Essentials of Stochastic Processes

Rick Durrett

Version Beta of the 2nd Edition
August 21, 2010

Contents

Chapter 1

Markov Chains

1.1 Definitions and Examples

The importance of Markov chains comes from two facts: (i) there are a large number of physical, biological, economic, and social phenomena that can be described in this way, and (ii) there is a well-developed theory that allows us to do computations. We begin with a famous example, then describe the property that is the defining feature of Markov chains

Example 1.1. Gambler's ruin. Consider a gambling game in which on any turn you win \$1 with probability $p = 0.4$ or lose \$1 with probability $1 - p = 0.6$. Suppose further that you adopt the rule that you quit playing if your fortune reaches $\$N$. Of course, if your fortune reaches \$0 the casino makes you stop.

Let X_n be the amount of money you have after n plays. I claim that your fortune, X_n has the "Markov property." In words, this means that given the current state, any other information about the past is irrelevant for predicting the next state X_{n+1}. To check this for the gambler's ruin chain, we note that if you are still playing at time n, i.e., your fortune $X_n = i$ with $0 < i < N$, then for any possible history of your wealth $i_{n-1}, i_{n-2}, \ldots i_1, i_0$

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \ldots X_0 = i_0) = 0.4$$

since to increase your wealth by one unit you have to win your next bet. Here we have used $P(B|A)$ for the conditional probability of the event B given that A occurs. Recall that this is defined by

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Turning now to the formal definition,

Definition 1.1. *We say that X_n is a discrete time* **Markov chain** *with* **transition matrix** $p(i,j)$ *if for any $j, i, i_{n-1}, \dots i_0$*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p(i,j) \tag{1.1}$$

This equation, also called the "Markov property" says that the conditional probability $X_{n+1} = j$ given the entire history $X_n = i, X_{n-1} = i_{n-1}, \dots X_1 = i_1, X_0 = i_0$ is the same as the conditional probability $X_{n+1} = j$ given only the previous state $X_n = i$. This is what we mean when we say that "any other information about the past is irrelevant for predicting X_{n+1}."

In formulating (1.1) we have restricted our attention to the **temporally homogeneous** case in which the **transition probability**

$$p(i,j) = P(X_{n+1} = j | X_n = i)$$

does not depend on the time n. Intuitively, the transition probability gives the rules of the game. It is the basic information needed to describe a Markov chain. In the case of the gambler's ruin chain, the transition probability has

$$\begin{aligned} &p(i, i+1) = 0.4, \quad p(i, i-1) = 0.6, \quad \text{if } 0 < i < N \\ &p(0,0) = 1 \qquad p(N,N) = 1 \end{aligned}$$

When $N = 5$ the matrix is

	0	**1**	**2**	**3**	**4**	**5**
0	1.0	0	0	0	0	0
1	0.6	0	0.4	0	0	0
2	0	0.6	0	0.4	0	0
3	0	0	0.6	0	0.4	0
4	0	0	0	0.6	0	0.4
5	0	0	0	0	0	1.0

or the chain by be represented pictorially as

$$\underset{1}{\hookrightarrow} \mathbf{0} \overset{.4}{\leftarrow} \mathbf{1} \underset{.6}{\overset{.4}{\leftrightarrows}} \mathbf{2} \underset{.6}{\overset{.4}{\leftrightarrows}} \mathbf{3} \underset{.6}{\overset{.4}{\leftrightarrows}} \mathbf{4} \underset{.6}{\rightarrow} \mathbf{5} \overset{1}{\hookleftarrow}$$

Example 1.2. Ehrenfest chain. This chain originated in physics as a model for two cubical volumes of air connected by a small hole. In the mathematical version, we have two "urns," i.e., two of the exalted trash cans of probability theory, in which there are a total of N balls. We pick one of the N balls at random and move it to the other urn.

