,

$$\pi_j = \begin{cases} c_2 \left(\frac{p}{1-p}\right)^j, & p \neq \frac{1}{2} \\ c_1, & p = \frac{1}{2} \end{cases}.$$

We need $\sum_{j=0}^{\infty} \pi_j = 1$ for π to be a stationary distribution. However, if $p = \frac{1}{2}$, no choice of c_1 could satisfy this condition. Thus, there is no stationary distribution in the case $p = \frac{1}{2}$.

Now just consider the case $p \neq \frac{1}{2}$. We obviously require that $c_2 > 0$. If $p > \frac{1}{2}$, then $p/(1-p) > 1$ and the sum

$$\sum_{j=0}^{\infty} c_2 \left(\frac{p}{1-p}\right)^j = \infty.$$

If, on the other hand, $p < \frac{1}{2}$, then

$$\sum_{j=0}^{\infty} c_2 \left(\frac{p}{1-p}\right)^j = c_2 \frac{1-p}{1-2p}.$$

Therefore, taking $c_2 = (1-2p)/(1-p)$ yields a stationary distribution of

$$\pi_j = \frac{1-2p}{1-p}\left(\frac{p}{1-p}\right)^j.$$

Thus, the chain is positive recurrent when $p < \frac{1}{2}$, which is hopefully intuitive. We also know that the chain is either null recurrent or transient if $p \geq \frac{1}{2}$. △

We want mathematical machinery that allows us to figure out when the chain of the previous example is either null recurrent or transient. The following theorem, which we will also make use of in Section 2.2, is helpful.

Theorem 1.76. *Let $X_n, n \geq 0$, be an irreducible Markov chain with state space S and let $i \in S$. Then X_n is transient if and only if there is a nonzero vector y satisfying the following: $y_i = 0$, $0 \leq y_j \leq 1$ for all j, and*

$$y_j = \sum_{k \neq i} p_{jk} y_k, \qquad \text{for } j \neq i. \tag{1.40}$$

Proof. Suppose first that X_n is transient. We will produce the vector y. Let $H_i = \inf\{n \geq 0 : X_n = i\}$ be the hitting time of state i, and let $y_j = P(H_i = \infty | X_0 = j)$ be the probability the chain never hits state i if it starts in state j. Clearly $y_i = 0$ and $0 \leq y_j \leq 1$ for all j. Also, since the chain is transient, $y_j > 0$ for some j (otherwise the chain would return to state i with a probability of one whenever it transitioned from state i). Using a first-step analysis, we see that for $j \neq i$,

$$\begin{aligned} y_j &= P(H_i = \infty | X_0 = j) = \sum_{k \neq i} P(H_i = \infty, X_1 = k | X_0 = j) \\ &= \sum_{k \neq i} P(H_i = \infty | X_0 = k) p_{jk} = \sum_{k \neq i} p_{jk} y_k, \end{aligned}$$

where we utilized the Markov property in the second to last equality. Thus, this direction is shown.

Conversely, suppose that there is such a vector y. Since y is nonzero, there is a j with $y_j > 0$ and so by repeatedly using the relation (1.40)

$$\begin{aligned} 0 < y_j &= \sum_{k_1 \neq i} p_{j,k_1} y_{k_1} = \sum_{k_1, k_2 \neq i} p_{j,k_1} p_{k_1,k_2} y_{k_2} = \cdots \\ &= \sum_{k_1, k_2, \ldots, k_m \neq i} p_{j,k_1} p_{k_1,k_2} \cdots p_{k_{m-1},k_m} y_{k_m} \\ &\leq \sum_{k_1, k_2, \ldots, k_m \neq i} p_{j,k_1} p_{k_1,k_2} \cdots p_{k_{m-1},k_m} = P_j(H_i > m). \end{aligned}$$

Letting $m \to \infty$ yields $P_j(H_i = \infty) > 0$, and by exercise 1.20 the chain is transient. □

Example 1.77. We return to Example 1.75 in order to determine the values of p for which the chain is transient. Recall that we already know that when $p < \frac{1}{2}$, the chain is positive recurrent. We will apply Theorem 1.76 with $i = 0$.

Equation (1.40) is

$$y_j = (1-p)y_{j-1} + py_{j+1},$$

for $j \geq 1$. The general solution to this difference equation is

$$y_j = \begin{cases} c_1 + c_2\left(\frac{1-p}{p}\right)^j, & \text{if } p \neq 1/2 \\ c_1 + c_2 j, & \text{if } p = 1/2 \end{cases},$$

which is valid for $j \geq 0$, and where c_1 and c_2 are unknown (see Appendix A.1). It is already clear that no matter what we choose for the constants c_1 and c_2, the formula $y_j = c_1 + c_2 j$ can not give a nonzero solution satisfying $y_0 = 0$ and $0 \leq y_j \leq 1$ for all j. Thus, we conclude that when $p = \frac{1}{2}$, the chain is not transient. Since it can not be positive recurrent (by Example 1.75), it must be null recurrent.

We turn to the $p \neq \frac{1}{2}$ case. Since $y_0 = 0$, the above formula becomes

$$y_j = c_1 - c_1\left(\frac{1-p}{p}\right)^j.$$

If $p < 1/2$, then $1 - p > p$ and there is no solution. Hence, in this case the chain is recurrent from Theorem 1.76. (Of course, we knew this already because we showed it was positive recurrent in this case!)

For the case $p > 1/2$, we have that $1 - p < p$, and a solution can be found by setting $c_1 = 1$, in which case

$$y_j = 1 - \left(\frac{1-p}{p}\right)^j,$$

is a solution. Thus, when $p > 1/2$, the chain is transient. △

We end this section with an important theorem stating that stationary distributions characterize how much time a positive recurrent chain spends in a given state. Importantly, the theorem holds for both periodic and aperiodic chains. For a proof, we point the interested reader to [31, Chapter 2.12].

Theorem 1.78. *Let $X_n, n \geq 0$, be an irreducible, positive recurrent Markov chain with state space S and unique stationary distribution π. For $i \in S$, let*

$$N_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}},$$

denote the number of visits to state i before time n. Then

$$P\left(\frac{N_i(n)}{n} \to \pi_i, \text{ as } n \to \infty\right) = 1.$$

Moreover, for any bounded function $f : S \to \mathbb{R}$,

$$P\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \sum_{i\in S} f(i)\pi_i, \text{ as } n \to \infty\right) = 1. \tag{1.41}$$

The theorem says that the time averages of a single realization of the Markov chain will converge (with probability one) to the "space averages" obtained by simply taking expectations with respect to the stationary distribution π. For example, suppose a random variable X_∞ has probability mass function $P(X_\infty = i) = \pi_i$. Then, by definition,

$$\sum_{i\in S} f(i)\pi_i = \mathbb{E}_\pi[f(X_\infty)].$$

Now another way to write (1.41) is

$$P\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \mathbb{E}_\pi[f(X_\infty)], \text{ as } n \to \infty\right) = 1.$$

Example 1.79. Consider the Markov chain with state space $\{1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & 0 \end{pmatrix}. \tag{1.42}$$

It is simply to check that the unique stationary distribution of this chain is $\pi = (\frac{3}{8}, \frac{1}{2}, \frac{1}{8})$. Therefore, for example, $\lim_{n\to\infty} P(X_n = 3) = \frac{1}{8}$. However, we can also approximate this value using Theorem 1.78. Let $N_3(n) = \sum_{k=0}^{n-1} 1_{\{X_k=3\}}$. Figure 1.1 plots $\frac{1}{n}N_3(n)$ versus n for one realization of the chain. Note that $\frac{1}{n}N_3(n)$ appears to be converging to $\frac{1}{8}$. △

Example 1.80 (Example 1.73 continued)**.** Consider the periodic chain with state space $\{1, 2, 3, 4\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The unique stationary distribution for the chain is $\pi = (\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6})$. Hence,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} 1_{\{X_n=1\}} = \frac{1}{6},$$

and the chain spends, on average, one sixth of its time in state 1. △

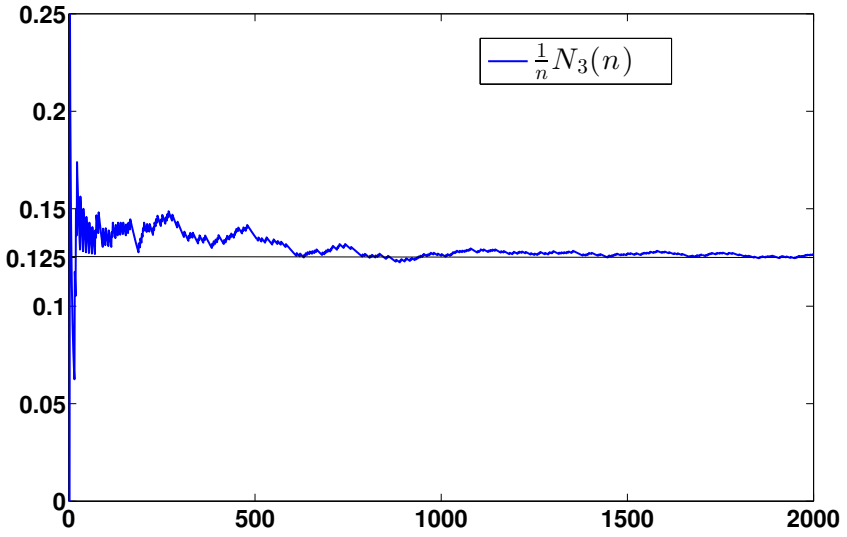


Figure 1.1: A single realization of the running average, $\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{X_k=3\}} = \frac{1}{n}N_3(n)$, for the amount of time spent in state 3 for the Markov chain with transition matrix (1.42). A line of height $\frac{1}{8} = 0.125$ has been added for reference.

1.8 Exercises

Exercise 1.1. Suppose that $p_{11}^{(n)}$ satisfies the recurrence relation

$$p_{11}^{(n+1)} = q + p_{11}^{(n)}(1-q-p),$$

where $0 \le p, q \le 1$, together with the initial condition $p_{11}^{(0)} = 1$. Show that for $n \ge 0$

$$p_{11}^{(n)} = \begin{cases} \frac{q}{q+p} + \frac{p}{q+p}(1-q-p)^n & \text{if } p+q > 0 \\ 1 & \text{if } p+q = 0. \end{cases}$$

Hint: start by assuming that $p_{11}^{(n)} = a + br^n$ for unknown a, b, and r. Plug this ansatz into the system and solve for the unknowns.

Exercise 1.2. Suppose there are two white and two green balls in two urns distributed so that each urn contains two balls. We say the system is in state i, $i = 0, 1, 2$, if there are i white balls in urn one. At each stage one ball is drawn at random from each urn and interchanged. Let X_n denote the state of the system after the nth draw.

(a) Argue that X_n is a Markov chain and give its transition matrix.

(b) Find $P(X_4 = 2 \mid X_0 = i)$ for each of $i \in \{0, 1, 2\}$.

Exercise 1.3. (**Success run chain.**) Suppose that Jake is shooting baskets in the school gym and is interested in the number of baskets he is able to make in a row.

Suppose that every shot will go in with a probability of $p \in (0,1)$, and the success or failure of each shot is independent of all other shots. Let X_n be the number of shots he has currently made in a row after n shots (so, for example, $X_0 = 0$ and $X_1 \in \{0,1\}$, depending upon whether or not he hit the first shot).

(a) Argue that X_n is a Markov chain. What is the state space of the chain?

(b) What are the one step transition probabilities of the chain?

Exercise 1.4. (Jukes-Cantor model of DNA mutations) Consider a single nucleotide on a strand of DNA. We are interested in modeling possible mutations to this single nucleotide. We say that X_n is in state 1, 2, 3, or 4, if the nucleotide is the base A, G, C, or T, respectively. We assume that there is a probability, $\rho \in (0,1)$, that between one time period and the next, we will observe a change in this base. If the base does change, we make the simple assumption that each of the other three bases are equally likely.

(a) What is the transition matrix for this Markov chain?

(b) If $\rho = 0.01$, what are (approximately): $p_{13}^{(10)}$, $p_{13}^{(100)}$, $p_{13}^{(1,000)}$, $p_{13}^{(10,000)}$? (You should use a computer for this problem.)

Exercise 1.5. Suppose that whether or not it rains tomorrow depends on previous weather conditions only through whether or not it is raining today. Assume that the probability it will rain tomorrow given it rains today is α and the probability it will rain tomorrow given it is not raining today is β. Supposing that the state space is $S = \{0,1\}$, where state 0 means it rains and state 1 means it does not rain on a given day, what is the transition matrix when we model this situation with a Markov chain. If we assume there is a 40% chance of rain today, what is the probability it will rain three days from now if $\alpha = 6/10$ and $\beta = 3/10$.

Exercise 1.6. (a) Show that the product of two stochastic matrices is stochastic.

(b) Show that for stochastic matrix P, and *any* row vector π, we have $\|\pi P\|_1 \leq \|\pi\|_1$, where $\|v\|_1 = \sum_i |v_i|$. Deduce that all eigenvalues, λ, of P must satisfy $|\lambda| \leq 1$. Hint: For the first part of (b) just write out each side and manipulate the expressions as needed. For the second part, just use the definition of a left-eigenvector and follow your nose.

Exercise 1.7. Let X_n denote a discrete time Markov chain with state space $S = \{1,2,3,4\}$, transition Matrix

$$P = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{5} & \frac{11}{20} \\ 0 & 0 & 0 & 1 \\ \frac{1}{6} & \frac{1}{7} & 0 & \frac{29}{42} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \end{pmatrix},$$

and initial distribution is $\alpha = \left(\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}\right)$.

(a) Suppose that

$$(\omega_0,\omega_1,\omega_2,\ldots,\omega_{10}) = (0.421,\ 0.794,\ 0.311,\ 0.528,\ 0.165,\ 0.602,\ 0.263,\ 0.654,\ 0.689,\ 0.748,\ 0.450)$$

is a sequence of 11 independent uniform$(0,1)$ random variables. Using these random variables (in the order presented above) and the construction of Section 1.3, what are X_n, $n \in \{0,1,\ldots,10\}$? Note, you are supposed to do this problem by hand (without the aid of a computer).

(b) Using Matlab, simulate a path of X_n up to time n =100 using the construction of Section 1.3. A helpful sample Matlab code has been provided on the course website that is intended to get you started (make sure you read my comments in the code). Play around with your script. Try different values of n and see the behavior of the chain. Turn in your MATLAB code along with a plot of a realization of your chain.

Exercise 1.8. Consider a chain with state space $\{0,1,2,3,4,5\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & \frac{7}{8} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

What are the communication classes? Which classes are closed? Which classes are recurrent and which are transient?

Exercise 1.9. Suppose that P is the transition matrix for an aperiodic Markov chain with finite state space S. Then for each state $i \in S$, there is an $N(i)$ for which $p_{ii}^{(n)} > 0$ for all $n \geq N(i)$.

Exercise 1.10. Consider a Markov chain with a finite state space S and transition matrix P. Suppose that the recurrent communication classes are $C_1, C_2, \ldots, C_r$, and that each C_k is irreducible and aperiodic. Let $\pi^{(k)}$ be the unique limiting stationary distribution for the Markov chain restricted to C_k (which is guaranteed to exist by Theorem 1.57) and let $\bar{\pi}^{(k)}$ be the vector whose ith element is zero if $i \notin C_k$ and is $\pi_i^{(k)}$ otherwise. Prove both the following.

(a) Let $a_i \geq 0$ and $\sum_{i=1}^r a_i = 1$. Then the linear combination

$$a_1\bar{\pi}^{(1)} + \cdots + a_r\bar{\pi}^{(r)}$$

is a stationary distribution for the full Markov chain with transition matrix P.

(b) All stationary distributions for the full Markov chain with transition matrix P can be written as a linear combination of the above form.

Exercise 1.11. Consider the Markov chain described in Problem 1.4 above. What is the stationary distribution for this Markov chain. Interpret this result in terms of the probabilities of the nucleotide being the different possible values for large times. Does this result make sense intuitively?

Exercise 1.12. Show that the success run chain of Problem 1.3 above is positive recurrent. What is the stationary distribution of this chain? *Using the stationary distribution*, what is the expected number of shots Jake will hit in a row.

Exercise 1.13. Let the transition matrix for a discrete time Markov chain be doubly stochastic (recall (1.7)). Show that the uniform distribution is the stationary distribution for the chain.

Exercise 1.14. Let X_n be the number of customers in line for some service at time n. During each time interval, we assume that there is a probability of p that a new customer arrives. Also, with probability q, the service for the first customer is completed and that customer leaves the queue. Assuming at most one arrival and at most one departure can happen per time interval, the transition probabilities are

$$p_{i,i-1} = q(1-p), \qquad p_{i,i+1} = p(1-q)$$

$$p_{ii} = 1 - q(1-p) - p(1-q), \qquad i > 0$$

$$p_{00} = 1 - p, \quad p_{01} = p.$$

(a) Argue why the above transition probabilities are the correct ones for this model.

(b) For which values of p and q is the chain null recurrent, positive recurrent, transient?

(c) For the positive recurrent case, give the limiting probability distribution π.
Hint: note that the equation for π_0 and π_1 are both different than the general nth term.

(d) Again in the positive recurrent case, using the stationary distribution you just calculated, what is the expected length of the queue in equilibrium? What happens to this average length as $p \to q$. Does this make sense?

Exercise 1.15. Prove Theorem 1.70 when the chain has a period of d.

Exercise 1.16. This problem has you redo the computation of Example 1.79, though with a different Markov chain. Suppose our state space is $\{1, 2, 3, 4\}$ and the transition matrix is

$$P = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{5} & \frac{11}{20} \\ 0 & 0 & 0 & 1 \\ \frac{1}{6} & \frac{1}{7} & 0 & \frac{29}{42} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \end{pmatrix},$$

which was the transition matrix of problem 1.7 above. Using Theorem 1.78, and MATLAB, estimate $\lim_{n\to\infty} P(X_n = 2)$. Make sure you choose a long enough path, and that you plot your output (to turn in). Compare your solution with the actual answer computed via the left eigenvector with an eigenvalue of one.

Exercise 1.17. You flip a fair coin until you flip four consecutive heads. Let τ be the total number of flips you perform. Find $E[\tau]$.

Exercise 1.18. You will need software for this problem to deal with the matrix manipulations. Consider a Markov chain X_n with state space $\{0, 1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & \frac{7}{8} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

Here the only recurrent class is $\{1, 2\}$. Suppose that $X_0 = 0$ and let

$$T = \inf\{n : X_n \in \{1, 2\}\}.$$

(a) What is $E[T]$?

(b) What is $P_0(X_T = 1)$? $P_0(X_T = 2)$? (Note that this is asking for the probabilities that when the chain enters the recurrent class, it enters into state 1 or 2.)

Exercise 1.19. (Taken from Lawler, [28]) You will need software for this problem to deal with the matrix manipulations. Let X_n and Y_n be independent Markov chains with state space $\{0, 1, 2\}$ and transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Suppose that $X_0 = 0$ and $Y_0 = 2$ and let

$$T = \inf\{n : X_n = Y_n\}.$$

A hint for all parts of this problem: consider the nine-state Markov chain $Z_n = (X_n, Y_n)$.

(a) Find $E(T)$.

(b) What is $P(X_T = 2)$?

(c) In the long run, what percentage of the time are both chains in the same state?

Exercise 1.20. Let $X_n, n \geq 0$, be a discrete time Markov chain with discrete state space. Suppose that $i \leftrightarrow j$. Let $H_i = \inf\{n \geq 0 : X_n = i\}$ and suppose that $P_j(H_i = \infty) > 0$. Show that i and j are transient.

Exercise 1.21. Suppose that $X_n \in \{1, 2, 3\}$ is a thee state Markov chain with transition matrix

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{7} & \frac{2}{7} & \frac{4}{7} \end{pmatrix}.$$

Suppose that the initial distribution is $\alpha = (\frac{1}{2}, \frac{1}{2}, 0)$. What is $P(X_0 = 1, X_1 = 3, X_2 = 2, X_3 = 2)$?

Exercise 1.22. Consider a Markov chain with state space $S = \{0, 1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} .5 & .5 & 0 & 0 & 0 & 0 \\ .3 & .7 & 0 & 0 & 0 & 0 \\ 0 & 0 & .1 & 0 & .9 & 0 \\ .25 & .25 & 0 & 0 & .25 & .25 \\ 0 & 0 & .7 & 0 & .3 & 0 \\ 0 & .2 & 0 & .2 & .2 & .4 \end{pmatrix} \end{array}$$

(a) What are the communication classes? Which ones are recurrent and which are transient?

(b) Suppose the system starts in state 0. What is the probability that it will be in state 0 at some time far into the future?

Exercise 1.23. Consider a Markov chain with state space $S = \{0, 1, 2, \dots\}$ and transition probabilities

$$p_{i,i+1} = \tfrac{2}{3}; \quad p_{i,0} = \tfrac{1}{3} \quad \text{for each of } i \in \{0, 1, 2, 3, \dots\}.$$

(a) Show that the chain is positive recurrent and give the stationary distribution, π. Specifically, find π_n for each n.

(b) Find the expected return times to state 0 and state 2. That is, give $\mathbb{E}_0[\tau_0]$ and $\mathbb{E}_2[\tau_2]$, where $\tau_i = \min\{n \geq 1 \mid X_n = i\}$.

(c) Find the distribution of τ_0. That is, find $P(\tau_0 = k)$ for each $k > 1$.

Exercise 1.24. Suppose a person is walking between three coffee shops and we let $X_n \in \{1, 2, 3\}$ give which coffee shop she is in at time n. We believe the motion of this walker can be described by a discrete time Markov chain with transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{7}{10} & \frac{2}{10} \end{pmatrix}.$$

Suppose that $X_0 = 1$. In order to simulate this Markov chain, I generated the following uniform random variables:

$$U_1 = 0.4505, \quad U_2 = 0.0838, \quad U_3 = 0.2290, \quad U_4 = 0.9133, \quad U_5 = 0.1524.$$

Using these uniform random variables together with Algorithm 1, find $X_1, X_2, \ldots, X_5$.

Exercise 1.25. A particular students' performance in school is highly dependent upon how they performed in the previous class (due to confidence issues). We are supposing that this student takes only one class at a time. Supposing the only grades possible are A, B, C, D, and F, we let $X_n \in \{A, B, C, D, F\}$ be the state of the Markov chain giving the grade of the nth class and assume the transition matrix is

$$P = \begin{array}{c} \\ A \\ B \\ C \\ D \\ F \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & F \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Note that F is a recurrent state and the student flunks out of school. Supposing school is ongoing for this student and that the student came to school as an "A" student, what is the expected amount of time before this student flunks out?

Exercise 1.26. A doctor has two drugs she uses to control nasal congestion, drug A and drug B. If a medicine works well on one patient, she automatically prescribes it for her next patient. (Assume she can tell immediately if the drug works well.) If it does not work well, she switches to the other drug for her next patient. Suppose that for any patient drug A works well with probability 1/2 and drug B works well with probability 3/4.

(a) Formulate a Markov chain model for this situation. That is, (*i*) describe the state space, (*ii*) specify clearly the interpretation of the random variables involved, and (*iii*) give the transition matrix.

(b) In the long run, what is the fraction of patients that receive drug A?

Exercise 1.27. A polymer is a string-like molecule composed of many repeated subunits, known as monomers. Polymers grow and shrink in length by having monomers either attach to the end of the polymer (which grows the polymer by one unit) or detach from the end of the polymer (which shrinks the polymer by one unit). We suppose that each polymer has a unique base monomer, from which the polymer is growing. We will call both the attachment of a monomer, and the detachment of a monomer an event. Let L_n denote the length of a given polymer after n events have occurred. That is, L_n is equal to the number of monomers connected to the base monomer plus one (for the base). We may model the length of the polymer as a discrete time Markov chain and we suppose that the probability that the next event is an attachment is always $p \in (0, 1)$ so long as $L_n > 1$. We can not lose the base monomer, so if $L_n = 1$ the next event is surely an attachment.

(a) Under what condition on p will L_n be transient? Prove that it is transient under this condition.

(b) Under what condition on p will L_n be positive recurrent? Prove that it is positive recurrent under this condition.

(c) Under what condition on p will the DTMC be null recurrent? Argue why it is null recurrent under this condition.

Exercise 1.28. A coffee loving mathematician will spend the day in four different coffee shops. Every hour the mathematician will transition to a new coffeeshop according to the following transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{2}{3} & \frac{1}{4} & 0 & \frac{1}{12} \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Suppose that the mathematician always starts the day in coffeeshop number 1. What is the probability the mathematician will be in coffee shop number 3 after the 9th transition. Use a computer to help you solve this problem.

Exercise 1.29. Give a state space and transition matrix for a Markov chain with period four. The chain should be "nontrivial" in that not all transition probabilities should be identically equal to one. Adding a diagram would be helpful.

Exercise 1.30. A student listens to music while studying and after each song will determine whether or not he wants to switch the station. This student transitions among three types of radio stations, 80s rock, 90s grunge, and Irish Folk, and does so according to the following transition matrix

$$P = \begin{array}{c} \text{80s rock} \\ \text{90s grunge} \\ \text{Irish Folk} \end{array} \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.8 & 0.05 & 0.15 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}.$$

What is the long run probability of listening to each type of music?

Exercise 1.31. There are two yards that a squirrel can choose to live in. Each yard has a single tree. The behavior of the squirrel is different in the two yards.

Behavior in yard one. When the squirrel is on the ground in yard #1, there is a probability of 1/2 that it will get scared in the next 1 minute (our time-unit), and run up a tree. If the squirrel is in the tree in yard #1, there is a probability of 1/3 that the squirrel will become brave and move to the ground in the next minute.

Behavior in yard two. When the squirrel is on the ground in yard #2, there is a probability of 1/4 that it will get scared in the next 1 minute, and run up the tree. If the squirrel is in the tree in yard #2, there is a probability of 2/3 that the squirrel will become brave and move to the ground in the next minute.

(a) The squirrel picks a yard to live in, each with equal probability. The squirrel then remains in that yard forever. What is the long run probability that the squirrel is on the ground? What is the long run probability that the squirrel is in a tree? Note that you must answer this question without knowing which yard the squirrel will enter with surety.

(b) Suppose that if the squirrel is on the ground in yard #1 and makes a sudden dash for the tree, then there is a probability of 1/10 that the squirrel will catch the eye of a bird of prey and be caught. Let τ_1 be the time the squirrel is caught. Assuming that the squirrel lives in yard #1, and begins the day on the ground, find $E[\tau_1]$.

(c) Suppose that if the squirrel is on the ground in yard #2 and makes a sudden dash for the tree, then there is a probability of 1/2 that the squirrel will catch the eye of a bird of prey and be caught. Let τ_2 be the time the squirrel is caught. Assuming that the squirrel lives in yard #2, and begins the day on the ground, find $E[\tau_2]$.

Chapter 2

Discrete Time Markov Chains in the Life Sciences

In this Chapter, we review some basic discrete time Markov chain models used in the life sciences. In Section 2.1 we discuss models of genetic inheritance, in Section 2.2 we discuss discrete time birth and death models, and in Section 2.3 we discuss branching processes. Both birth and death models and branching processes have many applications outside the life sciences.

2.1 Genetic Models

2.1.1 Mendelian inheritance

In the mid-19th century, an Augustinian friar named Gregor Mendel performed a number of experiments on pea plants that paved the way for our understanding of genetic inheritance. He is now known as the "father of modern genetics."

To explain and understand Mendel's experiments, some terminology is required. A cell with two sets of chromosomes, with one set coming from each parent, is called *diploid*. Chromosomes are long strands of organized DNA, and a *gene* is a region of DNA that encodes a functional RNA or protein product. Each gene can come in multiple different *alleles*, and different possible observable traits of the cell (or organism) depend upon which different types of alleles are present. For example, in his work on pea plants Mendel predicted that there were alleles for, among other things: tall and dwarf plants, round and wrinkled seeds, and yellow and green seeds.

To come to his conclusions, Mendel made a series of experiments and observations, which will be described here in the setting of tall and dwarf pea plants. First, Mendel observed that if he took tall plants that had been bred from a line of only tall plants, and cross-bred them with dwarf plants that had been bred from a line of only dwarf plants, then the resulting plants were all tall. Call this new generation of plants generation one. Next, he bred plants from generation one and produced a second generation of plants. In this second generation, roughly a quarter of the plants were now dwarfs, even though none of the plants from the first generation were dwarfs.

Mendel observed the above behavior of one trait dominating another in the first generation, followed by a 3:1 ratio in the second generation for all traits he had under consideration.

From these observations we can make the following conclusions, which still form the basic understanding we have today. Each pea plant has two genes, one from each parent, that determines each observable trait. Each gene can come in one of two alleles, call them A and a, with one being dominant, such as "tall" in the example above. Thus, when the gene pair is either AA or Aa (which is equivalent to aA, order does not matter), the dominant trait is observed, and only when the gene sequence is aa does the recessive trait emerge. The combination of alleles, AA, Aa, and aa are called *genotypes*. The observable traits are often called *phenotypes*.

We will use this knowledge to study a well known model of genetic inbreeding (see also [1] and [13]). Consider two randomly chosen individuals, who we will call the zeroth generation, that are mated to yield a first generation. Next, two offspring of opposite sex are randomly selected from this first generation and are mated. This gives rise to a second generation. Two individuals of opposite sex are randomly selected from this second generation and are used to produce a third generation. This processes then continues indefinitely.

We formulate a discrete time Markov chain in the following way. Let X_0 give the genotypes of the original individuals in the zeroth generation and for $n \geq 1$ let X_n give the genotypes of the randomly chosen pair from the nth generation. Note that the pair with genotype X_n are the parents of the entire $(n+1)$st generation. There are six possible values for X_n,

X_n	Genotypes	X_n	Genotypes
1	$AA \times AA$	4	$Aa \times aa$
2	$AA \times Aa$	5	$AA \times aa$
3	$Aa \times Aa$	6	$aa \times aa$

,

where we note that order of the genotypes of the individuals does not matter. That is, $AA \times Aa$ is the same as $Aa \times AA$. We must now find the associated transition probabilities for the chain.

First, note that if the genotype pair of a couple is $AA \times AA$, that is if both parents have genotype AA, then all of their offspring necessarily have genotype AA, and so $p_{11} = 1$. Similarly, we have $p_{66} = 1$.

Now consider the possibility that the genotype pairing is of type 2, that is $AA \times Aa$. Consider a specific offspring of this pairing and let $Z_1 \in \{A, a\}$ be the allele passed down to the offspring from the parent with genotype AA, and let $Z_2 \in \{A, a\}$ be the allele passed down to the offspring from the parent with genotype Aa. Finally, let Y

be the genotype of a randomly selected offspring. We then have

$$\begin{aligned}
P(Y = AA) &= P(Z_1 = A, Z_2 = A) = P(Z_1 = A)P(Z_2 = A) = 1 \times \tfrac{1}{2} = \tfrac{1}{2} \\
P(Y = Aa) &= P(Z_1 = A, Z_2 = a) + P(Z_1 = a, Z_2 = A) \\
&= P(Z_1 = A)P(Z_2 = a) + P(Z_1 = a)P(Z_2 = A) \\
&= 1 \times \tfrac{1}{2} + 0 \\
&= \tfrac{1}{2} \\
P(Y = aa) &= P(Z_1 = a, Z_2 = a) = P(Z_1 = a)P(Z_2 = a) = 0.
\end{aligned}$$

Thus, letting Y_1 and Y_2 be the genotypes of two randomly chosen offspring, we have that

$$\begin{aligned}
P(Y_1 \times Y_2 = AA \times AA) &= P(Y_1 = AA)P(Y_2 = AA) = \tfrac{1}{4} \\
P(Y_1 \times Y_2 = AA \times Aa) &= P(Y_1 = AA)P(Y_2 = Aa) = \tfrac{1}{4} \\
P(Y_1 \times Y_2 = Aa \times AA) &= P(Y_1 = Aa)P(Y_2 = AA) = \tfrac{1}{4} \\
P(Y_1 \times Y_2 = Aa \times Aa) &= P(Y_1 = Aa)P(Y_2 = Aa) = \tfrac{1}{4}.
\end{aligned}$$

Thus, recalling that $AA \times Aa = Aa \times AA$, we have that

$$p_{21} = \tfrac{1}{4}, \quad p_{22} = \tfrac{1}{2}, \quad \text{and} \quad p_{23} = \tfrac{1}{4}.$$

Continuing in this manner, the transition matrix for this six state Markov chain is

$$P = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.1}$$

where the states have been listed to the left of the transition matrix for reference. Note that the communication classes are $\{1\}$, $\{6\}$, and $\{2, 3, 4, 5\}$ with only classes $\{1\}$ and $\{6\}$ begin recurrent (since states 1 and 6 are absorbing).

We wish to calculate the relevant transition probabilities discussed in Section 1.6. Therefore, we reorder the states to be $1, 6, 2, 3, 4, 5$, yielding the transition matrix

$$\widetilde{P} = \begin{array}{c} 1 \\ 6 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \left(\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right).$$

Therefore, using the notation of Section 1.6,

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (I-Q)^{-1} = \begin{pmatrix} \frac{8}{3} & \frac{4}{3} & \frac{2}{3} & \frac{1}{6} \\ \frac{4}{3} & \frac{8}{3} & \frac{4}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{4}{3} & \frac{8}{3} & \frac{1}{6} \\ \frac{4}{3} & \frac{8}{3} & \frac{4}{3} & \frac{4}{3} \end{pmatrix}, \quad U = \begin{pmatrix} \frac{1}{4} & 0 \\ \frac{1}{16} & \frac{1}{16} \\ 0 & \frac{1}{4} \\ 0 & 0 \end{pmatrix},$$

and

$$(I-Q)^{-1}U = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Note also that

$$(I-Q)^{-1}\vec{1} = \begin{pmatrix} \frac{29}{6} \\ \frac{17}{3} \\ \frac{29}{6} \\ \frac{20}{3} \end{pmatrix}.$$

We may now make the following conclusions.

1. If $X_0 = 2$, i.e. genotype $AA \times Aa$, then the probability of eventual absorption into state 1, $AA \times AA$, is $3/4$ and the probability of absorption into state 6, $aa \times aa$, is $1/4$. This is seen in the first row of the matrix $(I-Q)^{-1}U$.

2. If $X_0 \in \{3, 5\}$, that is if the initial genotype pairing is of type $Aa \times Aa$ or $AA \times aa$, then there are equal probabilities of absorption in states 1 and 6. This should be intuitive via symmetry of the alleles A and a.

   However, the expected number of steps to achieve such absorption is not the same. For $X_0 = 3$, the expected number of steps to absorption is $\frac{17}{3} = 5.666\ldots$, whereas if $X_0 = 5$ the expected number of steps is $\frac{20}{3} = 6.666\ldots$. Note that the difference of 1 in the expected number of steps until absorption can be understood by observing that if the initial genotype is $AA \times aa$, then all offspring of the original pairing will have genotype Aa.

2.1.2 The Wright-Fisher Model

We consider another model from population genetics, which was first developed by Ronald Fisher, and later extended by Sewall Wright. In this model we assume the existence of N diploid (two copies of each gene) individuals. Thus, there are a total of $2N$ genes in the gene pool. We make the following assumptions:

1. The number of individuals remains constant, at N, from generation to generation.

2. The genes for any individual in the $(n+1)$st generation are randomly selected (with replacement) from the pool of genes in the nth generation.

Note that the last assumption allows us to disregard the individuals, and only consider the gene pool itself.

We suppose we have two alleles of the gene in question, which we denote by A and a. We let $X_n \in \{0, 1, \ldots, 2N\}$ denote the number of alleles of type A in the entire gene pool. Oftentimes A is assumed to be a *mutant* that has entered the population. We are interested in the probabilities associated with *fixation*, meaning when the system becomes homogeneous in A, which occurs when $X_n = 2N$ and A has overtaken the population, or in a, which occurs when $X_n = 0$ and the mutant has been erased from the population.

We build our Markov model by finding the transition probabilities. Supposing that $X_n = i$, for some $i \geq 0$, what is the probability that $X_{n+1} = j$? Because of our simplifying assumptions, we see that, conditioned on $X_n = i$, the value of X_{n+1} is a binomial(n, p) random variable with parameters $n = 2N$ and $p = i/(2N)$. Therefore,

$$p_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \quad \text{for } i, j \in \{0, \ldots, 2N\}. \tag{2.2}$$

Let $g_0(j)$ denote the probability that the chain is eventually absorbed by state 0, given that it started in state j. That is, $g_0(j)$ is the probability that the allele A is eventually erased from the population, given an initial abundance of j. We recall that we have methods, from Section 1.6, for the calculation of such probabilities. However, it is not clear how to use those methods here. Trying a first step analysis, we are led to the system of equations,

$$g_0(j) = 1 \cdot p_{j,0} + g_0(1)p_{j,1} + \cdots + g_0(2N-1)p_{j,2N-1}, \quad j \in \{1, 2, \ldots, 2N-1\}.$$

These equations do not look very promising. A different method is required to solve this problem elegantly.

We begin by noting that for each $i \in \{0, 1, \ldots, 2N\}$ we have

$$\mathbb{E}\left[X_{n+1} \;\middle|\; X_n = i\right] = i,$$

in which case $\mathbb{E}[X_{n+1}|X_n] = X_n$. Taking expectations of both sides yields

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n],$$

which is valid for all $n \geq 0$. Therefore, if $X_0 = j$ we have

$$\mathbb{E}_j[X_{n+1}] = \mathbb{E}_j[X_n] = \cdots = \mathbb{E}_j[X_1] = \mathbb{E}_j[X_0] = j,$$

for all $n \geq 0$. In particular, we now trivially have

$$\lim_{n\to\infty} \mathbb{E}_j[X_n] = j.$$

On the other hand, assuming that $X_0 = j$, the limit $\lim_{n\to\infty} X_n$ exists and is either state 0 with probability $g_0(j)$, or state $2N$ with probability $g_{2N}(j)$. We may use the following theorem, which we state without proof. (See [9] or Exercise 2.16 for a proof in our case.)

Theorem 2.1 (Bounded convergence theorem)**.** *Suppose that X_n is a sequence of random variables for which $|X_n| \leq M$ for some $M > 0$, and $\lim_{n\to\infty} X_n$ exists with a probability of one. Then,*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

Since in our case each X_n is bounded by $2N$ the theorem applies and we may conclude that

$$j = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] = 0 \cdot g_0(j) + 2N \cdot (1 - g_0(j)).$$

Solving yields

$$g_0(j) = \frac{2N - j}{2N} \quad \text{and so} \quad g_{2N}(j) = \frac{j}{2N}.$$

2.1.3 Phylogenetic Distance and the Jukes-Cantor Model

We present a simple model for the phylogenetic distance between two strands of DNA, the Jukes-Cantor model. The basic question is the following: given two strands of DNA, how many mutations have taken place to get from one to the other? Answering this type of question allows for the construction of a tree of life, showing which species are more, or less, related to others. In Exercise 2.18 we present another model for the phylogenetic distance between two strands of DNA, the Kimura two-paramter model.

We start by considering a single strand of DNA, which is molecule carrying most of the genetic instructions used in living cells. Each strand of DNA is a long sequence of just four nucleotides: adenine, guanine, cytosine, and thymine (A, G, C, and T).[1] We will consider the evolutionary history of a single strand of DNA as it is passed down through generations.

Instead of modeling the entire DNA strand, we begin by considering a single nucleotide on this strand of DNA and will let X_n give the state of the nucleotide in generation $n \in \{0, 1, 2, \dots\}$. Specifically, we say X_n is in state 1, 2, 3, or 4, if the nucleotide is the base A, G, C, or T, respectively. We assume that X_n is Markovian and that there is a probability, $\rho \in (0, 1)$, that between one generation and the next a change in the base takes place. Further, if the base does change, we assume that each of the other three bases are equally likely to be the next state. These assumptions lead to the following transition matrix for X_n,

$$P = \begin{pmatrix} 1-\rho & \rho/3 & \rho/3 & \rho/3 \\ \rho/3 & 1-\rho & \rho/3 & \rho/3 \\ \rho/3 & \rho/3 & 1-\rho & \rho/3 \\ \rho/3 & \rho/3 & \rho/3 & 1-\rho \end{pmatrix}.$$

Our first question is the following: after n generations, what is the probability that this nucleotide is the same value as it was in generation zero. That is, we want

[1] For example, a human cell is made of billions of nucleotides.

$p_{ii}^{(n)}$ for $i \in \{1,2,3,4\}$. Note that, by symmetry, this probability will be the same regardless of the initial state, and so we may simply find $p_{11}^{(n)}$.

Note that the first component of e_1P^n, where $e_1 = (1,0,0,0)$, gives the desired probability. The matrix P is doubly stochastic and so the left eigenvector of P with eigenvalue 1 is

$$(\tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4}).$$

The only other eigenvalue of P is $1-(4/3)\rho$, which has the three linearly independent left eigenvectors,

$$(\tfrac{1}{4},\ 0,\ 0,\ -\tfrac{1}{4}),\ (\tfrac{1}{4},\ 0,\ -\tfrac{1}{4},\ 0) \text{ and } (\tfrac{1}{4},\ -\tfrac{1}{4},\ 0,\ 0).$$

Therefore, we may write

$$e_1 = (\tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4}) + (\tfrac{1}{4},\ 0,\ 0,\ -\tfrac{1}{4}) + (\tfrac{1}{4},\ 0,\ -\tfrac{1}{4},\ 0) + (\tfrac{1}{4},\ -\tfrac{1}{4},\ 0,\ 0),$$

and conclude

$$e_1P^n = (\tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4},\ \tfrac{1}{4}) + \left(1-\tfrac{4}{3}\rho\right)^n \cdot \left(\tfrac{3}{4},\ -\tfrac{1}{4},\ -\tfrac{1}{4},\ -\tfrac{1}{4}\right).$$

Therefore,

$$p_{11}^{(n)} = \tfrac{1}{4} + \tfrac{3}{4}\left(1-\tfrac{4}{3}\rho\right)^n.$$

Note also that the probability that this particular nucleotide is not the same after n generations is

$$1 - p_{11}^{(n)} = \tfrac{3}{4} - \tfrac{3}{4}\left(1-\tfrac{4}{3}\rho\right)^n. \tag{2.3}$$

Now we reconsider the two strands of DNA. Suppose we observe that the fraction of nucleotides that are different is $\gamma \in [0,1]$. Assuming that the mutations of the nucleotides occur independently from one antother, the expected fraction of nucleotides that are different after n generations is given by (2.3). Therefore, we set

$$\gamma = \tfrac{3}{4} - \tfrac{3}{4}\left(1-\tfrac{4}{3}\rho\right)^n,$$

and solve for n

$$n = \frac{\ln(1-(4/3)\gamma)}{\ln(1-(4/3)\rho)},$$

which yields n as a function of γ, which is observable, and ρ. If we somehow know ρ, perhaps through other records, then we can figure out n. However, if ρ is unknown, then we obviously can not find n.

Note that according to our model, the number of mutations, per site, over n generations is a binomial random variable with parameters n and ρ. Hence, $n\rho$ gives the expected number of mutations that have taken place. Assuming that ρ is small, we may use a Taylor expansion of $\ln(1+x)$ to conclude

$$\ln(1-(4/3)\rho) = -(4/3)\rho + O(\rho^2),$$

which yields

$$n\rho = \frac{\ln(1-(4/3)\gamma)}{\ln(1-(4/3)\rho)}\rho \approx \frac{\ln(1-(4/3)\gamma)}{-(4/3)\rho}\rho = -\frac{3}{4}\ln\left(1-\frac{4}{3}\gamma\right).$$

The Jukes-Cantor *phylogenetic distance* is therefore defined as $d_{JC} = -\frac{3}{4}\ln(1-\frac{3}{4}\gamma)$.

Example 2.2. Suppose that two strands of DNA have been found, one from the distant past and one in the present. Suppose further that the strands differ in 22.3% of the nucleotide sites. Then, the Jukes-Cantor distance for the two strands is

$$d_{JC} = -\tfrac{3}{4}\ln\left(1 - \tfrac{4}{3} \times 0.223\right) = 0.26465.$$

Note that this value, giving the expected number of substitutions per site over the time period of interest, is higher than the observed 22.3% differences. This should be intuitive since d_{JC} takes into account mutations that have taken place, and then changed back. That is, it could be that the nucleotide changed from A to G and then back to A at different points in time.

If we also somehow know that the mutation rate is $\rho = .002$/generation, then the difference, in generations, between the strands can be estimated to be

$$n = d_{JC}\frac{1}{\rho} \approx \frac{0.26465}{.002} \approx 132 \text{ generations.}$$

△

2.2 Discrete Time Birth and Death models

In this section, we consider random walks on $\{0, 1, 2, \dots\}$ that can only undergo one of three transition types: increase by one, decrease by one, or stay the same. These Markov chains are useful in a variety of applications and are often used in modeling the size of some population. They are often called birth and death models for obvious reasons. We have seen similar Markov chains in Examples 1.14, 1.15, 1.16 and 1.45, and will revisit birth and death processes in the continuous time setting in Section 4.6.

More formally, we will consider a discrete time Markov chain on $S = \{0, 1, 2, \dots\}$ with transition probabilities

$$p_{ij} = \begin{cases} p_i, & \text{if } j = i+1 \\ q_i, & \text{if } j = i-1 \\ 1 - p_i - q_i, & \text{if } j = i, \\ 0, & \text{else} \end{cases},$$

where $p_i, q_i \geq 0$, $p_i + q_i \leq 1$, and $q_0 = 0$, which ensures that 0 is a lower bound for the chain. The values p_i and q_i are often called the *birth rate* and *death rate*, respectively. Note that the transition probabilities are allowed to be state dependent.

Example 2.3. The deterministically monotone chain of Example 1.11 is a birth process with $p_i = 1$ for all $i \geq 0$. △

Example 2.4. The Gambler's ruin of Example 1.45, which is random walk on $\{0, 1, \dots, N\}$ with absorbing boundaries, is a birth and death process with $p_i = p$ and $q_i = 1 - p$ for all $1 \leq i \leq N - 1$, and $p_0 = q_0 = p_N = q_N = 0$. △

Example 2.5 (Population size)**.** We develop a simple model for a population that can change due to single births and deaths. Let $N > 0$ be a positive integer and let $\lambda, \mu > 0$ satisfy $\lambda N + \mu N \leq 1$. We assume transition rates

$$p_i = \begin{cases} \lambda i & 0 \leq i \leq N-1 \\ 0 & \text{else} \end{cases}, \quad \text{and} \quad q_i = \begin{cases} \mu i & 0 \leq i \leq N \\ 0 & \text{else} \end{cases}.$$

The parameter N can be thought of as the carrying capacity for the system. Note that because $p_N = 0$, the natural state space for the model is the irreducible component $\{0, 1, \ldots, N\}$. △

Example 2.6 (Queueing Models)**.** Suppose that X_n represents the number of people in line for some service at time n. We suppose that people are arriving at a rate of λ, so that $p_i = \lambda$ for $i \geq 0$. Customers may also leave the line because they have been served. We can choose how to model the service times in a number of ways.