Let X_n be the number of balls in the "left" urn after the nth draw. It should be clear that X_n has the Markov property; i.e., if we want to guess the state at time $n+1$, then the current number of balls in the left urn X_n, is the only relevant information from the observed sequence of states $X_n, X_{n-1}, \dots X_1, X_0$. To check this we note that

$$P(X_{n+1} = i+1 | X_n = i, X_{n-1} = i_{n-1}, \dots X_0 = i_0) = (N-i)/N$$

since to increase the number we have to pick one of the $N-i$ balls in the other urn. The number can also decrease by 1 with probability i/N. In symbols, we have computed that the transition probability is given by

$$p(i, i+1) = (N-i)/N, \quad p(i, i-1) = i/N \qquad \text{for } 0 \le i \le N$$

with $p(i,j) = 0$ otherwise. When $N = 4$, for example, the matrix is

	0	**1**	**2**	**3**	**4**
0	0	1	0	0	0
1	1/4	0	3/4	0	0
2	0	2/4	0	2/4	0
3	0	0	3/4	0	1/4
4	0	0	0	1	0

In the first two examples we began with a verbal description and then wrote down the transition probabilities. However, one more commonly describes a k state Markov chain by writing down a transition probability $p(i,j)$ with

(i) $p(i,j) \geq 0$, since they are probabilities.

(ii) $\sum_j p(i,j) = 1$, since when $X_n = i$, X_{n+1} will be in some state j.

The equation in (ii) is read "sum $p(i,j)$ over all possible values of j." In words the last two conditions say: the entries of the matrix are nonnegative and each ROW of the matrix sums to 1.

Any matrix with properties (i) and (ii) gives rise to a Markov chain, X_n. To construct the chain we can think of playing a board game. When we are in state i, we roll a die (or generate a random number on a computer) to pick the next state, going to j with probability $p(i,j)$.

Example 1.3. Weather chain. Let X_n be the weather on day n in Ithaca, NY, which we assume is either: $1 =$ *rainy*, or $2 =$ *sunny*. Even though the weather is not exactly a Markov chain, we can propose a Markov chain model for the weather by writing down a transition probability

	1	**2**
1	.6	.4
2	.2	.8

Q. What is the long-run fraction of days that are sunny?

The table says, for example, the probability a rainy day (state 1) is followed by a sunny day (state 2) is $p(1,2) = 0.4$.

Example 1.4. Social mobility. Let X_n be a family's social class in the nth generation, which we assume is either $1 =$ *lower*, $2 =$ *middle*, or $3 =$ *upper*. In our simple version of sociology, changes of status are a Markov chain with the following transition probability

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

Q. Do the fractions of people in the three classes approach a limit?

Example 1.5. Brand preference. Suppose there are three types of laundry detergent, 1, 2, and 3, and let X_n be the brand chosen on the nth purchase. Customers who try these brands are satisfied and choose the same thing again with probabilities 0.8, 0.6, and

0.4 respectively. When they change they pick one of the other two brands at random. The transition probability is

	1	**2**	**3**
1	.8	.1	.1
2	.2	.6	.2
3	.3	.3	.4

Q. Do the market shares of the three product stabilize?

Example 1.6. Inventory chain. We will consider the consequences of using an s, S inventory control policy. That is, when the stock on hand at the end of the day falls to s or below we order enough to bring it back up to S which for simplicity we suppose happens at the beginning of the next day. Let D_{n+1} be the demand on day $n+1$. Introducing notation for the **positive part** of a real number,

$$x^+ = \max\{x, 0\} = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

then we can write the chain in general as

$$X_{n+1} = \begin{cases} (X_n - D_{n+1})^+ & \text{if } X_n > s \\ (S - D_{n+1})^+ & \text{if } X_n \le s \end{cases}$$

In words, if $X_n > s$ we order nothing, begin the day with X_n units. If the demand $D_{n+1} \le X_n$ we end the day with $X_{n+1} = X_n - D_{n+1}$. If the demand $D_{n+1} > X_n$ we end the day with $X_{n+1} = 0$. If $X_n \le s$ then we begin the day with S units, and the reasoning is the same as in the previous case.