(a) (Single server) If there is a single server, and that person always serves the first person in line, then we take $q_i = \mu > 0$ if $i \geq 1$, and $q_0 = 0$. We require that $\lambda + \mu \leq 1$.

(b) (K servers) If there are $K \geq 1$ servers then the first K people in line are being served. Hence, for some $\mu > 0$ satisfying $\lambda + K\mu \leq 1$ we take

$$q_i = \begin{cases} i\mu, & \text{if } i \leq K \\ K\mu, & \text{if } i \geq K \end{cases}.$$

There are other options for choosing how the people in line can be served, but we delay that discussion until later. △

Recurrence, transience, and stationary distributions

We begin by searching for a condition on the values p_i and q_i that will allow us to determine when a discrete time chain is recurrent, and when it is transient. Our tool will be Theorem 1.76. We restrict our attention to chains satisfying the following conditions, which guarantee $\{0, 1, \ldots\}$ is irreducible: (i) $p_i > 0$ for all $i \geq 0$, and (ii) $q_i > 0$ for all $i \geq 1$. The analysis carried out here will also be helpful in the study of continuous time birth and death models in Section 4.6.

Following Theorem 1.76, we set $i = 0$ as our reference state and look for a nonzero vector y satisfying

$$y_n = p_n y_{n+1} + q_n y_{n-1} + (1 - p_n - q_n) y_n, \quad n \geq 1,$$

$y_0 = 0$, and $0 \leq y_n \leq 1$ for all n. Rearranging the above relation yields

$$y_n - y_{n+1} = \frac{q_n}{p_n}(y_{n-1} - y_n). \tag{2.4}$$

Iterating (2.4) gives

$$y_n - y_{n+1} = \frac{q_1 \cdots q_n}{p_1 \cdots p_n}(y_0 - y_1) = -\frac{q_1 \cdots q_n}{p_1 \cdots p_n} y_1,$$

which is valid for $n \geq 1$ and where the final equality holds since $y_0 = 0$. Therefore, for $n \geq 1$,

$$y_{n+1} = y_{n+1} - y_0 = \sum_{k=0}^{n}(y_{k+1} - y_k) = y_1 \sum_{k=0}^{n} \frac{q_1 \cdots q_k}{p_1 \cdots p_k}, \tag{2.5}$$

where the $k = 0$ term in the sum is taken to be equal to 1. We can therefore find a solution, y, satisfying all the requirements of Theorem 1.76 if and only if the series in (2.5) converges, in which case we define

$$Z = \sum_{k=0}^{\infty} \frac{q_1 \cdots q_k}{p_1 \cdots p_k} < \infty, \tag{2.6}$$

where the $k = 0$ term in the sum is equal to one. We have shown the following.

Proposition 2.7. *An irreducible birth and death chain with birth rates $\{p_i\}$ and death rates $\{q_i\}$ is transient if and only if*

$$\sum_{k=1}^{\infty} \frac{q_1 \cdots q_k}{p_1 \cdots p_k} < \infty.$$

Example 2.8 (Example 2.6 continued)**.** Let $\lambda, \mu > 0$ satisfy $\lambda + \mu \leq 1$ and consider the single server queue with birth rates $p_i = \lambda$, for $i \geq 0$, and death rates $q_i = \mu$, for $i \geq 1$. We have

$$\sum_{k=1}^{\infty} \frac{q_1 \cdots q_k}{p_1 \cdots p_k} = \sum_{k=1}^{\infty} \left(\frac{\mu}{\lambda}\right)^k,$$

which converges if and only if $\mu < \lambda$. Thus, by Proposition 2.7 the single server queue is transient if and only if the arrival rate is strictly greater than the rate at which the server can work, and recurrent otherwise.

Fix an integer $K > 0$ and now suppose that $\lambda + K\mu \leq 1$. The K-server queuing model has birth rate $p_i = \lambda$ and death rate $q_i = \begin{cases} i\mu, & \text{if } i \leq K \\ K\mu, & \text{if } i \geq K \end{cases}$. Thus, for $n \geq K$ we have

$$\frac{q_1 \cdots q_n}{p_1 \cdots p_n} = \frac{q_1 \cdots q_K \cdot q_{K+1} \cdots q_n}{p_1 \cdots p_n} = \frac{K!}{K^K}\left(\frac{K\mu}{\lambda}\right)^n,$$

and the sum (2.6) converges, and the chain is transient, if and only if $K\mu < \lambda$. △

We now turn to the question of determining when a birth and death chain is positive recurrent. By Theorem 1.74, it is sufficient to find conditions that determine whether or not the model has a stationary distribution.

For a birth and death chain, π is a stationary distribution if and only if $\sum_{k=0}^{\infty} \pi_k = 1$ and the following conditions are satisfied,

$$\pi_k = \pi_{k-1} \cdot p_{k-1} + \pi_k \cdot (1 - p_k - q_k) + \pi_{k+1} \cdot q_{k+1}, \quad \text{if } k \geq 1 \tag{2.7}$$

$$\pi_0 = \pi_0 \cdot (1 - p_0) + \pi_1 \cdot q_1. \tag{2.8}$$

Rearranging equation (2.8) shows

$$q_1 \cdot \pi_1 - p_0 \cdot \pi_0 = 0,$$

while rearranging terms in equation (2.7) shows that for $k \geq 1$

$$q_{k+1}\pi_{k+1} - p_k\pi_k = q_k\pi_k - p_{k-1}\pi_{k-1}.$$

Iterating the above shows that if π satisfies (2.7) and (2.8), then for all $k \geq 0$

$$q_{k+1}\pi_{k+1} - p_k\pi_k = q_1\pi_1 - p_0\pi_0 = 0. \tag{2.9}$$

Solving (2.9) shows that for $k \geq 0$

$$\pi_{k+1} = \frac{p_k}{q_{k+1}}\pi_k,$$

and we conclude that for $k \geq 1$

$$\pi_k = \frac{p_{k-1}}{q_k}\pi_{k-1} = \cdots = \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k}\pi_0.$$

We therefore see that π can be made to satisfy conditions (2.7) and (2.8), while also being a probability vector, if and only if

$$\sum_{k=1}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} < \infty. \tag{2.10}$$

Specifically, when the sum converges we set

$$W = \sum_{k=0}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k},$$

where the $k = 0$ term in the sum is taken to be equal to one, and let $\pi_0 = W^{-1}$. We then have

$$\sum_{k=0}^{\infty} \pi_k = \pi_0 + \pi_0 \sum_{k=1}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} = \pi_0 W = 1,$$

and

$$\pi_k = \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} W^{-1}, \tag{2.11}$$

for $k \geq 1$. We collect thoughts with the following proposition.

Proposition 2.9. *An irreducible birth and death chain with birth rates* $\{p_i\}$ *and death rates* $\{q_i\}$ *is positive recurrent if and only if*

$$\sum_{k=1}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} < \infty.$$

In this case, $\pi_0 = \left(\sum_{k=0}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \right)^{-1}$, *and* $\pi_k = \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \pi_0$.

Example 2.10 (Example 2.6 continued)**.** Consider again the queuing network with a single server. In this case

$$\sum_{k=0}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} = \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^k = \frac{1}{1 - (\lambda/\mu)},$$

where the final equation holds provided that $\lambda < \mu$. Hence, when $\lambda < \mu$ the chain is positive recurrent and the equilibrium distribution is

$$\pi_k = \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^k, \text{ for } k \geq 0. \tag{2.12}$$

Combining this result with Example 2.6 shows that when $\lambda = \mu$ the chain is null current. △

If the natural state space of the birth and death chain is finite, but still irreducible, as in Example 2.5, the same analysis applies. Specifically, suppose that $S = \{0, 1, \ldots, N\}$ is an irreducible component of a birth and death chain. When searching for a stationary distribution, we still require that both (2.7) and (2.8) hold for $k \in \{0, 1, \ldots, N-1\}$. These equations alone imply that we still have

$$\pi_k = \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \pi_0, \qquad \text{for } k \leq N.$$

In order to ensure that π is a probability vector, we set

$$\pi_0 = \left(\sum_{k=0}^{N} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} \right)^{-1},$$

where we again take the $k = 0$ term in the sum to be equal to one. Note that the extra boundary condition for the finite state space chain, $\pi_N = \pi_N \cdot (1 - q_N) + \pi_{N-1} \cdot p_{N-1}$, is satisfied by this distribution.

Expected time until extinction

Suppose that $X_n, n \geq 0$, is a birth and death model in which $\{1, 2, \ldots, \}$ is an irreducible communication class and 0 is an absorbing state. For example, see Example

2.5. Suppose further that the probability of absorption into state 0 is equal to one. How long do we expect to wait before absorption?

Thus far, first step analysis has been useful, so we continue with it in the obvious manner. We let $H_0 = \inf\{n \geq 0 \ : \ X_n = 0\}$ and for $k \in \{0, 1, 2, \dots\}$ let $\tau_k = \mathbb{E}_k[H_0]$. We have that $\tau_0 = 0$, and by a first step analysis for $k \geq 1$ we have

$$\begin{aligned}
\tau_k = \mathbb{E}_k[H_0] &= \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell | X_0 = k) \\
&= \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell | X_1 = k+1) p_k + \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell | X_1 = k-1) q_k \\
&\quad + \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell | X_1 = k)(1 - p_k - q_k) \\
&= \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell - 1 | X_0 = k+1) p_k + \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell - 1 | X_0 = k-1) q_k \\
&\quad + \sum_{\ell=1}^{\infty} \ell P(H_0 = \ell - 1 | X_0 = k)(1 - p_k - q_k),
\end{aligned}$$

where we used time-homogeneity in the last step. By reindexing we have

$$\begin{aligned}
\tau_k &= \sum_{\ell=0}^{\infty} (\ell+1) P(H_0 = \ell | X_0 = k+1) p_k + \sum_{\ell=0}^{\infty} (\ell+1) P(H_0 = \ell | X_0 = k-1) q_k \\
&\quad + \sum_{\ell=0}^{\infty} (\ell+1) P(H_0 = \ell | X_0 = k)(1 - p_k - q_k) \\
&= p_k(1 + \tau_{k+1}) + q_k(1 + \tau_{k-1}) + (1 - p_k - q_k)(1 + \tau_k).
\end{aligned}$$

After some rearranging, we find that for $k \geq 1$

$$\tau_{k+1} = \tau_k + \frac{q_k}{p_k} \left(\tau_k - \tau_{k-1} - \frac{1}{q_k} \right). \tag{2.13}$$

For example, because $\tau_0 = 0$,

$$\tau_2 = \tau_1 + \frac{q_1}{p_1} \left(\tau_1 - \frac{1}{q_1} \right). \tag{2.14}$$

Similar to the analyses for transience and positive recurrence, we search for a usable pattern. For $k = 2$ in (2.13) we have

$$\begin{aligned}
\tau_3 &= \tau_2 + \frac{q_2}{p_2} \left(\tau_2 - \tau_1 - \frac{1}{q_2} \right) = \tau_2 + \frac{q_2}{p_2} \left(\frac{q_1}{p_1} \left(\tau_1 - \frac{1}{q_1} \right) - \frac{1}{q_2} \right) \\
&= \tau_1 + \frac{q_1}{p_1} \left(\tau_1 - \frac{1}{q_1} \right) + \frac{q_1 q_2}{p_1 p_2} \left(\tau_1 - \frac{1}{q_1} - \frac{p_1}{q_1 q_2} \right),
\end{aligned}$$

where we have used (2.14) multiple times. Let us think about the next step.

$$\tau_4 = \tau_3 + \frac{q_3}{p_3}\left(\tau_3 - \tau_2 - \frac{1}{q_3}\right).$$

We see that the first term, τ_3, was just calculated. We begin to see that a summation is forming for the general term. The new term in the sum comes from

$$\begin{aligned}\frac{q_3}{p_3}\left(\tau_3 - \tau_2 - \frac{1}{q_3}\right) &= \frac{q_3}{p_3}\left(\frac{q_1q_2}{p_1p_2}\left(\tau_1 - \frac{1}{q_1} - \frac{p_1}{q_1q_2}\right) - \frac{1}{q_3}\right)\\ &= \frac{q_1q_2q_3}{p_1p_2p_3}\left(\tau_1 - \frac{1}{q_1} - \frac{p_1}{q_1q_2} - \frac{p_1p_2}{q_1q_2q_3}\right),\end{aligned}$$

and the pattern is emerging. In general, we have for $m \geq 1$

$$\tau_m = \tau_1 + \sum_{k=1}^{m-1} \frac{q_1\cdots q_k}{p_1\cdots p_k}\left[\tau_1 - \frac{1}{q_1} - \sum_{i=2}^{k}\frac{p_1\cdots p_{i-1}}{q_1\cdots q_i}\right], \tag{2.15}$$

where we interpret the second summation as zero when $k < 2$. Therefore, if we can determine τ_1, we are done.

To determine τ_1, the expected amount of time needed to hit state zero conditioned upon an initial population of one, we change our model slightly. We let $Y_n, n \geq 0$, be a Markov chain with the same transition rates as our original model except that $\widetilde{p}_0 = P(Y_1 = 1|Y_0 = 0) = 1$. That is, if $Y_n = 0$ for some $n \geq 0$, then $Y_{n+1} = 1$ with probability one. Now let $T_0 = \inf\{n \geq 1 : Y_n = 0\}$ and note that $\mathbb{E}_0[T_0] = \tau_1 + 1$, where τ_1 is what we want.

The key observation to make is that computing $\mathbb{E}_0[T_0]$ is easy because we know from Theorem 1.72 that it is infinite if Y_n is null recurrent and is equal to $1/\tilde{\pi}_0$, where $\tilde{\pi}_0$ is the stationary distribution of Y_n, if Y_n is positive recurrent.

In the case that Y_n is positive recurrent we use our machinery from before and find that

$$\tilde{\pi}_0 = \frac{1}{1 + \sum_{k=1}^{\infty}\frac{p_0p_1\cdots p_{k-1}}{q_1q_2\cdots q_k}}.$$

Therefore,

$$\begin{aligned}\tau_1 &= \mathbb{E}[T_0] - 1 = \tilde{\pi}_0^{-1} - 1\\ &= \sum_{k=1}^{\infty}\frac{p_0p_1\cdots p_{k-1}}{q_1q_2\cdots q_k} = \frac{1}{q_1} + \sum_{k=2}^{\infty}\frac{p_1\cdots p_{k-1}}{q_1q_2\cdots q_k}.\end{aligned} \tag{2.16}$$

Plugging (2.16) into (2.15) yields

$$\tau_m = \tau_1 + \sum_{k=1}^{m-1}\left[\frac{q_1\cdots q_k}{p_1\cdots p_k}\sum_{i=k+1}^{\infty}\frac{p_1\cdots p_{i-1}}{q_1\cdots q_i}\right].$$

This formula is legitimately scary, however such is life. In the finite state space case, $S = \{0, 1, 2, \ldots, N\}$, the only things that change in the formulas for τ_1 and τ_m are that the infinities become N's.

Example 2.11 (Example 2.6 continued)**.** Consider a single server queue with arrival rate λ and service rate μ. Assume that $\lambda < \mu$. Then

$$\begin{aligned}\tau_1 &= \frac{1}{\mu} + \sum_{k=2}^{\infty} \frac{\lambda^{k-1}}{\mu^k} = \frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{\mu^k} = \frac{1}{\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{\mu^k} - 1 \right) \\ &= \frac{1}{\lambda} \left(\frac{1}{1 - (\lambda/\mu)} - 1 \right) \\ &= \frac{1}{\mu - \lambda}.\end{aligned}$$

Note that τ_1 tends to infinity as $\lambda \to \mu$. △

2.3 Branching Processes

The study of branching processes arose from a very natural question: why, even in a growing population, do some surnames die out? Specifically, in 1873 Francis Galton posed the following problem. Suppose that N adult males in a population each have different surnames. Suppose that in each generation the probability of a given individual to have $i \in \{0, 1, 2, \dots\}$ male children who survive to adulthood is p_i. What proportion of the surnames become extinct after r generations? It was Henry William Watson who helped solve the problem, and the study of branching processes had begun. Because it was Galton and Watson who first popularized the models introduced below, the are commonly referred to as Galton-Watson processes.

Branching processes have proven to be quite useful in a number of arenas, and their study has become a large subfield of probability theory.

2.3.1 Terminology and notation

We want a model for the size of a population satisfying the following assumptions.

1. Each individual independently produces a random number of offspring.
2. The distribution for the number of offspring is the same for each individual.
3. After reproducing, each individual dies.

We now turn the above assumptions into a precise mathematical model. For positive integers n and k, let $Y_k^{(n)}$ be independent and identically distributed random variables with range $\{0, 1, 2, \dots\}$ and probability mass function

$$P(Y_k^{(n)} = i) = p_i.$$

We will suppose throughout this section that $p_0 > 0$ and $p_0 + p_1 < 1$. Assuming X_0 is given, we define X_n recursively via the formula

$$X_n = \sum_{k=1}^{X_{n-1}} Y_k^{(n)}, \qquad \text{for } n \geq 1, \tag{2.17}$$

where any sum of the form $\sum_{k=1}^{0}$ is taken to be zero. Thus, zero is seen to be an absorbing state. The process X_n is called a branching process.

Note that $X_n, n \geq 0$, satisfies the three assumptions detailed above. Moreover, X_n satisfies the one-step Markov condition given in (1.2) since X_n is a function of only X_{n-1} and random variables that are independent from $\{X_0, X_1, \ldots, X_{n-1}\}$. Thus, by Proposition 1.4, X_n, also satisfies the Markov property of Definition 1.3.

It is not easy to compute the transition probabilities for this model, which are

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) = P(Y_1^{(n+1)} + Y_2^{(n+1)} + \cdots + Y_i^{(n+1)} = j). \tag{2.18}$$

Equation (2.18) can be solved using convolutions, but that solution is quite messy and we do not choose that route. Instead, we begin by simply computing how the mean of the process evolves in time, and then utilizing probability generating functions to solve the original problem considered by Galton and Watson.

2.3.2 Behavior of the mean

Let $\mu = \mathbb{E}[Y_1^{(1)}]$ denote the mean number of offspring produced per individual. We assume throughout that $\mu < \infty$. First note that from (2.17),

$$\mathbb{E}[X_n | X_{n-1} = i] = \mathbb{E}\left[\sum_{k=1}^{i} Y_k^{(n)}\right] = \sum_{k=1}^{i} \mathbb{E}\left[Y_k^{(n)}\right] = i \cdot \mu,$$

implying that for $n \geq 1$

$$E[X_n | X_{n-1}] = \mu \cdot X_{n-1}. \tag{2.19}$$

Taking expectations of (2.19) yields $\mathbb{E}[X_n] = \mu \cdot \mathbb{E}[X_{n-1}]$ for $n \geq 1$. Applying this relation repeatedly gives that for $n \geq 1$

$$\mathbb{E}[X_n] = \mu^n \mathbb{E}[X_0].$$

We can already draw some interesting conclusions from this calculation. Suppose that $\mu < 1$. Then, assuming $\mathbb{E}[X_0] < \infty$, we have that $\mathbb{E}[X_n] \to 0$, as $n \to \infty$. Moreover, we have the following crude estimate

$$\mathbb{E}[X_n] = \sum_{j=0}^{\infty} j P(X_n = j) \geq \sum_{j=1}^{\infty} P(X_n = j) = P(X_n \geq 1).$$

Therefore, if $\mu < 1$, we may conclude

$$P(X_n \geq 1) \leq \mathbb{E}[X_n] = \mu^n \mathbb{E}[X_0] \to 0, \qquad \text{as } n \to \infty.$$

Taking complements we find

$$\lim_{n \to \infty} P(X_n = 0) = 1. \tag{2.20}$$

Equation (2.20) says that the probability of the process having died out by time n goes to 1, as $n \to \infty$. However, this is not the same as saying "with probability one, the process eventually dies out," which we would write mathematically as

$$P\left(\lim_{n\to\infty} X_n = 0\right) = 1. \tag{2.21}$$

However, we can prove that equation (2.21) also holds in our current setting, which we do now.

Let $E_n = \{X_n = 0\}$ be the event that the process has died out by time n. Then, since 0 is an absorbing state, $E_n \subset E_{n+1}$, and $\{E_n\}$ is an increasing sequence of sets. Further,

$$\left\{\lim_{n\to\infty} X_n = 0\right\} = \bigcup_{i=1}^{\infty} E_i = \lim_{n\to\infty} \bigcup_{i=1}^{n} E_i = \lim_{n\to\infty} E_n$$

is the event that the process dies out at some time. By the continuity of probability functions (see [4]) we have

$$P\left(\lim_{n\to\infty} X_n = 0\right) = P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n) = \lim_{n\to\infty} P(X_n = 0) = 1. \tag{2.22}$$

Therefore, having (2.20) allows us to conclude that, with probability equal to one, the population eventually dies out. That is, (2.21) holds.

It is not yet entirely clear what happens when $\mu = 1$ or $\mu > 1$. In the case $\mu = 1$ we get the interesting result that $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all n, so the expected size of the population stays constant. When $\mu > 1$ the mean grows exponentially. However, this last fact does not imply that the process will not die out. To easily see this, simply note that there is always a non-zero probability that the population will go extinct in the next generation, regardless of how large the population is.

2.3.3 Probability of extinction

For $k, n \geq 0$ we define

$$a_n(k) = P_k(X_n = 0) = P(X_n = 0 \mid X_0 = k), \quad \text{and} \quad a(k) = \lim_{n\to\infty} a_n(k).$$

Note that by the argument in and around (2.22), we have

$$a(k) = \lim_{n\to\infty} P_k(X_n = 0) = P_k\left(\lim_{n\to\infty} X_n = 0\right).$$

and $a(k)$ is the probability that the population eventually dies out given an initial population of k.

If there are originally k people in the population, then for the entire population to die out, each of the k lineages (termed branches) must die out. Because the branches are all acting independently we may immediately conclude

$$a(k) = (a(1))^k,$$

and so it suffices to determine $a(1)$. We call $a(1)$ the *extinction probability* of the chain and denote it by a.

Let $H_0 = \inf\{n \geq 0 : X_n = 0\}$. We are interested in computing $a = P_1(H_0 < \infty)$. A first step analysis yields

$$a = \sum_{k=0}^{\infty} p_k P(H_0 < \infty | X_0 = k) = \sum_{k=0}^{\infty} p_k a(k) = \sum_{k=0}^{\infty} p_k a^k.$$

Hence, $a \in [0,1]$ satisfies the equation $a = \sum_{k=0}^{\infty} p_k a^k$, and if $p_k = 0$ for all but a finite number of k, then a is seen to be a root of a polynomial. However, polynomials can have multiple roots, and it is not yet clear how to uniquely identify a.

The key observation to make is that the function $\varphi(a) = \sum_{k=0}^{\infty} p_k a^k = \mathbb{E}[a^{Y_1^{(1)}}]$ is the probability generating function of the random variable $Y_1^{(1)}$. Since probability generating functions will play a large role in the remainder of this section, we pause for a mathematical aside to discuss them.

Mathematical aside (Probability generating functions)**.** The probability generating function of a random variable X is the function $\varphi(s) = \mathbb{E}[s^X]$. If we wish to highlight the random variable X in the notation, we will write φ_X.

We will assume that X takes values in $\{0, 1, 2, \dots\}$, in which case

$$\varphi(s) = \sum_{k=0}^{\infty} s^k P(X = k).$$

A few properties are immediate.

1. $\varphi(0) = P(X = 0)$ and $\varphi(1) = 1$.

2. The domain of convergence for the infinite sum includes $|s| \leq 1$. Therefore, the sum can be differentiated term-wise and for $|s| \leq 1$,

$$\varphi'(s) = \sum_{k=1}^{\infty} k s^{k-1} P(X = k), \quad \text{and} \quad \varphi''(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} P(X = k), \tag{2.23}$$

   where the second derivative can be taken term-wise so long as $\mathbb{E}[X] < \infty$.

3. From (2.23), we have

$$\varphi'(s) > 0, \quad \text{for } s \geq 0,$$

   so long as $P(X \geq 1) > 0$, and

$$\varphi''(s) > 0, \quad \text{for } s \geq 0,$$

   so long as $P(X \geq 2) > 0$. In this case, φ is strictly convex.

   Note that for the branching processes we are considering, we are assuming both $p_0 > 0$ and $p_0 + p_1 < 1$, showing that $\varphi_{Y_k^{(n)}}$ is strictly convex on $s \geq 0$.

4. Possibly defining the derivative as a limit from the left, we see

$$\varphi'(1) = \sum_{k=1}^{\infty} kP(X=k) = \mathbb{E}[X]$$
$$\varphi''(1) = \sum_{k=2}^{\infty} k(k-1)P(X=k) = \mathbb{E}[X^2] - \mathbb{E}[X].$$

5. Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables and that each of $\varphi_{X_i}(s)$ is finite for $0 \leq s \leq 1$. Then for $0 \leq s \leq 1$,

$$\varphi_{X_1+\cdots+X_n}(s) = \varphi_{X_1}(s)\varphi_{X_2}(s)\cdots\varphi_{X_n}(s).$$

This follows from a simple computation

$$\varphi_{X_1+\cdots+X_n}(s) = \mathbb{E}\left[s^{X_1+\cdots+X_n}\right] = \mathbb{E}\left[s^{X_1}\right]\cdots\mathbb{E}\left[s^{X_n}\right] = \varphi_{X_1}(s)\varphi_{X_2}(s)\cdots\varphi_{X_n}(s).$$

△

Returning to our study of branching processes. We see that the extinction probability a satisfies

$$a = \varphi_{Y_k^{(n)}}(a).$$

Of course, $a = 1$ is one solution to the the above equation. The question is whether or not there are more such solutions in the interval $[0, 1]$. To begin answering this question, we note that by the strict convexity of $\varphi_{Y_k^{(n)}}$, we may conclude that there can be at most two solutions to the equation on the interval $[0, 1]$. See Figure 2.1. Thus, we must only answer the following: (i) when will there be two solutions, and (ii) if there are two solutions, which solution provides the desired probability?

It will be useful to find the probability generating function of $X_n, n \geq 1$, given that $X_0 = 1$. We will show that, conditioned on $X_0 = 1$,

$$\varphi_{X_n}(s) = \varphi_{Y_1^{(1)}} \circ \varphi_{Y_1^{(1)}} \circ \cdots \circ \varphi_{Y_1^{(1)}} = \varphi^{(n)}_{Y_1^{(1)}} \quad (n \text{ compositions}). \tag{2.24}$$

First, note that (2.24) holds by definition in the case $n = 1$ (recall that we are assuming $X_0 = 1$). We show the general case by induction on n. By a first step analysis we have

$$\begin{aligned}\varphi_{X_n}(s) &= \sum_{k=0}^{\infty} P_1(X_n = k)s^k = \sum_{k=0}^{\infty}\left[\sum_{j=0}^{\infty} P_1(X_1 = j)P_1(X_n = k|X_1 = j)\right]s^k \\ &= \sum_{j=0}^{\infty} p_j\left[\sum_{k=0}^{\infty} P(X_{n-1} = k \mid X_0 = j)s^k\right] = \sum_{j=0}^{\infty} p_j\mathbb{E}_j\left[s^{X_{n-1}}\right].\end{aligned} \tag{2.25}$$

The expectation in the last sum above is the probability generating function of X_{n-1} conditioned upon $X_0 = j$. However, this term can be simplified.

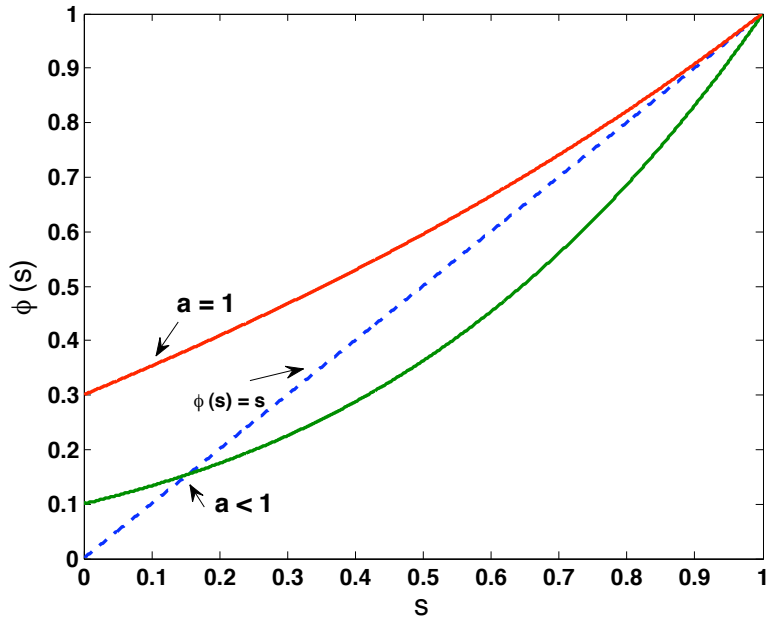


Figure 2.1: The different ways that the generating function can intersect the line $y = s$ on the interval $[0, 1]$. There are two possibilities: (a) one intersection at $s = 1$ or (b) one intersection at $s < 1$ and one intersection at $s = 1$. In the case that there is an intersection at a point strictly before 1, the point of intersection is the extinction probability, a.

For $\ell \in \{1, \ldots, j\}$, let $G_\ell^{(n-1)}$ denote the number of descendants at time $(n-1)$ of the ℓth individual of the original population. We see that, given an initial population of size j, we have $X_{n-1} = G_1^{(n-1)} + \cdots + G_j^{(n-1)}$. Further, the $G_\ell^{(n-1)}$ are independent. Note that $G_\ell^{(n-1)}$ has the same distribution as the random variable X_{n-1} given an initial population of $X_0 = 1$. In particular, we have that

$$\mathbb{E}_j \left[s^{G_\ell^{(n-1)}} \right] = \mathbb{E}_1 \left[s^{X_{n-1}} \right] = \varphi_{Y_1^{(1)}}^{(n-1)}(s), \tag{2.26}$$

where the final equality is from our inductive hypothesis.

Returning to the expectation in (2.25),

$$\mathbb{E}_j \left[s^{X_{n-1}} \right] = \mathbb{E}_j \left[s^{G_1^{(n-1)} + \cdots + G_j^{(n-1)}} \right] = \prod_{\ell=1}^{j} \mathbb{E}_j \left[s^{G_\ell^{(n-1)}} \right] = \left(\varphi_{Y_1^{(1)}}^{(n-1)}(s) \right)^j, \tag{2.27}$$

where the final equality is simply (2.26). Substituting (2.27) into (2.25) yields

$$\varphi_{X_n}(s) = \sum_{j=0}^{\infty} p_j \left(\varphi_{Y_1^{(1)}}^{(n-1)}(s) \right)^j = \varphi_{Y_1^{(1)}} \left(\varphi_{Y_1^{(1)}}^{(n-1)}(s) \right) = \varphi_{Y_1^{(1)}}^{(n)}(s),$$

giving the desired result.

We are ready to prove the following lemma.

Lemma 2.12. *Let $H_0 = \inf\{n \geq 0 : X_n = 0\}$ and $a = P_1(H_0 < \infty)$ be the extinction probability of the branching process. Then a is the smallest positive root of the equation $s = \varphi_{Y_1^{(1)}}(s)$.*

Proof. We already know that $a = \lim_{n\to\infty} a_n(1)$ must satisfy the equation $s = \varphi_{Y_1^{(1)}}(s)$. Let r denote any root of the equation contained in the interval $[0, 1]$. We will show that $a_n(1) = P_1(X_n = 0) \leq r$ for all $n \geq 0$. This will imply that $a = \lim_{n\to\infty} a_n(1) \leq r$, giving the result.

To prove the desired inequality, we first note that it is trivially true for $n = 0$ because $a_0(1) = 0$. We now proceed by induction on n. Assuming $a_{n-1}(1) \leq r$, we have

$$\begin{aligned} a_n(1) = P_1(X_n = 0) = \varphi_{X_n}(0) = \varphi^{(n)}_{Y_1^{(1)}}(0) = \varphi_{Y_1^{(1)}}(\varphi^{(n-1)}_{Y_1^{(1)}}(0)) \\ = \varphi_{Y_1^{(1)}}(\varphi_{X_{n-1}}(0)) = \varphi_{Y_1^{(1)}}(a_{n-1}(1)) \leq \varphi_{Y_1^{(1)}}(r) = r, \end{aligned}$$

where we have utilized (2.24) twice and the inequality holds because $\varphi_{Y_1^{(1)}}$ is a non-decreasing function. □

Henceforth we denote by φ the probability generating function for $Y_1^{(1)}$, $\varphi_{Y_1^{(1)}}$.

Example 2.13. Suppose that $p_0 = 0.3$, $p_1 = 0.6$, $p_2 = 0.05$, $p_3 = 0.05$. Then $\mu = 0.85$ and

$$\varphi(a) = 0.3 + 0.6a + 0.05a^2 + 0.05a^3.$$

Solving the equation $a = \varphi(a)$ yields $a \in \{1, 1.65\}$. Thus, the extinction probability is 1. △

Example 2.14. Suppose that $p_0 = 0.2$, $p_1 = 0.2$, $p_2 = 0.3$, $p_3 = 0.3$. Then $\mu = 1.7$ and

$$\varphi(a) = 0.2 + 0.2a + 0.3a^2 + 0.3a^3.$$

Solving the equation $a = \varphi(a)$ yields $a \in \{1, 0.291\}$. Thus, the extinction probability is 0.291. △

Example 2.15. Suppose that $p_0 = \frac{1}{4}$, $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$. Then $\mu = 1$ and

$$\varphi(a) = \tfrac{1}{4} + \tfrac{1}{2} \cdot a + \tfrac{1}{4} \cdot a^2.$$

Solving the equation $a = \varphi(a)$ yields $a \in \{1, 1\}$. Thus, the extinction probability is 1. △

We now establish the criteria for when $a < 1$.

Theorem 2.16. *For a branching process with $p_0 + p_1 < 1$ and $p_0 > 0$, the extinction probability a satisfies $a = 1$ if and only if $\mu = \mathbb{E}[Y_1^{(1)}] \leq 1$.*

Proof. Note that we have already proven that if $\mu < 1$, then $a = 1$. We now suppose that $\mu = 1$. In this case we know that

$$\varphi'(1) = \mathbb{E}[Y_1^{(1)}] = \mu = 1.$$

By the convexity of φ we must have $\varphi'(s) < 1$ for $s < 1$. Thus, for any $s < 1$ we have

$$1 - \varphi(s) = \int_s^1 \varphi'(s)ds < 1 - s.$$

That is, $\varphi(s) > s$. Therefore, there can be no root less than one and we have shown that in this case the extinction probability is one.

Now we consider the case $\mu > 1$. Then,

$$\varphi'(1) = \mu > 1.$$

Since we also know that $\varphi(1) = 1$, there must be an $s < 1$ with $\varphi(s) < s$. However, $\varphi(0) = p_0 > 0$, and so by continuity we have that there must be some $a \in (0, 1)$ for which $\varphi(a) = a$. Since convexity of φ guarantees there can be at most two roots, Lemma 2.12 shows that this minimal root is the extinction probability. □

The case $\mu = 1$ is quite interesting. For example we see that when $\mu = 1$,

$$\begin{aligned}\mathbb{E}[X_n \mid X_n > 0] &= \sum_{k=0}^{\infty} kP(X_n = k \mid X_n > 0) = \sum_{k=0}^{\infty} k \cdot \frac{P(X_n = k, X_n > 0)}{P(X_n > 0)} \\ &= \frac{1}{P(X_n > 0)} \sum_{k=1}^{\infty} kP(X_n = k) = \frac{1}{P(X_n > 0)}\mathbb{E}[X_n] \\ &= \frac{1}{P(X_n > 0)},\end{aligned}$$

which tends to ∞ as $n \to \infty$. Therefore, the expected value of the size of the population, conditioned on the population having survived to time n, goes to infinity.

Definition 2.17. The mean of the offspring distribution, $\mathbb{E}[Y_1^{(1)}] = \mu$, is called as the *criticality parameter*.

- If $\mu < 1$, then the process is called *subcritical*.
- If $\mu = 1$, then the process is called *critical*.
- If $\mu > 1$, then the process is called *supercritical*.

Note that we can now give the solution to Galton's original question, where there were N males worried about the longevity of their surnames. After n generations, the probability that a given surname has gone extinct is $a_n(1)$ and the probability that precisely j surnames have gone extinct is

$$\binom{N}{j} a_n(1)^j (1 - a_n(1))^{N-j}.$$

Because $a_n(1) = \varphi^{(n)}(0)$, these values can be solved for iteratively. If $a < 1$ we see that as $n \to \infty$ the probability that exactly j surnames eventually go extinct is

$$\binom{N}{j} a^j (1-a)^{N-j},$$

that is, the probability distribution is binomial(N, a).

We close this section with an example taken from [1].

Example 2.18. Suppose that a mutant gene appears in N individuals in a certain population, and suppose that the population grows according to a branching process. We suppose that the mean number of individuals produced by those with the mutant gene is μ, and that for some small $\varepsilon > 0$

$$\mu = 1 + \varepsilon.$$

What is the approximate probability that the mutant gene will become extinct?

Let a denote the extinction probability (when the population starts with one mutant). We note that $a \approx 1$ because (i) $\varphi'(1) = \mu = 1 + \varepsilon$, and (ii) φ is strictly convex. We will now approximate the value of a.

We begin by changing variables by defining θ via the equation $a = e^\theta$. Because $a \approx 1$, we know $\theta \approx 0$. We define $M(s) = \varphi(e^s)$, and note that $M(\theta) = \varphi(e^\theta) = \varphi(a) = a$. Next define $K(s) = \ln(M(s))$ and note that

$$K(\theta) = \ln(M(\theta)) = \ln(a) = \ln(e^\theta) = \theta. \tag{2.28}$$

We will use the first few terms in the Taylor expansion of K around $\theta \approx 0$, and use these, in conjunction with (2.28), to approximate θ, and hence a.

It is easy to check that $K(0) = 0$, $K'(0) = \mu$, and $K''(0) = \sigma^2 = \text{Var}(Y_1^{(1)})$. For example, we have

$$K(0) = \ln(M(0)) = \ln(\varphi(e^0)) = \ln(\varphi(1)) = \ln(1) = 0.$$

The calculations for $K'(0) = \mu$ and $K''(0) = \sigma^2 = \text{Var}(Y_1^{(1)})$ are left as Exercise 2.11.

Expanding the left hand side of (2.28) using a Taylor series, we have

$$\theta \approx K(0) + K'(0)\theta + \frac{1}{2}K''(0)\theta^2 = \mu \cdot \theta + \sigma^2 \cdot \frac{\theta^2}{2}.$$

Solving this equation, we find that

$$\theta \approx \frac{2}{\sigma^2}(1-\mu) = -\frac{2}{\sigma^2}\varepsilon.$$

Exponentiating both sides yields

$$a \approx e^{-2\varepsilon/\sigma^2}.$$

For an initial size of N mutants we have

$$a^N = P(\text{all mutant genes eventually go extinct}) \approx e^{-2N\varepsilon/\sigma^2}.$$

For example, in the case that the birth probabilities are Poisson, $m = \sigma^2 = 1 + \varepsilon$. Taking $\varepsilon = 0.01$, we have

$$a^N \approx e^{-2N(0.01)/1.01} \approx 0.98039^N.$$

The probability that the gene becomes established is $1 - a^N$. See Table 2.1. △

N	a^N	$1-a^N$
1	0.9804	0.0196
100	0.1380	0.8620
200	0.0191	0.9809
300	0.0026	0.9974

Table 2.1: Approximation of the probability that the gene goes extinct, a^N, or becomes established, $1-a^N$, under the assumptions that there are originally N mutant genes, each with a Poisson offspring distribution with mean 1.01.

2.4 Exercises

Exercise 2.1. Consider the genetics inbreeding problem with transition matrix (2.1). Compute $p_{3,i}$ for each $i \in \{1, 2, \ldots, 6\}$.

Exercise 2.2. Consider the single server queue. We found that when $\lambda < \mu$, the stationary distribution is given by (2.12). What is the expected length of the queue in equilibrium. What happens as $\lambda \to \mu$?

Exercise 2.3. Consider a birth-death process with $q_i = 1/4$, for all $i \geq 1$ with $q_0 = 0$. Suppose

$$p_i = \frac{1}{4} \cdot \frac{i+1}{i+2}, \quad i \geq 0.$$

Note that $p_i < q_i$ for all i, and $p_i \to q_i$, as $i \to \infty$. Is this chain transient, positive recurrent, or null recurrent?

Exercise 2.4. Consider a birth-death process with $q_i = 1/4$, for all $i \geq 1$ (with, as always $q_0 = 0$). Suppose that

$$p_i = \frac{1}{4} \cdot \frac{i+2}{i+1}, \quad i \geq 0.$$

Note that $p_i > q_i$ for all i, and $p_i \to q_i$, as $i \to \infty$. Is this chain transient, positive recurrent, or null recurrent?

Exercise 2.5. Consider a birth-death process with $q_i = 1/4$, for all $i \geq 1$ with $q_0 = 0$. Suppose

$$p_i = \frac{1}{4} \cdot \left(\frac{i+1}{i+2}\right)^2, \quad i \geq 0.$$

Note that $p_i < q_i$ for all i, and $p_i \to q_i$, as $i \to \infty$. Is this chain transient, positive recurrent, or null recurrent?

Exercise 2.6. Consider a birth-death process with $q_i = 1/4$, for all $i \geq 1$ with $q_0 = 0$. Suppose

$$p_i = \frac{1}{4} \cdot \left(\frac{i+2}{i+1}\right)^2, \quad i \geq 0.$$

Note that $p_i > q_i$ for all i, and $p_i \to q_i$, as $i \to \infty$. Is this chain transient, positive recurrent, or null recurrent?

Exercise 2.7. Consider a birth and death process with $q_i = 1/4$ if $i \geq 1$, and $q_0 = 0$, and

$$p_i = \frac{1}{4} \cdot \left(\frac{i+1}{i+2} \right)^2, \quad i \geq 0.$$

Note that this is the same chain as in Problem 2.5 above. We will again use Theorem 1.78 to estimate the stationary distribution. Simulate this process, with $X_0 = 0$, and average over the path to estimate $\pi_i = \lim_{n\to\infty} P(X_n = i)$, for $i \in \{0, 1, 2, 3, 4, 5\}$. Note that this problem is similar to (and in some ways easier, even though the state space is infinite) that of 1.16 of Chapter 3.

Exercise 2.8 (Lawler, 2006)**.** Given a branching process with the following offspring distributions, determine the extinction probability a.

1. $p_0 = 0.25,\ p_1 = 0.4,\ p_2 = 0.35.$
2. $p_0 = 0.5,\ p_1 = 0.1,\ p_3 = 0.4.$
3. $p_0 = 0.91,\ p_1 = 0.05,\ p_2 = 0.01, p_3 = 0.01,\ p_6 = 0.01,\ p_{13} = 0.01.$
4. $p_i = (1-q)q^i$, for some $q \in (0, 1)$.

Exercise 2.9. Consider again the branching process with $p_0 = 0.5,\ p_1 = 0.1,\ p_3 = 0.4$, and suppose that $X_0 = 1$. What is the probability that the population is extinct in the second generation ($X_2 = 0$), given that it did not die out in the first generation ($X_1 > 0$)?

Exercise 2.10 (Lawler, 2006)**.** Consider a branching process with $p_0 = \frac{1}{3},\ p_1 = \frac{1}{3},\ p_2 = \frac{1}{3}$. Find, with the aid of a computer, the probability that the population dies out in the first n steps for $n = 20, 100, 200, 1000, 1500, 2000, 5000$. Do the same with the values $p_0 = 0.35, p_1 = 0.33, p_2 = 0.32$, and then do it for $p_0 = 0.32,\ p_1 = 0.33$, and $p_2 = 0.35$.

Exercise 2.11. Let K be defined via (2.28). Show that $K'(0) = \mu = \mathbb{E}[Y_1^{(1)}]$ and $K''(0) = \sigma^2 = \mathrm{Var}(Y_1^{(1)})$.

Exercise 2.12. Consider Example 2.18 with $Y_1^{(1)}$ being a Poisson random variable with mean 1.01. With the aid of a computer, find a and compare with the approximate value $a \approx 0.98039$.

Exercise 2.13. Suppose a branching process has a Poisson offspring distribution

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

1. Supposing that $X_0 = 1$, find the mean of X_n, the size of the population at time n.
2. For $\lambda = 1.5$ and $\lambda = 2$, find the probability that the process eventually goes extinct.

Exercise 2.14. We consider a model of replication for polymer chains. We consider a polymer chain consisting of m nucleotides that will itself be copied, and whose offspring chains will be copied, etc. We assume a fixed probability of p that a single nucleotide is correctly copied during replication. Thus, the probability that the entire chain is copied correctly is p^m. We assume that the chain replicates at discrete times. We also suppose that during each time window, any polymer chain is destroyed with a probability of $1-q$, and survives with a probability of q. Therefore, a given polymer can yield zero (if it is destroyed), one (if it survives but is copied incorrectly), or two (survives and is replicated perfectly) exact replica chains in the next time step with respective probabilities

$$p_0 = 1 - q, \quad p_1 = q(1 - p^m), \quad p_2 = qp^m.$$

For a given p and q, determine the threshold for m, the size of the polymer, for which extinction of the exact chain is not assured. That is, give a condition on m for which the probability of extinction is less than one.

For a survival rate of $q = 0.8$, determine how long the chain has to be so that extinction is not guaranteed for the cases $p = 0.2,\ 0.4,\ 0.5,\ 0.9,\ 0.99$. Finally, using any formulas you have computed, what happens when $q < 1/2$? Interpret this result. Does it make sense?

Exercise 2.15. Consider a branching process with

$$p_0 = \frac{1}{2}, \quad p_1 = \frac{1}{4}, \quad \text{and} \quad p_2 = \frac{1}{4}.$$

For this chain, we have

$$\mu = p_1 + 2p_2 = \frac{3}{4},$$

and so extinction is gauranteed. Let $X_0 = 1$, and denote the time at which the chain goes extinct by T_{extinct}. Write a Matlab script that simulates individual realizations of this branching process. Use your script to estimate $E[T_{\text{extinct}}]$ via Monte Carlo.

Exercise 2.16. Prove Theorem 2.1 in the setting of the Wright-Fisher model. That is, let X_n be a Markov chain with transition probabilities (2.2). Let $\lim_{n\to\infty} X_n = X_\infty$, which is a random variable taking the value 0 with probability $g_0(j)$ and the value $2N$ with probability $g_{2N}(j)$. Show that $\lim_{n\to\infty} \mathbb{E}_j[X_n] = \mathbb{E}_j[X_\infty] = 0 \cdot g_0(j) + 2N \cdot g_{2N}(j)$.

Hint. First, use Jensen's inequality and an argument similar to that of the usual proof of Markov's inequality to show that for any $0 < \varepsilon < 1$,

$$|\mathbb{E}_j[X_n] - \mathbb{E}_j[X_\infty]| \le \mathbb{E}_j[|X_n - X_\infty|] \le \varepsilon + 2MP_j(|X_n - X_\infty| > \varepsilon).$$

Next argue that for $j \in \{1, \ldots, 2N-1\}$, we have

$$P_j(|X_n - X_\infty| > \varepsilon) = \sum_{k=1}^{2N-1} p_{j,k}^{(n)} \to 0, \qquad \text{as } n \to \infty.$$

Next, as ε was arbitrary, conclude the result.