Suppose now that an electronics store sells a video game system and uses an invenotry policy with $s = 1$, $S = 5$. That is, if at the end of the day, the number of units they have on hand is 1 or 0, they order enough new units so their total on hand at the beginning of the next day is 5. If we assume that

$k =$	0	1	2	3
$P(D_{n+1} = k)$	.3	.4	.2	.1

then we have the following transition matrix:

	0	**1**	**2**	**3**	**4**	**5**
0	0	0	.1	.2	.4	.3
1	0	0	.1	.2	.4	.3
2	.3	.4	.3	0	0	0
3	.1	.2	.4	.3	0	0
4	0	.1	.2	.4	.3	0
5	0	0	.1	.2	.4	.3

To explain the entries, we note that when $X_n \geq 3$ then $X_n - D_{n+1} \geq 0$. When $X_{n+1} = 2$ this is almost true but $p(2,0) = P(D_{n+1} = 2$ or $3)$. When $X_n = 1$ or 0 we start the day with 5 units so the end result is the same as when $X_n = 5$.

In this context we might be interested in:

Q. Suppose we make \$12 profit on each unit sold but it costs \$2 a day to store items. What is the long-run profit per day of this inventory policy? How do we choose s and S to maximize profit?

Example 1.7. Repair chain. A machine has three critical parts that are subject to failure, but can function as long as two of these parts are working. When two are broken, they are replaced and the machine is back to working order the next day. To formulate a Markov chain model we declare its state space to be the parts that are broken $\{0, 1, 2, 3, 12, 13, 23\}$. If we assume that parts 1, 2, and 3 fail with probabilities .01, .02, and .04, but no two parts fail on the same day, then we arrive at the following transition matrix:

	0	**1**	**2**	**3**	**12**	**13**	**23**
0	.93	.01	.02	.04	0	0	0
1	0	.94	0	0	.02	.04	0
2	0	0	.95	0	.01	0	.04
3	0	0	0	.97	0	.01	.02
12	1	0	0	0	0	0	0
13	1	0	0	0	0	0	0
23	1	0	0	0	0	0	0

If we own a machine like this, then it is natural to ask about the long-run cost per day to operate it. For example, we might ask:

Q. If we are going to operate the machine for 1800 days (about 5 years), then how many parts of types 1, 2, and 3 will we use?

Example 1.8. Branching processes. These processes arose from Francis Galton's statistical investigation of the extinction of family names. Consider a population in which each individual in the nth generation independently gives birth, producing k children (who are members of generation $n+1$) with probability p_k. In Galton's application only male children count since only they carry on the family name.

To define the Markov chain, note that the number of individuals in generation n, X_n, can be any nonnegative integer, so the state space is $\{0, 1, 2, \ldots\}$. If we let $Y_1, Y_2, \ldots$ be independent random variables with $P(Y_m = k) = p_k$, then we can write the transition probability as

$$p(i,j) = P(Y_1 + \cdots + Y_i = j) \quad \text{for } i > 0 \text{ and } j \geq 0$$

When there are no living members of the population, no new ones can be born, so $p(0,0) = 1$.

Galton's question, originally posed in the Educational Times of 1873, is

Q. What is the probability line of a man becomes extinct?, i.e., the process becomes absorbed at 0?

Reverend Henry William Watson replied with a solution. Together, they then wrote an 1874 paper entitled *On the probability of extinction of families*. For this reason, these chains are often called Galton-Watson processes.

Example 1.9. Wright–Fisher model. Thinking of a population of $N/2$ diploid individuals who have two copies of each of their chromosomes, or of N haploid individuals who have one copy, we consider a fixed population of N genes that can be one of two types: A or a. In the simplest version of this model the population at time $n+1$ is obtained by drawing with replacement from the population at time n. In this case if we let X_n be the number of A alleles at time n, then X_n is a Markov chain with transition probability

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

since the right-hand side is the binomial distribution for N independent trials with success probability i/N.