Exercise 2.17. Consider a branching process with offspring distribution

$$p_0 = 1 - p, \quad \text{and} \quad p_2 = p,$$

where $p \in [0, 1]$.

(a) For what values of p is extinction guaranteed?

(b) In the case that extinction may be avoided, calculate the probability of extinction as a function of the parameter p.

Exercise 2.18 (Kimura 2-parameter model, [21])**.** One problem with the Jukes-Cantor model of Section 2.1.3 is that it gives each mutation the same probability of happening. However, the nucleotides A and G are purines and the nucleotides C and T are pyrimidines and it is more likely for a purine to transition to the other purine and a pyrimidine to transition to the other pyrimidine. Mutations between A and G or between C and T are called *transitions* whereas the other mutations are called transversions. Suppose that the probability of a transition is α and the probability of a specific transversion, such as A $\to$ C, is β. Then the transition matrix for the model is

$$P = \begin{pmatrix} 1-\alpha-2\beta & \alpha & \beta & \beta \\ \alpha & 1-\alpha-2\beta & \beta & \beta \\ \beta & \beta & 1-\alpha-2\beta & \alpha \\ \beta & \beta & \alpha & 1-\alpha-2\beta \end{pmatrix},$$

where both α and β are small, and $\alpha > 2\beta$. This is the Kimura 2-parameter model. Note that the expected number of mutations per site over n generations is $n(\alpha+2\beta)$. We will derive the Kimura 2-parameter distance to approximate this value.

(a) Find the eigenvalues and left eigenvectors of P. Use them to determine the vector $(1, 0, 0, 0)P^n$ and, in particular, the probabilities $p_{12}^{(n)}, p_{13}^{(n)}$, and $p_{14}^{(n)}$.

(b) Consider two strands of DNA. Let γ_1 be the fraction of nucleotides that differ due to a transition difference (i.e. an A in one is a G in the other, or other such differences). Similarly, let γ_2 be the fraction differing due to a transversion difference. Why is it reasonable to set $\gamma_1 = p_{12}^{(n)}$ and $\gamma_2 = p_{13}^{(n)} + p_{14}^{(n)}$? Use the second equation to determine $2\beta n$ in a manner similar to how $n\rho$ was found for the Jukes-Cantor model.

(c) Substitute as needed by using $\gamma_2 = p_{13}^{(n)} + p_{14}^{(n)}$ in $\gamma_1 = p_{12}^{(n)}$ and solve for n as a function of α, β, γ_1, and γ_2. As in the Jukes-Cantor model, use an appropriate Taylor approximation of a logarithm to approximate $2(\alpha + \beta)n$ as a function of γ_1 and γ_2.

(d) Combine the previous two parts to conclude that

$$d_{K2P} = -\frac{1}{4}\ln\left((1-2\gamma_2)(1-2\gamma_1-\gamma_2)^2\right)$$

is a reasonable approximation to $n(\alpha+2\beta)$. The term d_{K2P} is the Kimura 2-parameter distance function.

Chapter 3

Renewal and Point processes

In this chapter we will study two classes of stochastic processes: renewal and point processes. The material in this chapter will play a critical role in the study of stochastic models of population processes found in Chapter 5.

3.1 Renewal Processes

A renewal process models occurrences of events happening at random times, where the times between the occurrences are determined from independent and identically distributed random variables. Specifically, let Y_n, $n \geq 1$, be a sequence of independent and identically distributed random variables which take non-negative values. We assume throughout that $P(Y_n = 0) < 1$. We also let Y_0 be a non-negative random variable, independent from $Y_n, n \geq 1$, though not necessarily of the same distribution. The range of these random variables could be discrete, perhaps $\{0, 1, 2, \dots\}$, or continuous, such as $[0, \infty)$. The random variables Y_n will be the inter-event times of the occurrences.

For each integer $n \geq 0$, we define the random variable S_n by

$$S_n = \sum_{i=0}^{n} Y_i. \tag{3.1}$$

For example,

$$S_0 = Y_0, \quad S_1 = Y_0 + Y_1, \quad \text{and} \quad S_2 = Y_0 + Y_1 + Y_2.$$

Note that the random variable S_n gives the time of the nth occurrence (we think of Y_0 as the "zeroth" occurrence). The sequence or random variables $\{S_n, n \geq 0\}$ is called a *renewal sequence*. The times S_n are called *renewal times*. The occurrences themselves are usually called the *renewals*.

The random variable Y_0 gives the time until the zeroth occurrence. If $P(Y_0 > 0) > 0$, then the process is called *delayed*. If, on the other hand, we have that $P(Y_0 = 0) = 1$, in which case $S_0 = Y_0 = 0$ with a probability of one, then the process is called *pure*.

Recalling that the indicator function, $1_A : \mathbb{R} \to \{0, 1\}$, is defined as

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases},$$

we define the counting process

$$N(t) = \sum_{n=0}^{\infty} 1_{[0,t]}(S_n) = \#\text{ of renewals up to and including time } t. \tag{3.2}$$

Note that if the process is pure, then $Y_0 = 0$ and $N(0) = 1$ with probability one. The counting function N will be our main object of study in this section. Note that N is function whose domain is all of $[0, \infty)$ even if Y_n are discrete random variables.

A few more definitions are required. If $P(Y_n < \infty) = 1$ for $n \geq 1$, then the renewal process is called *proper*. However, if $P(Y_n < \infty) < 1$, in which case $P(Y_n = \infty) > 0$ and $E[Y_n] = \infty$, then the process is called *defective*. Note that if the process is defective, then there will be a final renewal. In this case, $N(t)$ remains bounded with a probability of one, though the bound is a random variable. In fact, the bound can be determined from a geometric random variable with parameter $P(Y_n < \infty)$.

Our focus will be on trying to understand the large time behavior of the process N. We will derive both a law of large numbers, and a central limit theorem for $N(t)$, as $t \to \infty$. We point the interested reader to either [31] or [28] for a more complete treatment on renewal processes, including a study of the *renewal function*

$$U(t) \stackrel{\text{def}}{=} \mathbb{E}[N(t)] = \mathbb{E}\left[\sum_{n=0}^{\infty} 1_{[0,t]}(S_n)\right].$$

Example 3.1. Let $\{X_n, n \geq 0\}$ be an irreducible, positive recurrent, discrete time Markov chain with state space S. For $i \in S$, let

$$S_0 = \inf\{n \geq 0 \ : \ X_n = i\},$$

and for $n \geq 1$

$$S_n = \inf\{j > S_{n-1} \ : \ X_j = i\}.$$

Let $Y_0 = S_0$ and for $n \geq 1$ let

$$Y_n = S_n - S_{n-1}.$$

By the strong Markov property the random variables $Y_n, n \geq 1$, are independent and identically distributed. Thus, $\{S_n, n \geq 0\}$ is a sequence of renewal times. Since the Markov chain is recurrent, the renewal process is proper. Since the chain is positive recurrent, we know $E[Y_1] < \infty$.

Note that if $X_0 = i$, then the process is pure, whereas if $X_0 \neq i$ then the process is delayed. The process $N(t) = \sum_{k=0}^{\infty} 1_{[0,t]}(S_n)$ counts the number of times the process has returned to state i by time t. For example, if $S_0 = 4$, $S_1 = 10$, and $S_2 = 11$ are

the first three times the chain enters state i, then

$$N(t) = \begin{cases} 0 & 0 \le t < 4 \\ 1 & 4 \le t < 10 \\ 2 & 10 \le t < 11 \\ 3 & 11 \le t < 12, \end{cases}$$

with $N(t)$ as yet undetermined for $t \ge 12$. In particular, N is defined for all t, such as $N(10.5) = 2$.

If the Markov chain were irreducible and transient, instead of recurrent, then the associated renewal process defined above is defective, and there is a last time the process returns to state i. △

Example 3.2. At a certain nursing home, a particular resident will periodically come into contact with a particular virus. Suppose that the times between such contacts are well modeled by independent random variables with probability density function

$$f(x) = \begin{cases} \frac{2\sqrt{2}}{\pi(x^4+1)} & x > 0 \\ 0 & x \le 0 \end{cases} \tag{3.3}$$

where the units of t are in weeks. Letting Y_0 and $Y_n, n \ge 1$, have the density (3.3), S_n determined by (3.1) models the times at which the resident comes into contact with this virus. The process $N(t)$ defined in (3.2) gives the number of times the resident has come into contact with the virus by time t. △

3.1.1 The behavior of $N(t)$, as $t \to \infty$

Our goal in this section is to characterize the behavior of $N(t)$, as $t \to \infty$, where we preserve the notation from the previous section. We begin with a law of large numbers result.

Theorem 3.3. *Let $N(t)$ be the counting process (3.2) associated with the renewal sequence $\{S_n, n \ge 0\}$. Let $\mu = \mathbb{E}[Y_1]$. Then, with a probability of one,*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu},$$

where the right-hand side is interpreted as zero if $\mu = \infty$.

Proof. Note that if the process is defective, then (i) $\mu = \infty$, and (ii) $N(t)$ is uniformly bounded in t. In this case $1/\mu = 0$ and $\lim_{t\to\infty} N(t)/t = 0$. Thus, the result holds when the process is defective.

We now assume the process is proper. We begin by recalling the strong law of large numbers [4]. Suppose that $Z_n, n \ge 1$, are independent and identically distributed random variables with $\mathbb{E}[Z_1] \le \infty$. Then, with a probability of one

$$\frac{1}{n}(Z_1 + \cdots + Z_n) \to \mathbb{E}[Z_1], \qquad \text{as } n \to \infty.$$

Returning to our processes, the strong law of large numbers states that with a probability of one

$$\lim_{n\to\infty} \frac{Y_1 + \cdots + Y_n}{n} = \mu,$$

with the equality holding even if $\mu = \infty$. Thus, we may conclude that with probability one,

$$\frac{S_n}{n} = \frac{Y_0}{n} + \frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu, \text{ as } n \to \infty.$$

Noting that $N(t) \to \infty$, as $t \to \infty$, we have that

$$\frac{1}{N(t)} S_{N(t)} \to \mu, \text{ as } t \to \infty, \tag{3.4}$$

with a probability of one. By construction, for $N(t) \geq 1$ we have

$$S_{N(t)-1} \leq t < S_{N(t)}.$$

Therefore, so long as $N(t) \geq 2$,

$$\frac{S_{N(t)-1}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)}}{N(t)} \implies \frac{S_{N(t)-1}}{N(t)-1} \times \frac{N(t)-1}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)}}{N(t)}.$$

Applying (3.4), we see that

$$\frac{t}{N(t)} \to \mu \implies \frac{N(t)}{t} \to \frac{1}{\mu},$$

as $t \to \infty$, and the result is shown. □

Theorem 3.3 says that

$$N(t) = \frac{t}{\mu} + o(t), \text{ as } t \to \infty. \tag{3.5}$$

A natural next question is: what is the next term in the expansion? This is a *central limit theorem* type of question in that we are asking for the fluctuations around

$$N(t) - \frac{t}{\mu}.$$

We begin by defining

$$\sigma^2 \stackrel{\text{def}}{=} \text{Var}(Y_i),$$

for $i \geq 1$, and we assume that $\sigma^2 < \infty$.

Theorem 3.4. *As $t \to \infty$, the distribution of*

$$\frac{N(t) - \mu^{-1}t}{\sigma\mu^{-3/2}\sqrt{t}},$$

approaches that of a standard normal. That is,

$$\lim_{t\to\infty} P\left(\frac{N(t) - t\mu^{-1}}{\sqrt{t\sigma^2\mu^{-3}}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{3.6}$$

Note that Theorem 3.4 loosely states that

$$\frac{N(t) - \mu^{-1}t}{\sigma\mu^{-3/2}\sqrt{t}} \approx B, \tag{3.7}$$

where B is a Gaussian random variable with mean zero and variance one. Rearranging terms in (3.7) yields,

$$N(t) \approx \frac{t}{\mu} + \frac{\sigma}{\mu^{3/2}}\sqrt{t} \cdot B,$$

showing that we are indeed uncovering the next term in the expansion of $N(t)$.

Proof of Lemma 3.4. We begin by recalling the usual central limit theorem, see [4], for the discrete time process $S_n, n \geq 1$. It is

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt. \tag{3.8}$$

This expression will be useful in the derivation of the limit (3.6).

Instead of simply proving that (3.6) holds, which would not give much intuition, we will show *why* the process

$$\frac{N(t) - t\mu^{-1}}{\sqrt{t\sigma^2\mu^{-3}}}$$

is properly scaled to give the desired convergence. Specifically, we will search for two functions, h and g, for which

$$Z(t) \stackrel{\text{def}}{=} \frac{N(t) - h(t)}{g(t)} \Longrightarrow B, \quad \text{as } t \to \infty,$$

where B is a standard Gaussian and the convergence is in the sense of distributions (as in (3.6)).

Discovering the proper function h is straightforward. Theorem 3.3 implies that we must choose $h(t) = t\mu^{-1} + o(t)$ in order for $Z(t)$ to converge to a mean zero random variable as $t \to \infty$. Hence, we take $h(t) = t\mu^{-1}$ and turn our attention to the function $g(t)$ in the denominator.

Noting that we have the central limit theorem (3.8) for the process S_n, we would like to convert statements about N into equivalent statements pertaining to S_n. The key observations is that for any $n \geq 0$ we have that

$$N(t) \leq n \quad \text{is equivalent to} \quad S_n > t,$$

both of which state that the number of renewals by time t is less than or equal to n. Denoting by $\lfloor y \rfloor$ the greatest integer less than or equal to $y \in \mathbb{R}$, we have

$$\begin{aligned} P\left(\frac{N(t) - t\mu^{-1}}{g(t)} \leq x\right) &= P\left(N(t) \leq \lfloor t\mu^{-1} + g(t)x \rfloor\right) \\ &= P\left(S_{\lfloor t\mu^{-1} + g(t)x \rfloor} > t\right) \\ &= P\left(\frac{S_{\lfloor t\mu^{-1} + g(t)x \rfloor} - \mu\lfloor t\mu^{-1} + g(t)x \rfloor}{\sigma\sqrt{\lfloor t\mu^{-1} + g(t)x \rfloor}} > \frac{t - \mu\lfloor t\mu^{-1} + g(t)x \rfloor}{\sigma\sqrt{\lfloor t\mu^{-1} + g(t)x \rfloor}}\right). \end{aligned} \tag{3.9}$$

Because of (3.8) and (3.9), we see that we may conclude

$$P\left(\frac{N(t)-t\mu^{-1}}{g(t)} \le x\right) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt, \quad \text{as } t\to\infty,$$

so long as two conditions are met. The first condition is that

$$\lfloor t\mu^{-1}+g(t)x\rfloor \to \infty, \quad \text{as } t\to\infty,$$

and the second condition is that

$$\frac{t-\mu\lfloor t\mu^{-1}+g(t)x\rfloor}{\sigma\sqrt{\lfloor t\mu^{-1}+g(t)x\rfloor}} \to -x, \quad \text{as } t\to\infty.$$

After dropping the floor function (since the terms will be large from the first condition), the second condition is equivalent to

$$\frac{-\mu g(t)x}{\sigma\sqrt{t\mu^{-1}+g(t)x}} \to -x, \text{ as } t\to\infty,$$

which tells us that $g(t)$ must scale like $O(\sqrt{t})$. Therefore, writing $g(t)=\sqrt{t}K$, we see that the second condition is satisfied so long as K can be chosen in such a way that

$$\frac{-\mu\sqrt{t}Kx}{\sigma\sqrt{t\mu^{-1}+\sqrt{t}Kx}} = \frac{-\mu Kx}{\sigma\sqrt{\mu^{-1}+t^{-1/2}Kx}} \to -x, \quad \text{as } t\to\infty.$$

Since $t^{-1/2}Kx\to 0$ as $t\to\infty$, we arrive simply at the requirement

$$\frac{-\mu Kx}{\sigma u^{-1/2}} = -x,$$

which is equivalent to setting $K=\sigma u^{-3/2}$. We conclude that the second condition is satisfied if we let

$$g(t)=\sqrt{t\sigma^2\mu^{-3}}.$$

Note that in this case, condition (i) is also satisfied and the result is proved. □

Example 3.5 (Example 3.2 continued)**.** We again consider a resident in a nursing home who sporadically comes in contact with a particular virus. Specifically, let $Y_n, n\ge 0$, be independent random variables with density function (3.3), and let S_n be defined via (3.1). The random variables $S_n, n\ge 0$, give the times at which the resident comes in contact with the virus, and $N(t)=\sum_{k=0}^{\infty} 1_{[0,t]}(S_n)$ gives the number of times the patient has been exposed to the virus by time t, where the time scale is in weeks. Approximate the probability that this resident comes in contact with the virus at least 230 times over the next three years.

We have that

$$\mu = E[Y_1] = \int_0^\infty x\cdot\frac{2\sqrt{2}}{\pi(x^4+1)}dx = \frac{1}{\sqrt{2}},$$

and

$$E[Y_1^2] = \int_0^\infty x^2 \cdot \frac{2\sqrt{2}}{\pi(x^4+1)} dx = 1,$$

in which case

$$\sigma^2 = \mathsf{Var}(Y_1) = 1 - \tfrac{1}{2} = \tfrac{1}{2}.$$

Since there are 52 weeks in a year, we take $t = 156$. We have,

$$P(N(156) \geq 230) = P\left(\frac{N(156) - 156 \cdot \sqrt{2}}{\sqrt{156 \cdot \frac{1}{2} \cdot 2^{3/2}}} \geq \frac{230 - 156 \cdot \sqrt{2}}{\sqrt{156 \cdot \frac{1}{2} \cdot 2^{3/2}}} \right)$$

$$= P\left(\frac{N(156) - 156 \cdot \sqrt{2}}{\sqrt{156 \cdot \frac{1}{2} \cdot 2^{3/2}}} \geq 0.63 \right).$$

By Theorem 3.4, this probability is well approximated by $P(B \geq 0.63)$ where B is a standard normal random variable. Hence, $P(N(156) \geq 230) \approx 0.26$. △

Another result from renewal theory, and one which should not be surprising at this point, is called the *elementary renewal theorem*. It is stated below without proof.

Theorem 3.6. *Let $\mu = \mathbb{E}[Y_1]$. Then,*

$$\lim_{t\to\infty} t^{-1}\mathbb{E}[N(t)] = \frac{1}{\mu},$$

where the right hand side is interpreted as zero if $\mu = \infty$.

3.1.2 The renewal reward process

A renewal reward process is only a slight variation of the renewal process. We now suppose that at each renewal time, S_n, we are given a random reward, R_n, where R_n can be positive, negative, or zero. We assume that the random variables $\{R_n, n \geq 0\}$ are independent, and that the sequence $\{R_n, n \geq 1\}$ are also identically distributed. However, we do not assume that the R_n are necessarily independent of S_n. For example, they could be functions of the inter-event times Y_n. We then define the *renewal reward* process to be

$$R(t) = \sum_{i=0}^{N(t)-1} R_i = \sum_{i=0}^{\infty} R_i \cdot 1_{[0,t]}(S_i), \tag{3.10}$$

where we sum to $N(t) - 1$ since the we begin counting renewals at $i = 0$. Thus, $R(t)$ gives the total accumulated *reward* up to time t.

Example 3.7. Consider an insurance company with claims coming in at times S_n. The sizes of the claims are R_n and the cummulative amount of claims against the company by time t is $R(t)$. △

Example 3.8. Consider a metabolic network that requires the amino acid methionine to run properly. Methionine is ingested when the organism eats food. Letting S_n denote the times the organism eats, and R_n the amount of methionine ingested at time S_n, the process $R(t)$ defined via (3.10) is the total amount of methionine ingested by time t. △

Theorem 3.9. *If* $\mathbb{E}[|R_j|] < \infty$ *and* $\mathbb{E}[Y_1] = \mu < \infty$, *then*

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R_1]}{\mu}.$$

Proof. We have

$$\lim_{t\to\infty} \frac{R(t)}{t} = \lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{N(t)-1} R_i = \lim_{t\to\infty} \left[\frac{N(t)}{t} \cdot \frac{1}{N(t)} \sum_{i=0}^{N(t)-1} R_i \right].$$

From Theorem 3.3 we know that

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}.$$

From the usual law of large numbers applied to the random variables R_i we have

$$\lim_{t\to\infty} \frac{1}{N(t)} \sum_{i=0}^{N(t)-1} R_i = \mathbb{E}[R_1],$$

implying the result. □

Example 3.10. Suppose there are two animals, which we will call R and F, sharing an ecosystem. F is aggressive and will attempt to fight R whenever it can. We suppose that R only comes within range of F sporadically, and that the average number of days between times they are close is $d > 0$. We also suppose that R can expend energy each day in an effort to avoid detection by F. Specifically, by using an energy amount of $x \geq 0$ calories per day, the probability of actually being seen by F (if F is close to R) is $p(x)$. That is, the probability is a function of x. Finally, suppose that if R is seen, then it will have to fight and expend an amount of energy which is random, but has a mean value of $V > 0$. How much energy, x, should R expend each day in order to minimize its long term total energy expenditure?

To answer the question, we model the situation with a renewal reward process. For $n \geq 1$, let S_n be the nth time that R and F are close to each other. For $n \geq 1$ let $Y_n = S_n - S_{n-1}$, where we define $S_0 = Y_0 = 0$. We know that $\mathbb{E}[Y_i] = d$ for $i \geq 1$. Finally, let R_i be the amount of energy expended by R during the ith time that R and F are close to each other.

Of course, F does not necessarily see R every time they are close. Therefore, let $Z_i = 1$ if R is forced to fight during the ith encounter and zero otherwise. Note that $P(Z_i = 1) = p(x)$. Finally, let X_i be the random variable giving the amount of energy

needed by R to fight during the ith encounter if it is seen. Note that $E[X_i] = V$. Assuming X_i and Z_i are independent, we have

$$E[R_i] = E[X_i 1(Z_i = 1)] = E[X_i]P(Z_i = 1) = V \cdot p(x).$$

Now let $N(t)$ count the number of times R and F have been near each other by time t. The total amount of energy used by time $t \geq 0$ is then given by

$$R(t) = xt + \sum_{i=1}^{N(t)-1} R_i,$$

where we began counting at $i = 1$ since we took $S_0 = 0$, which implies $N(0) = 1$. The long run average energy usage for R is then

$$\frac{R(t)}{t} = x + \frac{1}{t}\sum_{i=1}^{N(t)-1} R_i \to x + \frac{\mathbb{E}[R_1]}{\mathbb{E}[Y_1]} = x + \frac{V}{d}p(x), \quad \text{as } t \to \infty.$$

For example, suppose that $p(x) = e^{-x}$ and that $V > d$. Then,

$$C(x) = x + \frac{V}{d}e^{-x}$$

is minimized at $\bar{x} = \ln(V/d)$, and $C(\bar{x}) = 1 + \ln(\frac{V}{d})$. △

3.2 Point Processes

The basic idea of a point process is to allow us to model a random distribution of points in a space, usually a subset of Euclidean space such as $\mathbb{R}$, $[0, \infty)$, or $\mathbb{R}^d$, for $d \geq 1$. Here are a few examples.

Example 3.11. Renewal processes distribute points, S_n, on $[0, \infty)$ so that the gaps between points are independent and identically distributed random variables. △

Example 3.12. The Poisson process, which will be the main object of our focus, is a renewal process which distributes points so gaps are i.i.d. exponential random variables. △

Some other examples could include the following:

1. the breakdown times of a certain part of a car,
2. the position of proteins on a cell membrane,
3. the positions and times of earthquakes in the next 100 years,
4. the locations of diseased deer in a given region of Wisconsin.

3.2.1 Preliminaries

We begin with an important mathematical notion, that of a *measure*.

Definition 3.13. Let E be a subset of Euclidean space, and let $\mathcal{F}$ be a σ-algebra of E (think of this as the collection of subsets of E). Then, $\mu : \mathcal{F} \to \mathbb{R}$ is a *measure* if the following three conditions hold

1. For $A \in \mathcal{F}$, i.e. for A a subset of E, $\mu(A) \geq 0$.
2. If $\{A_i\}$ are disjoint sets of $\mathcal{F}$, then
$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i).$$
3. $\mu(\emptyset) = 0$.

The concept of a measure generalizes the idea of length in one dimension, area in two, etc. In fact, one of the most important measures is that of *Lebesgue* measure, which is precisely length, area, volume, etc. For example, if μ is Lebesgue measure on $\mathbb{R}$, then for any $a < b$ we have $\mu([a, b]) = b - a$, and if $[a, b]$ and $[c, d]$ are disjoint, then
$$\mu([a, b] \cup [c, d]) = (b - a) + (d - c).$$
If μ is Lebesgue measure on $\mathbb{R}^2$, then for any region of space $A \subset \mathbb{R}^2$, we have $\mu(A) = \text{Area}(A)$.

Another class of measures you are already familiar with is the class of probability measures. Note that the axioms of probability are simply the three conditions laid out above, with the added restriction that $P(E) = 1$ (where E is the sample space).

We now turn to point processes and start with some notation, terminology, and definitions. We suppose that E is a subset of Euclidian space, $\mathbb{R}^d$ (or $[0, \infty), \mathbb{R}^2$, etc.). Similar to our study of the renewal process in the previous section, we want to be able to distribute the points throughout E, and have a compact notation that counts the number of points that fall in a given subset $A \subset E$. For renewal processes, we distributed the points by assuming independent gaps between them, and let $N(t)$ denote the counting process giving the number of points up to time t.

We assume that $\{X_n, n \geq 0\}$ are random elements of E, which represent points in the state space E. Next, we define the discrete (random, as it depends upon the point X_n) measure by
$$1_{X_n}(A) = \begin{cases} 1, & \text{if } X_n \in A, \\ 0, & \text{if } X_n \notin A \end{cases}. \tag{3.11}$$
Note, therefore, that 1_{X_n} is a function whose domain is $\mathcal{F}$, i.e. the subsets of E, whose range is $\{0, 1\}$, and that takes the value one whenever X_n is in the subset of interest. Next, we note that by taking the sum over n, we find the total number of the points contained in the set A. Therefore, we define the counting measure N by
$$N = \sum_n 1_{X_n},$$

so that for $A \subset E$,

$$N(A) = \sum_n 1_{X_n}(A),$$

simply gives the total number of points in $A \subset E$.

Definition 3.14. The function N is called a *point process* on E, and $\{X_n\}$ are called the *points*.

We note that as N depends explicitly on the values of the points, X_n, it is natural to call such an object a *random measure* since the points themselves are random.

We will make the (technical) running assumption that bounded regions of A must always contain a finite number of points with a probability of one. That is, for any bounded set A,

$$P(N(A) < \infty) = 1.$$

Example 3.15. For a renewal process, we have $E = [0, \infty)$, and the points are the renewal times $\{S_n\}_{n=0}^{\infty}$. The point process is

$$N = \sum_{n=0}^{\infty} 1_{S_n}.$$

Note that the notation for the counting process has changed from $N(t)$ to $N([0, t])$. △

Example 3.16. Consider modeling the positions, and illnesses, of sick deer in a given region. A good choice for a state space would be

$$E = \mathbb{R}^2 \times \{1, 2, \ldots, M\},$$

where the first component of E determines the deer's location and the second lists its ailment. The point process would then be

$$N = \sum_n 1_{\{(L_{n1}, L_{n2}), m\}},$$

where (L_{n1}, L_{n2}) represents the latitude and longitude of the deer, and m its ailment. △

An important statistic of a point process is the *mean measure*, or *intensity*, of the process, which is defined to be

$$\mu(A) = \mathbb{E}[N(A)],$$

giving the expected number of points found in the region A. We note that the intensity is commonly referred to as the *propensity* in the biosciences.

3.2.2 The Poisson process

Poisson processes are a special type of point process. We will begin with a formulation for a one-dimensional Poisson process that most people see in their first probability course. We will later generalize to higher dimensions and variable intensities.

Definition 3.17. N is a *counting process* if $N(0) = 0$ and N is constant except for jumps of $+1$.

For concreteness, we assume throughout that all processes are cadlag, which means they are right continuous with left limits for all $t > 0$.

We will consider processes that satisfy the following assumption. Recall that a function f is said to be $o(h)$, and written $f \in o(h)$ or $f = o(h)$, if $f(h)/h \to 0$, as $h \to 0$.

Assumption 3.18. *Let N be an integer valued process satisfying:*

1. *For some $\lambda > 0$,*

$$P(N(t+h) - N(t) = 1) = \lambda h + o(h), \text{ as } h \to 0,$$

   for any $t \geq 0$.

2. *For any $t \geq 0$,*

$$P(N(t+h) - N(t) \geq 2) = o(h), \text{ as } h \to 0.$$

3. *The process has independent increments. That is, if $t_0 < t_1 < \cdots < t_m$, then $N(t_k) - N(t_{k-1})$, $k = 1, \ldots, m$, are independent random variables.*

Definition 3.19. If the process N satisfies Assumption 3.18 with $N(0) = 0$, it is called a *homogeneous Poisson process* with *intensity*, *propensity*, or *rate*, $\lambda > 0$.

The following proposition characterizes the distribution of the random variable $N(t) - N(s)$, and makes clear why the process N is termed a Poisson process.

Proposition 3.20. *Let N satisfy Assumption 3.18. Then, N is a counting process and for any $t \geq s \geq 0$ and any $k \in \{0, 1, 2, \ldots\}$,*

$$P(N(t) - N(s) = k) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}.$$

Thus $N(t) - N(s)$ has a Poisson distribution with parameter $\lambda(t - s)$.

Proof. We will prove the proposition in the case $s = 0$. Hence, we must show that for any $k \geq 0$

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

For $n \geq 1$, break the interval $[0, t]$ into n subintervals of length t/n. We say a change has occurred in N in the interval $[a, b]$ if $N(s) - N(s-) \neq 0$ for $s \in [a, b]$. Define the two events

$$A_n = \{k \text{ subintervals contain exactly 1 change event and other } n-k \text{ contain zero}\}$$
$$B_n = \{N(t) = k \text{ and at least one of the subintervals contains 2 or more changes}\}.$$

The events A_n and B_n are disjoint and

$$P(N(t) = k) = P(A_n) + P(B_n).$$

Note that the left hand side does not depend upon n. We will show that $P(B_n) \to 0$ as $n \to \infty$, hence proving that events happen one at a time and that N is a counting process.

Recall Boole's inequality, which states that

$$P\left(\bigcup_i C_i\right) \leq \sum_i P(C_i),$$

for any set of events $\{C_i\}$. We have

$$\begin{aligned} P(B_n) &\leq P(\text{at least one subinterval has 2 or more events}) \\ &= P\left(\bigcup_{i=1}^{n}(\ i\text{th subinterval contains 2 or more})\right) \\ &\leq \sum_{i=1}^{n} P\left(\ i\text{th subinterval contains 2 or more}\right) \\ &= \sum_{i=1}^{n} o(t/n) = t\left[\frac{o(t/n)}{t/n}\right]. \end{aligned}$$

Thus, $P(B_n) \to 0$, as $n \to \infty$.

We turn to the limiting behavior of $P(A_n)$. From assumptions 1 and 2,

$$P(\ 0 \text{ change events occur in a given interval of length } h) = 1 - \lambda h - o(h).$$

Assumption 3 then gives

$$\begin{aligned} P(A_n) &= \binom{n}{k}\left[\lambda\frac{t}{n} + o(t/n)\right]^k \left[1 - \left(\frac{\lambda t}{n} - o(t/n)\right)\right]^{n-k} \\ &\approx \binom{n}{k}\left(\lambda\frac{t}{n}\right)^k \left[1 - \left(\frac{\lambda t}{n}\right)\right]^{n-k} \\ &= (\lambda t)^k \frac{n!}{(n-k)!k!}\left(\frac{1}{n}\right)^k \left(1 - \lambda t\left(\frac{1}{n}\right)\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-k}. \end{aligned}$$

As $n \to \infty$ we have the following three limits

$$\frac{n!}{(n-k)!}\left(\frac{1}{n}\right)^k = \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} \to 1$$

$$\left(1 - \lambda t\left(\frac{1}{n}\right)\right)^n \to e^{-\lambda t},$$

$$\left(1 - \frac{\lambda t}{n}\right)^{-k} \to 1.$$

Hence,

$$\lim_{n\to\infty} P(A_n) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

proving the result. □

Let S_k be the time of the kth change in the process N. First note that by Proposition 3.20,

$$P(S_1 > t) = P(N(t) = 0) = e^{-\lambda t}.$$

Hence, S_1 has an exponential distribution with parameter $\lambda > 0$. Further, and again by Proposition 3.20, for $k \geq 1$,

$$P\{S_k \leq t\} = P\{N(t) \geq k\} = 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \quad t \geq 0.$$

Differentiating to obtain the probability density function gives

$$f_{S_k}(t) = \begin{cases} \frac{1}{(k-1)!}\lambda^k t^{k-1} e^{-\lambda t} & t \geq 0 \\ 0 & t < 0, \end{cases}$$

and we see that S_k is has a gamma distribution with parameters k and λ. The above suggests the following theorem, whose proof can be found in Chapter 7 of [4].

Theorem 3.21. *Let $T_1 = S_1$ and for $k > 1$ let $T_k = S_k - S_{k-1}$. Then $T_1, T_2, \ldots$ are independent and exponentially distributed with parameter λ.*

Thus, we see that a homogeneous Poisson process with intensity $\lambda > 0$ is simply the counting process for a renewal process with inter-event times determined by independent exponential random variables with parameter λ. Note that Theorem 3.21 gives a method for the construction of a Poisson process from a sequence of i.i.d. exponential random variables.

Assumption 3.18 is equivalent to the following two conditions:

(*i*) for any $A \subset \mathbb{R}_{\geq 0}$ and $k \geq 0$, we have that

$$P(N(A) = k) = e^{-\lambda|A|}\frac{(\lambda|A|)^k}{k!},$$

where $|A|$ is the Lebesgue measure of A, and

(*ii*) If $A_1, \ldots, A_k$ are disjoint subsets of E, then $N(A_1), \ldots, N(A_k)$ are independent random variables.

(To see the equivalence, take $A = [t, t+h]$.) This equivalence allows us to generalize the notion of a Poisson process.

Definition 3.22. Let N be a point process on $E \subset \mathbb{R}^d$, $\mathcal{F}$ a σ-algebra of E, and μ a measure on E. We say that N is a *Poisson process with mean measure* μ, or a *Poisson random measure*, if the following two conditions hold:

(*i*) For $A \in \mathcal{F}$ and $k \geq 0$,

$$P(N(A) = k) = \begin{cases} \frac{e^{-\mu(A)}(\mu(A))^k}{k!}, & \text{if } \mu(A) < \infty \\ 0, & \text{if } \mu(A) = \infty \end{cases}.$$

(*ii*) If $A_1, \ldots, A_k$ are disjoint subsets of E, then $N(A_1), \ldots, N(A_k)$ are independent random variables.

Note that $E[N(A)] = \mu(A)$. One choice for the mean measure would be a multiple of Lebesgue measure: $\mu(A) = \lambda|A|$, where we recall that $|A| = \text{Area}(A)$ in two dimensions, $|A| = \text{Vol}(A)$ in three dimensions, etc. If $\lambda = 1$, then the measure is said to be *unit-rate*. When the mean measure is a multiple of Lebesgue measure, we call the process *homogeneous*.

With Definition 3.22 in hand we can generalize away from the homogeneous case.

Example 3.23. Suppose that for open intervals (a, b), the mean measure μ for a Poisson process is

$$\mu((a, b)) = \Lambda(b) - \Lambda(a),$$

for some non-decreasing, absolutely continuous function Λ. If Λ has density λ (i.e. if Λ is differentiable), then

$$\mu((a, b)) = \Lambda(b) - \Lambda(a) = \int_a^b \lambda(s)ds,$$

or, more generally,

$$\mu(A) = \int_A \lambda(s)ds,$$

for sets A. Note that, by construction, λ takes only non-negative values as Λ is non-decreasing. Further, we have that

$$P(N((a, b)) = k) = e^{-(\Lambda(b)-\Lambda(a))}\frac{(\Lambda(b) - \Lambda(a))^k}{k!} = e^{-(\int_a^b \lambda(s)ds)}\frac{(\int_a^b \lambda(s)ds)^k}{k!}, \quad (3.12)$$

and more generally,

$$P(N(A) = k) = e^{-\int_A \lambda(s)ds}\frac{\left(\int_A \lambda(s)ds\right)^k}{k!},$$

for any set $A \subset \mathbb{R}$. The function λ is usually termed the *rate, intensity*, or *propensity* function of the Poisson process, depending upon the specific scientific field in which the model is being considered. $\triangle$

We note that if the first assumption in Assumption 3.18 were changed to

$$P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h), \text{ as } h \to 0, \tag{3.13}$$

then it can be argued that the resulting process is equivalent to the non-homogeneous Poisson process just described. This is important from a modeling perspective, as it is usually an assumption of the form (3.13) that is the starting point of a mathematical model.

Example 3.24. Suppose we believe that the arrival times of frogs to a pond can be reasonably modeled by a Poisson process. We suppose that frogs are arriving at a rate of 3 per hour. What is the probability that no frogs will arrive in the next hour? What is the probability that 12 or less frogs arrive in the next five hours?

Let $N([0,t])$ be a Poisson process of rate 3. Then,

$$P(N([0,1]) = 0) = e^{-3\cdot 1}\frac{(3\cdot 1)^0}{0!} = e^{-3} = 0.04978.$$

Further,

$$P(N([0,5]) \leq 12) = \sum_{k=0}^{12} P(N([0,5]) = k) = \sum_{k=0}^{12} e^{-3\cdot 5}\frac{(3\cdot 5)^k}{k!} \approx 0.2676.$$

△

Example 3.25. We change the previous example by recognizing that it is unlikely that frogs would arrive anywhere according to a homogenous process. Instead, the rate of arrival should fluctuate throughout the day. Therefore, we change our model and suppose that the arrival of the frogs is modeled by an non-homogeneous Poisson process with intensity function

$$\lambda(t) = 3 + \sin(t/4),$$

where $t = 0$ is taken to be 8AM. Assuming it is exactly 8AM now, what is the probability that no frogs will arrive in the next hour? What is the probability that 12 or less frogs arrive in the next five hours?

We let $N([0,t])$ be the Poisson process with intensity $\lambda(t) = 3 + \sin(t/4)$. Then,

$$P(N([0,1]) = 0) = e^{-\int_0^1 (3+\sin(t/4))dt}\frac{\left(\int_0^1 (3+\sin(t/4))dt\right)^0}{0!} = e^{-(7-4\cos(1/4))} \approx 0.044.$$

Further,

$$\begin{aligned} P(N([0,5]) \leq 12\} = k) &= \sum_{k=0}^{12} P(N([0,5]) = k) \\ &= \sum_{k=0}^{12} e^{-\int_0^5 (3+\sin(t/4))dt}\frac{\left(\int_0^5 (3+\sin(t/4))dt\right)^k}{k!} \\ &\approx 0.1017. \end{aligned}$$

△

Example 3.26. Suppose we believe that bears in upper Wisconsin are distributed as a spatial Poisson process with rate $\lambda = 3$ per square mile. Suppose you are camping in northern Wisconsin tonight. What is the expected distance between you and the nearest bear?

Let R be the distance of the nearest bear from the campground and let $d(r)$ be a disc of radius r centered at the campground. Then,

$$P(R > r) = P(N(d(r)) = 0) = e^{-\lambda |d(r)|}.$$

Since

$$|d(r)| = \pi r^2,$$

we have

$$P(R > r) = e^{-3\pi r^2}.$$

Hence,

$$E[R] = \int_0^\infty r \cdot 6\pi r e^{-3\pi r^2} dr = \frac{1}{2 \cdot \sqrt{3}} \approx 0.2887 \text{ miles},$$

where we utilized that the density is $f_R(r) = 6\pi r e^{-3\pi r^2}$. △

3.2.3 Transformations of Poisson processes

Suppose that we let E_n^λ, $n \geq 1$, be i.i.d. exponential random variables with parameter $\lambda > 0$. Define

$$S_n^\lambda \stackrel{\text{def}}{=} \sum_{i=1}^n E_i^\lambda,$$

and let N^λ be the associated Poisson process,

$$N^\lambda([0,t]) \stackrel{\text{def}}{=} \sum_{n=1}^\infty 1_{[0,t]}(S_n^\lambda) = \sum_{n=1}^\infty 1_{\{S_n^\lambda \leq t\}} = \sum_{n=1}^\infty 1_{\{\sum_{i=1}^n E_i^\lambda \leq t\}}.$$

By the basic properties of exponential random variables, we know that E_i^λ has the same distribution as E_i/λ, where E_i are unit exponential random variables. Therefore,

$$N^\lambda([0,t]) = \sum_{n=1}^\infty 1_{\{\sum_{i=1}^n E_i^\lambda \leq t\}} \stackrel{\mathcal{D}}{=} \sum_{n=1}^\infty 1_{\{\sum_{i=1}^n E_i \leq \lambda t\}} = N([0, \lambda t]), \tag{3.14}$$

where $\stackrel{\mathcal{D}}{=}$ is equality in distribution and where N is a *unit-rate* Poisson process (since the exponential random variables used to construct it have a parameter of 1). The importance of the relation (3.14) can not be overstated, and we will interpret it in two different ways. First, it can be viewed as a *time-change*. That is, it shows that if a homogenous Poisson process with rate $\lambda > 0$ is desired, then it is sufficient to start with a *unit-rate* process and simply "move" along its time-frame at rate λ. If $\lambda > 1$ we move faster than unit speed, whereas if $\lambda < 1$, we move slower.

Second, (3.14) can be viewed as a spacial shifting of points. It shows that if the position of the points of a homogeneous process of rate $\lambda > 0$ are multiplied by λ, then the resulting point process is also Poisson, and it is, in fact, a homogeneous process of rate 1. Likewise, we could start with a unit-rate process and divide the position of each point by λ to get a homogeneous process with rate λ. This phenomenon will be explored further in the next subsection.

Both interpretations are important and will be returned to repeatedly. This is our first example of a transformation of a Poisson process, via time or space, yielding another Poisson process. In the next section we greatly expand our understanding of such transformations.

General transformations of Poisson processes

We return to the example at the end of the last subsection in which we transformed one Poisson process into another, though we now take a slightly different perspective. Let E_i denote independent, unit-exponential random variables, and let $S_n = \sum_{i=1}^{n} E_i$. Letting

$$N([0,t]) = \sum_{n=1}^{\infty} 1_{\{S_n \le t\}},$$

we know that N is a unit-rate Poisson process. For $\lambda > 0$, let $T : \mathbb{R} \to \mathbb{R}$ be defined via

$$T(x) = \frac{x}{\lambda}.$$

We then have that

$$N^{\lambda}([0,t]) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} 1_{\{T(S_n) \le t\}} = \sum_{n=1}^{\infty} 1_{\{S_n/\lambda \le t\}} = \sum_{n=1}^{\infty} 1_{\{\sum_{i=1}^{n} E_i/\lambda \le t\}}$$

is a homogeneous Poisson process with rate λ since $E_n/\lambda, n \ge 1$, are independent exponential random variables with a parameter of λ. Also, the mean measure of the process has changed during the transformation from

$$\mu((a,b)) = b - a,$$

to

$$\mu'((a,b)) = \lambda(b-a),$$

where $b > a$. Note that

$$\mu'((a,b)) = \lambda(b-a) = (\lambda b - \lambda a) = \mu((\lambda a, \lambda b)) = \mu(T^{-1}(a,b)).$$

Collecting thoughts, we see that moving the points around via the function, or transformation, T resulted in another Poisson process, and the new mean measure can be understood via the inverse of T and the original measure μ. We will see below that this is a general result, however we begin by building up our terminology.

For two euclidean spaces E and E', we assume the existence of some one-to-one function $T : E \to E'$. Note that the function T^{-1} induces a set mapping from the subsets of E' to those of E. That is, for $A' \subset E'$ we have

$$T^{-1}(A') = \{x \in E \ : \ T(x) \in A'\}.$$

Therefore, $T^{-1}(A')$ is simply the pre-image of A' under T.

We want to take the points of a general Poisson process, N, defined on E, apply T to them, and consider the resulting point process in E'. Note that because T is one-to-one, for any $A \subset E$ we have that $N(A) = N(T(A))$.

We will denote the mean measure of N by μ and the points associated with N as X_n. The goal is to be able to count the number of points, $T(X_n) \in E'$, in a given region $A' \subset E'$. Letting N' denote that counting process, we see that

$$\begin{aligned} N'(A') &= \sum_n 1_{T(X_n)}(A') = \sum_n 1_{\{T(X_n)\in A'\}} = \sum_n 1_{\{X_n \in T^{-1}(A')\}} \\ &= \sum_n 1_{X_n}(T^{-1}(A')) \\ &= N(T^{-1}(A')) \\ &= N \circ T^{-1}(A'). \end{aligned}$$

Further, we see the expected number of points is

$$\mu'(A') = \mu(T^{-1}(A')) = \mu \circ T^{-1}(A').$$

That is, once again, the mean measure is determined by T^{-1} and the original mean measure μ. The following proposition is incredibly useful and, even though the proof is rather straightforward, nearly miraculous.

Proposition 3.27. *Suppose that $T : E \to E'$ is a one-to-one mapping between Euclidean spaces such that if $B' \subset E'$ is bounded, then so is $T^{-1}(B') \subset E$. If N is Poisson process on E with mean measure μ and points $\{X_n\}$, then*

$$N' = N \circ T^{-1}$$

is a Poisson process on E' with points $\{T(X_n)\}$ and mean measure

$$\mu' = \mu \circ T^{-1}.$$

Proof. We need to show that the two properties of a Poisson process as given in Definition 3.22 are satsified. Firstly, we have for any $B' \subset E'$ and $k \geq 0$,

$$\begin{aligned} P(N'(B') = k) &= P(N(T^{-1}(B')) = k) \\ &= e^{-\mu(T^{-1}(B'))} \frac{(\mu(T^{-1}(B')))^k}{k!} \\ &= e^{-\mu'(B')} \frac{(\mu'(B'))^k}{k!}, \end{aligned}$$

where the second equality follows since N is a Poisson process with mean measure μ. Next, if $B'_1, \ldots, B'_m$ are disjoint, then so are $T^{-1}(B'_1), \ldots, T^{-1}(B'_m)$. Therefore, the random variables

$$\{N'(B'_1), \cdots, N'(B'_m)\} = \left\{N(T^{-1}(B'_1)), \cdots, N(T^{-1}(B'_m))\right\},$$

are independent. □

Example 3.28. Let $N = \sum_n 1_{X_n}$ be a homogeneous Poisson process with rate $\lambda = 1$ and state space $E = [0, \infty)$ (hence, $X_n - X_{n-1}$ are i.i.d. unit exponential random variables). In this case, the mean measure is $\mu([a, b]) = b - a$.

Let $T(x) = x^2$ and define the N' via

$$N'(A) = \sum_n 1_{X_n^2}(A),$$

where $A \subset [0, \infty)$. By Proposition 3.27, N' is a Poisson process on $[0, \infty)$ with mean measure

$$\mu'([0, t]) = \mu(T^{-1}([0, t])) = \mu\{x : x^2 \leq t\} = \mu\left([0, \sqrt{t}]\right) = \sqrt{t} = \int_0^t \tfrac{1}{2}\sqrt{s}\, ds.$$

Hence, recalling Example 3.23, the resulting process has intensity $\lambda(s) = \frac{1}{2}\sqrt{s}$, for $s \geq 0$. △

In the previous example, we started with a unit-rate Poisson processes, performed a transformation, and ended with a Poisson process with intensity $\lambda(s) = \frac{1}{2}\sqrt{s}$. This calculation raises the following natural question: given a specific nonnegative valued function $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, how do we transform the points of a unit-rate Poisson process in order to obtain one with intensity function λ?

We begin by defining the non-decreasing function m via

$$m(t) = \int_0^t \lambda(s) ds.$$

Next, we define a partial inverse function for m, denoted I, via

$$I(x) = \inf\{t \geq 0 : m(t) \geq x\}, \quad x \geq 0.$$

The function I is almost an inverse function because

$$m(I(x)) = \int_0^{I(x)} \lambda(s) ds = x, \tag{3.15}$$

which follows since $I(x)$ is defined to be the smallest $u \geq 0$ for which

$$\int_0^u \lambda(s) ds = x. \tag{3.16}$$

Example 3.29 (Example 3.28 continued)**.** Let

$$\lambda(s) = \begin{cases} \frac{1}{2}s^{-1/2} & s > 0 \\ 0 & \text{else} \end{cases}$$

and, for $t \geq 0$,

$$m(t) = \int_0^t \lambda(s)ds = \sqrt{t}.$$

Then,

$$I(x) = \inf\left\{t \geq 0 : \sqrt{t} \geq x\right\} = x^2.$$

Of course, in this case I is a true inverse of m since we have both

$$m(I(x)) = m(x^2) = x,$$
$$I(m(t)) = I(\sqrt{t}) = t,$$

for $x, t \geq 0$. △

The function I is not always a true inverse of m since we do not always have $I(m(x)) = x$. The issue is that λ may take the value zero for portions of its domain. For example, suppose $\lambda(s) = 0$ for all $s \in (u, x)$ where $0 \leq u < x$. This implies

$$m(u) = \int_0^u \lambda(s)ds = \int_0^u \lambda(s)ds + \int_u^x \lambda(s)ds = \int_0^x \lambda(s)ds = m(x).$$

Therefore,

$$I(m(x)) = \inf\{t : m(t) \geq m(x)\} \leq u < x.$$

Example 3.30. Let

$$\lambda(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & 1 \leq t < 2 \\ 1 & 2 \leq t < \infty \end{cases}.$$

in which case

$$m(t) = \begin{cases} t & 0 \leq t < 1 \\ 1 & 1 \leq t < 2 \\ t-1 & 2 \leq t < \infty \end{cases}$$

We have

$$I(x) = \inf\{t \geq 0 : m(t) \geq x\} = \begin{cases} x, & 0 < x < 1 \\ 1, & x = 1 \\ x+1, & 1 < x \end{cases}.$$

Thus, for example,

$$I(m(1.5)) = I(1) = 1,$$

whereas

$$I(m(2.5)) = I(1.5) = 1.5 + 1 = 2.5.$$

△

While it is not always true that $I(m(x)) = x$, we do always have the identity

$$I(m(x)) \leq x. \tag{3.17}$$

This follows because

$$I(m(x)) = \inf\{t \geq 0 : m(t) \geq m(x)\} \leq x,$$

as x is certainly one number (though not necessarily the smallest) that satisfies the inequality $m(t) \geq m(x)$.

Lemma 3.31. *The function I is strictly increasing.*

Proof. Suppose that $x < y$. Then,

$$\int_0^{I(x)} \lambda(s)ds = x < y = \int_0^{I(y)} \lambda(s)ds \implies I(x) < I(y).$$

□

Thus, the function I is a one-to-one transformation, and we will apply it to the points of a unit-rate Poisson process. Let $N = \sum_n 1_{X_n}$ be a unit-rate homogeneous Poisson process, with corresponding points X_n (so $X_n - X_{n-1}$ are i.i.d. unit exponential random variables). We then let

$$N^\lambda = \sum_n 1_{I(X_n)}, \tag{3.18}$$

which by Proposition 3.27 (and Lemma 3.31) is a Poisson process on $[0, \infty)$.

What is the mean measure? To find out, the following lemma is useful.

Lemma 3.32. *For all $x, t \geq 0$,*

$$I(x) \leq t \iff x \leq m(t).$$

Proof. First suppose that $I(x) \leq t$. We must show that $x \leq m(t)$. Applying m, which is non-decreasing, to both sides of the inequality $I(x) \leq t$ yields

$$m(I(x)) \leq m(t).$$

By (3.15) we have $m(I(x)) = x$, and so we conclude $x \leq m(t)$.

Next, we suppose that $x \leq m(t)$ and will show that $I(x) \leq t$. Apply I, which is strictly increasing, to $x \leq m(t)$ to find

$$I(x) \leq I(m(t)).$$

By (3.17) we know $I(m(t)) \leq t$, which gives us the desired implication. □

Returning to the mean measure for our new process, Lemma 3.32 shows that for any $t \geq 0$,

$$\{x : I(x) \leq t\} = \{x : x \leq m(t)\}. \tag{3.19}$$

Let μ be (unit-rate) Lebesgue measure. By Proposition 3.27, the mean measure, μ^λ, of our new process satisfies

$$\begin{aligned}\mu^\lambda([0,t]) &= \mu(I^{-1}([0,t])) = \mu(\{x : I(x) \leq t\}) = \mu(\{x : x \leq m(t)\}) \\ &= m(t) \\ &= \int_0^t \lambda(s)ds,\end{aligned}$$

confirming that the resulting process is a non-homogeneous Poisson process with intensity function λ.

Returning to N^λ defined in (3.18), we also see that

$$\begin{aligned}N^\lambda([0,t]) &= \sum_n 1_{I(X_n)}([0,t]) && \text{(definition of } N'\text{)} \\ &= \sum_n 1_{\{I(X_n)\leq t\}} \\ &= \sum_n 1_{\{X_n \leq m(t)\}} && \text{(Lemma (3.32))} \\ &= \sum_n 1_{X_n}([0,m(t)]) && (3.20) \\ &= N\left(\left[0, \int_0^t \lambda(s)ds\right]\right). && (3.21)\end{aligned}$$

Condensing notation, we see that if N^λ is a Poisson process with intensity λ, then

$$N^\lambda(t) = N\left(\int_0^t \lambda(s)ds\right), \tag{3.22}$$

where N is a unit-rate Poisson process. Thus, to get a non-homogeneous Poisson process with intensity function λ, it is enough to correctly modulate the speed at which the "clock" runs on a Poisson process with points determined by unit exponentials. This is called a *time-changed representation* for the non-homogeneous Poisson process and will play a critical role in our understanding of continuous time Markov chains in later chapters.

Simulating non-homogeneous Poisson processes

We turn to the issue of efficiently simulating a non-homogeneous Poisson process. We see that simulating such a process is equivalent to simulating the right hand side of (3.22), and the following strategy does just that.

1. Let E_1 be an exponential random variable with parameter one. This is the first point of the homogeneous Poisson point process.

2. Solve for the smallest t_1 that satisfies

$$\int_0^{t_1} \lambda(s)ds = E_1.$$

Using the notation from above, we see this is equivalent to setting $t_1 = I(E_1)$. Note that if the anti-derivative of λ is of a nice form, then solving this equation will be simple. The value t_1 is the first point of the non-homogeneous process.

3. Repeat. Let E_2 be an exponential random variable with parameter one. This is the second point of the homogeneous Poisson point process.

4. Solve for the smallest t_2 that satisfies

$$\int_{t_1}^{t_2} \lambda(s)ds = E_2.$$

The value t_2 is the second point of the non-homogeneous process.

5. Repeat until a desired number of points for the non-homogeneous process have been generated.

The following algorithm generates the points of a non-homogeneous Poisson process with intensity function λ. The points of the non-homogeneous process are denoted $t_n, n \geq 1$.

Algorithm. Set $t_0 = 0$. *Set* $n = 1$.

1. *Let* E_n *be an exponential random variable with parameter one, which is independent from all other random variables already generated.*

2. *Find the smallest* $u \geq 0$ *for which*

$$\int_{t_{n-1}}^{u} \lambda(s)ds = E_n.$$

Set $t_n = u$. *Note this is equivalent to solving*

$$\int_0^{u} \lambda(s)ds = E_1 + \cdots + E_n,$$

for u.

3. *Set* $n \leftarrow n + 1$.

4. *Return to step 1 or break.*

The above algorithm for simulating a non-homogeneous Poisson process is the core of future methods for the simulation of continuous time Markov chains.

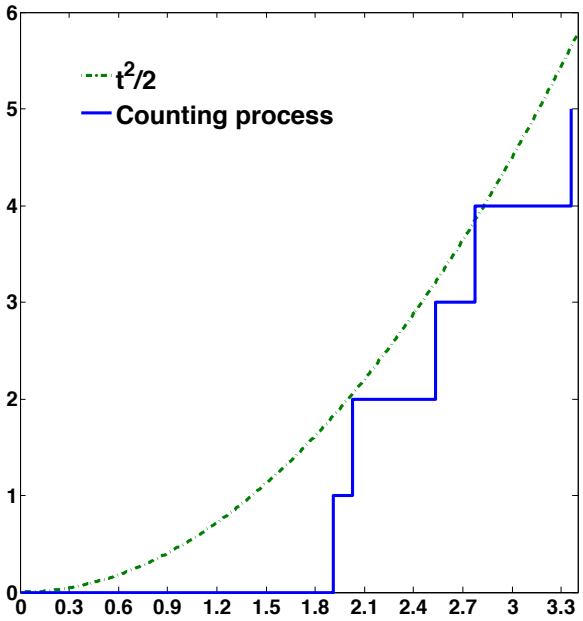


Figure 3.1: One realization of the counting process $N(t^2/2)$, blue curve, versus the plot of the deterministic function $t^2/2$, green curve. The data is generated in Example 3.33.

Example 3.33. Suppose that $\lambda(t) = t$ for all $t \geq 0$. We consider the problem of simulating the non-homogeneous Poisson process with intensity $\lambda(t)$:

$$N\left(\int_0^t \lambda(s)ds\right) = N\left(\int_0^t s\,ds\right) = N\left(\frac{1}{2}t^2\right),$$

where N is a unit-rate Poisson process. Supposing our stream of unit exponential random variables begins with $E = [1.8190, 0.2303, 1.1673, 0.6376, 1.7979]$, we have the following.

1. To find the first jump of our counting process we solve

$$\int_0^{t_1} \lambda(s)ds = \frac{1}{2}t_1^2 = 1.8190 \implies t_1 \approx 1.907.$$

2. To find the second jump, we solve

$$\int_{1.907}^{t_2} s\,ds = 0.2303 \implies \frac{1}{2}t_2^2 - \frac{1}{2}(1.907)^2 = 0.2303 \implies t_2 = 2.024.$$

Similarly, we simply could have solved for t_2 via

$$\int_0^{t_2} \lambda(s)ds = \frac{1}{2}t_2^2 = (1.8190 + 0.2303) \implies t_2 = 2.024.$$

3. Solving for t_3 :
$$\frac{1}{2}t_3^2 = (1.8190 + 0.2303 + 1.1673) \implies t_3 = 2.536.$$
4. Solving for t_4:
$$\frac{1}{2}t_4^2 = (1.8190 + 0.2303 + 1.1673 + 0.6376) \implies t_4 = 2.776.$$
5. Solving for t_5:
$$\frac{1}{2}t_5^2 = (1.8190 + 0.2303 + 1.1673 + 0.6376 + 1.7979) \implies t_5 = 3.362.$$

Note that the actual time between events, $t_n - t_{n-1}$, is (non-monotonically) reducing. Plots of (a) the counting process, and (b) the deterministic function $t^2/2$, are found in Figure 3.1. △

Differing time frames

Consider the non-homogeneous Poisson process with intensity function $\lambda(t)$,
$$N^\lambda(t) \stackrel{\text{def}}{=} N\left(\int_0^t \lambda(s)ds\right),$$
where N is a unit-rate Poisson process. Note that the variable t represents time. However, there is another time frame in the problem: that of the Poisson process N. Let
$$\tau(t) \stackrel{\text{def}}{=} \int_0^t \lambda(s)ds,$$
and note that the process N^λ can be written as
$$N^\lambda(t) = N\left(\int_0^t \lambda(s)ds\right) = N(\tau(t)).$$
We see that $\tau(t)$ give the current time of the Poisson process N. We can use this notation to simplify our calculations. For example, because N is a unit-rate Poisson process we have
$$P(N(\tau(t)+h) - N(\tau(t)) = 1) = h + o(h), \quad \text{as } h \to 0. \tag{3.23}$$
This tells us the following about the process N^λ,
$$\begin{aligned} P\big(N^\lambda(t+h) - N^\lambda(t) = 1\big) &= P\left(N\left(\int_0^{t+h} \lambda(s)ds\right) - N\left(\int_0^t \lambda(s)ds\right) = 1\right) \\ &= P\left(N\left(\int_t^{t+h} \lambda(s)ds + \tau(t)\right) - N(\tau(t)) = 1\right) \\ &= P(N(\lambda(t)h + o(h) + \tau(t)) - N(\tau(t)) = 1) \\ &= \lambda(t)h + o(h), \quad \text{as } h \to 0, \end{aligned} \tag{3.24}$$
where the second to last equality follows from calculus, and the final equality follows from (3.23).

Processes with random intensity

Suppose that $X(t)$ is some stochastic process defined for all time $t \geq 0$. For example, we could have that X_n, $n \geq 0$, is a discrete time Markov chain and

$$X(t) \stackrel{\text{def}}{=} X_{\lfloor t \rfloor},$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x. Note that in this case $X(t)$ is a step function, though in general this need not be the case.

We now consider the process which, conditioned upon the history of X, i.e. on $X(s), 0 \leq s \leq t$, behaves locally like a non-homogeneous Poisson process with intensity function $\lambda(X(t))$. That is, we want N^λ to satisfy the relations

$$\begin{aligned} &P\left(N^\lambda(t+h) - N^\lambda(t) = 1 \mid X_s, 0 \leq s \leq t\right) = \lambda(X(t))h + o(h), && \text{as } h \to 0 \\ &P\left(N^\lambda(t+h) - N^\lambda(t) \geq 2 \mid X_s, 0 \leq s \leq t\right) = o(h), && \text{as } h \to 0. \end{aligned} \tag{3.25}$$

From the results of the previous section, we believe we can model such a process via

$$N^\lambda(t) = N\left(\int_0^t \lambda(X(s))ds\right), \tag{3.26}$$

where N is a unit-rate Poisson process that is independent from X. Later we will discuss how the independence assumption can be weakened a bit. Note that conditioned upon $X_s, 0 \leq s \leq t$, the expected number of points for the process (3.26) in the interval $[0, t]$ is

$$\mathbb{E}\left[N^\lambda(t) \mid X_s, 0 \leq s \leq t\right] = \mathbb{E}\left[N\left(\int_0^t \lambda(X(s))ds\right) \,\middle|\, X_s, 0 \leq s \leq t\right] = \int_0^t \lambda(X(s))ds.$$

Taking expectations again shows that

$$\mathbb{E}[N^\lambda(t)] = \mathbb{E}\left[\int_0^t \lambda(X(s))ds\right] = \int_0^t \mathbb{E}[\lambda(X(s))]ds.$$

If λ is a nonlinear function, then it is not permissible to switch the expectation with λ, and so we do not have that the expected number of points is $\int_0^t \lambda(\mathbb{E}[X(s)])ds$.

We will show that the process (3.26) satisfies one of the modeling assumptions (3.25), leaving the second for a homework exercise. Denoting

$$\tau(t) = \int_0^t \lambda(X(s))ds,$$

which itself is a stochastic process, we have

$$\begin{aligned}
&P(N^\lambda(t+h) - N^\lambda(t) = 1 \mid X_s, 0 \le s \le t) \\
&= P\left(N\left(\int_0^{t+h} \lambda(X(s))ds\right) - N\left(\int_0^t \lambda(X(s))ds\right) = 1 \;\middle|\; X_s, 0 \le s \le t\right) \\
&= P\left(N\left(\int_t^{t+h} \lambda(X(s))ds + \tau(t)\right) - N\left(\tau(t)\right) = 1 \;\middle|\; X_s, 0 \le s \le t\right) \\
&= P\left(N\left(\lambda(X(t))h + o(h) + \tau(t)\right) - N\left(\tau(t)\right) = 1 \mid X_s, 0 \le s \le t\right) \\
&= \lambda(X(t))h + o(h),
\end{aligned}$$

valid as $h \to 0$, where the last equality follows from the independence of X and N. The remaining condition is left as a homework exercise.

Note that we did not strictly require that N and X be independent. Instead, we only require that for all $t \ge 0$, the increments $N(s+\tau(t)) - N(\tau(t))$ are independent from $X_s, 0 \le s \le t$. That is, loosely, we require that X can not look into the future behavior of N. This seems like a minor point, but it will have large modeling consequences with the first such example found in Example 3.35. In Example 3.36 we demonstrate what can go wrong if we do not have such an independence condition.

Example 3.34. Suppose X_n gives the number of people living in a valley at the beginning of year $n \ge 0$. We suppose that the number of births in this valley can be modeled via a Poisson process with local intensity $\lambda(X(n))$ for the entirety of year n. Letting $X(t) \stackrel{\text{def}}{=} X_{\lfloor t \rfloor}$ (which is the process attained by extending X_n to all of $\mathbb{R}_{\ge 0}$), we see that the number of births by time t can be modeled via

$$B(t) = N\left(\int_0^t \lambda(X(s))ds\right),$$

where N is a unit-rate Poisson process. Note that it would be reasonable to assume that $\lambda(\cdot)$ is a non-negative function with $\lambda(0) = 0$. △

Example 3.35. We give a general model for arrivals and departures. This formulation could be used to model a queue, the transcription and degradation of mRNA, etc. For concreteness, we choose the language of a queue.

We suppose that arrivals are taking place at a constant rate of $\lambda > 0$. Therefore, letting N_1 denote a unit-rate Poisson process, we define

$$A(t) = N_1\left(\int_0^t \lambda ds\right) = N_1(\lambda t),$$

to be our arrival process.

Let $X(t)$ be the number of people in the queue at time t and assume that departures are happening at a rate of $\mu(X(t))$. That is, we suppose that if $D(t)$ gives the

number of departures by time $t \geq 0$, then

$$\begin{aligned} &P(D(t+h) - D(t) = 1 \mid X_s, 0 \leq s \leq t) = \mu(X(t))h + o(h), \quad &&\text{as } h \to 0 \\ &P(D(t+h) - D(t) \geq 2 \mid X_s, 0 \leq s \leq t) = o(h), &&\text{as } h \to 0, \end{aligned} \tag{3.27}$$

Let N_2 be another unit-rate Poisson process that is independent of N_1 and define

$$D(t) = N_2\left(\int_0^t \mu(X(s))ds\right). \tag{3.28}$$

We will discuss below why D satisfies the conditions (3.27).

Noting that we must have $X(t) = X(0) + A(t) - D(t)$, we see that $X(t)$ is the solution to the stochastic equation

$$X(t) = X(0) + N_1(\lambda t) - N_2\left(\int_0^t \mu(X(s))ds\right).$$

Such an equation is an example of a *random time change representation*. Existence of a unique solution to the above equation can be shown by a "jump by jump" argument, which we will discuss in detail later in the course.

Note that N_2 and X are not independent. However, letting

$$\tau(t) = \int_0^t \mu(X(s))ds,$$

we see that $N_2(\tau(t) + s) - N_2(\tau(t))$, $s \geq 0$, is independent from $X_s, 0 \leq s \leq t$. This allows us to conclude that the conditions (3.27) are still valid in the usual manner. $\triangle$

Example 3.36. In each of the previous examples, we defined some process $\widetilde{N}$ as

$$\widetilde{N}(t) = N\left(\int_0^t X(s)ds\right),$$

where N is an independent unit-rate Poisson process and X is another stochastic process, in order to find a jump process satisfying the two conditions

$$\begin{aligned} &P(\widetilde{N}(t+h) - \widetilde{N}(t) = 1 \mid X_s, 0 \leq s \leq t) = \mu(X(t))h + o(h), \quad &&\text{as } h \to 0 \\ &P(\widetilde{N}(t+h) - \widetilde{N}(t) \geq 2 \mid X_s, 0 \leq s \leq t) = o(h), &&\text{as } h \to 0. \end{aligned}$$

For each example, we worried whether or not certain increments of the Poisson process N were independent of the stochastic process $X(s)$. This worry is justified, as we will demonstrate now.

Let E_i denote independent, exponential random variables with a parameter of one. Let N be the unit-rate Poisson process with points

$$S_n = \sum_{i=1}^{n} E_i,$$

where we take $S_0 = 0$. Now define $X(t) = S_{\lfloor t \rfloor + 1} - S_{\lfloor t \rfloor}$. That is, $X(t)$ gives the value of the waiting time of the current "gap" in the points. For example, for each $0 \leq s < 1$,

$$X(s) = S_1 - S_0 = E_1.$$

Finally, define

$$\widetilde{N}(t) = N\left(\int_0^t X(s)ds\right).$$

Note that $X(t)$ looks into the future of N. We have $\widetilde{N}(0) = 0$, and that for $0 \leq h < 1$

$$\widetilde{N}(h) = N(hE_1).$$

Thus,

$$P\left(\widetilde{N}(h) - \widetilde{N}(0) \geq 1 \mid X(0)\right) = \begin{cases} 1 & \text{if } h \geq 1 \\ 0 & \text{else} \end{cases},$$

showing this counting process *does not* satisfy the equations (3.25). △

3.3 Exercises

Exercise 3.1. Recall that for a renewal process if $P(Y_n < \infty) < 1$, then the process is called *defective*. Suppose that $P(Y_0 < \infty) = 1$ and argue why in the defective case $N(t)$ is bounded, with the bound given by a geometric random variable with a parameter of $p \stackrel{\text{def}}{=} 1 - P(Y_n < \infty)$. Give the distribution precisely in the event that the process is a pure renewal process.

Exercise 3.2. Show that for a renewal process

$$S_{N(t)-1} \leq t < S_{N(t)},$$

so long as $N(t) \geq 1$. Hint: draw a picture.

Exercise 3.3. Let E be an exponential random variable with parameter 1. For $\lambda > 0$, show that E/λ is an exponential random variable with parameter λ. That is, if E^λ is an exponential random variable with parameter $\lambda > 0$, show that $E^\lambda \stackrel{\mathcal{D}}{=} E/\lambda$, where E is a *unit* exponential random variable.

Exercise 3.4. Verify that 1_{X_n}, defined in (3.11), satisfies the three properties that make it a measure.

Exercise 3.5. Let $N(t)$ be a one-dimensional homogeneous Poisson process with rate $\lambda > 0$ and points S_n (i.e. jumps happen at the times S_n). We are assuming $N(0) = 0$. Suppose I tell you that $N(t) = 1$ for some $t > 0$. Find the conditional distribution of the time S_0. That is, find the distribution of the time of the first jump, S_0, conditioned upon knowing $N(t) = 1$. Hint: the distribution has a name, give it. Note how the answer depends upon λ.

Exercise 3.6. Let $N(t)$ be a one-dimensional homogeneous Poisson process with rate $\lambda > 0$ and points S_n (i.e. jumps happen at the times S_n). We are assuming $N(0) = 0$. Suppose that I tell you $N(t) = n$ for some $t > 0$ and $n \geq 1$. Find the conditional distribution of the time S_0. That is, find the distribution of the time of the *first* jump conditioned upon knowing $N(t) = n$ for some $n \geq 1$. Note how the answer depends upon λ.

Exercise 3.7. Let N be a unit-rate Poisson process with associated points $S_n, n \geq 0$. That is, for $t \geq 0$,

$$N([0,t]) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} 1_{\{S_n \leq t\}},$$

where S_0 and $S_n - S_{n-1}, n \geq 1$, are independent unit exponential random variables. Let $T : \mathbb{R}_{\geq 0} \to \mathbb{R}^2_{\geq 0}$ be defined via $T(x) = (x, x^2)$. Describe the resulting counting process $\widetilde{N}$ when the points S_n are transformed by T. Be sure to give the state space and mean measure of the resulting process.

Exercise 3.8. Let N be a non-homogeneous Poisson process with local intensity $\lambda(t) = t^2$. Write a Matlab code that simulates this process until 500 jumps have taken place. Next, write a script that simulates this process up to a time of 15. You are required to turn in both of your codes, and three plots from each.

Exercise 3.9. Students at a boarding school can be in one of three states: sad, neutral, or happy. If they are sad, they do not want to make many phone calls to each other. If they are neutral, they make some phone calls to each other, and if they are happy, they tend to make lots of phone calls to each other. Suppose that the state of the student body changes each day according to a discrete time Markov chain with state 1 being sad, state 2 being neutral, and state 3 being happy, and that the transition matrix is given by

$$P = \begin{pmatrix} .1 & .8 & .1 \\ .3 & .1 & .6 \\ .1 & .4 & .5 \end{pmatrix}.$$

Let $X(t) \in \{1, 2, 3\}$ denote the state of this Markov chain at time t, noting that it is a step function and constant each day. Now suppose we believe that the number of calls to the local cell phone tower can be modeled as a time non-homogeneous Poisson process with local intensity

$$\lambda(t) = \lambda(X(t)) = \begin{cases} 10, & \text{if } X(t) = 1 \\ 33, & \text{if } X(t) = 2 \\ 56, & \text{if } X(t) = 3 \end{cases},$$

where the units of t are days. Assuming that the Markov chain starts day 1 in state 1, approximate the probability that the cell tower receives more calls on day two than day one **and** that it receives more calls on day 3 than day 2. Solve this problem by simulating the model $n = 10^2, 10^3$, and 10^4 times and averaging (that is, give

three different answers based upon the different choices of n). Note that you will be applying to law of large numbers to conclude that these values constitute an estimate for the desired probability.

Next, answer the same question for 4 days (that is, estimate the probability of increasing number of calls for the first 4 days), and then answer the same question for 5 days.

Exercise 3.10. Suppose that $X(t)$ is a stochastic process and that $\widetilde{N}$ satisfies (3.26), where N is a unit-rate Poisson process independent from X. Show that

$$P\left(\widetilde{N}(t+h) - \widetilde{N}(t) \geq 2 \mid X_s, 0 \leq s \leq t\right) = o(h), \text{ as } h \to 0.$$

Exercise 3.11. Let X_n be a discrete time Markov chain with state space $S = \{0, 1, 2, \dots\}$ and transition probabilities

$$p_{i,j} = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = 0 \\ 0 & \text{else} \end{cases}.$$

For $n \geq 0$, let S_n be the $(n+1)$st time the chain hits zero. Let $Y_0 = S_0$ and for $n \geq 1$, let $Y_n = S_n - S_{n-1}$.

(a) What is the distribution of Y_n for $n \geq 1$?

(b) Argue that $\{S_n\}$ is a renewal sequence. Under what conditions would the sequence be delayed, and under what conditions would the sequence be pure?

(c) Let $X_0 = 1$ and suppose that $p = \frac{1}{4}$ and let $N(t) = \sum_{n=0}^{\infty} 1_{[0,t]}(S_n)$. Analytically approximate

$$P(N(1000) \leq 761.5).$$

(d) Still supposing that $X_0 = 1$ and $p = \frac{1}{4}$, write a Matlab code that uses Monte Carlo to approximate $p = P(N(1000) \leq 761.5)$. Use the (standard) central limit theorem to provide a 95% confidence interval with width 0.02.

Exercise 3.12. An entrepreneur has just opened a restaurant. She believes that if she spends x dollars per day on food preparation and cleaning, then her daily revenue will follow a Poisson distribution with a parameter of $100\sqrt{x}$. However, health inspectors check on the restaurant sporadically. Specifically, they arrive after random amounts of time that are well modeled by exponential random variables with parameter $\lambda = 1/4$. If the health inspector shows up, there is a probability of

$$p(x) = e^{-x/4000}$$

that the restaurant will receive a fine. Finally, the expected value of the fine is known to be $\frac{2{,}000{,}000}{(1+x)}$.

(a) For what value of x will this entrepeneur maximize her long run profits? What will her average daily profit be?

(b) If the average fine is pushed up to

$$\frac{200{,}000{,}000}{1+x},$$

is it still possible for her to make a profit? If so, what is the value x that maximizes profit, and what will her average daily profit be?

(c) Answer the same question as part (b) if the average fine is pushed to

$$\frac{2{,}000{,}000{,}000}{1+x}.$$

Exercise 3.13. Suppose that a class of students are observing a meteor shower between the hours of 11pm and 3am. Suppose that meteors are appearing according to a Poisson process with intensity $\lambda = 4$ per hour. find the following:

(a) The probability they see more than 3 meteors in the first hour.

(b) The probability they see zero meteors in the first hour, but at least 11 meteors in the final three hours.

(c) Given that there were 13 meteors seen all night, what is the probability there were no meteors seen in the first hour?

Exercise 3.14. Redo all parts of Exercise 3.13 under the assumption that the occurrences of meteors is well modeled by a Poisson process with rate function

$$\lambda(s) = 3.5 + \sqrt{s}, \quad s \geq 0.$$

Exercise 3.15. Consider a forrest that is inhabited by deer. Specifically, suppose that the forest is rectangular and is 10 miles wide (north-south) and 20 miles long (east-west). Suppose that the positions of the deer are well modeled by a spatial Poisson point process with mean measure

$$\mu(A) = \iint_A \lambda(x,y)\,dx\,dy,$$

where $(0,0)$ is the southwest corner of the forest, and

$$\lambda(x,y) = \begin{cases} \exp\left\{-\left(\frac{(x-10)^2}{100} + \frac{(y-5)^2}{25}\right)\right\}, & 0 \leq x \leq 20,\ 0 \leq y \leq 10 \\ 0, & \text{else} \end{cases}.$$

For a region of the forrest A, we let $N(A)$ be the number of deer in that region.

(a) What is $P(N(A_1) = 3)$, where $A_1 = [2,4] \times [7,8]$?

(b) What is the probability there are no deer within one mile of the edge of the forest?

Exercise 3.16. In this problem, we verify (3.22) directly. A non-homogeneous Poisson process with intensity function $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a Poisson process with mean measure satisfying

$$\mu([a,b)) = \int_a^b \lambda(s)ds, \tag{3.29}$$

for all $0 \leq a \leq b$ and any $k \in \{0, 1, 2, \dots\}$. (See (3.12).) Let N be a unit-rate Poisson process on $[0, \infty)$ and define

$$N^\lambda(t) = N\left(\int_0^t \lambda(s)\, ds\right).$$

Verify directly that N^λ satisfies the two conditions of Definition 3.22 with $E = [0, \infty)$ and mean measure (3.29). That is, show that $N^\lambda(A)$ has the correct Poisson distribution and that N^λ has independent increments.

Exercise 3.17. Let N be a unit-rate Poisson process with associated points $S_n, n \geq 0$. That is, for $t \geq 0$,

$$N([0,t]) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} 1_{\{S_n \leq t\}},$$

where S_0 and $S_n - S_{n-1}, n \geq 1$, are independent unit exponential random variables. Find a function f so that $f(S_n), n \geq 1$ are the points of a Poisson process on $[0, \infty)$ with mean measure satisfying

$$\mu([a,b)) = b^4 - a^4,$$

for any $0 \leq a \leq b$.

Exercise 3.18. In this exercise we will prove that if Y is a unit Poisson process, then for any $T > 0$ we have

$$\sup_{0 \leq t \leq T} |\tfrac{1}{V} Y(Vt) - t| \to 0, \qquad \text{as } V \to \infty, \tag{3.30}$$

with a probability of one.

(a) Let Y be a unit Poisson process. Prove that for any $t \geq 0$,

$$\lim_{V \to \infty} \tfrac{1}{V} Y(Vt) = t,$$

with a probability of one. (Hint: $Y(Vt)$ is a renewal process.)

(b) Let $f_n(t), n \geq 1$, be a sequence of non-decreasing functions. Suppose that for some $T > 0$ we have $f_n(t) \to f(t)$, as $n \to \infty$, for all $0 \leq t \leq T$ (pointwise convergence). Then we have the following stronger result

$$\sup_{0 \leq t \leq T} |f_n(t) - f(t)| \to 0, \quad \text{as} \quad n \to \infty.$$

Use this fact, which you do not need to prove, and part (a) to conclude (3.30) holds with a probability of one.

Chapter 4

Continuous Time Markov Chains

In Chapter 1, we considered stochastic processes that were discrete in both time and space, and that satisfied the Markov property: conditional on the present, the future of the process is independent from the past. Here we generalize such models by allowing for time to be continuous. As before, we will always take our state space to be either finite or countably infinite. A good mental image to have when first encountering continuous time Markov chains is simply a discrete time Markov chain in which transitions can happen at any time.

Throughout this chapter, we make the assumption that our processes are *cadlag*, meaning they are right-continuous with left hand limits. That is, we assume that for each $t \geq 0$, $\lim_{s \to t^-} X(s)$ exists and that $\lim_{s \to t^+} X(s) = X(t)$. This condition implies that if a transition occurs "at time t," then we take $X(t)$ to be the new state so that $X(t) \neq \lim_{s \to t^-} X(s)$.

4.1 The basics

We begin with the definition of a continuous time Markov chain.

Definition 4.1. The continuous time process $X(t), t \geq 0$, taking values in the discrete state space S is said to be a *continuous time Markov chain* (CTMC) if for any $t > s$ and any $j \in S$,

$$P(X(t) = j \mid X(r), 0 \leq r \leq s) = P(X(t) = j \mid X(s)). \tag{4.1}$$

The process is said to be *time-homogeneous* if

$$P(X(t) = j \mid X(s) = k) = P(X(t-s) = j \mid X(0) = k) \tag{4.2}$$

for any $0 \leq s \leq t$ and any states $j, k \in S$.

As in the discrete time setting, the assumption in Definition 4.1 can be stated succinctly as: conditioned upon the present (time s), the future (time $t > s$) is independent from the past.

Example 4.2. Suppose the state space of a continuous time Markov chain is $S = \{1, 2\}$. We assume there can only be transitions between the two states, so that we do not allow transitions of the form $(1) \to (1)$ or $(2) \to (2)$. Graphically, we therefore only allow

$$(1) \rightleftarrows (2).$$

If we were to model the dynamics via a discrete time Markov chain, the transition matrix would simply be

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and the dynamics would be quite trivial: the process would begin in state 1 or 2, and then deterministically transition between the two states.

At this point, we do not know how to understand the dynamics in the continuous time setting. In particular, we do not yet know the distribution for the times between transitions. △

Motivated by Example 4.2, our first question is: how long will a time-homogeneous continuous time Markov chain remain in a given state? Explicitly, suppose $X(0) = x \in S$ and let T_x denote the time we transition away from state x. To find the distribution of T_x, we let $s, t \geq 0$ and find

$$\begin{aligned}
&P(T_x > s + t \mid T_x > s) \\
&\quad = P(X(r) = x \text{ for } r \in [s, s+t] \mid X(r) = x \text{ for } r \in [0, s]) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{since } P(A \cap B|B) = P(A|B)) \\
&\quad = P(X(r) = x \text{ for } r \in [s, s+t] \mid X(s) = x) \qquad\qquad (\text{Markov property}) \\
&\quad = P(X(r) = x \text{ for } r \in [0, t] \mid X(0) = x) \qquad\qquad (\text{time homogeneity}) \\
&\quad = P(T_x > t).
\end{aligned}$$

Hence, T_x satisfies the loss of memory property, and is therefore exponentially distributed.[1] We denote the parameter of the exponential holding time for state x as $\lambda(x)$ so that for any $t \geq 0$,

$$P(T_x \leq t) = 1 - e^{-\lambda(x)t}.$$

Of course, we have $\mathbb{E}[T_x] = \frac{1}{\lambda(x)}$. Thus, the larger $\lambda(x)$, representing the *rate out* of state x, the smaller the expected time for the transition to occur.

Example 4.3. We return to Example 4.2, though now we assume the rate from state 1 to state 2 is $\lambda(1) > 0$, and the rate from state 2 to state 1 is $\lambda(2) > 0$. We commonly incorporate these parameters into the model by placing them next to the transition arrow in the graph:

$$(1) \underset{\lambda(2)}{\overset{\lambda(1)}{\rightleftarrows}} (2).$$

[1] Recall that the exponential random variable is the only continuous random variable with this property.

The dynamics of the model are now clear. Assuming $X(0) = 1$, the process will remain in state 1 for an exponentially distributed amount of time, with parameter $\lambda(1)$, at which point it will transition to state 2. The process will remain in state 2 for an exponentially distributed amount of time with parameter $\lambda(2)$, at which time it will transition to state 1. This process then continuous indefinitely. △

Example 4.3 is deceptively simple as it is clear that when the process transitions out of state 1, it must go to state 2, and vice versa. However, consider the process with states 1, 2, and 3 with transition graph

$$(1) \rightleftarrows (2) \rightleftarrows (3).$$

Even if you are told the holding time parameter for state 2, without further information you can not figure out wether you transition to state 1 or state 3 after you leave state 2. Therefore, we see we need to understand the *transition probabilities* associated with the process, which we do now.

We begin with the following proposition stating that where the chain goes is independent from how long it takes to transition.

Proposition 4.4. *Conditioned upon $X(0) = x$, the random variable $X(T_x)$ is independent from T_x.*

Proof. It is sufficient to show that for any $s \geq 0$ and any $y \in S$

$$P(X(T_x) = y | T_x > s, X(0) = x) = P(X(T_x) = y | X(0) = x).$$

We have

$$\begin{aligned} P(X(T_x) = y | T_x > s, X(0) = x) &= P(X(T_x) = y | X(r) = x, 0 \leq r \leq s) \\ &= P(X(T_x) = y | X(s) = x) \\ &= P(X(T_x) = y | X(0) = x), \end{aligned}$$

where the second equality follows from the Markov property and the third equality follows from time-homogeneity. □

For each $y \neq x$ in S, we define

$$p_{xy} = P(X(T_x) = y \mid X(0) = x).$$

We next define

$$\lambda(x, y) = \lambda(x) p_{xy}.$$

Since T_x is exponential with parameter $\lambda(x)$, we have that

$$P(T_x < h) = 1 - e^{-\lambda(x)h} = \lambda(x)h + o(h), \text{ as } h \to 0.$$

Combining the above, for $y \neq x$ we have

$$\begin{aligned} P(X(h) = y \mid X(0) = x) &= P(T_x < h, X(T_x) = y \mid X(0) = x) + o(h) \\ &= \lambda(x) h p_{xy} + o(h) \\ &= \lambda(x, y) h + o(h), \end{aligned} \tag{4.3}$$

as $h \to 0$, where the $o(h)$ in the first equality represents the probability of seeing two or more jumps (each with an exponential distribution) in the time window $[0, h]$. Therefore, $\lambda(x, y)$ yields the *local rate*, or intensity, of transitioning from state x to state y. Note that for $x \in S$

$$\sum_{y \neq x} \lambda(x, y) = \sum_{y \neq x} \lambda(x) p_{xy} = \lambda(x).$$

We also have

$$P(X(h) = x \mid X(0) = x) = 1 - \lambda(x)h + o(h). \tag{4.4}$$

By time-homogeneity and (4.3)-(4.4), for any $t \geq 0$ we have

$$P(X(t + h) = y | X(t) = x) = \lambda(x, y)h + o(h) \tag{4.5}$$
$$P(X(t + h) = x | X(t) = x) = 1 - \lambda(x)h + o(h) \tag{4.6}$$

Similarly to our consideration of the Poisson process, it can be argued that any time homogeneous process satisfying the local conditions (4.5) and (4.6) also satisfies the Markov property (4.1). This is not surprising as the conditions (4.5)-(4.6) only make use of the current state of the system and ignore the entire past. This leads to an equivalent definition of a continuous time Markov chain that incorporates all the relevant parameters of the model and is probably the most common definition in the literature.

Definition 4.5. A time-homogeneous *continuous time Markov chain* with transition rates $\lambda(x, y)$ is a stochastic process $X(t), t \geq 0$, taking values in a finite or countably infinite state space S satisfying

$$P(X(t + h) = x \mid X(t) = x) = 1 - \lambda(x)h + o(h)$$
$$P(X(t + h) = y \mid X(t) = x) = \lambda(x, y)h + o(h),$$

where $y \neq x$, and $\lambda(x) = \sum_{y \neq x} \lambda(x, y)$.

When only the local rates $\lambda(x, y)$ are given, then the transition probabilities of the chain can be recovered via the identity

$$p_{xy} = \frac{\lambda(x, y)}{\lambda(x)} = \frac{\lambda(x, y)}{\sum_{y \neq x} \lambda(x, y)}.$$

Example 4.6. Let N be a Poisson process with intensity $\lambda > 0$. As N satisfies

$$P(N(t + h) = j + 1 \mid N(t) = j) = \lambda h + o(h)$$
$$P(N(t + h) = j \mid N(t) = j) = 1 - \lambda h + o(h),$$

we see that it is a time-homogeneous continuous time Markov chain. Note also that the Poisson process is the continuous time version of the deterministically monotone chain from Example 1.11. △

Example 4.7. Consider again the three state Markov chain

$$\textcircled{1} \underset{\lambda(2,1)}{\overset{\lambda(1,2)}{\rightleftarrows}} \textcircled{2} \underset{\lambda(3,2)}{\overset{\lambda(2,3)}{\rightleftarrows}} \textcircled{3},$$

where the local transition rates have been placed next to their respective arrows. Note that the holding time in state two is an exponential random variable with a parameter of

$$\lambda(2) = \lambda(2,1) + \lambda(2,3),$$

and the probability that the chain enters state 1 after leaving state 2 is

$$p_{21} = \frac{\lambda(2,1)}{\lambda(2,1) + \lambda(2,3)},$$

whereas the probability that the chain enters state 3 after leaving state 2 is

$$p_{23} = \frac{\lambda(2,3)}{\lambda(2,1) + \lambda(2,3)}.$$

This chain could then be simulated by sequentially computing holding times and transitions. △

An algorithmic construction of a general continuous time Markov chain should now be apparent, and will involve two building blocks. The first will be a stream of unit exponential random variables used to construct our holding times, and the second will be a discrete time Markov chain, denoted X_n, with transition probabilities p_{xy} that will be used to determine the sequence of states. Note that for this discrete time chain we necessarily have that $p_{xx} = 0$ for each x. We also explicitly note that the discrete time chain, X_n, is different than the continuous time Markov chain, $X(t)$, and the reader should be certain to clarify this distinction. The discrete time chain is often called the *embedded* chain associated with the process $X(t)$.

Algorithm 2. (Algorithmic construction of continuous time Markov chain)
Input:

- Let X_n, $n \geq 0$, be a discrete time Markov chain with transition matrix Q. Let the initial distribution of this chain be denoted by α so that $P(X_0 = k) = \alpha_k$.
- Let $E_n, n \geq 0$, be a sequence of independent unit exponential random variables.

Algorithmic construction:

1. Select $X(0) = X_0$ according to the initial distribution α.
2. Let $T_0 = 0$ and define $W(0) = E_0/\lambda(X(0))$, which is exponential with parameter $\lambda(X(0))$, to be the waiting time in state $X(0)$.
3. Let $T_1 = T_0 + W(0)$, and define $X(t) = X(0)$ for all $t \in [T_0, T_1)$.

4. Let X_1 be chosen according to the row associated with X_0 in the transition matrix Q, and define $W(1) = E_1/\lambda(X_1)$.

5. Let $T_2 = T_1 + W(1)$ and define $X(t) = X_1$ for all $t \in [T_1, T_2)$.

6. Continue process.

Note that two random variables will be needed at each iteration of Algorithm 2, one to compute the holding time, and one to compute the next state of the discrete time Markov chain. In the biology/chemistry context, the algorithm implicit in the above construction is typically called the *Gillespie algorithm*, after Dan Gillespie. However, it (and its natural variants) is also called, depending on the field, the *stochastic simulation algorithm*, *kinetic Monte Carlo*, *dynamic Monte Carlo*, the *residence-time algorithm*, the *n-fold way*, or the *Bortz-Kalos-Liebowitz algorithm*. This algorithm has been discovered many times and plays a critical role in many branches of science.

As the future of the process constructed in Algorithm 2 only depends upon the current state of the system, and the current holding time is exponentially distributed, it satisfies the Markov property (4.1). Further, for $y \neq x$ we have

$$\begin{aligned} P(X(h) = y \mid X(0) = x) &= P(X(T_1) = y, T_1 \leq h \mid X(0) = x) + o(h) \\ &= \lambda(x) h p_{xy} + o(h) \\ &= \lambda(x, y) h, \end{aligned}$$

showing we also get the correct local intensities.

One useful way to think about the construction in Algorithm 2 is in terms of alarm clocks:

1. When the chain enters state x, independent "alarm clocks" are placed at each state y, and the yth is programed to go off after an exponentially distributed amount of time with parameter $\lambda(x, y)$.

2. When the first alarm goes off, the chain moves to that state, all alarm clock are discarded, and we repeat the process.

Note that to prove that this algorithm is, in fact, equivalent to the algorithmic construction above, you need to recall that the minimum of exponential random variables with parameters $\lambda(x, y)$ is itself exponentially distributed with parameter

$$\lambda(x) = \sum_y \lambda(x, y),$$

and that it is the yth that went off with probability

$$\frac{\lambda(x, y)}{\sum_{j \neq x} \lambda(x, j)} = \frac{\lambda(x, y)}{\lambda(x)}.$$

See Propositions B.3 and B.4.

We close this section with three examples.

Example 4.8. We consider a random walker on $S = \{0, 1, \dots\}$. We suppose the transition intensities are

$$\begin{aligned} \lambda(i, i+1) &= \lambda \\ \lambda(i, i-1) &= \mu, \qquad \text{if } i > 0, \end{aligned}$$

and $\lambda(0, -1) = 0$. Therefore, the probability of the embedded discrete time Markov chain transitioning up if the current state is $i \neq 0$, is $\lambda/(\lambda+\mu)$, whereas the probability of transitioning down is $\mu/(\lambda + \mu)$. When $i \neq 0$, the holding times will always be exponentially distributed with a parameter of $\lambda + \mu$. △

Example 4.9. We generalize Example 4.8 by allowing the transition rates to depend upon the current state of the system. As in the discrete time setting this leads to a birth and death process.

For $i \in \{0, 1, \dots, \}$ let

$$\begin{aligned} \lambda(i, i+1) &= B(i) \\ \lambda(i, i-1) &= D(i), \end{aligned}$$

where we take $D(0) = 0$. Note that the transition rates are now state dependent, and may even be unbounded as $i \to \infty$. Common choices for the rates include

$$\begin{aligned} B(i) &= \lambda i \\ D(i) &= \mu i, \end{aligned}$$

for some scalar $\lambda, \mu > 0$. Another common model would be to assume a population satisfies a logistical growth model,

$$\begin{aligned} B(i) &= ri \\ D(i) &= \frac{r}{K} i^2. \end{aligned}$$

where K is the *carrying capacity*.