In this model the states 0 and N that correspond to fixation of the population in the all a or all A states are absorbing states, so it is natural to ask:

Q1. Starting from i of the A alleles and $N-i$ of the a alleles, what is the probability that the population fixates in the all A state?

To make this simple model more realistic we can introduce the possibility of mutations: an A that is drawn ends up being an a in the next generation with probability u, while an a that is drawn ends up being an A in the next generation with probability v. In this case the probability an A is produced by a given draw is

$$\rho_i = \frac{i}{N}(1-u) + \frac{N-i}{N}v$$

but the transition probability still has the binomial form

$$p(i,j) = \binom{N}{j}(\rho_i)^j(1-\rho_i)^{N-j}$$

If u and v are both positive, then 0 and N are no longer absorbing states, so we ask:

Q2. Does the genetic composition settle down to an equilibrium distribution as time $t \to \infty$?

As the next example shows it is easy to extend the notion of a Markov chain to cover situations oin which the future evolution is independent of the past when we know the last two states.

Example 1.10. Two-stage Markov chains. In a Markov chain the distribution of X_{n+1} only depends on X_n. This can easily be generalized to case in which the distribution of X_{n+1} only depends on (X_n, X_{n-1}). For a concrete example consider a basketball player who makes a shot with the following probabilities:

1/2 if he has missed the last two times
2/3 if he has hit one of his last two shots
3/4 if he has hit both of his last two shots

To formulate a Markov chain to model his shooting, we let the states of the process be the outcomes of his last two shots: $\{HH, HM, MH, MM\}$

where M is short for miss and H for hit. The transition probability is

	HH	**HM**	**MH**	**MM**
HH	3/4	1/4	0	0
HM	0	0	2/3	1/3
MH	2/3	1/3	0	0
MM	0	0	1/2	1/2

To explain suppose the state is HM, i.e., $X_{n-1} = H$ and $X_n = M$. In this case the next outcome will be H with probability 2/3. When this occurs the next state will be $(X_n, X_{n+1}) = (M, H)$ with probability 2/3. If he misses an event of probability 1/3, $(X_n, X_{n+1}) = (M, M)$.

The Hot Hand is a phenomenon known to everyone who plays or watches basketball. After making a couple of shots, players are thought to "get into a groove" so that subsequent successes are more likely. Purvis Short of the Golden State Warriors describes this more poetically as

> "You're in a world all your own. It's hard to describe. But the basket seems to be so wide. No matter what you do, you know the ball is going to go in."

Unfortunately for basketball players, data collected by Tversky and Gliovich (*Chance* vol. 2 (1989), No. 1, pages 16–21) shows that this is a misconception. The next table gives data for the conditional probability of hitting a shot after missing the last three, missing the last two, ... hitting the last three, for nine players of the Philadelphia 76ers: Darryl Dawkins (403), Maurice Cheeks (339), Steve Mix (351), Bobby Jones (433), Clint Richardson (248), Julius Erving (884), Andrew Toney (451), Caldwell Jones (272), and Lionel Hollins (419). The numbers in parentheses are the number of shots for each player.

$P(H\|3M)$	$P(H\|2M)$	$P(H\|1M)$	$P(H\|1H)$	$P(H\|2H)$	$P(H\|3H)$
.88	.73	.71	.57	.58	.51
.77	.60	.60	.55	.54	.59
.70	.56	.52	.51	.48	.36
.61	.58	.58	.53	.47	.53
.52	.51	.51	.53	.52	.48
.50	.47	.56	.49	.50	.48
.50	.48	.47	.45	.43	.27
.52	.53	.51	.43	.40	.34
.50	.49	.46	.46	.46	.32

In fact, the data supports the opposite assertion: after missing a player is more conservative about the shots that they take and will hit more frequently.