Analogously to Example 3.35, if we let $X(t)$ denote the state of the system at time t, we have that $X(t)$ solves the stochastic equation

$$X(t) = X(0) + Y_1 \left(\int_0^t B(X(s)) ds \right) - Y_2 \left(\int_0^t D(X(s)) ds \right), \tag{4.7}$$

where Y_1 and Y_2 are independent unit-rate Poisson processes. As in Example 3.35, it is now an exercise to show that the solution to (4.7) satisfies the correct local intensity relations of Definition 4.5. For example, denoting

$$A(t) = Y_1 \left(\int_0^t B(X(s)) ds \right), \quad \text{and} \quad D(t) = Y_2 \left(\int_0^t D(X(s)) ds \right),$$

we see that

$$\begin{aligned} P(X(t+h) = x+1 \mid X(t) = x) &\\ &= P(A(t+h) - A(t) = 1, D(t+h) - D(t) = 0 \mid X(t) = x) + o(h) \\ &= B(x)h(1 - D(x)h) + o(h) \\ &= B(x)h + o(h). \end{aligned}$$

△

4.2 Explosions

We demonstrate a behavior that is not possible in the discrete time setting: explosions. Recall that in Algorithm 2, which constructs a continuous time Markov chain, the value T_n represents the time of the nth transition of the chain. Therefore, the chain so constructed is only defined up until the (random) time

$$T_\infty \stackrel{\text{def}}{=} \lim_{n\to\infty} T_n.$$

If $T_\infty < \infty$, then we say that an *explosion* has happened.

Definition 4.10. If

$$P_i(T_\infty = \infty) = P(T_\infty = \infty \mid X(0) = i) = 1, \quad \text{for all } i \in S,$$

than we will say the process is *non-explosive*. Otherwise we will say the process is *explosive*.

Note that a process could be explosive even if

$$P_i(T_\infty = \infty) = 1,$$

for some $i \in S$; see Example 4.13. It is not too difficult to construct an explosive process.

Proposition 4.11. *Suppose that $\{E_n\}$, $n \geq 1$, are independent exponential random variables with respective parameters λ_n. Then,*

$$P\left(\sum_{n=1}^{\infty} E_n < \infty\right) = 1 \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty.$$

Proof. We will prove one direction of the implication (the one we will use). For the other direction, see [31, Section 5.1]. We suppose that $\sum_n \frac{1}{\lambda_n} < \infty$. Because $\sum_n E_n \geq 0$ and

$$\mathbb{E}\left[\sum_n E_n\right] = \sum_n \mathbb{E}[E_n] = \sum_n \frac{1}{\lambda_n} < \infty,$$

we may conclude that $\sum_n E_n < \infty$ with probability one. □

Thus, we see that we can construct an explosive birth process by requiring that the parameters of the exponential holding times satisfy $\sum_n 1/\lambda(X_n) < \infty$.

Example 4.12 (An explosive process)**.** Consider a pure birth process in which the embedded discrete time Markov chain is the deterministically monotone chain of Example 1.11. Suppose that the holding time parameter in state i is $\lambda(i)$. Finally,

let $X(t)$ denote the state of the continuous time process at time t. The stochastic equation satisfied by X is

$$X(t) = X(0) + N\left(\int_0^t \lambda(X(s))ds\right),$$

where N is a unit-rate Poisson process. Suppose that $\lambda(n) = \lambda n^2$ for some $\lambda > 0$ and that $X(0) = 1$. Then the nth holding time is determined by an exponential random variable with parameter λn^2, which we denote by E_n. Since

$$\sum_n \frac{1}{\lambda n^2} < \infty,$$

we may conclude by Proposition 4.11 that

$$P\left(\sum_n E_n < \infty\right) = 1,$$

and the process is explosive. The stochastic equation for this model is

$$X(t) = X(0) + N\left(\lambda \int_0^t X(s)^2 ds\right),$$

and should be compared with the deterministic ordinary differential equation

$$x'(t) = \lambda x^2(t) \quad \Longleftrightarrow \quad x(t) = x(0) + \lambda \int_0^t x(s)^2 ds,$$

which also explodes in finite time. △

Example 4.13. Consider a continuous time Markov chain with state space

$$S = \{-2, -1, 0, 1, 2, \dots\}.$$

We suppose that the graph of the model is

$$\text{-2} \underset{1}{\overset{1}{\rightleftarrows}} \text{-1} \overset{2}{\leftarrow} \text{0} \overset{1}{\rightarrow} \text{1} \overset{1}{\rightarrow} \text{2} \overset{2^2}{\rightarrow} \text{3} \overset{3^2}{\rightarrow} \cdots,$$

where, in general, the intensity of the transition $n \to n+1$, for $n \geq 1$, is $\lambda(n) = n^2$. From the previous example, we know this process is explosive. However, if $X(0) \in \{-2, -1\}$, then the probability of explosion is zero, whereas if $X(0) = 0$, the probability of explosion is $1/3$. △

The following proposition characterizes the most common ways in which a process is non-explosive. A full proof can be found in [31].

Proposition 4.14. *For any $i \in S$,*

$$P_i(T_\infty < \infty) = P_i\left(\sum_n \frac{1}{\lambda(X_n)} < \infty\right), \tag{4.8}$$

and therefore, the continuous time Markov chain is non-explosive iff

$$P_i\left(\sum_n \frac{1}{\lambda(X_n)} = \infty\right) = 1,$$

for every $i \in S$. In particular,

(1) If $\lambda(i) \leq c$ for all $i \in S$ for some $c > 0$, then the chain is non-explosive.

(2) If S is a finite set, then the chain is non-explosive.

(3) If $T \subset S$ are the transient states of the embedded discrete time chain $\{X_n\}$ and if

$$P_i(X_n \in T, \text{ for every } n \geq 1) = 0,$$

for every $i \in S$, then the chain is non-explosive.

Proof. The equivalence of the probabilities in (4.8) is shown in [31, Section 5.2]. We will focus on the three conditions implying non-explosiveness.

For (1), simply note that

$$\sum_n \frac{1}{\lambda(X(n))} \geq \sum_n \frac{1}{c} = \infty.$$

To show (2), we note that if the state space is finite, we may simply take $c = \max\{\lambda_i\}$, and apply (1).

We will now show (3). If $P_i(X_n \in T, \text{ for every } n \geq 1) = 0$, then entry into T^c is assured. There must, therefore, be a state $i \in T^c$, which is hit infinitely often (note that this value can be different for different realizations of the process). Let the infinite sequence of times for which $X_n = i$ be denoted by $n_j, j \geq 1$. Then,

$$\sum_n \frac{1}{\lambda(X_n)} \geq \sum_j \frac{1}{\lambda(X_{n_j})} = \sum_j \frac{1}{\lambda(i)} = \infty,$$

showing the result. □

We will henceforth have a running assumption that unless otherwise explicitly stated, all processes considered in this chapter are non-explosive.

4.3 Forward Equation, Backward Equation, and the Generator Matrix

In this section, we consider non-explosive continuous time Markov chains and we focus on transition probabilities of the form

$$P_{ij}(t) = P(X(t) = j \mid X(0) = i),$$

where $i, j \in S$ and $t \geq 0$? We first derive the Kolmogorov forward equations. We have

$$\begin{aligned}
P'_{ij}(t) &= \lim_{h\to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} \\
&= \lim_{h\to 0} \frac{1}{h} \left(P(X(t+h) = j \mid X(0) = i) - P(X(t) = j \mid X(0) = i)\right) \\
&= \lim_{h\to 0} \frac{1}{h} \Bigg(\sum_{y\in S} P(X(t+h) = j \mid X(t) = y, X(0) = i) P(X(t) = y \mid X(0) = i) \\
&\qquad - P(X(t) = j \mid X(0) = i) \Bigg).
\end{aligned}$$

However,

$$\begin{aligned}
&\sum_{y\in S} P(X(t+h) = j \mid X(t) = y, X(0) = i) P(X(t) = y \mid X(0) = i) \\
&\quad = P(X(t+h) = j \mid X(t) = j, X(0) = i) P(X(t) = j \mid X(0) = i) \\
&\qquad + \sum_{y\neq j} P(X(t+h) = j \mid X(t) = y, X(0) = i) P(X(t) = y \mid X(0) = i) \qquad (4.9) \\
&\quad = (1 - \lambda(j)h) P_{ij}(t) + \sum_{y\neq j} \lambda(y, j) h P_{iy}(t) + o(h), \qquad (4.10)
\end{aligned}$$

and so

$$\begin{aligned}
P'_{ij}(t) &= \lim_{h\to 0} \frac{1}{h} \left((1 - \lambda(j)h - 1) P_{ij}(t) + \sum_{y\neq j} \lambda(y, j) P_{iy}(t) h + o(h) \right) \\
&= -\lambda(j) P_{ij}(t) + \sum_{y\neq j} \lambda(y, j) P_{iy}(t).
\end{aligned}$$

Thus,

$$P'_{ij}(t) = -\lambda(j) P_{ij}(t) + \sum_{y\neq j} P_{iy}(t) \lambda(y, j). \qquad (4.11)$$

These are the *Kolmogorov forward equations* for the process. In the biology literature this system of equations is often termed the *chemical master equation*.

We point out that there was a small mathematical "slight of hand" in the above calculation. To move from (4.9) to (4.10), we had to assume that

$$\sum_y P_{iy}(t)o_y(h) = o(h),$$

where we write $o_y(h)$ to show that the size of the error can depend upon the state y. This condition is satisfied for all systems we will consider.

Definition 4.15. Let $X(t)$ be a continuous time Markov chain on some state space S with transition intensities $\lambda(i,j) \geq 0$. Recalling that

$$\lambda(i) = \sum_{j \neq i} \lambda(i,j),$$

The matrix

$$A_{ij} = \begin{cases} -\lambda(i), & \text{if } i = j \\ \lambda(i,j), & \text{if } i \neq j \end{cases} = \begin{cases} -\sum_j \lambda(i,j), & \text{if } i = j \\ \lambda(i,j), & \text{if } i \neq j \end{cases}$$

is called the *generator*, or *infinitesimal generator*, or *generator matrix* of the Markov chain.

We see that the Kolmogorov forward equations (4.11) can be written as the matrix differential equation

$$P'(t) = P(t)A,$$

since

$$\begin{aligned}(P(t)A)_{ij} &= \sum_y P_{iy}(t)A_{yj} = P_{ij}A_{jj} + \sum_{y \neq j} P_{iy}A_{yj} \\ &= -\lambda(j)P_{ij}(t) + \sum_{y \neq j} P_{iy}\lambda(y,j).\end{aligned}$$

At least formally, this system can be solved

$$P(t) = P(0)e^{tA} = e^{tA},$$

where e^{tA} is the matrix exponential and we used that $P(0) = I$, the identity matrix. recall that the matrix exponential is defined by

$$e^{At} \overset{\text{def}}{=} \sum_{k=0}^{\infty} \frac{t^n A^n}{n!}.$$

This solution is always valid in the case that the state space is finite.

We make the following observations pertaining to the generator A:

1. The elements on the main diagonal are all strictly negative.

2. The elements off the main diagonal are non-negative.

3. Each row sums to zero.

We also point out that given a state space S, the infinitesimal generator A completely determines the Markov chain as it contains all the local information pertaining to the transitions: $\lambda(i,j)$. Thus, it is sufficient to characterize a chain by simply providing a state space, S, and generator, A.

Example 4.16. A molecule transitions between states 0 and 1. The transition rates are $\lambda(0,1) = 3$ and $\lambda(1,0) = 1$. The generator matrix is

$$A = \begin{pmatrix} -3 & 3 \\ 1 & -1 \end{pmatrix}.$$

△

Example 4.17. Consider a mathematician wandering between three coffee shops with graphical structure

$$A \underset{\lambda_1}{\overset{\mu_1}{\rightleftarrows}} B \underset{\lambda_2}{\overset{\mu_2}{\rightleftarrows}} C.$$

The infinitesimal generator of this process is

$$A = \begin{pmatrix} -\mu_1 & \mu_1 & 0 \\ \lambda_1 & -(\lambda_1+\mu_2) & \mu_2 \\ 0 & \lambda_2 & -\lambda_2 \end{pmatrix},$$

and the transition matrix for the embedded Markov chain is

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \lambda_1/(\lambda_1+\mu_2) & 0 & \mu_2/(\lambda_1+\mu_2) \\ 0 & 1 & 0 \end{pmatrix}.$$

△

Example 4.18. For a unit-rate Poisson process, we have

$$A = \begin{pmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & 0\dots \\ 0 & 0 & -1 & 1 \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

△

If we are given an initial distribution of α and want the probability of being in state j at time t, then we simply note

$$P_\alpha(X(t) = j) \overset{\text{def}}{=} \sum_i P(X(t) = j \mid X(0) = i)P(X(0) = i) = \sum_i \alpha_i P_{ij} = (\alpha P(t))_j.$$

Thus,

$$\alpha P(t) = P_\alpha(t) = \alpha e^{tA}. \tag{4.12}$$

Backward equation

Before attempting to solve a system using Kolmogorov's forward equations, we introduce another set of equations, called *Kolmogorov's backward equations*, which are valid for all continuous time Markov chains. The derivation below follows that of [31].

We begin by finding an integral equation satisfied by $P_{ij}(t)$. We will then differentiate it to get the backward equations.

Proposition 4.19. *For all $i, j \in S$ and $t \geq 0$, we have*

$$P_{ij}(t) = \delta_{ij} e^{-\lambda(i)t} + \int_0^t \lambda(i) e^{-\lambda(i)s} \sum_{k \neq i} Q_{ik} P_{kj}(t-s) ds,$$

where, as usual,

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

is the Kronecker delta function, and Q is the transition matrix of the embedded discrete time Markov chain.

Proof. Conditioning on the first jump time of the chain, T_1, we have

$$\begin{aligned} &P(X(t) = j \mid X(0) = i) \\ &\qquad = P(X(t) = j, T_1 > t \mid X(0) = i) + P(X(t) = j, T_1 \leq t \mid X(0) = i). \end{aligned}$$

We handle these terms separately. For the first term on the right hand side of the above equation, a first transition has not been made. Thus, $X(t) = j$ iff $j = i$ and does so with a probability of one. That is,

$$\begin{aligned} &P(X(t) = j, T_1 > t \mid X(0) = i) \\ &\qquad = P(X(t) = j \mid T_1 > t, X(0) = i) P(T_1 > t \mid X(0) = i) \\ &\qquad = \delta_{ij} P_i(T_1 > t) \\ &\qquad = \delta_{ij} e^{-\lambda(i)t}. \end{aligned}$$

For the second term, we will condition on the time of the first jump happening in $(s, s+\Delta)$, for small Δ (we will eventually take $\Delta \to 0$). As the holding time is exponential with parameter $\lambda(i)$, this event has probability

$$\int_s^{s+\Delta} \lambda(i) e^{-\lambda(i)r} dr = \lambda(i) e^{-\lambda(i)s} \Delta + O(\Delta^2).$$

We let $s_n = nt/N$ for some large N, denote $\Delta = t/N$, and see

$$\begin{aligned}
P(X(t) = j, T_1 \le t \mid X(0) = i) &= \sum_{n=0}^{N-1} P(X(t) = j, T_1 \in (s_n, s_{n+1}) \mid X(0) = i) \\
&= \sum_{n=0}^{N-1} P(X(t) = j \mid X(0) = i, T_1 \in (s_n, s_{n+1})) P(T_1 \in (s_n, s_{n+1}) \mid X(0) = i) \\
&= \sum_{n=0}^{N-1} P(X(t) = j \mid X(0) = i, T_1 \in (s_n, s_{n+1})) \left[\lambda(i) e^{-\lambda(i) s_n} \Delta + O(\Delta^2)\right] \\
&= \sum_{n=0}^{N-1} \lambda(i) e^{-\lambda(i) s_n} \sum_{k \neq i} P(X(t) = j, X_1 = k \mid X(0) = i, T_1 \in (s_n, s_{n+1})) \Delta + O(\Delta) \\
&= \sum_{n=0}^{N-1} \lambda(i) e^{-\lambda(i) s_n} \sum_{k \neq i} \Big[P(X(t) = j \mid X_1 = k, X(0) = i, T_1 \in (s_n, s_{n+1})) \\
&\qquad \times P(X_1 = k \mid X(0) = i, T_1 \in (s_n, s_{n+1})) \Big] \Delta + O(\Delta) \\
&\approx \sum_{n=0}^{N-1} \lambda(i) e^{-\lambda(i) s_n} \sum_{k \neq i} Q_{ik} P_{kj}(t - s_n) \Delta + O(\Delta) \\
&\to \int_0^t \lambda(i) e^{-\lambda(i) s} \sum_{k \neq i} Q_{ik} P_{kj}(t - s) ds,
\end{aligned}$$

as $\Delta \to 0$. Combining the above shows the result. □

Proposition 4.20. *For all $i, j \in S$, we have that $P_{ij}(t)$ is continuously differentiable and*

$$P'(t) = AP(t), \tag{4.13}$$

which in component form is

$$P'_{ij}(t) = \sum_k A_{ik} P_{kj}(t).$$

The system of equations (4.13) is called the *Kolmogorov backwards equations*. Note that the difference with the forward equations is the order of the multiplication of $P(t)$ and A. However, the solution of the backwards equation is once again seen to be

$$P(t) = e^{tA},$$

agreeing with previous results.

Proof. Use the substitution $u = t - s$ in the integral equation to find that

$$\begin{aligned} P_{ij}(t) &= \delta_{ij} e^{-\lambda(i)t} + \int_0^t \lambda(i) e^{-\lambda(i)s} \sum_{k \neq i} Q_{ik} P_{kj}(t-s) ds \\ &= \delta_{ij} e^{-\lambda(i)t} + \int_0^t \lambda(i) e^{-\lambda(i)(t-u)} \sum_{k \neq i} Q_{ik} P_{kj}(u) ds \\ &= e^{-\lambda(i)t} \left[\delta_{ij} + \int_0^t \lambda(i) e^{\lambda(i)u} \sum_{k \neq i} Q_{ik} P_{kj}(u) ds \right]. \end{aligned}$$

Differentiating yields

$$\begin{aligned} P'_{ij}(t) &= -\lambda(i) e^{-\lambda(i)t} \left[\delta_{ij} + \int_0^t \lambda(i) e^{\lambda(i)u} \sum_{k \neq i} Q_{ik} P_{kj}(u) ds \right] \\ &\quad + e^{-\lambda(i)t} \cdot \lambda(i) e^{\lambda(i)t} \sum_{k \neq i} Q_{ik} P_{kj}(t) \\ &= -\lambda(i) P_{ij}(t) + \lambda(i) \sum_{k \neq i} Q_{ik} P_{kj}(t) \\ &= \sum_k (-\lambda(i) \delta_{ik} P_{kj}(t)) + \sum_k \lambda(i) Q_{ik} P_{kj}(t) \\ &= \sum_k (-\lambda(i) \delta_{ik} + \lambda(i) Q_{ik}) P_{kj}(t) \\ &= \sum_k A_{ik} P_{kj}(t). \end{aligned}$$

□

Both the forward and backward equations can be used to solve for the associated probabilities as the next example demonstrates.

Example 4.21. We consider a two state, $\{0, 1\}$, continuous time Markov chain with generator matrix

$$A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

We will use both the forwards and backwards equations to solve for $P(t)$.

Approach 1: Backward equation. While we want to compute $P_{ij}(t)$ for each pair $i, j \in \{0, 1\}$, we know that

$$P_{00}(t) + P_{01}(t) = P_{10}(t) + P_{11}(t) = 1,$$

for all $t \geq 0$, and so it is sufficient to solve just for $P_{00}(t)$ and $P_{10}(t)$.

The backwards equation is $P'(t) = AP(t)$, yielding the equations

$$\begin{aligned} P'_{00}(t) &= \lambda[P_{10}(t) - P_{00}(t)] \\ P'_{10}(t) &= \mu[P_{00}(t) - P_{10}(t)]. \end{aligned}$$

We see that

$$\mu P'_{00}(t) + \lambda P'_{10}(t) = 0 \implies \mu P_{00}(t) + \lambda P_{10}(t) = c.$$

We know that $P(0) = I$, so we see that

$$\mu P_{00}(0) + \lambda P_{10}(0) = c \iff \mu = c.$$

Thus,

$$\mu P_{00}(t) + \lambda P_{10}(t) = \mu \implies \lambda P_{10}(t) = \mu - \mu P_{00}(t).$$

Putting this back into our differential equations above we have that

$$P'_{00}(t) = \mu - \mu P_{00}(t) - \lambda P_{00}(t) = \mu - (\mu + \lambda) P_{00}(t).$$

Solving, with $P_{00}(t) = 1$ yields

$$P_{00}(t) = \frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} e^{-(\mu+\lambda)t}.$$

Of course, we also have that

$$\begin{aligned} P_{01}(t) &= 1 - P_{00}(t) \\ P_{10}(t) &= \frac{\mu}{\lambda} - \frac{\mu}{\lambda}\left(\frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} e^{-(\mu+\lambda)t}\right) = \frac{\mu}{\mu + \lambda} - \frac{\mu}{\mu + \lambda} e^{-(\mu+\lambda)t}. \end{aligned}$$

Approach 2: Forward equation. This is easier. We want to solve

$$P'(t) = P(t)A.$$

We now get

$$\begin{aligned} P'_{00}(t) &= -P_{00}(t)\lambda + P_{01}(t)\mu = -P_{00}(t)\lambda + (1 - P_{00}(t))\mu = \mu - (\lambda + \mu)P_{00}(t) \\ P'_{10}(t) &= -\lambda P_{10}(t) + \mu P_{11}(t) = -\lambda P_{10}(t) + \mu(1 - P_{10}(t)) = \mu - (\lambda + \mu)P_{00}(t), \end{aligned}$$

and the solutions above follow easily.

Note that, as in the discrete time setting, we have that

$$\lim_{t\to\infty} P(t) = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix},$$

yielding a common row vector which can be interpreted as a limiting distribution. $\triangle$

There is a more straightforward way to make the above computations: simply solve the matrix exponential.

Example 4.22 (Computing matrix exponentials)**.** Suppose that A is an $n \times n$ matrix with n distinct eigenvectors. Then, letting D be a diagonal matrix consisting of the eigenvalues of A, we can decompose A into

$$A = QDQ^{-1},$$

where Q consists of the eigenvectors of A (ordered similarly to the order of the eigenvalues in D). In this case, we get the very nice identity

$$e^{At} = \sum_{n=0}^{\infty} \frac{t^n (QDQ^{-1})^n}{n!} = Q \left(\sum_{n=0}^{\infty} \frac{t^n D^n}{n!} \right) Q^{-1} = Q e^{Dt} Q^{-1},$$

where e^{Dt}, because D is diagonal, is a diagonal matrix with diagonal elements $e^{\lambda_i t}$ where λ_i is the ith eigenvalue.

Example 4.23. We now solve the above problem using the matrix exponential. Supposing, for concreteness, that $\lambda = 3$ and $\mu = 1$, we have that the generator matrix is

$$A = \begin{pmatrix} -3 & 3 \\ 1 & -1 \end{pmatrix}$$

It is easy to check that the eigenvalues are $0, -4$ and the associated eigenvalues are $[1, 1]^t$ and $[-3, 1]^t$. Therefore,

$$Q = \begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} 1/4 & 3/4 \\ -1/4 & 1/4 \end{pmatrix},$$

and

$$e^{tA} = \begin{pmatrix} 1/4 + (3/4)e^{-4t} & 3/4 - (3/4)e^{-4t} \\ 1/4 - (1/4)e^{-4t} & 3/4 + (1/4)e^{-4t} \end{pmatrix}.$$

You should note that

$$\lim_{t \to \infty} e^{tA} = \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix},$$

which has a common row. Thus, for example, in the long run, the chain will be in state zero with a probability of $1/4$. △

4.4 Stationary Distributions

In this section we will parallel our treatment of stationary distributions for discrete time Markov chains. We will aim for intuition, as opposed to attempting to prove everything, and point the interested reader to [31] and [29] for the full details of the proofs.

4.4.1 Classification of states

We start by again classifying the states of our process. Viewing a continuous time Markov chain as an embedded discrete time Markov chain with exponential holding times makes the classification of states, analogous to Section 1.5 in the discrete time setting, easy. We will again denote our state space as S.

Definition 4.24. The communication classes of the continuous time Markov chain $X(t), t \geq 0$, are the communication classes of the embedded Markov chain $X_n, n \geq 0$. If there is only one communication class, we say the chain is *irreducible*; otherwise it is said to be *reducible*.

Noting that $X(t)$ will return to a state i infinitely often if and only if the embedded discrete time chain does motivates the following.

Definition 4.25. State $i \in S$ is called *recurrent* for $X(t), t \geq 0$, if i is recurrent for the embedded discrete time chain $X_n, n \geq 0$. Otherwise, i is *transient*.

Definition 4.26. Let T_1 denote the first jump time of the continuous time chain. We define

$$\tau_i \stackrel{\text{def}}{=} \inf\{t \geq T_1 \ : \ X(t) = i\},$$

and set $m_i = \mathbb{E}_i[\tau_i]$. We say that state i is *positive recurrent* if $m_i < \infty$.

Note that, perhaps surprisingly, we *do not* define i to be positive recurrent if i is positive recurrent for the discrete time chain. In Example 4.33 we will demonstrate that i may be positive recurrent for X_n, while not for $X(t)$.

As in the discrete time setting, recurrence, transience, and positive recurrence are class properties.

Note that the concept of periodicity no longer plays a role, or even makes sense to define, as time is no longer discrete. In fact, if $P(t)$ is the matrix with entries $P_{ij}(t) = P(X(t) = j \mid X(0) = i)$ for an irreducible continuous time chain, then for every $t > 0$, $P_{ij}(t)$ has strictly positive entries because there is necessarily a path between i and j, and a non-zero probability of moving along that path in time $t > 0$.

4.4.2 Invariant measures

Recall that equation (4.12) states that if the initial distribution of the process is α, then $\alpha P(t)$ is the vector whose ith component gives the probability that $X(t) = i$. We therefore define an invariant measure in the following manner.

Definition 4.27. A measure $\eta = \{\eta_j, j \in S\}$ on S is called *invariant* if for all $t > 0$

$$\eta P(t) = \eta.$$

If this measure is a probability distribution, then it is called a *stationary distribution*.

The following theorem gives us a nice way to find stationary distributions of continuous time Markov chains.

Theorem 4.28. *Let $X(t)$ be an irreducible and recurrent continuous time Markov chain with generator matrix A. Then the following statements are equivalent:*

1. *$\eta A = 0$;*
2. *$\eta P(t) = \eta$, for all $t \geq 0$.*

Proof. The proof of this fact is easy in the case of a finite state space, which is what we will assume here. Recall Kolmogorov's backward equation

$$P'(t) = AP(t).$$

Assume that $\eta A = 0$. Multiplying the backwards equation on the left by η shows

$$0 = \eta AP(t) = \eta P'(t) = \frac{d}{dt}\eta P(t).$$

Thus,

$$\eta P(t) = \eta P(0) = \eta,$$

for all $t \geq 0$.

Now assume that $\eta P(t) = \eta$ for al $t \geq 0$. Then, for all $h > 0$, we have

$$\eta P(h) = \eta \quad \Longrightarrow \quad \eta(P(h) - I) = 0 \quad \Longrightarrow \quad \eta\frac{(P(h) - I)}{h} = 0.$$

Taking $h \to 0$ now shows that

$$0 = \eta P'(0) = \eta A,$$

where we have used that $P'(0) = A$, which follows from either the forward or backward equations.

The interchange above of differentiation with summation can not in general be justified in the infinite dimensional setting, and different proof is needed and we refer the reader to [29, Section 3.5]. □

Theorem 4.29. *Suppose that $X(t)$ is irreducible and recurrent. Then $X(t)$ has an invariant measure η, which is unique up to multiplicative factors. Moreover, for each $k \in S$, we have*

$$\eta_k = \frac{\pi_k}{\lambda(k)},$$

where π is the unique invariant measure of the embedded discrete time Markov chain X_n. Finally, η satisfies

$$0 < \eta_j < \infty, \quad \forall j \in S,$$

and if $\sum_i \eta_i < \infty$ then η can normalize by $1/\sum_k \eta_k$ to give a stationary distribution.

Proof. By Theorem 4.28, we must only show that there is a solution to $\eta A = 0$, satisfying all the desired results, if and only if there is an invariant measure to the discrete time chain. We first recall that π was an invariant measure for a discrete time Markov chain if and only if $\pi Q = \pi$, where Q is the transition matrix. By Theorem 1.72, such a π exists, and is unique up to multiplicative constants, if X_n is irreducible and recurrent.

Recall that if $j \neq k$, then $A_{jk} = \lambda(j)Q_{jk}$ and that $A_{jj} = -\lambda(j)$. We now simply note that

$$\eta' A = 0 \iff \sum_j \eta_j A_{jk} = 0, \quad \forall k \iff \sum_{j \neq k} \eta_j \lambda(j) Q_{jk} - \eta_k \lambda(k) = 0.$$

However, this holds if and only if

$$\sum_{j \neq k} \eta_j \lambda(j) Q_{jk} = \eta_k \lambda(k) \iff \pi Q = \pi, \quad \text{where } \pi_k \stackrel{\text{def}}{=} \lambda(k)\eta_k.$$

That is, the final equation (and hence all the others) holds if and only if π is invariant for the Markov matrix Q. Such a π exists, and satisfies all the desired properties, by Theorem 1.72. Further, we see the invariant measure of the continuous time Process satisfies $\eta_k = \pi_k/\lambda(k)$, as desired. □

Example 4.30. Consider the continuous time Markov chain with generator matrix

$$A = \begin{pmatrix} -5 & 3 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 2 & 1 & -4 & 1 \\ 0 & 2 & 2 & -4 \end{pmatrix}.$$

The unique left eigenvector of A with eigenvalue 0, i.e. the solution to $\eta A = 0$, normalized to sum to one is

$$\eta = \left(\frac{14}{83}, \frac{58}{83}, \frac{6}{83}, \frac{5}{83} \right).$$

Further, note that the transition matrix for the embedded discrete time Markov chain is

$$P = \begin{pmatrix} 0 & \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

Solving for the stationary distribution of the embedded chain, i.e. solving $\pi P = \pi$, yields

$$\pi = \left(\frac{35}{86}, \frac{29}{86}, \frac{6}{43}, \frac{5}{43} \right).$$

Finally, note that

$$\begin{aligned}(\eta_1\lambda(1), \eta_2\lambda(2), \eta_3\lambda(3), \eta_4\lambda(4)) &= \left(5 \cdot \tfrac{14}{83}, \tfrac{58}{83}, 4 \cdot \tfrac{6}{83}, 4 \cdot \tfrac{5}{83}\right) \\ &= \left(\tfrac{70}{83}, \tfrac{58}{83}, \tfrac{24}{83}, \tfrac{20}{83}\right) \\ &= \tfrac{172}{83}\left(\tfrac{35}{86}, \tfrac{29}{86}, \tfrac{6}{43}, \tfrac{5}{43}\right) \\ &= \tfrac{172}{83}\pi,\end{aligned}$$

as predicted by the theory. △

We now consider the positive recurrent case. We recall that $m_i = \mathbb{E}_i[\tau_i]$, the expected first return time to state i. The following result should not be surprising at this point. See [29] for a proof.

Theorem 4.31. *Let A be the generator matrix for an irreducible continuous time Markov chain. Then the following are equivalent*

1. *Every state is positive recurrent.*
2. *Some state is positive recurrent.*
3. *A is non-explosive and has an invariant distribution η.*

Definition 4.32. We call the continuous time Markov chain $X(t), t \geq 0$, *ergodic* if it is irreducible and positive recurrent.

The following example shows that positive recurrence of the embedded discrete time Markov chain $X_n, n \geq 0$, does not guarantee that $X(t), t \geq 0$, is positive recurrent.

Example 4.33. We consider a continuous time Markov chain whose embedded discrete time Markov chain has state space $S = \{0, 1, 2, \dots\}$ and transition matrix

$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ q & 0 & p & 0 & \cdots \\ q & 0 & 0 & p & \\ \vdots & & \ddots & & \end{pmatrix},$$

where $p + q = 1$. This is the "success run chain" and we showed in Problem 1.12 that the discrete time chain is positive recurrent. Let $\lambda(i)$ be the holding time parameter for state i of the associated continuous time Markov chain, and let E_m, $m \geq 0$, denote a sequence of independent unit exponential random variables, which are also independent of the embedded discrete time Markov chain. Finally, assuming that $X_0 = 0$, let T_1 denote the first return time to state 0 of the *embedded chain*. For example, if $T_1 = 3$, then $X_0 = 0, X_1 = 1, X_2 = 2$, and $X_3 = 0$. More generally,

we have $X_0 = 0, X_1 = 1, \ldots, X_{T_1-1} = T_1 - 1$, and $X_{T_1} = 0$. For $m < T_1$, we let $W(m) = E_m/\lambda(m)$ be the holding time in state m. We have

$$\begin{aligned} m_0 = \mathbb{E}_0[\tau_0] &= \mathbb{E}_0\left[\sum_{m=0}^{T_1-1} W(m)\right] \\ &= \mathbb{E}\left[\sum_{m=0}^{\infty} W(m)1_{\{m<T_1\}}\right] \\ &= \sum_{m=0}^{\infty} \mathbb{E}[W(m)1_{\{m<T_1\}}]. \end{aligned}$$

However, we know that the holding times and the embedded chain are independent. Thus, as $1_{\{m<T_1\}}$ is simply a statement pertaining to the embedded chain,

$$\mathbb{E}[W(m)1_{\{m<T_1\}}] = \mathbb{E}[W(m)] \cdot \mathbb{E}[1_{\{m<T_1\}}] = \frac{1}{\lambda(m)} P_0(m < T_1).$$

Combining the above,

$$\begin{aligned} m_0 &= \sum_{m=0}^{\infty} \frac{1}{\lambda(m)} P_0(m < T_1) \\ &= \frac{1}{\lambda(0)} + \sum_{m=1}^{\infty} \frac{1}{\lambda(m)} P_0(m < T_1). \end{aligned}$$

For $m \geq 1$,

$$P(m < T_1) = \sum_{n=m+1}^{\infty} P(T_1 = n) = \sum_{n=m+1}^{\infty} p^{n-2}q = qp^{m-1} \sum_{n=0}^{\infty} p^n = p^{m-1}.$$

Thus,

$$m_0 = \frac{1}{\lambda(0)} + \sum_{m=1}^{\infty} \frac{1}{\lambda(m)} p^{m-1}.$$

Of course, we have not chosen $\lambda(m)$ yet. Taking $\lambda(m) = p^m$, we see

$$m_0 = \frac{1}{\lambda(0)} + \sum_{m=1}^{\infty} \frac{1}{p^m} p^{m-1} = 1 + \sum_{m=1}^{\infty} \frac{1}{p} = \infty.$$

So, X_n is positive recurrent, but $X(t)$ is not. △

The following example, taken from [29], shows two things. First, it demonstrates that a transient chain *can* have an invariant measure. Further, it even shows stranger behavior is possible: a transient chain can have an *invariant distribution*! Of course, the previous theorems seem to suggest that this is not possible. However, there is a catch: the chain could be explosive. In fact, if a transient chain is shown to have a stationary distribution, then the chain must be explosive for otherwise Theorem 4.31 is violated.

Example 4.34. Consider a discrete time random walker on $S = \{0, 1, 2, \dots\}$. Suppose that the probability of moving to the right is $p > 0$ and to the left is $q = 1 - p$. To convert this into a continuous time chain, we suppose that $\lambda(i)$ is the holding time parameter in state i. More specifically, we assume $X(t)$ is a continuous time Markov chain with generator matrix A satisfying

$$A = \begin{pmatrix} -\lambda(0)p & \lambda(0)p & 0 & 0 & 0 & \cdots \\ q\lambda(1) & -\lambda(1) & p\lambda(1) & 0 & 0 & \cdots \\ 0 & q\lambda(2) & -\lambda(2) & p\lambda(2) & 0 & \cdots \\ 0 & 0 & q\lambda(3) & -\lambda(3) & p\lambda(3) & \\ \vdots & \ddots & & \ddots & & \end{pmatrix}$$

We know that this chain is transient if $p > q$ since the discrete time chain is. We now search for an invariant measure satisfying

$$\eta A = 0,$$

which in component form is

$$\begin{aligned} -\lambda(0)p\eta_0 + q\lambda(1)\eta_1 &= 0 \\ \lambda(i-1)p\eta_{i-1} - \lambda(i)\eta_i + \lambda(i+1)q\eta_{i+1} &= 0 \qquad i > 0. \end{aligned}$$

We will confirm that η satisfying

$$\eta(i) = \frac{1}{\lambda(i)} \left(\frac{p}{q}\right)^i,$$

is a solution. The case $i = 0$ is easy to verify

$$\lambda(0)p\eta_0 = \lambda(0)p\frac{1}{\lambda(0)} = p = q\lambda(1)\frac{1}{\lambda(1)}\frac{p}{q} = q\lambda(1)\eta_1.$$

The $i > 0$ case follows similarly.

Therefore, there is always an invariant measure, regardless of the values p and q. Taking $p > q$ and $\lambda(i) = 1$ for all i, we see that the resulting continuous time Markov chain is transient, and has an invariant measure

$$\eta(i) = \left(\frac{p}{q}\right)^i,$$

which can not be normalized to provide an invariant *distribution*.

Now, consider the case when $p > q$, with $1 < p/q < 2$, and take $\lambda(i) = 2^i$. Define $\alpha \stackrel{\text{def}}{=} p/q < 2$. Then,

$$\sum_{i=0}^{\infty} \eta(i) = \sum_{i=0}^{\infty} \left(\frac{\alpha}{2}\right)^i = \frac{1}{1 - \alpha/2} = \frac{2}{2 - \alpha} < \infty,$$

Therefore, we can normalize to get a stationary distribution. Since we already know this chain is transient, we have shown that it must, in fact, explode. $\triangle$

4.4.3 Limiting distributions and convergence

We have found conditions for the existence of a unique stationary distribution to a continuous time Markov chain: irreducibility and positive recurrence (i.e. *ergodicity*). As in the discrete time case, there is still the question of convergence. The following is proven in [29].

Theorem 4.35. *Let $X(t)$ be an ergodic continuous time Markov chain with unique invariant distribution η. Then, for all $i, j \in S$,*

$$\lim_{t\to\infty} P_{ij}(t) = \eta_j.$$

Example 4.36. Let $S = \{0, 1\}$ with transition rates $\lambda(0, 1) = 3$ and $\lambda(1, 0) = 1$. Then the generator matrix is

$$A = \begin{pmatrix} -3 & 3 \\ 1 & -1 \end{pmatrix}.$$

Solving directly for the left eigenvector of A with eigenvalue 0 yields

$$\pi = (\tfrac{1}{4},\ \tfrac{3}{4}),$$

which agrees with the result found in Example 4.23. △

As in the discrete time setting, we have an ergodic theorem. For a proof, see [31, Section 5.5].

Theorem 4.37. *Suppose that $X(t)$ is an irreducible, positive recurrent continuous time Markov chain with unique stationary distribution η. Then, for any initial condition, and any $i \in S$,*

$$P\left(\frac{1}{t}\int_0^t 1_{\{X(s)=i\}} ds \to \eta_i, \quad as\ t \to \infty\right) = 1.$$

Moreover, for any bounded function $f : S \to \mathbb{R}$ we have

$$P\left(\frac{1}{t}\int_0^t f(X(s))ds \to \bar{f}, \quad as\ t \to \infty\right) = 1,$$

where

$$\bar{f} = \sum_{j\in S} \eta_j f(j) = \mathbb{E}_\eta[f(X_\infty)],$$

where X_∞ has distribution η.

Thus, as in the discrete time setting, we see that η_i gives the proportion of time spent in state i over long periods of time. This gives us an algorithmic way to sample from the stationary distribution: simulate a single long trajectory and average over it.

4.5 The Generator, Revisited

In this section, we let $X(t), t \geq 0$, be a time-homogeneous continuous time Markov chain with discrete state space S and generator matrix A.

Consider a function $f : S \to \mathbb{R}$. Since f is simply a mapping from the discrete space S to $\mathbb{R}$, we can represent f as a column vector whose ith component is equal to $f(i)$. For example, if $S = \{1, 2, 3\}$ and $f(1) = -2, f(2) = \pi$, and $f(3) = 100$, then we take

$$f = \begin{bmatrix} -2 \\ \pi \\ 100 \end{bmatrix}.$$

As A is a matrix, it is possible to consider Af, which is itself a column vector, and hence a function from S to $\mathbb{R}$.

If the initial distribution for our Markov chain is α, then for any f we have that

$$\begin{aligned} \mathbb{E}_\alpha[f(X(t))] &= \sum_{j \in S} P_\alpha(X(t) = j) f(j) \\ &= \sum_{j \in S} \left(\sum_{i \in S} P(X(t) = j \mid X(0) = i) P_\alpha(X(0) = i) \right) f(j) \\ &= \sum_{i \in S} \alpha_i \left(\sum_{j \in S} P_{ij}(t) f(j) \right) \\ &= \sum_{i \in S} \alpha_i (P(t)f)_i \\ &= \alpha P(t) f. \end{aligned} \tag{4.14}$$

Now recall that the forward equation stated that $P'(t) = P(t)A$. Integrating this equation yields

$$P(t) = I + \int_0^t P(s) A ds,$$

and multiplication on the right by f gives

$$P(t)f = f + \int_0^t P(s) A f ds. \tag{4.15}$$

Multiplying (4.15) on the left by α yields

$$\alpha P(t) f = \alpha f + \int_0^t \alpha P(s)(Af) ds,$$

which combined with (4.14) gives

$$\begin{aligned} \mathbb{E}_\alpha[f(X(t))] &= \mathbb{E}_\alpha[f(X(0))] + \int_0^t \mathbb{E}_\alpha[(Af)(X(s))] \, ds \\ &= \mathbb{E}_\alpha[f(X(0))] + \mathbb{E}_\alpha \left[\int_0^t (Af)(X(s)) \, ds \right]. \end{aligned} \tag{4.16}$$

Equation 4.16 is a version of *Dynkin's formula*. For a more formal derivation in the Markov process setting, see [12, Section 1.1].

Example 4.38. We will re-derive the mean and variance of a Poisson process using Dynkin's formula. Let $X(t)$ be a Poisson process with intensity $\lambda > 0$. The state space is $S = \{0, 1, 2, \dots\}$ and for any function $f : S \to \mathbb{R}$

$$(Af) = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \ddots & & \ddots & \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ f(2) \\ \vdots \end{pmatrix} = \begin{pmatrix} -\lambda f(0) + \lambda f(1) \\ -\lambda f(1) + \lambda f(2) \\ -\lambda f(2) + \lambda f(3) \\ \vdots \end{pmatrix}.$$

Thus, for any $i \in \{0, 1, 2, \dots\}$

$$(Af)(i) = \lambda(f(i+1) - f(i)).$$

Letting $f(i) = i$, and taking $X(0) = 0$ we have

$$\begin{aligned} \mathbb{E}[f(X(t))] = \mathbb{E}[X(t)] &= 0 + \int_0^t \mathbb{E}[(Af)(X(s))]ds \\ &= \int_0^t \mathbb{E}\left[\lambda\big(f(X(s)+1) - f(X(s))\big)\right] ds \\ &= \lambda \int_0^t ds \\ &= \lambda t. \end{aligned}$$

In order to find the second moment, we let $g(i) = i^2$ and repeat the previous steps,

$$\begin{aligned} \mathbb{E}[g(X(t))] = \mathbb{E}[X(t)^2] &= 0 + \int_0^t \mathbb{E}\left[(Af)(X(s))\right] ds \\ &= \int_0^t \mathbb{E}\left[\lambda\big(g(X(s)+1) - g(X(s))\big)\right] ds \\ &= \lambda \int_0^t \mathbb{E}\big[X(s)^2 + 2X(s) + 1 - X(s)^2\big] ds \\ &= \lambda \int_0^t \mathbb{E}\big[2X(s) + 1\big] ds = \lambda \int_0^t (2\lambda s + 1) ds = \lambda^2 t^2 + \lambda t. \end{aligned}$$

Therefore, the variance is

$$\mathsf{Var}(X(t)) = \mathbb{E}\left[X(t)^2\right] - \left(\mathbb{E}[X(t)]\right)^2 = \lambda t,$$

as expected. △

Both the mean and variance of a time-homogeneous Poisson process are well known. However, the above method is quite general and is useful in myriad applications.

Example 4.39. Let $b > 0$ and consider a pure birth process on $S = \{1, 2, \dots\}$ with growth rate $\lambda(i) = bi$ and initial condition $X(0) = 1$. For $f : S \to \mathbb{R}$, we have that

$$(Af)(i) = bi\big(f(i+1) - f(i)\big), \tag{4.17}$$

for all $i \in S$, where A is the generator for the continuous time chain.

For $f(i) = i$, we have that

$$\begin{aligned}\mathbb{E}[f(X(t))] = \mathbb{E}[X(t)] &= 1 + \int_0^t \mathbb{E}\left[(Af)(X(s))\right] ds \\ &= 1 + \int_0^t \mathbb{E}\left[bX(s)\big(f(X(s)+1) - f(X(s))\big)\right] ds \\ &= 1 + b\int_0^t \mathbb{E}[X(s)] ds.\end{aligned}$$

Therefore, defining $g(t) = \mathbb{E}[X(t)]$, we see that

$$g'(t) = bg(t), \quad g(0) = 1.$$

Thus,

$$g(t) = \mathbb{E}[X(t)] = e^{bt}.$$

This result should be compared with the solution to the deterministic linear growth model $x'(t) = bx(t)$, which yields the same solution. △

Let e_i be the row vector with a one in the ith component, and zeros elsewhere. We see from (4.14) that for all $t \geq 0$

$$e_i P(t) f = \mathbb{E}_i[f(X(t))]. \tag{4.18}$$

In words, the ith component of the vector $P(t)f$ gives $\mathbb{E}_i[f(X(t))]$. Next, note that

$$\begin{aligned}(Af)(i) = e_i(Af) = e_i(P'(0)f) &= e_i \lim_{h\to 0} \frac{1}{h}\left(P(h)f - P(0)f\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(e_i P(h) f - e_i f\right) \\ &= \lim_{h\to 0} \frac{\mathbb{E}_i[f(X(h))] - f(i)}{h},\end{aligned} \tag{4.19}$$

where we used (4.18) in the final equality. Taking $f(i) = 1_{\{i=j\}}$ for some $j \in S$, i.e. $f = e_j^T$, we see that (4.19) yields

$$A_{ij} = \lim_{h\to 0} \frac{1}{h}(P(X(h) = j \mid X(0) = i)) = \lambda(i, j),$$

when $i \neq j$, and

$$A_{jj} = \lim_{h\to 0} \frac{1}{h}(P(X(h) = j \mid X(0) = j) - 1) = -\lambda(j),$$

for the diagonal elements. Therefore, (4.19) could be taken as an alternative definition of the generator for a Markov process, though one which views the generator as an operator and not simply as a matrix that stores the transition intensities.

Example 4.40. Consider a store in which customers arrive at rate $\lambda > 0$ and depart at rate $\mu X(t)$, where $X(t)$ is the number of customers at time t. For $i \geq 0$ we have

$$\begin{aligned}(Af)(i) &= \lim_{h\to 0} \frac{\mathbb{E}_i[f(X(h))] - f(i)}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\Big[f(i+1)P_i(X(h) = i+1) + f(i-1)P_i(X(h) = i-1) \\ &\qquad\qquad + f(i)P_i(X(h) = i) - f(i) + o(h)\Big] \\ &= \lim_{h\to 0} \frac{1}{h}\Big[f(i+1)\lambda h + f(i-1)\mu i h + f(i)(1 - \lambda h - \mu i h) - f(i) + o(h)\Big] \\ &= \lambda(f(i+1) - f(i)) + \mu i(f(i-1) - f(i)).\end{aligned}$$

So, for example, taking $f(y) = y$ to be the identity, and $X(0) = x$, we have that

$$\begin{aligned}\mathbb{E}[X(t)] = \mathbb{E}[f(X(t))] &= \mathbb{E}[X(0)] + \mathbb{E}\left[\int_0^t (Af)(X(s))ds\right] \\ &= x + \mathbb{E}\left[\int_0^t \big(\lambda(X(s) + 1 - X(s))\big) + \mu X(s)\big(X(s) - 1 - X(s)\big)ds\right] \\ &= x + \int_0^t (\lambda - \mu\mathbb{E}[X(s)])ds.\end{aligned}$$

Setting $g(t) = \mathbb{E}[X(t)]$, we see that $g(0) = x$ and $g'(t) = \lambda - \mu g(t)$. Solving this initial value problem yields the solution

$$\mathbb{E}[X(t)] = xe^{-\mu t} + \frac{\lambda}{\mu}(1 - e^{-\mu t}).$$

The second moment, and hence the variance, of the process can be calculated in a similar manner. △

4.6 Continuous Time Birth and Death Processes

We revisit the topic of birth and death process, though now in the setting of continuous time. As in Section 2.2, our state space is $S = \{0, 1, 2, \dots\}$. For each $n \in \{0, 1, 2, \dots\}$, the transition rates are

$$\lambda(n, n+1) = b_n \quad \text{and} \quad \lambda(n, n-1) = d_n,$$

where $b_n, d_n \geq 0$, $d_0 = 0$, and $\lambda(n, j) = 0$, when $|j - n| \geq 2$. The generator matrix is

$$A = \begin{pmatrix} -b_0 & b_0 & 0 & 0 & 0 & \cdots \\ d_1 & -(b_1 + d_1) & b_1 & 0 & 0 & \cdots \\ 0 & d_2 & -(b_2 + d_2) & b_2 & 0 & \cdots \\ 0 & 0 & d_3 & -(b_3 + d_3) & b_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \end{pmatrix},$$

whereas the generator as an operator (see Section 4.5) satisfies

$$Af(n) = b_n(f(n+1) - f(n)) + d_n(f(n-1) - f(n)), \tag{4.20}$$

where $f : \{0, 1, 2, \dots\} \to \mathbb{R}$.

We begin with examples, many of which are analogous to those in the discrete time setting.

Example 4.41. The Poisson process is a birth-death process with $b_n \equiv \lambda$, for some $\lambda > 0$, and $d_n \equiv 0$. △

Example 4.42. A pure birth process with $b_n \geq 0$, and $d_n \equiv 0$ is an example of a birth and death process. △

Example 4.43 (Queueing Models)**.** We suppose that arrivals of customers are occurring at a constant rate of $\lambda > 0$. That is, we assume that $b_n \equiv \lambda$. However, departures occur when a customer has been served. There are a number of natural choices for the model of the service times.

(a) (Single server) If there is a single server, and that person always serves the first person in line, then we take $d_n = \mu > 0$, if $n \geq 1$, and $d_0 = 0$.

(b) (k servers) If there are $k \geq 1$ servers, and the first k people in line are always being served, then for some $\mu > 0$ we take

$$d_n = \begin{cases} n\mu, & \text{if } n \leq k \\ k\mu, & \text{if } n \geq k \end{cases}.$$

(c) (∞ servers) If we suppose that there are an infinite number of servers, then $d_n = n\mu$ for some $\mu > 0$.

△

Example 4.44 (Population Models)**.** Suppose that $X(t)$ represents the number of individuals in a certain population at time $t \geq 0$. Assuming the rates of both reproduction and death are proportional to population size we have

$$b_n = \lambda n \quad \text{and} \quad d_n = \mu n,$$

for some $\lambda, \mu > 0$. △

Example 4.45 (Population with immigration)**.** Consider the previous system except $b_n = \lambda n + \nu$ for some $\nu > 0$, representing an inflow due to immigration. Now 0 is no longer an absorbing state. △

Example 4.46 (Fast growing population)**.** Consider a population that grows at a rate proportional to the square of the number of individuals. Assuming no deaths, we have for some $\lambda > 0$ that

$$b_n = \lambda n^2, \quad \text{and} \quad \mu_n = 0.$$

We saw in Example 4.12 that this population grows so fast that it reaches an infinite population in finite time with a probability of one. △

Returning to a general system, consider the embedded discrete time Markov chain of a continuous time birth and death process. The transition probabilities of this chain are

$$p_{n,n+1} = p_n \stackrel{\text{def}}{=} \frac{b_n}{b_n + d_n}$$
$$q_{n,n-1} = q_n \stackrel{\text{def}}{=} \frac{d_n}{b_n + d_n}.$$

The following proposition follows directly from Proposition 2.7.

Proposition 4.47. *A continuous time birth and death process is transient if and only if*

$$\sum_{k=1}^{\infty} \frac{d_1 \cdots d_k}{b_1 \cdots b_k} < \infty.$$

Proof. From Proposition 2.7, the embedded chain, and hence the continuous time chain, is transient if and only if

$$\sum_{k=1}^{\infty} \frac{q_1 \cdots q_n}{p_1 \cdots p_k} < \infty.$$

Noting that

$$\sum_{k=1}^{\infty} \frac{q_1 \cdots q_n}{p_1 \cdots p_k} = \sum_{k=1}^{\infty} \frac{d_1 \cdots d_k}{b_1 \cdots b_k},$$

completes the proof. □

Similarly to the discrete time case, we can now conclude that the single server queue is transient if and only if $\mu < \lambda$, and that the k server queue is transient if and only if $k\mu < \lambda$. For the infinite server queue we have

$$\sum_{k=1}^{\infty} \frac{d_1 \cdots d_k}{p_1 \cdots p_k} = \sum_{k=1}^{\infty} k! \left(\frac{\mu}{\lambda}\right)^k = \infty,$$

for any choice of $\mu, \lambda > 0$. Thus, the infinite server queue is always recurrent.

We turn to the question of positive recurrence and stationary distributions. We know that a stationary distribution η must satisfy $\eta A = 0$, which in component form is

$$\eta_0 b_0 = \eta_1 d_1$$
$$(b_k + d_k)\eta_k = b_{k-1}\eta_{k-1} + d_{k+1}\eta_{k+1}, \quad \text{for} \quad k \geq 1.$$

Noting that these are the same equations as (2.7) and (2.8), we can conclude that such an η exists and can be made into a probability vector if and only if

$$\sum_{k=1}^{\infty} \frac{b_0 b_1 \cdots b_{k-1}}{d_1 \cdots d_k} < \infty.$$

The following is analogous to Proposition 2.9.

Proposition 4.48. *There exists a stationary distribution for a continuous time birth and death chain if and only if*

$$\sum_{k=1}^{\infty} \frac{b_0 b_1 \cdots b_{k-1}}{d_1 \cdots d_k} < \infty.$$

In this case,

$$\eta_0 = \left(\sum_{k=0}^{\infty} \frac{b_0 b_1 \cdots b_{k-1}}{d_1 \cdots d_k} \right)^{-1},$$

where the $k = 0$ term in the above sum is taken to be equal to one, and for $k \geq 1$,

$$\eta_k = \frac{b_0 \cdots b_{k-1}}{d_1 \cdots d_k} \eta_0.$$

For example, for the single server queue we have

$$\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^k = \left(1 - \frac{\lambda}{\mu} \right)^{-1},$$

provided $\lambda < \mu$, and in this case

$$\eta_k = \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^k.$$

The expected length of the queue in equilibrium is

$$\sum_{k=0}^{\infty} k \eta_k = k \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^k = \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu} \right)^{-1} = \frac{\lambda}{\mu - \lambda},$$

which grows to infinity as λ approaches μ.

For the infinite server queue we have

$$\sum_{k=0}^{\infty} \frac{b_0 \cdots b_{k-1}}{d_1 \cdots d_k} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k = e^{\lambda/\mu}.$$

Therefore, a stationary distribution exists, and since we already know the chain is recurrent we may conclude it is positive recurrent. Note that the stationary distribution is Poisson(λ/μ), and

$$\eta_k = e^{-\lambda/\mu} \frac{(\lambda/\mu)^k}{k!}, \quad \text{for } k \geq 0. \tag{4.21}$$

In the next chapter, we will see why many models from chemistry and biology have stationary distributions that are Poisson.

We close by noting that the generator as described in (4.20) can be utilized to compute the moments of a birth and death process in the case that the intensities are linear functions of the state.

Example 4.49. Consider linear birth and death process with transition rates

$$b_n = \lambda n, \quad d_n = \mu n, \quad \text{and} \quad d_0 = 0,$$

where $\lambda, \mu > 0$. For $f : \{0, 1, 2, \dots\} \to \mathbb{R}$, we have

$$(Af)(n) = \lambda n(f(n+1) - f(n)) + \mu n(f(n-1) - f(n)),$$

for $n \geq 0$. Taking $f(n) = n$ and $X(0) = x$ we have

$$\begin{aligned}
\mathbb{E}[X(t)] = \mathbb{E}[f(X(t))] &= \mathbb{E}[f(X(0))] + \mathbb{E}\left[\int_0^t (Af)(X(s))ds\right] \\
&= f(x) + \mathbb{E}\Bigg[\int_0^t \lambda X(s)(f(X(s)+1) - f(X(s))) \\
&\qquad + \mu X(s)(f(X(s)-1) - f(X(s)))ds\Bigg] \\
&= x + \mathbb{E}\left[\int_0^t \lambda X(s)(X(s)+1-X(s)) + \mu X(s)\big(X(s)-1-X(s)\big)ds\right] \\
&= x + (\lambda - \mu)\int_0^t \mathbb{E}[X(s)]ds.
\end{aligned}$$

Solving this integral equation yields

$$\mathbb{E}[X(t)] = xe^{(\lambda-\mu)t}. \tag{4.22}$$

Solving for the second moment, and hence the variance, is done in a similar manner. △

4.7 Exercises

Exercise 4.1. Consider a continuous time Markov chain with state space $\{1, 2, 3, 4\}$ and generator matrix

$$A = \begin{pmatrix} -3 & 2 & 0 & 1 \\ 0 & -2 & \frac{1}{2} & \frac{3}{2} \\ 1 & 1 & -4 & 2 \\ 1 & 0 & 0 & -1 \end{pmatrix}.$$

Write a Matlab code that simulates a path of this chain. To do so, use the construction provided in the notes (i.e. simulate the embedded chain and holding times sequentially). Using this code and assuming that $X(0) = 1$, estimate $\mathbb{E}[X(3)]$ by averaging over 10,000 such paths. Note that you will need to make sure you break your "for" or "while" loop after you see that the time will go beyond $T = 3$, without updating the state for that step.

Exercise 4.2. In Example 4.13, it was stated that if $X(0) = 0$, then the probability of an explosion was $1/3$. Why is that?

Exercise 4.3. Suppose that $\alpha > 1$. Prove that any birth process for which $\lambda(n) = n^\alpha$ is explosive.

Exercise 4.4. Consider a continuous time Markov chain with $S = \{1, 2\}$ and

$$\lambda(1, 2) = 2, \quad \lambda(2, 1) = 3.$$

Find $P(t)$, the matrix whose i, jth entry gives $P(X(t) = j \mid X(0) = i)$.

Exercise 4.5. Let A be the generator matrix for an irreducible continuous time Markov chain with finite state space S. Let $\lambda_0 > 0$ be such that $|A_{ij}| < \lambda_0$ for all $i, j \in S$. Now define

$$P = \frac{1}{\lambda_0} A + I,$$

where I is the identity matrix.

(a) Show that P is a transition matrix for an irreducible, aperiodic discrete time Markov chain on the state space S.

(b) Using part (a), show that A has a unique left eigenvector with eigenvalue 0 that is a probability vector and that all other eigenvalues have negative real part.

Exercise 4.6 (Taken from Durrett, [9])**.** Consider a hemoglobin molecule that can carry either one oxygen molecule or one carbon monoxide molecule. Suppose that the oxygen and carbon dioxide molecules are arriving one at a time according to rates λ_1 and λ_2 respectively. Suppose further that if the hemoglobin molecule is free when one of the molecules arrives it attaches to the hemoglobin molecule for an exponential amount of time with parameter μ_1 (for oxygen) and μ_2 (for carbon dioxide). Formulate a continuous time Markov chain for this system with state space $\{O, E, C\}$, where O represents an attached oxygen molecule, C represents an attached carbon dioxide molecule, and E represents an empty hemoglobin molecule. Next, find the long run fraction of time the hemoglobin molecule is in each of its states.

Exercise 4.7. For Example 4.39, verify that the generator of the process satisfies equation (4.17).

Exercise 4.8. Using Dynkin's formula, calculate $\mathsf{Var}(X(t))$ of the linear birth process of Example 4.39.

Exercise 4.9. Using Dynkin's formula, calculate $\mathsf{Var}(X(t))$ of the linear birth and death process of Example 4.49.

Exercise 4.10. Let $X(t), t \geq 0$, be a continuous time Markov chain with generator matrix

$$A = \begin{bmatrix} -5 & 3 & 2 \\ 2 & -3 & 1 \\ 3 & 4 & -7 \end{bmatrix}.$$

(a) Find Kolmogorov's forward equations for this model.

(b) Find Kolmogorov's backward equations for this model.

(c) Find the stationary distribution for this chain.

(d) Find the stationary distribution for the embedded discrete time Markov chain.

Exercise 4.11. Let $X(t), t \geq 0$, be a continuous time birth and death process with intensities

$$\lambda(n, n+1) = 7(n+1)^2 \quad \text{and} \quad \lambda(n, n-1) = 4n^3,$$

for $n \geq 0$. Determine if the chain is transient, recurrent, or positive recurrent. If the chain is positive recurrent, give the stationary distribution.

Exercise 4.12. Let $X(t), t \geq 0$, be a continuous time birth and death process with intensities

$$\lambda(n, n+1) = 7(n+1)^3 \quad \text{and} \quad \lambda(n, n-1) = 4n^3,$$

for $n \geq 0$. Determine if the chain is transient, recurrent, or positive recurrent. If the chain is positive recurrent, give the stationary distribution.

Exercise 4.13. Let $X(t), t \geq 0$, be a continuous time birth and death process with intensities

$$\lambda(n, n+1) = 3(n+1)^3 \quad \text{and} \quad \lambda(n, n-1) = 4n^3,$$

for $n \geq 0$. Determine if the chain is transient, recurrent, or positive recurrent. If the chain is positive recurrent, give the stationary distribution.

Chapter 5

Continuous Time Markov Chains in biochemistry

5.1 Models of biochemical reaction systems

We introduce the most common stochastic model for biochemical reaction systems. These models are used extensively in cell biology, with applications ranging from gene interaction and protein regulatory networks, to virology, to neural networks. This chapter is adapted from [3].

5.1.1 The basic model

It is useful to understand that a biochemical reaction *system* consists of two parts: (i) a reaction *network*, and (ii) a choice of *dynamics*. The network is a static object that consists of a triple of sets:

(i) *species*, $\mathcal{S}$, which are the chemical components whose counts we wish to model dynamically,

(ii) *complexes*, $\mathcal{C}$, which are non-negative linear combinations of the species that describe how the species can interact, and

(iii) *reactions*, $\mathcal{R}$, which describe how to convert one such complex to another.

For example, if in our system we have only three species, which we denote by A, B, and C, and the only transition type we allow is the merging of an A and a B molecule to form a C molecule, then we may depict this network by the directed graph

$$A + B \to C.$$

For this very simply model our network consists of species $\mathcal{S} = \{A, B, C\}$, complexes $\mathcal{C} = \{A + B,\ C\}$, and reactions $\mathcal{R} = \{A + B \to C\}$.

Example 5.1. Suppose there are two forms of a given protein: "active" and "inactive." Denote by A the active form of the protein and denote by B the inactive form.

We suppose that there are only two types of transitions that can take place in the model: an active protein can become inactive, and an inactive protein can become active. However, we further suppose that an inactive protein B is required to catalyze the inactivation of an active protein A. That is, we suppose that the two possible reactions can be depicted in the following manner

$$A + B \to 2B, \tag{R1}$$
$$B \to A, \tag{R2}$$

where, for example, the reaction (R1) captures the idea that both an A and a B molecule are required for the deactivation of an A molecule and the result of such a reaction is a net gain of one molecule of B and a net loss of one molecule of A.

We again see that there are three sets of objects necessary to give a full description of the above network. First, we need a set of species, which in this case is just $\mathcal{S} = \{A, B\}$. We require a directed graph in which the vertices are linear combinations of the species. These linear combinations are the complexes, which for this model is the set $\mathcal{C} = \{A + B,\ 2B,\ B,\ A\}$. Finally, we associate the edges of the graph with the reactions, $\mathcal{R} = \{A + B \to 2B,\ B \to A\}$. △

Definition 5.2. A chemical reaction network is a triple $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ where

(i) $\mathcal{S} = \{S_1, \ldots, S_N\}$ is the set of species,

(ii) $\mathcal{C}$ is the set of complexes, consisting of linear combinations of the species with non-negative integer coefficients,

(iii) $\mathcal{R} = \{y_k \to y_k' : y_k, y_k' \in \mathcal{C} \text{ and } y_k \neq y_k'\}$ is the set of reactions. We assume there are $M > 0$ reactions.

The notation we use throughout is to write the kth reaction as

$$\sum_{i=1}^{M} y_{ki} S_i \to \sum_{i=1}^{M} y_{ki}' S_i, \tag{5.1}$$

where the vectors $y_k, y_k' \in \mathbb{Z}_{\geq 0}^N$ are associated with the source and product complex, respectively. For example, if the kth reaction is $S_1 + S_2 \to S_3$, we have, $y_{k1} = 1, y_{k2} = 1, y_{k3} = 0$ and $y_{k1}' = 0, y_{k2}' = 0, y_{k3}' = 1$. Note that we abuse notation slightly by writing $y_k \to y_k'$ as opposed to (5.1). We define $\zeta_k := y_k' - y_k \in \mathbb{Z}^N$ to be the *reaction vectors* of the network. For example, the reaction vector for the reaction $S_1 + S_2 \to S_3$ is

$$\zeta_k = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

It is most common to forgo formally giving each of the three sets necessary for a reaction network, as it is easier to simply give the directed graph implied by the reaction network. For example, the network

$$S + E \rightleftarrows C \to S + P, \qquad E \rightleftarrows \emptyset, \tag{5.2}$$

corresponds to the reaction network with $\mathcal{S} = \{S, E, C, P\}, \mathcal{C} = \{S + E,\ C,\ S + P,\ E,\ \emptyset\}$, and $\mathcal{R} = \{S + E \to C,\ C \to S + E,\ C \to S + P,\ E \to \emptyset,\ \emptyset \to E\}$.

Note that the empty set appearing in (5.2) is a valid complex. It is used to model the inflow or outflow (or degradation) of molecules.

Having a notion of a reaction network in hand, we turn to the question of how to model the dynamical behavior of the counts of the different species.

Returning to Example 5.1 for the time being, let $X_1(t)$ and $X_2(t)$ be random variables giving the numbers of molecules of type A and B present in the system at time t, respectively. Denote by $R_1(t)$ and $R_2(t)$ the counting processes determining the number of times reactions (R1) and (R2) have occurred by time t. Clearly, X satisfies

$$X(t) = X(0) + R_1(t)\begin{pmatrix} -1 \\ 1 \end{pmatrix} + R_2(t)\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

From the results of Chapter 3, the counting processes R_1 and R_2 can be specified by specifying their respective intensity functions. For the time being, we delay the conversation regarding what the appropriate form for these intensity functions should be and simply denote them by λ_1 and λ_2. The process X then satisfies the stochastic equation

$$X(t) = X(0) + Y_1\left(\int_0^t \lambda_1(X(s))ds\right)\begin{pmatrix} -1 \\ 1 \end{pmatrix} + Y_2\left(\int_0^t \lambda_2(X(s))ds\right)\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

where Y_1, Y_2 are independent, unit Poisson processes.

Returning to the general reaction network of Definition 5.2, for each reaction $y_k \to y_k' \in \mathcal{R}$ we specify an intensity function $\lambda_k : \mathbb{Z}_{\geq 0}^N \to \mathbb{R}_{\geq 0}$. The number of times that the kth reaction occurs by time t can then be represented by the counting process

$$R_k(t) = Y_k\left(\int_0^t \lambda_k(X(s))ds\right),$$

where the Y_k are independent unit Poisson processes. The state of the system then satisfies the equation $X(t) = X(0) + \sum_k R_k(t)\zeta_k$, or

$$X(t) = X(0) + \sum_{k=1}^{M} Y_k\left(\int_0^t \lambda_k(X(s))ds\right)\zeta_k \tag{5.3}$$

where the sum is over the reaction channels. (Recall that $\zeta_k = y_k' - y_k$.) The representation (5.3) was developed by Thomas Kurtz, see, for example, [12, Chapter 6] or [27].

5.1.2 Intensity functions: mass-action kinetics

We specify the intensity functions, or *kinetics* of the model. The minimal assumption that can be put on the kinetics is that it is *stoichiometrically admissible*, which simply says that $\lambda_k(x) = 0$ if $x_i < y_{ki}$ for any $i \in \{1, \ldots, N\}$. Stoichiometric admissibility ensures that reactions can only occur if there are sufficient molecules to produce the

source complex and guarantees that the process remains within $\mathbb{Z}_{\geq 0}^N$ for all time. The most common choice of intensity function λ_k is that of stochastic *mass-action kinetics*. The stochastic form of the law of mass-action says that for some constant κ_k, termed the *reaction rate constant*, the rate of the kth reaction should be

$$\lambda_k(x) = \kappa_k \prod_{i=1}^{N} y_{ki}! \binom{x_i}{y_{ki}} = \kappa_k \prod_{i=1}^{N} \frac{x_i!}{(x_i - y_{ki})!}. \tag{5.4}$$

Note that the rate is proportional to the number of distinct subsets of the molecules present that can form the inputs for the reaction. This assumption reflects the idea that the system is well-stirred. The reaction rate constants are typically placed next to the arrow in the reaction diagram. The following table gives a representative list of reactions with their respective intensities under the assumption of mass-action kinetics,

Reaction	Intensity Function
$\emptyset \xrightarrow{\kappa_1} S_1$	$\lambda_1(x) = \kappa_1$
$S_1 \xrightarrow{\kappa_2} S_2$	$\lambda_2(x) = \kappa_2 x_1$
$S_1 + S_2 \xrightarrow{\kappa_3} S_3$	$\lambda_3(x) = \kappa_3 x_1 x_2$
$2S_1 \xrightarrow{\kappa_4} S_2$	$\lambda_4(x) = \kappa_4 x_1(x_1 - 1)$

where similar expressions hold for intensity functions of higher order reactions.

5.1.3 Example: Gene transcription and translation

We give a series of stochastic models for gene *transcription* and *translation*. Transcription is the process by which the information encoded in a section of DNA is transferred to a piece of messenger RNA (mRNA). Next, this mRNA is translated by a ribosome, yielding proteins. We will give a series of three examples, with the first, Example 5.3, only including basic transcription, translation, and degradation of both mRNA and proteins. Next, in Example 5.4, we allow for the developed proteins to *dimerize*. Finally, in Example 5.5, we allow the resulting dimer to inhibit the production of the mRNA, and hence the protein and dimers themselves. This inhibition is an example of a *negative feedback loop* in that the protein product (i.e. the dimer) inhibits the rate of its own production. We note that each of our models leaves out many components, such as the RNA polymerase that is necessary for transcription, and the ribosomes that are critical for translation. Instead, we will assume that the abundances of such species are fixed and have been incorporated into the rate constants. More complicated, and realistic, models can of course incorporate both ribosomes and RNA polymerase.

Example 5.3. Consider a model of transcription and translation consisting of three species: $\mathcal{S} = \{G, M, P\}$, representing Gene, mRNA, and Protein, respectively. We

suppose there are four possible transitions in our model:

$$\begin{aligned} &R1) && G \xrightarrow{\kappa_1} G + M && \text{(Transcription)} \\ &R2) && M \xrightarrow{\kappa_2} M + P && \text{(Translation)} \\ &R3) && M \xrightarrow{d_M} \emptyset && \text{(Degradation of mRNA)} \\ &R4) && P \xrightarrow{d_P} \emptyset && \text{(Degradation of protein)} \end{aligned}$$

where, as usual, the reaction rate constants for the different reactions have been placed above the reaction arrows.

We denote by $X(t) = (X_1(t), X_2(t), X_3(t))^T \in \mathbb{Z}^3_{\geq 0}$ the vector giving the numbers of genes, mRNA molecules, and proteins at time t, respectively. The four reaction channels have reaction vectors

$$\zeta_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \zeta_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \zeta_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \quad \zeta_4 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix},$$

and respective intensities $\kappa_1,\ \kappa_2 X_1(t),\ d_M X_1(t),\ d_P X_2(t)$. The stochastic equation governing $X(t)$ is

$$\begin{aligned} X(t) = X(0) &+ Y_1(\kappa_1 t)\zeta_1 + Y_2\left(\kappa_2 \int_0^t X_2(s)ds\right)\zeta_2 + Y_3\left(d_M \int_0^t X_2(s)ds\right)\zeta_3 \\ &+ Y_4\left(d_P \int_0^t X_3(s)ds\right)\zeta_4, \end{aligned} \tag{5.5}$$

where Y_i, $i \in \{1, 2, 3, 4\}$, are independent unit Poisson processes, and we have assumed that $X_1(t) \equiv 1$. Note that the rate of reaction 3 is zero when $X_2(t) = 0$ and the rate of reaction 4 is zero when $X_3(t) = 0$. Therefore, non-negativity of the numbers of molecules is assured. See Figure 5.1 for a single realization of the stochastic model together with the associated deterministic model (see Section 5.4 for a definition of the deterministic model of a chemical reaction system). △

Example 5.4. We continue the previous example but now allow for the possibility that the protein dimerizes via the reaction $2P \xrightarrow{\kappa_3} D$. The degradation of the dimer is allowed by the reaction $D \xrightarrow{d_d} \emptyset$. The set of species for the model is now $\mathcal{S} = \{G, M, P, D\}$ and, keeping all other notation the same as in Example 5.3, we let $X_4(t)$ denote the number of dimers at time t. The stochastic equation for this model

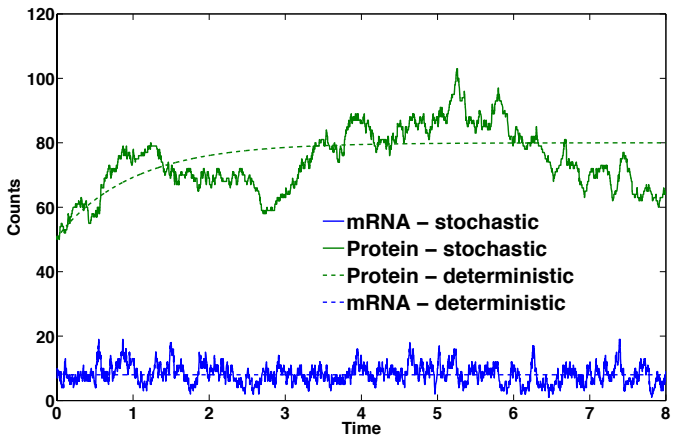


Figure 5.1: A single trajectory of the model in Example 5.3 with choice of rate constants $\kappa_1 = 200$, $\kappa_2 = 10$, $d_M = 25$, and $d_p = 1$. The associated deterministic model with the same choice of rate constants is overlain (dashed).

is

$$\begin{aligned} X(t) =& X(0) + Y_1\left(\kappa_1 t\right)\begin{bmatrix}0\\1\\0\\0\end{bmatrix} + Y_2\left(\kappa_2\int_0^t X_2(s)ds\right)\begin{bmatrix}0\\0\\1\\0\end{bmatrix} \\ &+ Y_3\left(d_M\int_0^t X_2(s)ds\right)\begin{bmatrix}0\\-1\\0\\0\end{bmatrix} + Y_4\left(d_P\int_0^t X_3(s)ds\right)\begin{bmatrix}0\\0\\-1\\0\end{bmatrix} \\ &+ Y_5\left(\kappa_3\int_0^t X_3(s)(X_3(s)-1)ds\right)\begin{bmatrix}0\\0\\-2\\1\end{bmatrix} + Y_6\left(d_d\int_0^t X_4(s)ds\right)\begin{bmatrix}0\\0\\0\\-1\end{bmatrix}, \end{aligned}$$

where $Y_k, k \in \{1, \ldots, 6\}$, are independent unit Poisson processes. See Figure 5.2 for a single realization of the stochastic process together with the associated deterministic model. △

Example 5.5. Continuing the previous examples, we now allow for the dimer to interfere with, or *inhibit*, the production of the mRNA. Specifically, we assume the dimer can bind to the segment of DNA being translated, at which point no mRNA can be produced. Because the resulting dimers inhibit their own production (through the mRNA), this is an example of *negative feedback*. We must now explicitly model the gene to be in one of two states: bound and unbound. We let G remain the notation for the unbound gene, and use B to represent a bound gene. Let $X_5(t)$ give the number of bound genes at time t. Note that $X_1 + X_5$ gives the total number of genes. We continue to assume that $X_1(t) + X_5(t) \equiv 1$. Now the set of species is

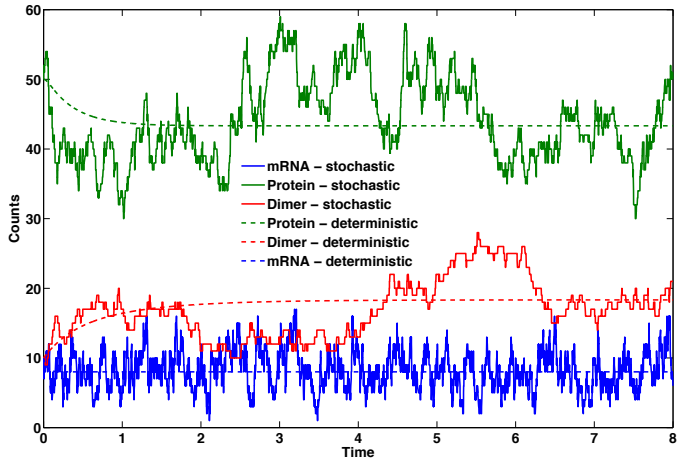


Figure 5.2: A single trajectory of the model in Example 5.4 with choice of rate constants $\kappa_1 = 200$, $\kappa_2 = 10$, $d_M = 25$, $d_p = 1$, $\kappa_3 = 0.01$, and $d_d = 1$. The associated deterministic model with the same choice of rate constants is overlain (dashed).

$\mathcal{S} = \{G, M, P, D, B\}$ and we must add the reactions corresponding to binding and unbinding to our model,

$$G + D \underset{\kappa_{\text{off}}}{\overset{\kappa_{\text{on}}}{\rightleftarrows}} B,$$

where $\kappa_{\text{on}}, \kappa_{\text{off}} > 0$ are the reaction rate constants. The stochastic equations are now

$$\begin{aligned} X(t) =& X(0) + Y_1\left(\kappa_1 \int_0^t X_1(s)ds\right)\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix} + Y_2\left(\kappa_2 \int_0^t X_2(s)ds\right)\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix} \\ &+Y_3\left(d_M \int_0^t X_2(s)ds\right)\begin{bmatrix}0\\-1\\0\\0\\0\end{bmatrix} + Y_4\left(d_P \int_0^t X_3(s)ds\right)\begin{bmatrix}0\\0\\-1\\0\\0\end{bmatrix} \\ &+Y_5\left(\kappa_5 \int_0^t X_3(s)(X_3(s)-1)ds\right)\begin{bmatrix}0\\0\\-2\\1\\0\end{bmatrix} + Y_6\left(d_d \int_0^t X_4(s)ds\right)\begin{bmatrix}0\\0\\0\\-1\\0\end{bmatrix} \\ &+Y_7\left(\kappa_{\text{on}} \int_0^t X_4(s)X_1(s)ds\right)\begin{bmatrix}-1\\0\\0\\-1\\1\end{bmatrix} + Y_8\left(\kappa_{\text{off}} \int_0^t X_5(s)ds\right)\begin{bmatrix}1\\0\\0\\1\\-1\end{bmatrix}. \end{aligned}$$

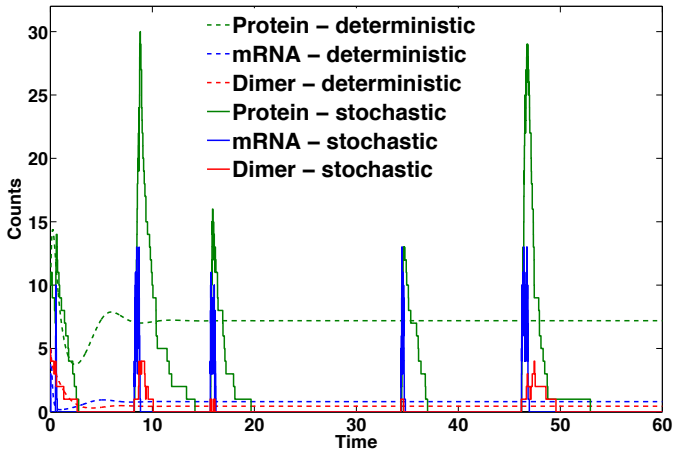


Figure 5.3: A single trajectory of the model in Example 5.5 with choice of rate constants $\kappa_1 = 200$, $\kappa_2 = 10$, $d_M = 25$, $d_p = 1$, $\kappa_3 = 0.01$, $d_d = 1$, $k_{\text{on}} = 2$, $k_{\text{off}} = 0.1$. The associated deterministic model with the same choice of rate constants is overlain (dashed). Note that the negative feedback loop has allowed for strikingly different dynamics between the two models.

Note that the rate of the first reaction has changed to incorporate the fact that mRNA molecules will only be produced when the gene is free. We note that this example can be easily modified to have the dimer only slow the rate of production, or even raise the rate of production. If the rate of production is raised, then this would be an example with *positive feedback*. See Figure 5.3 for a single realization of the stochastic process modeled above (i.e. with the negative feedback) together with the associated deterministic model. Note the strikingly different behavior between the stochastic and deterministic model. △

5.1.4 Example: Virus kinetics

The following model of viral kinetics was first developed in [33] by Srivastava et al., and subsequently studied by Haseltine and Rawlings in [20], and Ball et al., in [5].

Example 5.6 (Viral infection)**.** The model includes four time-varying species: the viral genome (G), the viral structural protein (S), the viral template (T), and the secreted virus itself (V). We denote these as species 1, 2, 3, and 4, respectively, and let $X_i(t)$ denote the number of molecules of species i at time t. The model has six reactions,

$$R_1: \quad T \xrightarrow{1} T + G, \qquad R_2: \quad G \xrightarrow{0.025} T, \qquad R_3: \quad T \xrightarrow{1000} T + S,$$

$$R_4: \quad T \xrightarrow{0.25} \emptyset, \qquad R_5: \quad S \xrightarrow{2} \emptyset, \qquad R_6: \quad G + S \xrightarrow{7.5\times10^{-6}} V,$$

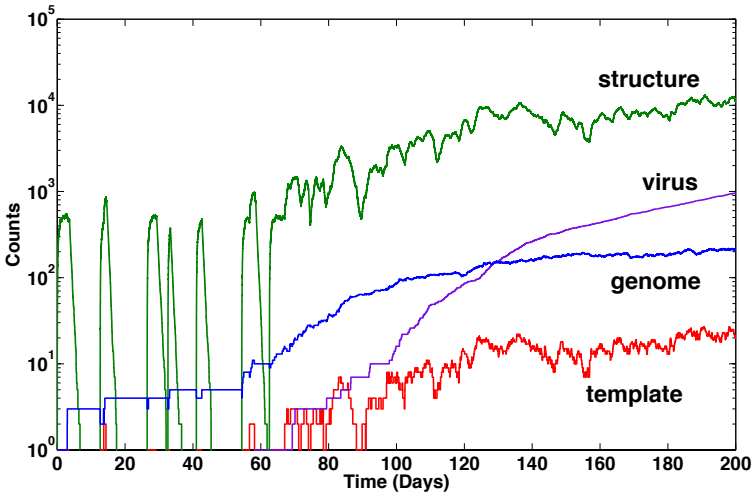


Figure 5.4: A single trajectory of the model in Example 5.6. Note that the y-axis uses a log scale.

where the units of time are in days. The stochastic equations for this model are

$$\begin{aligned}
X_1(t) &= X_1(0) + Y_1\left(\int_0^t X_3(s)ds\right) - Y_2\left(0.025\int_0^t X_1(s)ds\right) \\
&\quad - Y_6\left(7.5\cdot 10^{-6}\int_0^t X_1(s)X_2(s)ds\right) \\
X_2(t) &= X_2(0) + Y_3\left(1000\int_0^t X_3(s)ds\right) - Y_5\left(2\int_0^t X_2(s)ds\right) \\
&\quad - Y_6\left(7.5\cdot 10^{-6}\int_0^t X_1(s)X_2(s)ds\right) \\
X_3(t) &= X_3(0) + Y_2\left(0.025\int_0^t X_1(s)ds\right) - Y_4\left(0.25\int_0^t X_3(s)ds\right) \\
X_4(t) &= X_4(0) + Y_6\left(7.5\cdot 10^{-6}\int_0^t X_1(s)X_2(s)ds\right).
\end{aligned} \tag{5.6}$$

Note that the rate constants of the above model vary over several orders of magnitude, which will in turn cause a large variation in the molecular counts of the different species. See Figure 5.4 for a single realization of this process. △

5.1.5 Example: Enzyme kinetics

We consider a standard model in which an enzyme catalyzes the conversion of some substrate to a product.

Example 5.7. Let S be a substrate, E be an enzyme, SE be an enzyme-substrate

complex, and P be a product and consider the reaction network

$$S + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} SE \overset{\kappa_3}{\to} P + E,$$

which is a slightly simplified version of (5.2). Letting X_1, X_2, X_3, X_4 be the processes giving the counts of species S, E, SE, and P, respectively, the stochastic equations for this model are

$$\begin{aligned}
X_1(t) &= X_1(0) - Y_1\left(\int_0^t \kappa_1 X_1(s)X_2(s)ds\right) + Y_2\left(\int_0^t \kappa_2 X_3(s)ds\right)\\
X_2(t) &= X_2(0) - Y_1\left(\int_0^t \kappa_1 X_1(s)X_2(s)ds\right) + Y_2\left(\int_0^t \kappa_2 X_3(s)ds\right) + Y_3\left(\int_0^t \kappa_3 X_3(s)ds\right)\\
X_3(t) &= X_3(0) + Y_1\left(\int_0^t \kappa_1 X_1(s)X_2(s)ds\right) - Y_2\left(\int_0^t \kappa_2 X_3(s)ds\right) - Y_3\left(\int_0^t \kappa_3 X_3(s)ds\right)\\
X_4(t) &= X_4(0) + Y_3\left(\int_0^t \kappa_3 X_3(s)ds\right).
\end{aligned}$$

△

5.1.6 Generator of the process and the forward equations

We derive the generator, A, for the process (5.3). For $i \in \mathbb{Z}^N_{\geq 0}$ and a bounded function f we have

$$\begin{aligned}
\mathbb{E}_i[f(X(h))] &= \left(\sum_{k=1}^M f(i+\zeta_k)P_i(X(h) = i+\zeta_k)\right) + f(i)P_i(X(h) = i) + o(h)\\
&= \left(\sum_{k=1}^M f(i+\zeta_k)\lambda_k(i)h + o(h)\right) + f(i)\left(1 - \lambda(i)h + o(h)\right) + o(h),
\end{aligned}$$

where the first $o(h)$ term incorporates the probability of having more than one reaction in the time interval of size h. Since $\lambda(i) = \sum_k \lambda_k(i)$ we have

$$\begin{aligned}
\mathbb{E}_i[f(X(h))] &= \left(\sum_{k=1}^M f(i+\zeta_k)\lambda_k(i)h + o(h)\right) + f(i)\left(1 - \sum_{k=1}^M \lambda_k(i)h + o(h)\right) + o(h)\\
&= \sum_{k=1}^M \lambda_k(i)(f(i+\zeta_k) - f(i))h + f(i) + o(h),
\end{aligned}$$

where we have collected the $o(h)$ terms. Therefore, the generator satisfies

$$(Af)(i) = \lim_{h\to 0} \frac{\mathbb{E}_i[f(X(h))] - f(i)}{h} = \sum_{k=1}^M \lambda_k(i)(f(i+\zeta_k) - f(i)). \tag{5.7}$$

The generator can be viewed as giving us the correct chain-rule for our stochastic equations. For example, consider the related ODE system

$$\dot{x}(t) = \sum_{k=1}^M \lambda_k(x)\zeta_k. \tag{5.8}$$

Then, for any smooth $f : \mathbb{R}^N \to \mathbb{R}$, we have

$$\frac{d}{dt} f(x(t)) = \sum_k \lambda_k(x(t))\zeta_k \cdot \nabla f(x(t)),$$

in which case

$$\begin{aligned} f(x(t)) &= f(x(0)) + \int_0^t \left(\frac{d}{ds} f(x(s)) \right) ds \\ &= f(x(0)) + \int_0^t \sum_k \lambda_k(x(s))\zeta_k \cdot \nabla f(x(s)) ds \end{aligned}$$

For the stochastic equation Dynkin's formula (4.16) yields

$$\mathbb{E}[f(X(t))] = \mathbb{E}[f(X(0))] + \mathbb{E}\left[\int_0^t \sum_k \lambda_k(X(s))(f(X(s) + \zeta_k) - f(X(s))) ds \right]$$

Noting that

$$\lim_{h \to 0} \frac{f(x + h\zeta_k) - f(x)}{h} = \nabla f(x) \cdot \zeta_k$$

completes the analogy. Note that letting $x_i(t)$ be the solution to the ode (5.8) with initial condition i, it would be fair to call the operator

$$(Bf)(i) = \lim_{h \to 0} \frac{f(x_i(h)) - f(i)}{h} = \sum_k \lambda_k(i)\zeta_k \cdot \nabla f(i)$$

the generator of the deterministic process.

We turn to the forward equations, which are

$$\begin{aligned} \frac{d}{dt} P_i(X(t) = j) &= \sum_{y \neq j} \lambda(y, j) P_i(X(t) = y) - \lambda(j) P_i(X(t) = j) \\ &= \sum_{k=1}^M \lambda_k(j - \zeta_k) P_i(X(t) = j - \zeta_k) - \sum_{k=1}^M \lambda_k(j) P_i(X(t) = j), \end{aligned}$$

or, more generally,

$$\frac{d}{dt} P_\alpha(X(t) = j) = \sum_k \lambda_k(j - \zeta_k) P_\alpha(X(t) = j - \zeta_k) - \sum_k \lambda_k(x) P_\alpha(X(t) = j), \quad (5.9)$$

where α is the initial distribution. Equation (5.9) is often called the *chemical master equation* in the biological literature and is probably the most well known equation in those settings.

We can use the forward equation (5.9) to find the system of equations that must be satisfied by any stationary distribution. Setting the left hand side of (5.9) to zero, we see a stationary distribution ν satisfies

$$\sum_{k=1}^M \lambda_k(x - \zeta_k)\nu(x - \zeta_k) = \sum_{k=1}^M \lambda_k(x)\nu(x). \quad (5.10)$$

Solving for such a ν is, in general, non-trivial. In fact, it is an open problem to even classify those networks for which a solution exists. Later, we will see a large class of systems for which the equations can be solved.

5.1.7 Population processes

We point out that systems of the form (5.3) are quite general. In particular, they are utilized to model many types of population processes, and not just models from chemistry.

Consider a continuous time Markov chain taking values in $\mathbb{R}^N$ that can undergo transitions in directions $\{\zeta_\ell\}$, where each $\zeta_\ell \in \mathbb{R}^N$. Suppose that the intensity of the chain in direction ζ_ℓ is given by the function λ_ℓ, in which case

$$P(X(t+\Delta t) - X(t) = \zeta_l \,|\, X_s, 0 \le s \le t) = \lambda_l(X(t))\Delta t + o(\Delta t). \tag{5.11}$$

If we write

$$X(t) = X(0) + \sum_l \zeta_l R_l(t)$$

where $R_l(t)$ is the number of jumps of the chain in direction ζ_l at or before time t, then (5.11) implies

$$P(R_l(t+\Delta t) - R_l(t) = 1 \,|\, X_s, 0 \le s \le t) = \lambda_l(X(t))\Delta t + o(\Delta t), \quad \zeta_l \in \mathbb{R}^N.$$

R_l is a *counting process* with intensity $\lambda_l(X(t))$ and so following Chapter 3 we have

$$X(t) = X(0) + \sum \zeta_l Y_l \left(\int_0^t \lambda_l(X(s))ds \right), \tag{5.12}$$

where the Y_l are independent unit Poisson processes. This equation, also a random time change, has a unique solution by the same jump by jump argument used in Section 5.1.1, provided $\sum_l \lambda_l(x) < \infty$ for all x. Of course, as we know from Section 4.2, unless we add additional assumptions, we cannot rule out the possibility that the solution only exists up to some finite time. For example, if $d = 1$ and $\lambda_1(k) = (1+k)^2$, the solution of

$$X(t) = Y_1 \left(\int_0^t (1 + X(s))^2 ds \right)$$

hits infinity in finite time.

We present two examples that are not chemical in nature, but instead are population processes.