1.2 Multistep Transition Probabilities

The transition probability $p(i,j) = P(X_{n+1} = j | X_n = i)$ gives the probability of going from i to j in one step. Our goal in this section is to compute the probability of going from i to j in $m > 1$ steps:

$$p^m(i,j) = P(X_{n+m} = j | X_n = i)$$

As the notation may already suggest, p^m will turn out to the be the mth power of the transition matrix, see Theorem 1.1.

To warm up we recall the transition probability of the social mobility chain:

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

and consider the following concrete question:

Q1. *Your parents were middle class (state 2). What is the probability that you are in the upper class (state 3) but your children are lower class (state 1)?*

Solution. Intuitively, the Markov property implies that starting from state 2 the probability of jumping to 3 and then to 1 is given by

$$p(2,3)p(3,1)$$

To get this conclusion from the definitions, we note that using the definition of conditional probability,

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

By the Markov property (1.1) the last expression is

$$P(X_2 = 1 | X_1 = 3) \cdot P(X_1 = 3 | X_0 = 2) = p(2,3)p(3,1)$$

Moving on to the real question:

Q2. *What is the probability your children are lower class (1) given your parents were middle class (2)?*

Solution. To do this we simply have to consider the three possible states for Wednesday and use the solution of the previous problem.

$$\begin{aligned}
P(X_2 = 1|X_0 = 2) &= \sum_{k=1}^{3} P(X_2 = 1, X_1 = k|X_0 = 2) = \sum_{k=1}^{3} p(2,k)p(k,1) \\
&= (.3)(.7) + (.5)(.3) + (.2)(.2) = .21 + .15 + .04 = .21
\end{aligned}$$

There is nothing special here about the states 2 and 1 here. By the same reasoning,

$$P(X_2 = j|X_0 = i) = \sum_{k=1}^{3} p(i,k)\, p(k,j)$$

The right-hand side of the last equation gives the (i,j)th entry of the matrix p is multiplied by itself.

To explain this, we note that to compute $p^2(2,1)$ we multiplied the entries of the second row by those in the first column:

$$\begin{pmatrix} . & . & . \\ .3 & .5 & .2 \\ . & . & . \end{pmatrix} \begin{pmatrix} .7 & . & . \\ .3 & . & . \\ .2 & . & . \end{pmatrix} = \begin{pmatrix} . & . & . \\ .40 & . & . \\ . & . & . \end{pmatrix}$$

If we wanted $p^2(1,3)$ we would multiply the first row by the third column:

$$\begin{pmatrix} .7 & .2 & .1 \\ . & . & . \\ . & . & . \end{pmatrix} \begin{pmatrix} . & . & .1 \\ . & . & .2 \\ . & . & .4 \end{pmatrix} = \begin{pmatrix} . & . & .15 \\ . & . & . \\ . & . & . \end{pmatrix}$$

When all of the computations are done we have

$$\begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} \begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} = \begin{pmatrix} .57 & .28 & .15 \\ .40 & .39 & .21 \\ .34 & .40 & .26 \end{pmatrix}$$

All of this becomes much easier if we use a scientific calculator like the T1-83. Using 2nd-MATRIX we can access a screen with NAMES, MATH, EDIT at the top. Selecting EDIT we can enter the matrix into the computer as say [A]. The selecting the NAMES we can enter $[A] \wedge 2$ on the computation line to get A^2. If we use this procedure to compute A^{20} we get a matrix with three rows that agree in the first six decimal places with

$$.468085 \quad .340425 \quad .191489$$

Later we will see that A^n converges to a matrix with all three rows equal to $(22/47, 16/47, 9/47)$.