Example 5.8. We build a model for the behavior of predator-prey interactions. We denote the predator by F (foxes) and the prey by R (rabbits). We now consider what would make a reasonable model. We first note that because rabbits will reproduce, we have a transition of the general form

$$R \xrightarrow{\kappa_1} 2R.$$

This yields an intensity function of the form $\lambda_1(x) = \kappa_1 x$, and simply assumes the growth rate is proportional to the population size. We also have reproduction of the foxes. However, note that the rate of reproduction should be a function of the number of rabbits. Thus, we have

$$F \overset{\kappa_2 g(R)}{\to} 2F.$$

where $g(R)$ is some function of the number of rabbits. Said differently, we are claiming that the intensity of this transition should be of the form $\kappa_2 X_F g(X_R)$, where X is the state of the system giving the numbers of each animal. It seems plausible that g should be non-decreasing and $g(0) = 0$. For ease, we take $g(R) = R$, though do not try to provide a good argument for why. That is, we are just choosing something. If data is provided for an actual model, more could be said about the function g. Next, it should be that interactions between rabbits and foxes decrease the rabbit population. That is, we have a transition of the form

$$R + F \overset{\kappa_3}{\to} F.$$

Finally, we have death by natural causes

$$R \overset{\kappa_4}{\to} \emptyset, \quad F \overset{\kappa_5}{\to} \emptyset.$$

This example is famous for producing oscillations. For example, when we choose

$$\kappa_1 = 10, \quad \kappa_2 = 0.01, \quad \kappa_3 = 0.01, \quad \kappa_4 = 0.01, \quad \kappa_5 = 10,$$

and an initial condition of $X_R(0) = X_F(0) =$1,000, we get dynamics as exemplified in Figure 5.5. △

Example 5.9. We consider a basic stochastic model for the transmission of a disease in a population. We suppose that there are three types of people in the population:

1. Those that are susceptible to infection, denoted S.
2. Those that are infected, denote I.
3. Those that are recovered, and no longer susceptible, denote R.

Such models are commonly referred to as SIR models.

Letting $X = (X_S, X_I, X_R)^T$, natural transitions, including rates, for the model are

Transition	Rate
$\emptyset \to S$	$\lambda_1(X) = \kappa_1(X_S + X_I + X_R)$
$S + I \to 2I$	$\lambda_2(X) = \kappa_2 X_S X_I$
$I \to R$	$\kappa_3 X_I$
$S \to \emptyset$	$\kappa_4 X_S$
$I \to \emptyset$	$\kappa_5 X_I$
$R \to \emptyset$	$\kappa_6 X_R$

.

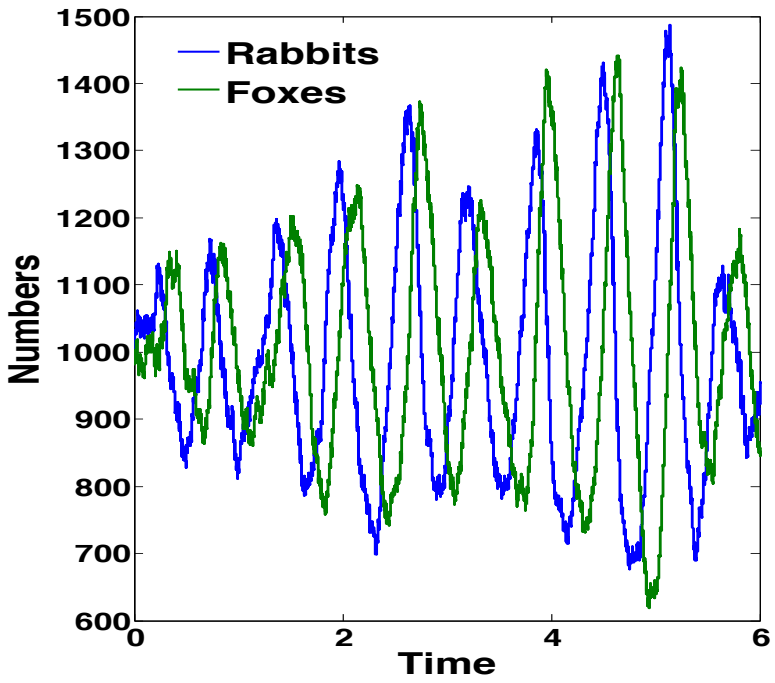


Figure 5.5: The numbers of Rabbits and Foxes as a function of time for the predator-prey model of Example 5.8. The oscillatory behavior of this model is apparent.

Note the the first rate says that the rate of growth of the susceptible population is proportional to the size of the entire population (i.e. only healthy children are born). Such models are extensively studied in population heath where the goal is to find the best treatment so that the disease goes extinct with some high probability. Of course, for each disease the parameters will be different and must be estimated. Natural conditions would be that $\kappa_5 \geq \kappa_4$ and $\kappa_5 \geq \kappa_6$. Depending on the time-scale of the problem, it may be natural to assume that $\kappa_1 = \kappa_4 = \kappa_5 = \kappa_6 = 0$, leaving only the system

$$S + I \overset{\kappa_2}{\to} 2I, \quad I \overset{\kappa_3}{\to} R.$$

For example, if the infection being modeled is a cold then the time-frame may only be a few weeks. Obviously, this is not the case for a longer term infection such as HIV.

There are many variants to this model, including sophisticated models that take space (i.e. countries, rates of air travel, etc.) into account. △

5.2 Simulation

Three algorithms for the numerical simulation of sample paths of stochastically modeled chemical reaction systems are provided below. In Section 5.2.1, the "Gillespie algorithm" or stochastic simulation algorithm is presented. This is simply Algorithm 2 from Section 4.1 for simulating the embedded discrete time Markov chain. It is

redundant to give the algorithm here, but we present it for completeness. Next, in Section 5.2.2, the algorithm for simulating the random time change representation (5.3) is given. This method often goes by the name "the next reaction method." Finally, in Section 5.2.3 a natural approximate algorithm is presented. The method is known as τ-leaping, and is simply Euler's method applied to the random time change representation (5.3). Throughout this section all random variables generated are assumed to be independent of each other and all previous random variables.

5.2.1 The stochastic simulation algorithm

The stochastic simulation algorithm, or *Gillespie's algorithm*, is simply Algorithm 2 of Section 4.1 applied in the chemical kinetic context. See [16, 17] for historical references.

Algorithm 3. (Gillespie Algorithm)
Initialize: Set the initial number of molecules of each species and set $t = 0$.

1. Calculate the intensity function, λ_k, for each reaction.
2. Set $\lambda_0 = \sum_{k=1}^{M} \lambda_k$.
3. Generate two independent uniform(0,1) random numbers r_1 and r_2.
4. Set $\Delta = \ln(1/r_1)/\lambda_0$ (equivalent to drawing an exponential random variable with parameter λ_0).
5. Find $\mu \in [1, \ldots, M]$ such that

$$\frac{1}{\lambda_0}\sum_{k=1}^{\mu-1}\lambda_k < r_2 \leq \frac{1}{\lambda_0}\sum_{k=1}^{\mu}\lambda_k,$$

   which is equivalent to choosing from reactions $1, \ldots, M$, with the kth reaction having probability λ_k/λ_0.
6. Set $X(t + \Delta) = X(t) + \zeta_\mu$.
7. Set $t = t + \Delta$.
8. Return to step 1 or quit.

Note that Algorithm 3 uses two random numbers per step. The first is used to find *when* the next reaction occurs and the second is used to determine *which* reaction occurs at that time. That is, the second random variable (in step 5) simulates the next step of the embedded chain.

5.2.2 The next reaction method

The algorithm below simulates the random time change representation (5.3). The method is usually termed the *next reaction method*. See [15, 2]. The formal algorithm we present follows that of [2].

In Algorithm 4 below, the variable T_k will represent the value of the integrated intensity function, $\int_0^t \lambda_k(X(s))ds$, where t is the current time. Further, the variable P_k will represent the first jump time of the process Y_k after time T_k. That is,

$$P_k = \inf\{s > T_k \ : \ Y_k(s) > Y_k(T_k)\}.$$

Note that if T_k happens to be equal to a jump time of Y_k, then $P_k - T_k$ is an exponential random variable with a parameter of one. Continuing, for each $k \in \{1, \ldots, M\}$ we will set Δt_k to be the solution to the equation

$$\int_t^{t+\Delta t_k} \lambda(X(s))ds = P_k - T_k.$$

Under the assumption that no other reaction takes place before $t + \Delta t_k$, we see

$$\int_t^{t+\Delta t_k} \lambda(X(s))ds = \Delta t_k \lambda_k(X(t)) \implies \Delta t_k = (P_k - T_k)/\lambda_k(X(t)).$$

Finding the minimum of these Δt_k then yields both the reaction (given by the index of the minimum) and the time until the next reaction takes place.

Algorithm 4. (Next Reaction Method)
Initialize: Set the initial number of molecules of each species. Set $t = 0$. For each $k \in \{1, \ldots, M\}$, set $P_k = \ln(1/r_k)$, where r_k are independent uniform(0,1) random variables, and set $T_k = 0$.

1. Calculate the intensity function, λ_k, for each reaction.
2. For each k, set
$$\Delta t_k = \begin{cases} (P_k - T_k)/\lambda_k, & \text{if } \lambda_k \neq 0 \\ \infty, & \text{if } \lambda_k = 0 \end{cases}.$$
3. Set $\Delta = \min_k\{\Delta t_k\}$, and let μ be the index where the minimum is achieved.
4. Set $X(t + \Delta) = X(t) + \zeta_\mu$.
5. For each $k \in \{1, \ldots, M\}$, set $T_k = T_k + \lambda_k \cdot \Delta$.
6. Set $P_\mu = P_\mu + \ln(1/r)$, where r is uniform(0,1).
7. Set $t = t + \Delta$.
8. Return to step 1 or quit.

Note that after initialization the Next Reaction Method only demands one random number to be generated per step.

Time dependent intensity functions

Due to changes in temperature and/or volume, the rate constants of a chemical system may change in time. Therefore, the intensity functions will no longer be constant between reactions. That is, we may have $\lambda_k(t) = \lambda_k(X(t), t)$, and the random time change representation is

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s), s) ds \right) \zeta_k, \tag{5.13}$$

where the Y_k are independent, unit rate Poisson processes. The next reaction method as presented in Algorithm 4 is easily modified to incorporate this time dependence. The only step that would change is step 2., which becomes:

2. For each k, find Δt_k satisfying

$$\int_t^{t+\Delta t_k} \lambda_k(X(s), s) ds = P_k - T_k.$$

Note, in particular, that the integral ranges from t to $t + \Delta t_k$. The remainder of the algorithm stays the same.

5.2.3 Euler's method

We briefly review Euler's method, termed tau-leaping in the chemical kinetic literature [18], as applied to the models (5.3). The basic idea of tau-leaping is to hold the intensity functions fixed over a time interval $[t_n, t_n + h]$ at the values $\lambda_k(X(t_n))$, where $X(t_n)$ is the current state of the system, and, under this assumption, compute the number of times each reaction takes place over this period. Analogously to (5.3), a path-wise representation of Euler tau-leaping defined for all $t \geq 0$ can be given through a random time change of Poisson processes:

$$Z_h(t) = Z_h(0) + \sum_k Y_k \left(\int_0^t \lambda_k(Z_h \circ \eta(s)) ds \right) \zeta_k, \tag{5.14}$$

where the Y_k are as before, and $\eta(s) \stackrel{\text{def}}{=} \left\lfloor \frac{s}{h} \right\rfloor h$. Thus, $Z_h \circ \eta(s) = Z_h(t_n)$ if $t_n \leq s < t_{n+1}$. Noting that

$$\int_0^{t_{n+1}} \lambda_k(Z_h \circ \eta(s)) ds = \sum_{i=0}^{n} \lambda_k(Z_h(t_i))(t_{i+1} - t_i)$$

explains why this method is called Euler tau-leaping.

The following algorithm simulates (5.14) up to a time of $T > 0$. Below and in the sequel, for $x \geq 0$ we will write Poisson(x) to denote a sample from the Poisson distribution with parameter x, with all such samples being independent of each other and of all other sources of randomness used.

Algorithm 5. (Euler tau-leaping)
Initialize: Fix $h > 0$. Set $Z_h(0) = x_0$, $t_0 = 0$, $n = 0$ and repeat the following until $t_n = T$:

1. Set $t_{n+1} = t_n + h$. If $t_{n+1} \geq T$, set $t_{n+1} = T$ and $h = T - t_n$.
2. For each k, let $\Lambda_k = \text{Poisson}(\lambda_k(Z_h(t_n))h)$.
3. Set $Z_h(t_{n+1}) = Z_h(t_n) + \sum_k \Lambda_k \zeta_k$.
4. Set $n \leftarrow n + 1$.

Several improvements and modifications have been made to the basic algorithm described above over the years. Some concern adaptive step-size selection along a path [7, 19]. Others focus on ensuring non-negative population values [2, 6, 8, 34]. We also note that it is straightforward to incorporate time dependence of the intensity functions λ_k in the above algorithm; simply change step 2 to read

$$\Lambda_k = \text{Poisson}\left(\int_t^{t+h} \lambda_k(Z(t_n), s)ds\right),$$

or

$$\Lambda_k = \text{Poisson}\left(\lambda_k(Z(t_n), t_n)h\right),$$

with the choice depending upon the specific problem.

Historically, the time discretization parameter for Euler's method has been τ, leading to the name "τ-leaping methods." We break from this tradition so as not to confuse τ with a stopping time, and we denote our time-step by h to be consistent with much of the numerical analysis literature.

5.3 First order reaction networks

We briefly discuss *first order reaction networks*. A system is said to be a first order reaction network if each intensity function λ_k is linear. In the chemical context with mass-action kinetics, the system is linear if and only if each reaction is either unary, $S_i \to *$, where "$*$" could refer to any linear combination of the species, or of the form $\emptyset \to *$. Note that in this case we have the identity

$$\mathbb{E}[\lambda_k(X(s))] = \lambda_k(\mathbb{E}[X(s)]),$$

for each k. Therefore, Dynkin's formula gives

$$\begin{aligned}\mathbb{E}[X(t)] &= \mathbb{E}[X(0)] + \sum_k \zeta_k \int_0^t \mathbb{E}\left[\lambda_k(X(s))\right] ds \\ &= \mathbb{E}[X(0)] + \sum_k \zeta_k \int_0^t \lambda_k(\mathbb{E}[X(s)])ds.\end{aligned}$$

This gives a very easy method for solving for the means: just solve the associated ordinary differential equation implied by the above integral equation.

Example 5.10. We consider a model of gene transcription and translation

$$G \xrightarrow{2} G + M, \quad M \xrightarrow{10} M + P, \quad M \xrightarrow{k} \emptyset, \quad P \xrightarrow{1} \emptyset, \tag{5.15}$$

where a single gene is being translated into mRNA, which is then being transcribed into proteins. The final two reactions represent degradation of the mRNA and protein molecules, respectively. Assuming that there is a single gene copy, the stochastic equation for this model is

$$\begin{aligned} X^k(t) = X^k(0) &+ Y_1(2t) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + Y_2 \left(\int_0^t 10 X_1^k(s) ds \right) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &+ Y_3 \left(\int_0^t k X_1^k(s) ds \right) \begin{pmatrix} -1 \\ 0 \end{pmatrix} + Y_4 \left(\int_0^t X_2^k(s) ds \right) \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \end{aligned} \tag{5.16}$$

where $X_1^k(t)$ and $X_2^k(t)$ give the number of mRNA and protein molecules at time t, respectively, and Y_i, $i \in \{1, 2, 3, 4\}$, are independent unit-rate Poisson processes. The differential equations governing the mean values are

$$\begin{aligned} \dot{x}_1(t) &= 2 - k x_1(t) \\ \dot{x}_2(t) &= 10 x_1(t) - x_2(t). \end{aligned}$$

Solving this system yields

$$\begin{aligned} x_1(t) &= x_1(0) e^{-kt} + \frac{2}{k} \left(1 - e^{-kt} \right) \\ x_2(t) &= e^{-t} \left(x_2(0) - 10 \frac{k x_1(0) - 2k}{k\,(1-k)} \right) + 10\, \frac{x_1(0) e^{-kt} k + 2 - 2\,k - 2\,e^{-k\,t}}{k\,(1-k)}. \end{aligned}$$

△

For the calculation of second moments of first-order systems, see [14].

We note that if λ_k is non-linear, then $\mathbb{E}[\lambda_k(X(s))] \neq \lambda_k(\mathbb{E}[X(s)])$, and the mean value of the stochastic process *does not* satisfy the ordinary differential equation

$$\dot{x}(t) = \sum_k \lambda_k(x(t)) \zeta_k. \tag{5.17}$$

In the next section we will show, however, when a scaled version of the stochastic process can be shown to be well approximated by the solution to the differential equation (5.17).

Stationary distributions can also be calculated for a class of first order reaction networks.

Definition 5.11. A chemical reaction network, $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, is called *weakly reversible* if for any reaction $y_k \to y_k' \in \mathcal{R}$, there is a sequence of directed reactions beginning with y_k' as a source complex and ending with y_k as a product complex. That is, there exist complexes $y_1, \ldots, y_r$ such that $y_k' \to y_1, y_1 \to y_2, \ldots, y_r \to y_k \in \mathcal{R}$. A network is called *reversible* if $y_k' \to y_k \in \mathcal{R}$ whenever $y_k \to y_k' \in \mathcal{R}$.

Note, for example, that the reaction network

$$S + E \rightleftarrows SE \rightleftarrows P + E, \quad E \rightleftarrows \emptyset,$$

is reversible, and hence weakly reversible, whereas the network

$$S + E \rightleftarrows SE \to P + E, \quad E \rightleftarrows \emptyset,$$

is neither reversible nor weakly reversible. The network

$$\begin{array}{ccc} A & \longrightarrow & B \\ & \nwarrow \quad \swarrow & \\ & C & \end{array} \tag{5.18}$$

is weakly reversible, but not reversible.

A proof of the following theorem can be found in [3].

Theorem 5.12. *Let $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a chemical reaction network and let $\{\kappa_k\}$ be a choice of rate constants. Suppose that the network is weakly reversible and first order. Then, there is an equilibrium, $c \in \mathbb{R}^N_{>0}$, for the associated deterministic model with mass-action kinetics and rate constants $\{\kappa_k\}$. Further, the stochastic model with mass-action kinetics with parameters $\{\kappa_k\}$ has a stationary distribution consisting of the product of Poisson distributions,*

$$\pi(x) = \prod_{i=1}^{N} \frac{c_i^{x_i}}{x_i!} e^{-c_i}, \qquad x \in \mathbb{Z}^N_{\geq 0}. \tag{5.19}$$

If $\mathbb{Z}^N_{\geq 0}$ is irreducible, then (5.19) is the unique stationary distribution, whereas if $\mathbb{Z}^N_{\geq 0}$ is not irreducible then the unique stationary distribution on any irreducible component, $\Gamma \subset \mathbb{Z}^N_{\geq 0}$, is given by the product-form stationary distribution

$$\pi_\Gamma(x) = M_\Gamma \prod_{i=1}^{N} \frac{c_i^{x_i}}{x_i!}, \qquad x \in \Gamma,$$

and $\pi_\Gamma(x) = 0$ otherwise, where M_Γ is a positive normalizing constant.

Example 5.13. Consider the reaction network

$$\emptyset \underset{\mu}{\overset{\lambda}{\rightleftarrows}} S,$$

where $\lambda, \mu > 0$. This network is weakly reversible, and is first-order. The equilibrium of the corresponding deterministically modeled system is $c = \lambda/\mu$ and the state space is all of $\mathbb{Z}_{\geq 0}$. Hence, the stationary distribution of the stochastic model is Poisson with parameter $\lambda/\mu > 0$. Note that this model is identical to the simplest model of an infinite server queue, whose stationary distribution was solved in (4.21). $\triangle$

Example 5.14. Consider the network

$$S_1 \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} S_2,$$

where $\kappa_1, \kappa_2 > 0$. Suppose that $X_1(0) + X_2(0) = n$ so that $X_1(t) + X_2(t) = n$ for all $t \geq 0$. This system is weakly reversible and is first-order. An equilibrium for the deterministically modeled system is

$$c = \left(\frac{\kappa_2}{\kappa_1 + \kappa_2}, \frac{\kappa_1}{\kappa_1 + \kappa_2} \right),$$

and the product-form stationary distribution for the stochastic system is therefore

$$\pi(x) = M \frac{c_1^{x_1}}{x_1!} \frac{c_2^{x_2}}{x_2!},$$

where $M > 0$ is a normalizing constant. Using that $X_1(t) + X_2(t) = n$ for all t yields

$$\pi_1(x_1) = M \frac{c_1^{x_1}}{x_1!} \frac{c_2^{n-x_1}}{(n-x_1)!} = \frac{M}{x_1!(n-x_1)!} c_1^{x_1} (1-c_1)^{n-x_1},$$

for $0 \leq x_1 \leq n$. After setting $M = n!$, we see that X_1 is binomially distributed. Similarly,

$$\pi_2(x_2) = \binom{n}{x_2} c_2^{x_2} (1-c_2)^{n-x_2},$$

for $0 \leq x_2 \leq n$. Note that due to the conservation relation, X_1 and X_2 are *not* independent under this stationary distribution. △

5.4 Relationship with Deterministic Model: the Law of Large Numbers

We will explore the relationship between the stochastic and deterministic models for biochemical systems. Recall that the associated ordinary differential equation for the stochastic process (5.3) is

$$\dot{x}(t) = \sum_k \hat{\lambda}_k(x(t)) \zeta_k,$$

where $\hat{\lambda}_k$ is deterministic mass-action kinetics. That is,

$$\hat{\lambda}_k(x) = \hat{\kappa}_k \prod_i x_i^{y_{ki}},$$

where $\hat{\kappa}_k$ is the rate constant, and we recall that y_{ki} is the number of molecules required of species S_i for the kth reaction.

5.4.1 Conversion of rate constants

To understand the relationship between the stochastic and deterministic models, we must first note that they differ in how they are representing the abundance of each species. In the stochastic model, the abundance is an integer representing the number of molecules present. However, in the usual deterministic model it is the concentration of the species that are being modeled, for example in moles per liter. Therefore, we begin trying to understand the relationship between the two models by explicitly taking the volume into account in the stochastic system, and will do so by introducing a scaling parameter V, which is defined to be the volume multiplied by Avogadro's number.

Consider the stochastic rate constants, κ_k, used in mass-action kinetics. Explicitly thinking of V as proportional to the volume of the system, we see that for a binary reaction the rate constant κ_k should satisfy

$$\kappa_k = \frac{1}{V}\hat{\kappa}_k,$$

for some constant $\hat{\kappa}_k$. This follows since κ_k is assumed to be proportional to the probability that a *particular* pair of molecules interact within a small time window, and this probability should intuitively scale inversely with the volume. For unary reactions the volume should not affect the associated probability of a reaction, and so $\kappa_k = \hat{\kappa}_k$. In general, we have

$$\kappa_k = V^{-(\sum_i y_{ki}-1)}\hat{\kappa}_k. \tag{5.20}$$

See, for example, [23].

We will show the relation (5.20) in a second way that follows [36]. Note that if x gives the concentration of a particular species in moles per unit volume, then Vx gives the total number of molecules present. We first consider the zeroth order reaction

$$\emptyset \to S.$$

The rate of change induced by this reaction for the deterministic model is $\hat{\kappa}\ Ms^{-1}$. Thus, if X represents the number of molecules of species S, this reaction increases X at a rate of

$$\hat{\kappa}V$$

molecules per second. Because the stochastic rate law is κ molecules per second, we have that

$$\kappa = V\hat{\kappa}.$$

Now consider the first order reaction

$$S \to *,$$

where "$*$" can represent any linear combination of the species. The deterministic rate is $\hat{\kappa}x\ Ms^{-1}$. Multiplying the rate $\hat{\kappa}x$ by V, we see that X, the number of molecules of S, is changing at a rate of

$$\hat{\kappa}xV,$$

molecules per second. However, since $X = xV$, we have that X is changing at a rate $\hat{\kappa}X$. Since the stochastic rate law for this reaction is κX molecules per second, we see $\kappa = \hat{\kappa}$.

Finally, consider the binary reaction

$$S_1 + S_2 \to *.$$

The deterministic rate is $cx_1x_2\ Ms^{-1}$, implying the rate of change is $\hat{\kappa}x_1x_2V$ molecules per second. Since $X_1 = x_1V$ and $X_2 = x_2V$ give the number of molecules of S_1 and S_2, respectively, we see that the rate of change of the X_i is

$$\hat{\kappa}x_1x_2V = \hat{\kappa}\frac{1}{V}X_1X_2,$$

implying $\kappa = V^{-1}\hat{\kappa}$.

The above arguments again confirm the relation (5.20) for the most common types of reactions. Similar arguments work for all higher order reactions.

5.4.2 The classical scaling

The following scaling argument is commonly referred to as the classical scaling. It shows how to understand the relationship between the stochastic and deterministic models. See [23] or [26] for technical details and full statements of the requisite theorems.

Again let V be the volume of the system multiplied by Avogadro's numbers. If X is the solution to the stochastic system, which counts the numbers of molecules, then

$$X^V(t) \stackrel{\text{def}}{=} X(t)/V,$$

gives the concentration of the different species in moles per unit volume. Therefore, the stochastic equation governing the *concentrations* is

$$X^V(t) = X^V(0) + \sum_k \frac{1}{V}Y_k\left(\int_0^t \lambda_k(VX^V(s))ds\right)\zeta_k,$$

where we divided each instance of X by V.

Letting λ_k be stochastic mass-action kinetics, and using the relation (5.20), we have

$$\begin{aligned}\lambda_k(X) &= V^{-(\sum_i y_{ki}-1)}\hat{\kappa}_k\left(\prod_i y_{ki}!\right)\prod_i \binom{X_i}{y_{ki}} \\ &= \hat{\kappa}_k V\prod_i V^{-y_{ki}}X_i\cdots(X_i - y_{ki}+1) \\ &\approx V\hat{\kappa}_k\prod_i (X_i^V)^{y_{ki}},\end{aligned}$$

where the approximation is valid for large V. For vectors u, v, define

$$u^v \stackrel{\text{def}}{=} \prod_i u_i^{v_i},$$

where we take $0^0 = 1$. Now, the stochastic equation governing X^V is

$$\begin{aligned} X^V(t) &= X^V(0) + \sum_k \frac{1}{V} Y_k \left(\int_0^t \lambda_k(VX^V(s))ds \right) \zeta_k \\ &\approx X^V(0) + \sum_k \frac{1}{V} Y_k \left(V \int_0^t \hat{\kappa}_k X^V(s)^{y_k} ds \right) \zeta_k. \end{aligned} \tag{5.21}$$

Recalling that

$$\lim_{V\to\infty} \frac{1}{V} Y(Vu) = u,$$

for any unit-rate Poisson process Y, we see that in the limit as $V \to \infty$, X^V satisfies the integral equation

$$x(t) = x(0) + \sum_k \zeta_k \int_0^t \hat{\kappa}_k x(s)^{y_k} ds, \tag{5.22}$$

which in differential form is

$$\dot{x}(t) = \sum_k \hat{\lambda}_k(x(t))\zeta_k,$$

where $\hat{\lambda}_k$ is deterministic mass-action kinetics with rate parameter $\hat{\kappa}_k$. That is, for large V, we have that $X^V \approx x$, and X^V is well approximated by the solution to the deterministic model.

We were quite loose with the above scaling. The correct technical result, which can be found in [23] states that if the rate constants for the stochastic model satisfy the above scaling assumptions, and if $X^V(0) = O(1)$, then for any $\epsilon > 0$ and any $t > 0$

$$\lim_{V\to\infty} P\left\{ \sup_{s\le t} |X^V(s) - x(s)| > \epsilon \right\} = 0.$$

See also [12, 24, 26]. For practical purposes, this result say that if $X^V(0) = O(1)$ for some large V, and if the systems satisfies mass-action kinetics, then it is plausible to use the deterministic model, with deterministic mass-action kinetics, as an approximation for the stochastic model that governs the *concentrations* of the molecular abundances. This result has been used for decades to justify the use of ordinary differential equation models in chemistry and, more generally, population processes (though such models were used well before such a rigorous justification was available).

Example 5.15. Consider the dimerization of a certain protein

$$2P \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} D.$$

Note that we have the conservation relation $2X_D + X_P = M$, for some $M > 0$, where X_D and X_P give the numbers of dimers and proteins, respectively. The stochastic model for this system is

$$X_P(t) = X_P(0) - 2Y_1\left(\int_0^t \kappa_1 X_P(s)(X_P(s)-1)ds\right) + 2Y_2\left(\int_0^t \kappa_2 \frac{1}{2}(M - X_P(s))ds\right),$$

where we made use of the conservation relation. The corresponding deterministic model is

$$x_p(t) = x_p(0) - 2\int_0^t \hat{\kappa}_1 x_p(s)^2 ds + 2\int_0^t \hat{\kappa}_2 \frac{1}{2}(\hat{M} - x_p(s))ds, \tag{5.23}$$

where $\hat{\kappa}_2$ and $\hat{M}$ are the normalized rate constant and conservation relation, respectively. Taking expectations of the stochastic version yields

$$\begin{aligned}\mathbb{E}[X_P(t)] &= \mathbb{E}[X_P(0)] - 2\int_0^t \kappa_1 \mathbb{E}[X_P(s)(X_P(s)-1)]ds + 2\int_0^t \kappa_2 \frac{1}{2}(M - \mathbb{E}[X_P(s)])ds \\ &= \mathbb{E}[X_P(0)] - 2\int_0^t \kappa_1 (\mathbb{E}[X_P(s)])^2 ds + 2\int_0^t \kappa_2 \frac{1}{2}(M - \mathbb{E}[X_P(s)])ds \quad (5.24) \\ &\quad - 2\int_0^t \kappa_1 (\mathsf{Var}(X_P(s)) - \mathbb{E}[X_P(s)])ds. \quad (5.25)\end{aligned}$$

Note that the portion of the above integral equation given by (5.24) is the same form as (5.23). Further, the part given by (5.25) is, in general, non-zero.[1] Therefore, we see that the equation for the mean of the stochastic process is *not* the same as the equation for the associated deterministic model. In general, only linear systems have the property that the mean of the stochastic process satisfies the equations of the associated deterministic model. △

5.5 Exercises

Exercise 5.1. Consider the reaction network

$$\emptyset \xrightarrow{\kappa_1} S_1 \xrightarrow{\kappa_2} S_2 \xrightarrow{\kappa_3} \emptyset.$$

Supposing that $(X_1(0), X_2(0)) = (x_1, x_2) \in \mathbb{Z}^2_{\geq 0}$ are fixed, use Dynkin's formula to compute the expectations $E[X_1(t)]$, $E[X_2(t)]$.

Exercise 5.2. Consider the reaction network for a pure birth process

$$S \xrightarrow{\lambda} 2S,$$

where $\lambda > 0$. Supposing that $X(0) = x \geq 1$ is fixed, compute $E[X(t)]$ and $\mathsf{Var}(X(t))$.

[1] To see this, take $X_P(0) = 10$ with a probability of one.

Exercise 5.3. Consider the reaction network for a birth and death process

$$\emptyset \xleftarrow{\mu} S \xrightarrow{\lambda} 2S,$$

where $\lambda, \mu > 0$. Compute $E[X(t)]$ and $\mathsf{Var}(X(t))$.

Exercise 5.4. Consider the reaction network

$$\emptyset \underset{\mu}{\overset{\lambda}{\rightleftarrows}} S,$$

where $\lambda, \mu > 0$. Derive the equations for the time evolution of the mean of the process by noting that $E[X(t)] = \sum_{k=0}^{\infty} kP(X(t) = k)$, differentiating, and using the Kolmogorov forward equation (5.9).

Exercise 5.5. Simulate the model of foxes and rabbits in Example 5.8 in order to output a plot similar to that in Figure 5.5. Do this two ways, using Gillespie's algorithm and the next reaction method.

Exercise 5.6. Consider the model found in Example 5.4 with $\kappa_1 = 200, \kappa_2 = 10, d_M = 25, d_p = 1, \kappa_3 = 0.01$, and $d_d = 1$, and $X(0) = (1, 10, 50, 10)^T$. For this problem you will estimate $E[X_4(5)]$ and $\mathsf{Var}(X_4(5))$ in two different ways.

1. Write a script that implements Algorithm 3. Average over 10,000 sample paths to estimate both $E[X_4(5)]$ and $\mathsf{Var}(X_4(5))$.

2. Write a script that implements Algorithm 4. Average over 10,000 sample paths to estimate both $E[X_4(5)]$ and $\mathsf{Var}(X_4(5))$.

Exercise 5.7. Let $\{\kappa_{A\to B}, \kappa_{B\to C}, \kappa_{C\to A}\}$ be the rate constants for the reaction network (5.18). Find the stationary distribution of this model under the assumption that $x_A(0) + x_B(0) + x_C(0) = n$ for some $n \geq 1$.

Chapter 6

Brownian motion and stochastic integration

6.1 Brownian Motions

We wish to construct a process, W, satisfying the following four properties.

1. $W(0) = 0$;
2. For any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$, the random variables $W(t_1) - W(s_1), \ldots, W(t_n) - W(s_n)$ are independent;
3. For any $s < t$, the random variable $W(t) - W(s)$ is normal with mean zero and variance $\lambda(t - s)$.
4. The function $t \to W(t)$ is a continuous function of t.

Any process satisfying the four conditions above is termed a *Brownian motion* or *Wiener process*, with variance parameter $\lambda > 0$. A *standard* Brownian motion is one in which $\lambda = 1$.

There are a number of ways to construct such a process, with the most common method using symmetric random walks. We choose a different method. Consider a homogeneous Poisson process, Y_λ, with rate $\lambda > 0$. Viewed as a counting process, the holding time is exponential with mean $\mu = 1/\lambda$ and variance $\sigma^2 = 1/\lambda^2$. Therefore, from Theorem 3.4

$$\frac{Y_\lambda(Vt) - \lambda Vt}{\lambda^{-1}\lambda^{3/2}\sqrt{Vt}} = \frac{Y_\lambda(Vt) - \lambda Vt}{\lambda^{1/2}\sqrt{Vt}} \approx N(0,1),$$

implying

$$V^{-1/2}\left[Y_\lambda(Vt) - V\lambda t\right] \approx N(0, \lambda t).$$

Letting Y be a unit rate Poisson process, the above is equivalent to

$$V^{-1/2}\left[Y(V\lambda t) - \lambda Vt\right] \approx N(0, \lambda t).$$

Define

$$W^{(V)}(t) \stackrel{\text{def}}{=} V^{-1/2}\left[Y(V\lambda t) - \lambda V t\right].$$

The following four properties of the process $W^{(V)}$ all follow from the corresponding properties of the Poisson process.

1. $W^{(V)}(0) = 0$;
2. For any $s_1 \le t_1 \le s_2 \le t_2 \le \cdots \le s_n \le t_n$, the random variables $W^{(V)}(t_1) - W^{(V)}(s_1), \ldots, W^{(V)}(t_n) - W^{(V)}(s_n)$ are independent;
3. For any $s < t$, the random variable $W^{(V)}(t) - W^{(V)}(s)$ is approximately normal with mean zero and variance $\lambda(t-s)$.
4. $W^{(V)}$ is constant except for jumps of size $1/\sqrt{V}$.

It can be shown that as $V \to \infty$, the above process converges to a process, W,[1] written $W^{(V)} \Rightarrow W$, satisfying the conditions of a Brownian motion. Note that in this scaling/limit above, a standard Brownian motion arises from scaling a unit-rate Poisson process.

Note that if we wish the process to start at $x \in \mathbb{R}$ as opposed to zero, then item 1 becomes $W(0) = x$ and all other items remain the same. Further, note the important fact that for any $s < t$,

$$W(t) - W(s)$$

has a normal distribution with mean zero and variance $\lambda(t-s)$, and so, for example,

$$\mathbb{E}[W(t) - W(s)] = 0, \quad \text{and} \quad \mathbb{E}\left[|W(t) - W(s)|^2\right] = \lambda(t-s).$$

We will not rigorously study Brownian motions in these notes.

6.1.1 Markov property and generator of a Brownian motion

The Brownian motion W has theMarkov property by condition (ii) in its definition. That is, the future behavior of the process only depends upon its current value. We may therefore ask: is there a generator for the process?

Let W be a Brownian motion with variance parameter σ^2. We attempt to find a generator in the sense of (4.19):

$$(Af)(x) \stackrel{\text{def}}{=} \lim_{h \to 0} \frac{\mathbb{E}_x f(W(h)) - f(x)}{h}.$$

Letting $\rho = N(0, \sigma^2 h)$, we note

$$\begin{aligned}
\mathbb{E}[\rho] &= 0 \\
\mathbb{E}[\rho^2] &= \sigma^2 h \\
\mathbb{E}[\rho^3] &= 0 \\
\mathbb{E}[\rho^4] &= O(h^2).
\end{aligned}$$

[1] The type of convergence is beyond the scope of this book. Technically, it is convergence in distribution and we write $W^{(V)} \Rightarrow W$.

Assuming $W(0) = x$, we have that

$$W(h) = W(0) + W(h) - W(0) = x + \rho,$$

and so

$$\begin{aligned}(Af)(x) &= \lim_{h\to 0} \frac{\mathbb{E}_x f(W(h)) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{\mathbb{E}_x f(x+\rho) - f(x)}{h}.\end{aligned}$$

Taking a Taylor expansion of f (we are assuming that f is smooth enough for all the derivatives needed to exist) yields

$$(Af)(x) = \lim_{h\to 0} \frac{\mathbb{E}_x[f'(x)\rho + (1/2)f''(x)\rho^2 + (1/3!)f'''(x)\rho^3 + \cdots]}{h} = \frac{1}{2}\sigma^2 f''(x).$$

Said differently, the operator A is the second derivative times $\sigma^2/2$:

$$A = \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2}. \tag{6.1}$$

We return to our construction of the Brownian motion as a limit of a properly scaled Poisson processes to see if we can understand this generator in a different way. Let Y be a unit rate Poisson process. Then, $Y(\sigma^2\cdot)$ is a Poisson process with rate σ^2. We have that for large V,

$$\frac{1}{\sqrt{V}}\left[Y(V\sigma^2 t) - \sigma^2 V t\right] \approx W(\sigma^2 t),$$

where W is a standard Brownian motion. Consider now the generator of the process

$$Z^V(t) = Z^V(0) + \frac{1}{\sqrt{V}}\left[Y(V\sigma^2 t) - \sigma^2 V t\right].$$

Denoting the generator of Z^V by A_Z, we have

$$(A_Z f)(x) \stackrel{\text{def}}{=} \lim_{h\to 0} \frac{1}{h}\left[\mathbb{E}_x f(Z^V(h)) - f(x)\right].$$

Further,

$$\begin{aligned}\mathbb{E}_x[f(Z^V(h))] &= f\left(x + \frac{1}{\sqrt{V}}\left[1 - \sigma^2 V h\right]\right) V\sigma^2 h \\ &\quad + f\left(x - \sigma^2\sqrt{V}h\right)(1 - V\sigma^2 h) + o(h) \\ &= f\left(x + \frac{1}{\sqrt{V}}\right) V\sigma^2 h \\ &\quad + \left(f(x) - f'(x)\sqrt{V}\sigma^2 h + o(h)\right)(1 - V\sigma^2 h) + o(h) \\ &= f\left(x + \frac{1}{\sqrt{V}}\right) V\sigma^2 h + f(x)(1 - V\sigma^2 h) - f'(x)\sqrt{V}\sigma^2 h + o(h).\end{aligned}$$

Therefore,

$$\begin{aligned}(A_Z f)(x) &= \lim_{h\to 0}\frac{1}{h}\left[\mathbb{E}_x[f(Z^V(h))] - f(x)\right] \\ &= f(x+1/\sqrt{V})V\sigma^2 - f(x)V\sigma^2 - f'(x)\sqrt{V}\sigma^2 \\ &= V\sigma^2\left(f(x+1/\sqrt{V}) - f(x)\right) - \sqrt{V}\sigma^2 f'(x).\end{aligned}$$

Note that this generator can be understood by first considering the jump portion,

$$V\sigma^2\left(f(x+1/\sqrt{V}) - f(x)\right),$$

followed by the deterministic portion $-\sqrt{V}\sigma^2 f'(x)$ (recall Section 5.1.6). Taking a Taylor approximation of $A_Z f$ now yields:

$$\begin{aligned}(A_Z f)(x) &= V\sigma^2\left(\frac{f'(x)}{\sqrt{V}} + \frac{1}{2}\frac{1}{V}f''(x) + \frac{1}{3!}\frac{1}{V^{3/2}}f'''(x) + O(V^{-2})\right) - \sqrt{V}\sigma^2 f'(x) \\ &= \frac{1}{2}\sigma^2 f''(x) + O(V^{-1/2}) \\ &\approx (Af)(x),\end{aligned}$$

where A is the generator for the Brownian motion as given by (6.1). Thus, not unexpectedly, the generator of the Brownian motion can be obtained via the generator of the scaled Poisson process simply by truncating the Taylor expansion of $A_Z f$.

6.2 Integration with Respect to Brownian Motion

Before discussing how to integrate with respect to Brownian motion, we consider how to integrate with respect to a more standard function. Consider two functions, g and U. We will discuss what we mean by "integration of g with respect to U." That is, we will define

$$\int_0^t g(x)dU(x).$$

Letting $t_i = it/n$, this integral basically means

$$\int_0^t g(x)dU(x) = \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} g(x)dU(x) \approx \sum_{i=0}^{n-1} g(t_i)(U(t_i+\Delta t) - U(t_i)), \tag{6.2}$$

where $\Delta t = t_{i+1} - t_i$, where the definition of the integral is in the limit $\Delta t \to 0$. The most common type of such integration is when U is *absolutely continuous*. If you are not sure what this means, just think of U as being differentiable, with derivative (or

density), u. In this case, (6.2) yields

$$\begin{aligned}\int_0^t g(x)dU(x) &\approx \sum_{i=0}^{n-1} g(t_i)(U(t_i+\Delta t)-U(t_i)) \\ &\approx \sum_{i=0}^{n-1} g(t_i)u(t_i)\Delta t \\ &\approx \int_0^t g(t_i)u(t_i)dt.\end{aligned}$$

As discussed in the previous section, a Brownian motion is not differentiable, and it can be shown to not be absolutely continuous. However, you can still make sense of

$$\int_0^t g(s)dW(s) = \lim_{\Delta t\to\infty}\sum_{i=0}^{n-1} g(t_i)(W(t_i+\Delta t)-W(t_i)), \tag{6.3}$$

where g is some function and W is a (standard) Brownian motion, in a similar manner. This integral is termed an Itô integral. The reason that constructing such an integral is actually difficult is because $W(t_i+\Delta)-W(t_i)\approx O(\sqrt{\Delta})$, and so

$$\int_0^t g(s)dW(s) \approx \sum_{i=0}^{n-1} g(t_i)\sqrt{\Delta t}\frac{(W(t_i+\Delta t)-W(t_i))}{\sqrt{\Delta t}},$$

appears to blow up as $\Delta t\to 0$ since $n = 1/\Delta t$. However, using the properties of a Brownian motion,

$$\begin{aligned}\mathbb{E}\left(\sum_{i=0}^{n-1} g(t_i)(W(t_i+\Delta t)-W(t_i))\right)^2 &= \sum_{i=0}^{n-1} g(t_i)^2\mathbb{E}(W(t_i+\Delta t)-W(t_i))^2 \\ &= \sum_{i=0}^{n-1} g(t_i)^2\Delta t \\ &\approx \int_0^t g(s)^2 ds,\end{aligned} \tag{6.4}$$

where the first equality holds because all the cross terms are zero since for $j>i$

$$\mathbb{E}(W(t_{j+1})-W(t_j))(W(t_{i+1})-W(t_i)) = \mathbb{E}(W(t_{j+1})-W(t_j))\mathbb{E}(W(t_{i+1})-W(t_i)) = 0.$$

Thus, the right hand side of (6.3) at least does not blow up as $\Delta t\to 0$. To see that such an integral is actually well defined, with all technical details provided, see, for example, [30]. However, the basic construction implied above is correct and, as implied by equation 6.4 above,

$$\mathbb{E}\left(\int_0^t g(s)dW(s)\right)^2 = \int_0^t g(s)^2 ds,$$

which is called the Itô isometry. Even more generally, if $g(s,\omega)$ depends upon W only up through time s (that is, $g(s,\omega)$ is contained in $\mathcal{F}_s$, or is $\mathcal{F}_s$ measurable), then the Itô isometry still holds,

$$\mathbb{E}\left(\int_0^t g(s,\omega)dW(s)\right)^2 = \int_0^t \mathbb{E}[g(s,\omega)^2]ds.$$

This follows from only a slight reworking of (6.4)

$$\begin{aligned}\mathbb{E}\left(\sum_{i=0}^{n-1} g(t_i,\omega)(W(t_i+\Delta t)-W(t_i))\right)^2 &= \sum_{i=0}^{n-1}\mathbb{E}[g(t_i,\omega)^2]\mathbb{E}(W(t_i+\Delta t)-W(t_i))^2\\ &= \sum_{i=0}^{n-1}\mathbb{E}g(t_i,\omega)^2\Delta t\\ &\approx \int_0^t \mathbb{E}[g(s,\omega)^2]ds,\end{aligned}$$

where the first equality holds since $g(t_i,\omega)$ and $W(t_i+\Delta)-W(t_i)$ are independent. For example,

$$\mathbb{E}\left(\int_0^t W(s)dW(s)\right)^2 = \int_0^t \mathbb{E}[W(s)^2]ds = \int_0^t s\,ds = \frac{t^2}{2}.$$

We will not give a comprehensive introduction to stochastic integration. The interested reader is instead pointed towards [28]. However, we will point out a few things, and solve two integrals explicitly. First, we note that

$$\mathbb{E}\left[\int_0^t g(s,\omega)dW(s)\right] \approx \sum_i \mathbb{E}[g(t_i,\omega)]\mathbb{E}[(W(t_{i+1})-W(t_i))] = 0,$$

and so all such integrals have a mean of zero. This is in stark departure from standard Riemannian integration where $g(s)>0$ implies $\int_0^t g(s)ds > 0$.