To explain our interest in A^m we will now prove:

Theorem 1.1. *The m step transition probability $P(X_{n+m} = j|X_n = i)$ is the mth power of the transition matrix p.*

The key ingredient in proving this is the **Chapman–Kolmogorov equation**

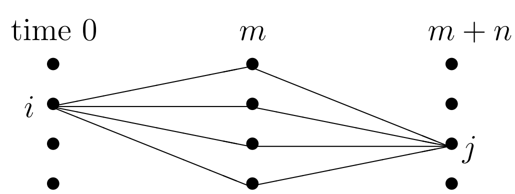
$$p^{m+n}(i,j) = \sum_k p^m(i,k)\, p^n(k,j) \tag{1.2}$$

Once this is proved, Theorem 1.1 follows, since taking $n = 1$ in (4.1), we see that

$$p^{m+1}(i,j) = \sum_k p^m(i,k)\, p(k,j)$$

That is, the $m+1$ step transition probability is the m step transition probability times p. Theorem 1.1 now follows.

Why is (4.1) true? To go from i to j in $m+n$ steps, we have to go from i to some state k in m steps and then from k to j in n steps. The Markov property implies that the two parts of our journey are independent. □



Proof of (4.1). We do this by combining the solutions of Q1 and Q2. Breaking things down according to the state at time m,

$$P(X_{m+n} = j|X_0 = i) = \sum_k P(X_{m+n} = j, X_m = k|X_0 = i)$$

Using the definition of conditional probability as in the solution of Q1,

$$
\begin{aligned}
P(X_{m+n}=j, X_m=k|X_0=i) &= \frac{P(X_{m+n}=j, X_m=k, X_0=i)}{P(X_0=i)} \\
&= \frac{P(X_{m+n}=j, X_m=k, X_0=i)}{P(X_m=k, X_0=i)} \cdot \frac{P(X_m=k, X_0=i)}{P(X_0=i)} \\
&= P(X_{m+n}=j|X_m=k, X_0=i) \cdot P(X_m=k|X_0=i)
\end{aligned}
$$

By the Markov property (1.1) the last expression is

$$= P(X_{m+n}=j|X_m=k) \cdot P(X_m=k|X_0=i) = p^m(i,k)p^n(k,j)$$

and we have proved (4.1). □

Having established (4.1), we now return to computations.

Example 1.11. Gambler's ruin. Suppose for simplicity that $N = 4$ in Example 1.1, so that the transition probability is

	0	**1**	**2**	**3**	**4**
0	1.0	0	0	0	0
1	0.6	0	0.4	0	0
2	0	0.6	0	0.4	0
3	0	0	0.6	0	0.4
4	0	0	0	0	1.0

To compute p^2 one row at a time we note:

$p^2(0,0) = 1$ and $p^2(4,4) = 1$, since these are absorbing states.

$p^2(1,3) = (.4)^2 = 0.16$, since the chain has to go up twice.

$p^2(1,1) = (.4)(.6) = 0.24$. The chain must go from 1 to 2 to 1.

$p^2(1,0) = 0.6$. To be at 0 at time 2, the first jump must be to 0.

Leaving the cases $i = 2, 3$ to the reader, we have

$$p^2 = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ .6 & .24 & 0 & .16 & 0 \\ .36 & 0 & .48 & 0 & .16 \\ 0 & .36 & 0 & .24 & .4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Using a calculator one can easily compute

$$p^{20} = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ .87655 & .00032 & 0 & .00022 & .12291 \\ .69186 & 0 & .00065 & 0 & .30749 \\ .41842 & .00049 & 0 & .00032 & .58437 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

0 and 4 are absorbing states. Here we see that the probability of avoing absorption for 20 steps is 0.00054 from state 3, 0.00065 from state 2, and 0.00081 from state 1. Later we will see that

$$\lim_{n\to\infty} p^n = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 57/65 & 0 & 0 & 0 & 8/65 \\ 45/65 & 0 & 0 & 0 & 20/65 \\ 27/65 & 0 & 0 & 0 & 38/65 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

1.3 Classification of States

We begin with some important notation. We are often interested in the behavior of the chain for a fixed initial state, so we will introduce the shorthand

$$P_x(A) = P(A|X_0 = x)$$

Later we will have to consider expected values for this probability and we will denote them by E_x.

Let $T_y = \min\{n \geq 1 : X_n = y\}$ be the time of the **first return** to y (i.e., being there at time 0 doesn't count), and let

$$\rho_{yy} = P_y(T_y < \infty)$$

be the probability X_n returns to y when it starts at y. Intuitively, the Markov property implies that the probability X_n will return at least twice to y is ρ_{yy}^2, since after the first return, the chain is at y, and the Markov property implies that the probability of a second return following the first is again ρ_{yy}.

To show that the reasoning in the last paragraph is valid, we have to introduce a definition and state a theorem.

Definition 1.2. *We say that T is a* **stopping time** *if the occurrence (or nonoccurrence) of the event "we stop at time n," $\{T = n\}$ can be determined by looking at the values of the process up to that time: $X_0, \ldots, X_n$.*

To see that T_y is a stopping time note that

$$\{T_y = n\} = \{X_1 \neq y, \ldots, X_{n-1} \neq y, X_n = y\}$$

and that the right-hand side can be determined from $X_0, \ldots, X_n$.

Since stopping at time n depends only on the values $X_0, \ldots, X_n$, and in a Markov chain the distribution of the future only depends on the past through the current state, it should not be hard to believe that the Markov property holds at stopping times. This fact can be stated formally as:

Theorem 1.2. Strong Markov property. *Suppose T is a stopping time. Given that $T = n$ and $X_T = y$, any other information about $X_0, \ldots X_T$ is irrelevant for predicting the future, and X_{T+k}, $k \geq 0$ behaves like the Markov chain with initial state y.*

Why is this true? To keep things as simple as possible we will show only that

$$P(X_{T+1} = z | X_T = y, T = n) = p(y, z)$$

Let V_n be the set of vectors $(x_0, \ldots, x_n)$ so that if $X_0 = x_0, \ldots, X_n = x_n$, then $T = n$ and $X_T = y$. Breaking things down according to the values of $X_0, \ldots, X_n$ gives

$$P(X_{T+1} = z, X_T = y, T = n) = \sum_{x \in V_n} P(X_{n+1} = z, X_n = x_n, \ldots, X_0 = x_0)$$
$$= \sum_{x \in V_n} P(X_{n+1} = z | X_n = x_n, \ldots, X_0 = x_0) P(X_n = x_n, \ldots, X_0 = x_0)$$

where in the second step we have used the multiplication rule

$$P(A \cap B) = P(B|A)P(A)$$

For any $(x_0, \ldots, x_n) \in A$ we have $T = n$ and $X_T = y$ so $x_n = y$. Using the Markov property, (1.1), and recalling the definition of V_n shows the above

$$P(X_{T+1} = z, T = n, X_T = y) = p(y, z) \sum_{x \in V_n} P(X_n = x_n, \ldots, X_0 = x_0)$$
$$= p(y, z) P(T = n, X_T = y)$$

Dividing both sides by $P(T = n, X_T = y)$ gives the desired result. □

Definition 1.3. *Let $T_y^1 = T_y$ and for $k \geq 2$ let*

$$T_y^k = \min\{n > T_y^{k-1} : X_n = y\} \tag{1.3}$$

be the time of the kth return to y.

The strong Markov property implies that the conditional probability we will return one more time given that we have returned $k - 1$ times is ρ_{yy}. This and induction implies that

$$P_y(T_y^k < \infty) = \rho_{yy}^k \tag{1.4}$$

At this point, there are two possibilities:

(i) $\rho_{yy} < 1$: The probability of returning k times is $\rho_{yy}^k \to 0$ as $k \to \infty$. Thus, eventually the Markov chain does not find its way back to y. In this case the state y is called **transient**, since after some point it is never visited by the Markov chain.