We now solve two examples explicitly. First, for constant $\sigma > 0$, we have that

$$\int_0^t \sigma dW(s) = \sigma\sum_i (W(t_{i+1})-W(t_i)) = \sigma W(t).$$

This agrees with our intuition that comes from integrating deterministic functions:

$$\int_0^t \sigma df(s) = \sigma(f(t)-f(0)).$$

Second, we consider the integral above: $\int_0^t W(s)dW(s)$. A first (incorrect) guess would be to argue as follows: since for differentiable f we have

$$\int_0^t f(s)df(s) = \int_0^t f(s)f'(s)ds = \frac{1}{2}(f(t)^2-f(0)^2),$$

it must be that

$$\int_0^t W(s)dW(s) = \frac{1}{2}W(t)^2.$$

However, we can instantly see this is incorrect by simply checking the moments of $(1/2)W(t)^2$:

$$\mathbb{E}\left[\frac{1}{2}W(t)^2\right] = \frac{1}{2}t \neq 0$$
$$\mathbb{E}\left(\frac{1}{2}W(t)^2\right)^2 = \frac{1}{4}\mathbb{E}[W(t)^4] = \frac{1}{4}3t^2 \neq \frac{1}{2}t^2.$$

Thus, it has incorrect first and second moments, and so we must be more careful. Let

$$Z_{\Delta t}(t) = \sum_i W(t_i)(W(t_{i+1}) - W(t_i)),$$

where $\Delta t = t_{i+1} - t_i$. Then,

$$\begin{aligned} Z_{\Delta t}(t) =& \sum_i \frac{1}{2}(W(t_{i+1}) + W(t_i))(W(t_{i+1}) - W(t_i)) \\ & - \sum_i \frac{1}{2}(W(t_{i+1}) - W(t_i))(W(t_{i+1}) - W(t_i)) \\ =& \sum_i \frac{1}{2}(W(t_{i+1})^2 - W(t_i)^2) - \sum_i \frac{1}{2}(W(t_{i+1}) - W(t_i))^2 \\ =& \frac{1}{2}W(t)^2 - \frac{1}{2}\sum_i (W(t_{i+1}) - W(t_i))^2. \end{aligned} \tag{6.5}$$

Note that we have recovered our first guess of: $W(t)^2/2$, though there is now a correction in the form of the sum on the far right hand side of (6.5)? Let

$$Q_i = (W(t_{i+1}) - W(t_i))^2.$$

Q_i has a mean of Δt and variance of $O(\Delta t^2)$. Therefore, the random variable $\sum_i Q_i$ has a mean of t and a variance of $O(\Delta)$. Thus, for an appropriate constant C,

$$\frac{\sum_i Q_i - t}{C\sqrt{\Delta}} \approx N(0,1).$$

In particular,

$$\sum_i Q_i - t = O(\sqrt{\Delta t}) \to 0, \text{ as } \Delta t \to 0,$$

implying

$$\sum_i Q_i \to t, \text{ as } \Delta t \to 0.$$

Hence,

$$\frac{1}{2}\sum_i Q_i \to \frac{1}{2}t.$$

Collecting the above shows that

$$\int_0^t W(s)dW(s) \approx Z_{\Delta t}(t) \to \frac{1}{2}W(t)^2 - \frac{1}{2}t, \text{ as } \Delta t \to 0.$$

Hence, we conclude

$$\int_0^t W(s)dW(s) = \frac{1}{2}W(t)^2 - \frac{1}{2}t.$$

Note that, as expected, we have

$$\mathbb{E}\left[\frac{1}{2}W(t)^2 - \frac{1}{2}t\right] = 0,$$

and

$$\begin{aligned}\mathbb{E}\left[\int_0^t W(s)dW(s)\right]^2 = \mathbb{E}\left[\frac{1}{2}W(t)^2 - \frac{1}{2}t\right]^2 &= \frac{1}{4}\mathbb{E}[W(t)^4] - \frac{1}{2}t\cdot\mathbb{E}[W(t)^2] + \frac{1}{4}t^2 \\ &= \frac{3}{4}t^2 - \frac{1}{2}t^2 = \frac{1}{2}t^2.\end{aligned}$$

We now construct another process from a Brownian motion that is in many ways equivalent to the one constructed above, and will be of use to us. Let W be a standard Brownian motion and consider the process

$$Z(t) \stackrel{\text{def}}{=} W\left(\int_0^t g(s,\omega)^2 ds\right),$$

where, again, $g(s,\omega)$ may depend upon W, but only up until time $\tau = \int_0^t g(s,\omega)ds$. That is, it is contained within $\mathcal{F}_\tau$. The above is a *time-changed* Brownian motion.We have

$$Z(t+h) - Z(t) = W\left(\int_t^{t+h} g(s,\omega)^2 ds + \int_0^t g(s,\omega)^2 ds\right) - W\left(\int_0^t g(s,\omega)^2 ds\right),$$

which by the independent increments of W is approximately normal with mean zero and variance $g(t,\omega)^2 h$, which is exactly the same distribution as the infinitesimal increment

$$\int_0^{t+h} g(s,\omega)dW(s) - \int_0^t g(s,\omega)dW(s) \approx g(t,\omega)(W(t+h) - W(t)).$$

This implies that the two processes

$$\int_0^t g(s,\omega)dW(s) \quad \text{and} \quad W\left(\int_0^t g(s,\omega)^2 ds\right) \tag{6.6}$$

are distributionally equivalent (thought not equal for a *given* Brownian path). The representation on the left of (6.6) is termed an *Itô* integral, whereas the process on the right is the time changed process, and can be traced back to Wolfgang Doeblin.

6.3 Diffusion and Linear Noise Approximations

We are in position to give two approximations to the process (5.3) which use Brownian motions.

6.3.1 Diffusion approximation

Define the function F via

$$F(x) = \sum_k \hat{\kappa}_k x^{y_k} \zeta_k, \tag{6.7}$$

which is deterministic mass-action kinetics. Returning to (5.21), the scaled model satisfies

$$X^V(t) = X^V(0) + \sum_k \frac{1}{V} Y_k \left(V \int_0^t \hat{\kappa}_k X(s)^{y_k} ds \right) \zeta_k,$$

which, after centering the counting processes[2] yields

$$X^V(t) = X^V(0) + \sum_k \frac{1}{V} \left(Y_k \left(V \int_0^t \hat{\kappa}_k X^V(s)^{y_k} ds \right) - V \int_0^t \hat{\kappa}_k X^V(s)^{y_k} ds \right) \zeta_k + \int_0^t F(X^V(s)) ds.$$

Using that

$$\frac{1}{\sqrt{V}} \left[Y_k(Vu) - Vu \right] \approx W_k(u), \tag{6.8}$$

where W is a standard Brownian motion, we then have that

$$X^V(t) \approx X^V(0) + \int_0^t F(X^V(s)) ds + \sum_k \frac{1}{\sqrt{V}} W_k \left(\int_0^t \hat{\kappa}_k X^V(s)^{y_k} ds \right) \zeta_k,$$

where the W_k are independent standard Brownian motions. This implies that a good approximation to X^V would be the process Z^V satisfying

$$Z^V(t) = X^V(0) + \int_0^t F(Z^V(s)) ds + \sum_k \frac{1}{\sqrt{V}} W_k \left(\int_0^t \hat{\kappa}_k Z^V(s)^{y_k} ds \right) \zeta_k.$$

Considering (6.6), an equivalent way to represent Z^V is via the Itô representation

$$Z^V(t) = Z^V(0) + \int_0^t F(Z^V(s)) ds + \sum_k \frac{1}{\sqrt{V}} \zeta_k \int_0^t \sqrt{\hat{\kappa}_k Z^V(s)^{y_k}} dW_k(s).$$

This equation is often represented in differential form

$$dZ^V(t) = F(Z^V(t)) dt + \sum_k \frac{1}{\sqrt{V}} \zeta_k \sqrt{\hat{\kappa}_k Z^V(t)^{y_k}} dW_k(s). \tag{6.9}$$

[2] The centered version of $Y(u)$ is $Y(u) - u$. that is, it arrises simply by subtracting off the mean.

Note that the portion of the system that is stochastic, often termed the "noise" in the system, is $O(1/\sqrt{V})$, and hence assumed small. This equation is known as the *Langevin* approximation in the biology literature, and as the *diffusion* approximation in probability. There is actually an issue related to this approximation that is still not completely resolved in the chemical setting pertaining to the non-negativity of the system. Therefore, perhaps a more reasonable representation would be

$$dZ^V(t) = F(Z^V(t))dt + \sum_k \frac{1}{\sqrt{V}}\zeta_k \sqrt{\hat{\kappa}_k [Z^V(t)^{y_k}]^+} dW_k(s),$$

where $[x]^+ = \max\{x, 0\}$.

Note that there is no limit taking place in the derivation of the diffusion approximation. In fact, the system satisfying (6.9) converges to the solution of the deterministic process with mass-action kinetics in the limit $V \to \infty$.

Example 6.1. Consider the system

$$A \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} B.$$

Letting X_A^V, X_B^V denote the normalized abundances of the species A and B, respectively, we have that $X_A^V(t) + X_B^V(t) = M$, for some $M > 0$, and

$$X_A^V(t) = X_A^V(0) + \frac{1}{V}Y_2\left(V\int_0^t \kappa_2(M - X_A^V(s))ds\right) - \frac{1}{V}Y_1\left(V\int_0^t \kappa_1 X_A^V(s)ds\right).$$

Therefore, the diffusion approximation is the solution to

$$\begin{aligned} Z^V(t) = Z^V(0) &+ \kappa_2\int_0^t (M - Z^V(s))ds - \kappa_1\int_0^t Z^V(s)ds \\ &+ \frac{1}{\sqrt{V}}W_2\left(\int_0^t \kappa_2[M - Z^V(s)]^+ ds\right) - \frac{1}{\sqrt{V}}W_1\left(\int_0^t \kappa_1[Z^V(s)]^+ ds\right), \end{aligned}$$

or, equivalently, the solution to the stochastic differential equation

$$\begin{aligned} dZ^V(t) =& \kappa_2(M - Z^V(t))dt - \kappa_1\int_0^t Z^V(s)ds \\ &+ \frac{1}{\sqrt{V}}\sqrt{\kappa_2[M - Z^V(t)]^+}dW_1(t) - \frac{1}{\sqrt{V}}\sqrt{\kappa_1[Z^V(t)]^+}dW_2(t). \end{aligned}$$

△

6.3.2 Linear noise approximation

Let $x(t)$ be the solution to the limiting deterministic system (5.22), and recall that F is the deterministic kinetics defined in (6.7). Since by (6.8), $V^{-1/2}[Y_k(Vu) - Vu]$

is approximately a Brownian motion,

$$\begin{aligned}
L^V(t) &\stackrel{\text{def}}{=} \sqrt{V}(X^V(t) - x(t)) \\
&= L^V(0) + \sqrt{V}\left(\sum_k \frac{1}{V} Y_k \left(V \int_0^t \hat{\lambda}_k(X^V(s))ds \right) \zeta_k - \int_0^t F(x(s))ds \right) \\
&= L^V(0) + \sum_k \frac{1}{\sqrt{V}} \left[Y_k \left(V \int_0^t \hat{\lambda}_k(X^V(s))ds \right) - V \int_0^t \hat{\lambda}_k(X^V(s))ds \right] \zeta_k \\
&\qquad + \int_0^t \sqrt{V}(F(X^V(s)) - F(x(s)))ds \\
&\approx L^V(0) + \sum_k W_k \left(\int_0^t \hat{\lambda}_k(x(s))ds \right) \zeta_k + \int_0^t \nabla F(x(s)) \cdot L^V(s)ds.
\end{aligned}$$

The limit as V goes to infinity gives $L^V \Rightarrow L$ where

$$L(t) = L(0) + \sum_k W_k \left(\int_0^t \hat{\lambda}_k(x(s))ds \right) \zeta_k + \int_0^t \nabla F(x(s)) \cdot L(s)ds. \tag{6.10}$$

For more details, see [22, 25, 35] and Chapter 11 of [12]. Note that an alternative representation of (6.10) is

$$L(t) = L(0) + \sum_k \zeta_k \int_0^t \sqrt{\hat{\lambda}_k(x(s))} dW_k(s) + \int_0^t \nabla F(x(s)) \cdot L(s)ds,$$

where now positivity of the term in the square root is guaranteed as x is the solution to the deterministic model and stays positive for all time (so long as each $x_i(t) > 0$). The above limit suggests the approximation

$$X^V(t) \approx \hat{X}_V(t) \stackrel{\text{def}}{=} x(t) + \frac{1}{\sqrt{V}} L(t),$$

which is often called the *linear noise approximation* to X^V, and is used quite extensively. Note that, once again, the "noise" scales like $1/\sqrt{V}$.

Example 6.2. Consider the system

$$A \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} B.$$

The ordinary differential equation governing $x_A(t)$, the concentration of A, is

$$\dot{x}_A(t) = F(x(t)) \stackrel{\text{def}}{=} \kappa_2(M - x_A(t)) - \kappa_1 x_A(t), \tag{6.11}$$

where $M = x_A(t) + x_B(t)$. Therefore,

$$F'(x) = -\kappa_2 - \kappa_1.$$

Assuming $X_A^V(0) = x_A(0)$, the equation for L is then

$$L(t) = \int_0^t \sqrt{\kappa_2(M - x_A(s))}dW_1(s) - \int_0^t \sqrt{\kappa_1 x_A(s)}dW_2(s) - (\kappa_1 + \kappa_2)\int_0^t L(s)ds,$$

or

$$dL(t) = \sqrt{\kappa_2(M - x_A(t))}dW_1(t) - \sqrt{\kappa_1 x_A(t)}dW_2(t) - (\kappa_1 + \kappa_2)L(s)dt.$$

Solving this equation yields,

$$L(t) = \int_0^t e^{-(\kappa_1+\kappa_2)(t-s)}\sqrt{\kappa_2(M - x_A(s))}dW_1(s) - \int_0^t e^{-(\kappa_1+\kappa_2)(t-s)}\sqrt{\kappa_1 x_A(s)}dW_2(s).$$

Finally, we set

$$\hat{X}_V(t) = x_A(t) + \frac{1}{\sqrt{V}}L(t).$$

△

6.4 Numerical methods for stochastic differential equations

Consider the stochastic equation

$$X(t) = X(0) + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dW(s), \tag{6.12}$$

where $X \in \mathbb{R}$, $b : \mathbb{R} \to \mathbb{R}$, and $\sigma : \mathbb{R} \to \mathbb{R}_{\geq 0}$. The differential form of the above equation is

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t),$$

where we do not use the notation $dX(t)/dt$ since W is not differentiable, and hence $dW(t)/dt$ is problematic; therefore, it is better to think in terms of the differentials dX or dW. Note, however, that this is all just notation and no matter what we write, we mean that X satisfies the integral equation above.

The most common numerical method, by far, to solve for the solution to (6.12) is Euler's method. That is, we use

$$\begin{aligned} X(t+h) &= X(t) + \int_t^{t+h} b(X(s))ds + \int_t^{t+h} \sigma(X(s))dW(s) \\ &\approx X(t) + b(X(t))h + \sigma(X(t))(W(t+h) - W(t)) \\ &\overset{\text{dist}}{=} X(t) + b(X(t))h + \sigma(X(t))\sqrt{h}\rho, \end{aligned}$$

where $\rho \sim N(0,1)$. This observation leads to the following algorithm. In the algorithm below, all random variables generated are assumed to be independent of all previous random variables. The constructed process will be denoted by Z so as to differentiate it and the exact process X.

Algorithm 6. Fix $Z(0)$ and $h > 0$. Set $n = 0$. Repeat the following steps.

1. Let $\rho \sim N(0,1)$.
2. Set
$$Z((n+1)h) = Z(nh) + b(Z(nh))h + \sigma(Z(nh))\sqrt{h}\rho$$

Note that one way to represent Z (at least at the points nh), is as the solution to

$$Z(t) = Z(0) + \int_0^t b(Z \circ \eta(s))ds + \int_0^t \sigma(Z \circ \eta(s))dW(s),$$

where $\eta(s) = nh$ for $nh \leq s < (n+1)h$. This follows since if Z is so generated then

$$\begin{aligned} Z((n+1)h) &= Z(nh) + \int_{nh}^{(n+1)h} b(Z(nh))ds + \int_{nh}^{(n+1)h} \sigma(Z(nh))dW(s) \\ &= Z(nh) + b(Z(nh))h + \sigma(Z(nh))(W((n+1)h) - W(nh)), \end{aligned}$$

and $W((n+1)h) - W(nh) \sim N(0,h)$.

Note also that (6.12) is distributionally equivalent to the time changed representation

$$X(t) = X(0) + \int_0^t b(X(s))ds + W\left(\int_0^t \sigma^2(X(s))ds\right). \tag{6.13}$$

This can be seen by noting that for (6.13), Euler's method reduces to

$$Z(t) = Z(0) + \int_0^t b(Z \circ \eta(s))ds + W\left(\int_0^t \sigma^2(Z \circ \eta(s))ds\right),$$

yielding

$$\begin{aligned} Z((n+1)h) = Z(nh) &+ \int_{nh}^{(n+1)h} b(Z(nh))ds \\ &+ W\left(\int_{nh}^{(n+1)h} \sigma^2(Z \circ \eta(s))ds + \int_0^{nh} \sigma^2(Z \circ \eta(s))ds\right) \\ &- W\left(\int_0^{nh} \sigma^2(Z \circ \eta(s))ds\right) \\ = Z(nh) &+ b(Z(nh))h + W\left(\sigma^2(Z(nh))h + \int_0^{nh} \sigma^2(Z \circ \eta(s))ds\right) \\ &- W\left(\int_0^{nh} \sigma^2(Z \circ \eta(s))ds\right), \end{aligned}$$

where

$$W\left(\sigma^2(Z(nh))h + \int_0^{nh} \sigma^2(Z \circ \eta(s))ds\right) - W\left(\int_0^{nh} \sigma^2(Z \circ \eta(s))ds\right),$$

is normally distributed with mean zero and variance $\sigma^2(Z(nh))h$.

Appendix A

Solving linear difference equations

A.1 Linear Difference Equations

Difference equations come up naturally in the study of discrete time Markov chains, and it is important that we are able to solve them.

For real valued constants a and b consider the following system of equations,

$$f(n) = af(n-1) + bf(n+1), \quad \text{valid for } A < n < B, \tag{A.1}$$

where f is a function defined on the integers $\{A, \ldots, B\}$, and the value B can be chosen to be infinity.

Note that if f_1 and f_2 are two solutions of (A.1), then for real numbers c_1 and c_2, the function $g = c_1 f_1 + c_2 f_2$ is also a solution. This can be easily checked. Let $A < n < B$ be an integer and let f_1 and f_2 be two solutions to (A.1). Then,

$$\begin{aligned} g(n) &= c_1 f_1(n) + c_2 f_2(n) \\ &= c_1 \left[a f_1(n-1) + b f_1(n+1) \right] + c_2 \left[a f_2(n-1) + b f_2(n+1) \right] \\ &= a g(n-1) + b g(n+1). \end{aligned}$$

It is a fact that if f_1 is not a multiple constant of f_2, then functions of the form $c_1 f_1 + c_2 f_2$ make up all possible solutions to (A.1). (This last statment is true because the solution space to (A.1) is a two-dimensional vector space, but we will not dwell on this.) Our goal is therefore clear: find two solutions to (A.1) that do not differ by a multiple constant. Once we have two such solutions, we can solve for our specific solution, i.e. determine the constants c_1 and c_2 above, via any boundary conditions provided.

Solving equations of the form (A.1) turns out to be relatively straightforward. We will simply look for solutions of the form $f(n) = \alpha^n$, for some $\alpha \neq 0$. By plugging α^n into equation (A.1) we find that $f(n) = \alpha^n$ is a solution if and only if α solves the following algebraic equation,

$$\alpha^n = a\alpha^{n-1} + b\alpha^{n+1}, \quad \text{valid for} \quad A < n < B.$$

After dividing by α^{n-1}, we have $\alpha = a + b\alpha^2$. Solving the quadratic for α yields

$$\alpha = \frac{1 \pm \sqrt{1-4ba}}{2b}. \tag{A.2}$$

There are three cases that need handling based upon whether or not the discriminant, $1-4ba$, is positive, negative, or zero.

Case 1: If $1-4ba > 0$, we find two real solutions, α_1 and α_2, and see that the general solution to the difference equation (A.1) is

$$c_1\alpha_1^n + c_2\alpha_2^n,$$

with c_1 and c_2 found depending upon any boundary conditions given.

Case 2: If $1-4ba < 0$, then the roots are complex and the general solution is found by switching to polar coordinates. That is, we let $\alpha = re^{i\theta}$ be one solution of (A.2) and find

$$f(n) = r^n e^{in\theta} = r^n \cos(n\theta) \pm ir^n \sin(n\theta), \tag{A.3}$$

is a solution to (A.1). However, this implies that the real and imaginary parts of (A.3) are also solutions to (A.1). Therefore, the general solution to (A.1) is

$$c_1 r^n \cos(n\theta) + c_2 r^n \sin(n\theta),$$

with c_1 and c_2 found depending upon any boundary conditions given.

Case 3: If $1-4ba = 0$, then there is only one solution to (A.2). Therefore, we have only found one solution to (A.1), namely $f_1(n) = (1/2b)^n$. However, let $f_2(n) = n(2b)^{-n}$. We will check that f_2 is also a solution. For $A < n < B$ we have

$$\begin{aligned}
af_2(n-1) + bf_2(n+1) &= a(n-1)(2b)^{-(n-1)} + b(n+1)(2b)^{-(n+1)} \\
&= \left(\frac{1}{2b}\right)^n \left[a(n-1)2b + b(n+1)\frac{1}{2b}\right] \\
&= \left(\frac{1}{2b}\right)^n \left[(n-1)\frac{1}{2} + (n+1)\frac{1}{2}\right] \qquad (\text{since, } 4ab = 1) \\
&= \left(\frac{1}{2b}\right)^n n \\
&= f_2(n).
\end{aligned}$$

Note that f_2 is not a multiple constant of f_1. Thus, when $4ab = 1$, the general form of the solution to (A.1) is

$$f(n) = c_1\left(\frac{1}{2b}\right)^n + c_2 n\left(\frac{1}{2b}\right)^n,$$

with c_1 and c_2 found depending upon the boundary conditions given.

Example A.1. Find a function $f(n)$ satisfying

$$f(n) = 2f(n-1) + \tfrac{1}{10}f(n+1), \qquad 0 < n < \infty,$$

with $f(0) = 8$ and $f(1) = 2$.

Comparing the above equation with (A.1), we have $a = 2$, $b = \frac{1}{10}$, $A = 0$ and $B = \infty$. Finding α via (A.2) yields

$$\alpha = 5 \pm \sqrt{5},$$

and the general solution to the difference equation is

$$f(n) = c_1 \left(5 + \sqrt{5}\right)^n + c_2 \left(5 - \sqrt{5}\right)^n.$$

Using the boundary conditions yields

$$\begin{aligned} 8 &= f(0) = c_1 + c_2 \\ 2 &= f(1) = c_1(5 + \sqrt{5}) + c_2(5 - \sqrt{5}), \end{aligned}$$

which has solution $c_1 = 4 - \frac{19\sqrt{5}}{5}$, $c_2 = 4 + \frac{19\sqrt{5}}{5}$. Thus, the particular solution is

$$f(n) = \left(4 - \frac{19\sqrt{5}}{5}\right)\left(5 + \sqrt{5}\right)^n + \left(4 + \frac{19\sqrt{5}}{5}\right)\left(5 - \sqrt{5}\right)^n.$$

□

The most form of difference equation we will see in this book are of the form

$$f(n) = q \cdot f(n-1) + p \cdot f(n+1), \quad \text{with } p + q = 1, \quad p, q \geq 0.$$

Supposing that $p \neq q$, the roots of the quadratic formula (A.2) can be found:

$$\frac{1 \pm \sqrt{1 - 4qp}}{2p} = \frac{1 \pm \sqrt{(q-p)^2}}{2p} = \frac{1 \pm |q-p|}{2p} = \left\{1, \frac{q}{p}\right\}.$$

Thus, the general solution when $p \neq \frac{1}{2}$ is

$$f(n) = c_1 + c_2 \left(\frac{q}{p}\right)^n.$$

For the case that $p = q = \frac{1}{2}$, the only root is 1. Hence, when $p = q = \frac{1}{2}$, the general solution is

$$f(n) = c_1 + c_2 n.$$

A.2 Exercises

Problems A.1 through A.3 require basic linear algebra, which is not explicitly covered in these notes. They are provided to ensure that you have the minimum required knowledge of linear algebra. If the concepts being presented are foreign to you, I strongly encouraged you to consult a text on basic linear algebra, for example, [11], and get in touch with your instructor.

Some of the problems require you to use MATLAB. You must always turn in your MATLAB code when a problem is solved using MATLAB.

Exercise A.1. Let

$$A = \begin{bmatrix} -2 & 1 \\ 2 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad v = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

Calculate (by hand) the following:

(a) A^2.

(b) AB.

(c) BA.

(d) Av.

Exercise A.2. Redo Problem A.1 using MATLAB. Sample code showing how to multiply two matrices is provided on the course website.

Exercise A.3. Find the eigenvalues and left eigenvectors of the matrix A in Problem A.1. Do this in two ways: by hand and by using MATLAB.

Exercise A.4. Solve the following difference equation by finding a formula for $f(n)$, valid on $n = 0, \ldots, 100$.

$$f(n) = \tfrac{2}{5} f(n-1) + \tfrac{3}{5} f(n+1), \quad n = 1, \ldots, 99,$$

and $f(0) = 1, f(1) = 0$.

Appendix B

Some probability basics

In this section, we collect a few topics from probability theory that are most pertinent to the subject matter of this text. Namely, we discuss exponential random variables, transformations of random variables, and Monte Carlo methods. We assume the reader already has a basic knowledge of probability theory at the level of [4] or [32].

B.1 Exponential random variables

A random variable X has an *exponential distribution* with parameter $\lambda > 0$ if it has probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}.$$

We write $X \sim \text{Exp}(\lambda)$.

If $X \sim \text{Exp}(\lambda)$, then

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{and} \quad \mathsf{Var}(X) = \frac{1}{\lambda^2}.$$

The exponential random variable satisfies the memoryless property.

Proposition B.1 (Memoryless property)**.** *Let $X \sim Exp(\lambda)$, then for any $s, t \geq 0$,*

$$P(X > (s+t) \mid X > t) = P(X > s). \tag{B.1}$$

One important role the exponential random variable plays in these notes is as the inter-event time of Poisson processes.

Proposition B.2. *Consider a Poisson process with rate $\lambda > 0$. Let T_i be the time between the ith and $i+1$st events. Then $T_i \sim Exp(\lambda)$.*

The following properties of independent exponential random variables play a large role in simulation methods for continuous time Markov chains.

Proposition B.3. *Suppose that $X_i \sim Exp(\lambda_i)$, $1 \leq i \leq n$, are independent. Then*

$$X_0 = \min_{1 \leq i \leq n} \{X_i\} \sim Exp(\lambda_0), \quad \textit{where} \quad \lambda_0 = \sum_{i=1}^{n} \lambda_i.$$

Proof. Let $X_0 = \min_i\{X_i\}$. Set $\lambda_0 = \sum_{i=1}^n \lambda$. Then,

$$P(X_0 > t) = P(X_1 > t, \ldots, X_n > t) = \prod_{i=1}^{n} P(X_i > t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-\lambda_0 t},$$

where the second equality follows by the independence of the random variables. □

Proposition B.4. *Suppose that $X_i \sim Exp(\lambda_i)$, $1 \leq i \leq n$, are independent. Let $j = argmin_{1 \leq i \leq n}\{X_i\}$ be the index of the smallest X_i. Then, for $k \in \{1, \ldots, n\}$,*

$$P(j = k) = \frac{\lambda_k}{\sum_{i=1}^n \lambda_i}.$$

Proof. We first consider the case of $n = 2$. Let $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$ be independent. Then,

$$P(X < Y) = \iint_{0<x<y} \lambda e^{-\lambda x} \mu e^{-\mu y} dx dy = \int_0^\infty \int_0^y \lambda e^{-\lambda x} \mu e^{-\mu y}\, dx\, dy = \frac{\lambda}{\mu + \lambda}.$$

Returning to the general case, let $Y_k = \min_{i \neq k}\{X_i\}$. By Proposition B.3, $Y_k \sim \text{Exp}\left(\sum_{i \neq k} \lambda_i\right)$. We have already proven the case when $n = 2$, so we may conclude

$$P(j = k) = P(X_k < Y_k) = \frac{\lambda_k}{\lambda_k + \sum_{i \neq k} \lambda_i} = \frac{\lambda_k}{\sum_{i=1}^n \lambda_i},$$

which is the desired result. □

One interpretation of the above two propositions is the following. If you have n alarm clocks, with the ith set to go off after an $\text{Exp}(\lambda_i)$ amount of time, then Proposition B.3 tells you when the first will go off, and Proposition B.4 tells you which one will go off at that time.

B.2 Transformations of random variables

Most software packages have very good and efficient methods for the generation of pseudo-random numbers that are uniformly distributed on the interval $(0, 1)$. These pseudo-random numbers are so good that we will take the perspective throughout these notes that they are, in fact, truly uniformly distributed over $(0, 1)$. We would then like to be able to construct all other random variables as transformations, or functions, of these uniform random variables. The method for doing so will depend upon whether or not the desired random variable is continuous or discrete. In the continuous case, Theorem B.5 will often be used, whereas in the discrete case Theorem B.7 will be used.

Theorem B.5. *Let U be uniformly distributed on the interval $(0,1)$ and let F be an invertible distribution function. Then $X = F^{-1}(U)$ has distribution function F.*

Before proving the theorem, we show how it may be used in practice.

Example B.6. Suppose we want to generate an exponential random variable with parameter $\lambda > 0$. The distribution function on $[0,\infty)$ for this random variable is

$$F(t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

Therefore, $F^{-1} : [0,1) \to \mathbb{R}_{\geq 0}$ is given by

$$F^{-1}(u) = -\frac{1}{\lambda}\ln(1-u), \quad 0 \leq u \leq 1.$$

If U is uniform$(0,1)$, then so is $1-U$. Thus, to simulate a realization of $X \sim \text{Exp}(\lambda)$, we first simulate U from uniform$(0,1)$, and then set

$$x = -\frac{1}{\lambda}\ln(U) = \ln(1/U)/\lambda.$$

□

Proof. (of Theorem B.5) Let $U \sim \text{Unif}[0,1]$ and set $X = F^{-1}(U)$. Then,

$$P(X \leq t) = P(F^{-1}(U) \leq t) = P(U \leq F(t)) = F(t).$$

□

Theorem B.7. *Let U be uniformly distributed on the interval $(0,1)$. Suppose that $p_k \geq 0$ for each $k \in \{0,1,\dots,\}$, and that $\sum_{k=0}^{\infty} p_k = 1$. Define*

$$q_k = \sum_{i=0}^{k} p_i,$$

and let $X = \min\{k : q_k \geq U\}$. Then, $P(X = k) = p_k$.

Proof. Taking $q_{-1} = 0$, we have

$$P(X = k) = P(q_{k-1} < U \leq q_k) = q_k - q_{k-1} = p_k$$

for any $k \in \{0,1,\dots\}$. □

In practice, the above theorem is typically used by repeatedly checking whether or not $U \leq \sum_{i=0}^{k} p_i$, and stopping the first time the inequality holds. We note that the theorem is stated in the setting of an infinite state space, though the analogous theorem holds in the finite state space case.

B.3 Estimation via simulation: Monte Carlo

Let X be a random variable and suppose that we can generate independent copies of X via simulation. How can we estimate $\mu = \mathbb{E}[X]$?

The law of large numbers provides the answer. Generate a sequence of independent realizations of the random variable, $X_1, X_2, \ldots$, and then set

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{B.2}$$

for some $n \geq 1$. We call $\hat{\mu}_n$ an *estimator*. In fact, $\hat{\mu}_n$ is an *unbiased estimator*,

$$\mathbb{E}[\hat{\mu}_n] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mu.$$

By the strong law of large numbers we know that

$$P\left(\lim_{n\to\infty} \hat{\mu}_n = \mu\right) = 1.$$

Knowing that $\hat{\mu}_n \to \mu$, as $n \to \infty$, does not tell us how large of an n we need in practice. This brings us to the next logical question: how good is the estimate for a given, finite n. To answer this question, we will apply the central limit theorem.

Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$. We know from the central limit theorem that the distribution of the standardized random variable

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\hat{\mu}_n - \mu)$$

is approximately that of a Gaussian with mean zero and variance one. Therefore, if $Z \sim N(0,1)$ is a standard normal random variable and $z \in \mathbb{R}$,

$$\begin{aligned} P\left(-z \leq Z \leq z\right) &\approx P\left(-z \leq \frac{\sqrt{n}}{\sigma}(\hat{\mu}_n - \mu) \leq z\right) \\ &= P\left(-\frac{\sigma z}{\sqrt{n}} \leq (\hat{\mu}_n - \mu) \leq \frac{\sigma z}{\sqrt{n}}\right) \\ &= P\left(\hat{\mu}_n - \frac{\sigma z}{\sqrt{n}} \leq \mu \leq \hat{\mu}_n + \frac{\sigma z}{\sqrt{n}}\right). \end{aligned}$$

In words, the above says that the probability that the true value, μ, is within $\pm\sigma z/\sqrt{n}$ of the estimator $\hat{\mu}_n$ is approximately $P\left(-z \leq N(0,1) \leq z\right)$. The interval $(\mu - \sigma z/\sqrt{n}, \mu + \sigma z/\sqrt{n})$ is called our *confidence interval* and the probability $P\left(-z \leq Z \leq z\right)$ is our *confidence*.

We now turn to finding the value z for a desired level of confidence. Again letting $Z \sim N(0,1)$, we let

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}dt,$$

and have

$$P(-z \le Z \le z) = P(Z \le z) - P(Z \le -z) = \Phi(z) - (1 - \Phi(z)) = 2\Phi(z) - 1.$$

Therefore, if for some $\delta > 0$ we want to have a probability of $1-\delta$ that the true value is in the constructed confidence interval, then we must choose z so that

$$2\Phi(z) - 1 = 1 - \delta, \quad \text{or} \quad \Phi(z) = 1 - \frac{\delta}{2}.$$

For example, if $\delta = 0.1$, so that a 90% confidence interval is desired, then we want

$$\Phi(z) = 1 - 0.05 = 0.95,$$

and $z = 1.65$. If, on the other hand, we want $\delta = 0.05$, so that a 95% confidence interval is desired, then

$$\Phi(z) = 1 - 0.025 = 0.975$$

and $z = 1.96$.

There is a major problem with the preceding arguments. If we do not know μ, then we most likely do not know σ. Hence, we can not provide the confidence interval. We must estimate σ from our independent samples.

Theorem B.8. *Let $X_1, \dots, X_n$ be independent and identical samples with mean μ and variance σ^2. Let*

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2,$$

where $\hat{\mu}_n$ is defined in (B.2). Then,

$$\mathbb{E}[\hat{\sigma}_n^2] = \sigma^2.$$

Proof. We have

$$\begin{aligned}
(n-1)\mathbb{E}[\hat{\sigma}_n^2] &= \mathbb{E}\left[\sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2\right] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] - 2\mathbb{E}\left[\hat{\mu}_n \sum_{i=1}^{n} X_i\right] + n\mathbb{E}\left[\hat{\mu}_n^2\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2] - 2\mathbb{E}\left[\hat{\mu}_n n \hat{\mu}_n\right] + n\mathbb{E}\left[\hat{\mu}_n^2\right] \\
&= n\mathbb{E}[X^2] - n\mathbb{E}\left[\hat{\mu}_n^2\right].
\end{aligned}$$

However,

$$\mathbb{E}\left[\hat{\mu}_n^2\right] = \mathsf{Var}(\hat{\mu}_n) + \left(\mathbb{E}\left[\mu_n\right]\right)^2 = \mathsf{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) + \mu^2 = \frac{1}{n}\sigma^2 + \mu^2.$$

Therefore,

$$\frac{n-1}{n}\mathbb{E}[\hat{\sigma}_n^2] = \mathbb{E}[X^2] - \mathbb{E}[\hat{\mu}_n^2] = \left(\sigma^2 + \mu^2\right) - \left(\frac{1}{n}\sigma^2 + \mu^2\right) = \frac{n-1}{n}\sigma^2,$$

completing the proof. □

Therefore, we can use

$$\hat{\sigma}_n = \sqrt{\hat{\sigma}_n^2}$$

as an estimate for the standard deviation in the confidence interval and

$$\left[\hat{\mu}_n - \frac{\hat{\sigma}_n z}{\sqrt{n}}, \quad \hat{\mu}_n + \frac{\hat{\sigma}_n z}{\sqrt{n}}\right]$$

is an approximate $(1-\delta)100\%$ confidence interval for $\mu = \mathbb{E}[X]$.

We note that there are two sources of error in the development of the above confidence interval: the accuracy of the central limit theorem, and the statistical error in $\hat{\sigma}_n$. We will not explore these errors in these notes.

We have the following algorithm for producing a confidence interval for an expectation when the number of realizations, n, is fixed.

Algorithm for producing a confidence interval when n is fixed.

1. Select n, the number of experiments to be run, and $\delta > 0$.

2. Perform n independent replications of the experiment, obtaining the observations $X_1, X_2, \ldots, X_n$ of the random variable X.

3. Compute the sample mean and sample variance

$$\hat{\mu}_n = \frac{1}{n}(X_1 + \cdots + X_n), \quad \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2.$$

4. Select z such that $\Phi(z) = 1-\delta/2$. Then an approximate $(1-\delta)100\%$ confidence interval for $\mu = \mathbb{E}[X]$ is

$$\left[\hat{\mu}_n - \frac{\hat{\sigma}_n z}{\sqrt{n}}, \quad \hat{\mu}_n + \frac{\hat{\sigma}_n z}{\sqrt{n}}\right].$$

If a level of precision is desired, and n is allowed to depend upon δ and a tolerance ϵ, then the following algorithm is most useful.

Algorithm for producing confidence intervals to a given tolerance.

1. Select $\delta > 0$, determining the desired confidence, and $\epsilon > 0$ giving the desired precision. Select z so that $\Phi(z) = 1 - \delta/2$.

2. Perform independent replications of the experiment, obtaining the observations $X_1, X_2, \ldots, X_n$ of the random variable X, until

$$\frac{\hat{\sigma}_n z}{\sqrt{n}} < \epsilon.$$

3. Report

$$\hat{\mu}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

and the $(1-\delta)100\%$ confidence interval for $\hat{\mu} = \mathbb{E}[X]$,

$$\left[\hat{\mu}_n - \frac{\hat{\sigma}_n z}{\sqrt{n}}, \quad \hat{\mu}_n + \frac{\hat{\sigma}_n z}{\sqrt{n}}\right] \approx [\mu - \epsilon, \mu + \epsilon]\,.$$

There is normally a minimal number of samples generated, n_0 say, before one checks whether or not $\hat{\sigma}_n z/\sqrt{n} < \epsilon$. This is to avoid the possibility of an artificially low $\hat{\sigma}_n$ being produced from only a few samples.

B.4 Exercises

Some of the problems require you to use MATLAB. You must always turn in your MATLAB code when a problem is solved using MATLAB.

Exercise B.1. Verify the memoryless property, Equation (B.1), for exponential random variables.

Exercise B.2. MATLAB exercise. Perform the following tasks using Matlab. Report your findings.

1. Using a FOR LOOP, use the etime command to time how long it takes Matlab to generate 100,000 exponential random variables with a parameter of 1/10 using the built-in exponential random number generator.

2. Again using a FOR LOOP, use the etime command to time how long it takes Matlab to generate 100,000 exponential random variables with parameter 1/10 using the transformation method given in Theorem B.5.

Exercise B.3. MATLAB exercise. Let X be a random variable taking values on $\{-10,\ 0,\ 1,\ 4,\ 12\}$ with probability mass function

$$P(X=-10) = \tfrac{1}{5}, \quad P(X=0) = \tfrac{1}{8}, \quad P(X=1) = \tfrac{1}{4},$$
$$P(X=4) = \tfrac{1}{3}, \quad P(X=12) = \tfrac{11}{120}.$$

Using Theorem B.7, generate N independent copies of X and use them to estimate $\mathbb{E}[X]$ via

$$\mathbb{E}[X] \approx \frac{1}{N}\sum_{i=1}^{N} X_{[i]},$$

where $X_{[i]}$ is the ith independent copy of X and $N \in \{100, 10^3, 10^4, 10^5\}$. Compare the result for each N to the actual expected value. A helpful sample MATLAB code has been provided on the course website.

Bibliography

[1] Linda Allen, *Stochastic Processes with Applications to Biology*, Pearson, New Jersey, 2003.

[2] David F. Anderson, *Incorporating postleap checks in tau-leaping*, J. Chem. Phys. **128** (2008), no. 5, 054103.

[3] David F. Anderson and Thomas G. Kurtz, *Stochastic analysis of biochemical systems*, Springer, 2015.

[4] David F. Anderson, Timo Seppäläinen, and Benedek Valko, *Introduction to probability lecture notes*, UW-Madison, 2015.

[5] Karen Ball, Thomas G. Kurtz, Lea Popovic, and Greg Rempala, *Asymptotic analysis of multiscale approximations to reaction networks*, Ann. Appl. Prob. **16** (2006), no. 4, 1925–1961.

[6] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold, *Avoiding negative populations in explicit poisson tau-leaping*, J. Chem. Phys. **123** (2005), 054104.

[7] ———, *Efficient step size selection for the tau-leaping simulation method*, J. Chem. Phys. **124** (2006), 044109.

[8] Abhijit Chatterjee and Dionisios G. Vlachos, *Binomial distribution based τ-leap accelerated stochastic simulation*, J. Chem. Phys. **122** (2005), 024112.

[9] Rick Durrett, *Probability: Theory and examples*, 4th ed., Cambridge University Press, 2010.

[10] ———, *Essentials of stochastic processes*, 2 ed., Springer-Verlag New York, 2012.

[11] C. Henry Edwards and David E. Penney, *Differential equations and linear algebra*, 3rd ed., Prentice Hall, 2008.

[12] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, New York, 1986.

[13] William Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, John Wiley & Sons, 1968.

[14] Chetan Gadgil, Chang Hyeong Lee, and Hans G. Othmer, *A stochastic analysis of first-order reaction networks*, Bull. Math. Bio. **67** (2005), 901–946.

[15] M.A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A **105** (2000), 1876–1889.

[16] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys. **22** (1976), 403–434.

[17] _____, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. **81** (1977), no. 25, 2340–2361.

[18] _____, *Approximate accelerated simulation of chemically reaction systems*, J. Chem. Phys. **115** (2001), no. 4, 1716–1733.

[19] D. T. Gillespie and Linda R. Petzold, *Improved leap-size selection for accelerated stochastic simulation*, J. Chem. Phys. **119** (2003), no. 16, 8229–8234.

[20] Eric L. Haseltine and James B. Rawlings, *Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics*, J. Chem. Phys. **117** (2002), no. 15, 6959–6969.

[21] Motoo Kimura, *A simple method for estimating evolutionary rates of base substitutions through comparative studeis of nucleotide sequences*, J. Mol. Evol. **16** (1980), 111–120.

[22] Thomas G. Kurtz, *Limit theorems for sequences of jump Markov processes approximating ordinary differential processes*, J. Appl. Probability **8** (1971), 344–356. MR MR0287609 (44 #4812)

[23] Thomas G. Kurtz, *The relationship between stochastic and deterministic models for chemical reactions*, J. Chem. Phys. **57** (1972), no. 7, 2976–2978.

[24] _____, *Strong approximation theorems for density dependent Markov chains*, Stoch. Proc. Appl. **6** (1977/78), 223–240.

[25] Thomas G. Kurtz, *Strong approximation theorems for density dependent Markov chains*, Stochastic Processes Appl. **6** (1977/78), no. 3, 223–240. MR 57 #4344

[26] Thomas G. Kurtz, *Representations of Markov processes as multparameter time changes*, Ann. Prob. **8** (1980), no. 4, 682–715.

[27] _____, *Approximation of population processes*, CBMS-NSF Reg. Conf. Series in Appl. Math.: 36, SIAM, 1981.

[28] Gregory F. Lawler, *Introduction to Stochastic Processes*, 2 ed., Chapman & hall, Baco Raton, FL, 2006.

[29] J. R. Norris, *Markov chains*, Cambridge University Press, 1997.

[30] Bernt Øksendal, *Stochastic differential equations: An introduction with applications*, Springer, 2003.

[31] Sidney I. Resnick, *Adventures in Stochastic Processes*, 1st ed., Birkhäuser, 1992.

[32] Sheldon Ross, *A First Course in Probability*, 8th ed., Pearson, 2008.

[33] R. Srivastava, L. You, J. Summers, and J. Yin, *Stochastic vs. deterministic modeling of intracellular viral kinetics*, J. Theoret. Biol. **218** (2002), no. 3, 309–321.

[34] T. Tian and K. Burrage, *Binomial leap methods for simulating stochastic chemical kinetics*, J. Chem. Phys. **121** (2004), 10356.

[35] N. G. van Kampen, *A power series expansion of the master equation*, Canad. J. Phys. **39** (1961), 551–567. MR MR0128921 (23 #B1958)

[36] D. J. Wilkinson, *Stochastic modelling for systems biology*, Chapman and Hall/CRC Press, 2006.

Index

n-step transition probability, 12

absorbing state, 19
accessible state, 17
alleles, 61
aperiodic, 20
axonal transport, 8

branching process, 75, 76
Brownian motion, 184

cadlag, 100, 123
Chapman-Kolmogorov equations, 13
chemical master equation, 133, 168
chemical reaction network, 159
class property, 21
closed set, 19
communication class, 18
    continuous time Markov chain, 141
confidence interval, 204
continuous time Markov chain, 123, 126
counting process, 100
criticality parameter for branching process, 82

deterministically monotone Markov chain, 6
diploid, 61
discrete time Markov chain, 2
doubly stochastic matrix, 5
Dynkin's formula, 149

elementary renewal theorem, 95
embedded discrete time Markov chain, 127
estimator
    Monte Carlo, 204
Euler's method, 174
explosion, 130
explosive process, 130
exponential random variable, 201

First order reaction networks, 175

gene, 61
generator for a Markov process, 150
generator matrix, 134
genotype, 62
Gillespie algorithm, 128
Gillespie's Algorithm, 172

homogeneous process, 4

infinitesimal generator, 134
intensity function, 103
invariant measure, 34
    continuous time Markov chain, 141
irreducible
    continuous time Markov chain, 141
irreducible DTMC, 19
Itô integral, 191
Itô isometry, 189

Jukes-Cantor distance, 67
Jukes-Cantor model, 66

kinetic Monte Carlo, 128
kinetics, 160
Kolmogorov forward equations, 133

Lebesgue measure, 98
Linear noise approximation, 194

Markov matrix, 5
Markov property
    continuous time, 123
    discrete time, 2
mass-action kinetics
    stochastic, 161
mean recurrence time, 45

measure, 98
Mendelian genetics, 61

negative feedback, 163
Next reaction method, 173
null recurrent state, 45

one-step transition probability, 4

period of a Markov chain, 20
phenotype, 62
phylogenetic distance, 66
    Jukes-Cantor, 67
point process, 99
Poisson process
    homogeneous, 103
    homogeneous, one-dimensional, 100
Poisson process with mean measure, 103
polymer, 58
positive feedback, 165
positive recurrence
    for a continuous time Markov chain, 144
positive recurrent
    continuous time Markov chain, 141
positive recurrent state, 45
probability generating function, 78
probability generating functions, 78
product complex, 159
propensity function, 103

queuing models, 69

random measure, 99
random time change, 117, 169
random walk, 7, 8
rate constant, 161
reaction vector, 159
recurrent state, 22
    continuous time Markov chain, 141
reducible DTMC, 19
renewal function, 90
renewal process, 89
    defective, 90
    delayed, 89
    proper, 90
    pure, 89
renewal reward process, 95
renewal times, 89

simulation
    continuous time Markov chain, 127
    discrete time Markov chain, 10
source complex, 159
stationary distribution, 34
    continuous time Markov chain, 141
stochastic matrix, 5
stochastic process, 1
Stochastic simulation algorithm, 172
stochastic simulation algorithm, 128
stoichiometrically admissible kinetics, 160
stopping time, 9
strong Markov property, 9

tau-leaping, 174
Time dependent intensity functions, 174
time homogeneous process, 4
transcription, 161
transient state, 22
    continuous time Markov chain, 141
transition matrix, 5
translation, 161

virus model, 165

weakly reversible network, 176
Wiener process, 184
Wright-Fisher model, 64