(ii) $\rho_{yy} = 1$: The probability of returning k times $\rho_{yy}^n = 1$, so the chain returns to y infinitely many times. In this case, the state y is called **recurrent**, it continually recurs in the Markov chain.

To understand these notions, we turn to our examples, beginning with

Example 1.12. Gambler's ruin. Consider, for concreteness, the case $N = 4$.

	0	**1**	**2**	**3**	**4**
0	1	0	0	0	0
1	.6	0	.4	0	0
2	0	.6	0	.4	0
3	0	0	.6	0	.4
4	0	0	0	0	1

We will show that eventually the chain gets stuck in either the bankrupt (0) or happy winner (4) state. In the terms of our recent definitions, we will show that states $0 < y < 4$ are transient, while the states 0 and 4 are recurrent.

It is easy to check that 0 and 4 are recurrent. Since $p(0,0) = 1$, the chain comes back on the next step with probability one, i.e.,

$$P_0(T_0 = 1) = 1$$

and hence $\rho_{00} = 1$. A similar argument shows that 4 is recurrent. In general if y is an **absorbing state**, i.e., if $p(y,y) = 1$, then y is a very strongly recurrent state – the chain always stays there.

To check the transience of the interior states, $1, 2, 3$, we note that starting from 1, if the chain goes to 0, it will never return to 1, so the probability of never returning to 1,

$$P_1(T_1 = \infty) \geq p(1,0) = 0.6 > 0$$

Similarly, starting from 2, the chain can go to 1 and then to 0, so

$$P_2(T_2 = \infty) \geq p(2,1)p(1,0) = 0.36 > 0$$

Finally for starting from 3, we note that the chain can go immediately to 4 and never return with probability 0.4, so

$$P_3(T_3 = \infty) \geq p(3,4) = 0.4 > 0$$

Generalizing from our experience with the gambler's ruin chain, we come to a general result that will help us identify transient states.

Definition 1.4. *We say that x* **communicates with** *y and write $x \to y$ if there is a positive probability of reaching y starting from x, that is, the probability*

$$\rho_{xy} = P_x(T_y < \infty) > 0$$

Note that the last probability includes not only the possibility of jumping from x to y in one step but also going from x to y after visiting several other states in between. The following property is simple but useful. Here and in what follows, lemmas are a means to prove the more important conclusions called theorems. To make it easier to locate things, theorems and lemmas are numbered in the same sequence.

Lemma 1.1. *If $x \to y$ and $y \to z$, then $x \to z$.*

Proof. Since $x \to y$ there is an m so that $p^m(x, y) > 0$. Similarly there is an n so that $p^n(y, z) > 0$. Since $p^{m+n}(x, z) \geq p^m(x, y)p^n(y, z)$ it follows that $x \to z$. □

Theorem 1.3. *If $\rho_{xy} > 0$, but $\rho_{yx} < 1$, then x is transient.*

Proof. Let $K = \min\{k : p^k(x, y) > 0\}$ be the smallest number of steps we can take to get from x to y. Since $p^K(x, y) > 0$ there must be a sequence $y_1, \ldots y_{K-1}$ so that

$$p(x, y_1)p(y_1, y_2) \cdots p(y_{K-1}, y) > 0$$

Since K is minimal all the $y_i \neq y$ (or there would be a shorter path), and we have

$$P_x(T_x = \infty) \geq p(x, y_1)p(y_1, y_2) \cdots p(y_{K-1}, y)(1 - \rho_{yx}) > 0$$

so x is transient. □

We will see later that Theorem 1.3 allows us to to identify all the transient states when the state space is finite. An immediate consequence of Theorem 1.3 is

Lemma 1.2. *If x is recurrent and $\rho_{xy} > 0$ then $\rho_{yx} = 1$.*

Proof. If $\rho_{yx} < 1$ then Lemma 1.3 would imply x is transient. □

In some cases it is easy to identify recurrent states.

Example 1.13. Social mobility. Recall that the transition probability is

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

To begin we note that no matter where X_n is, there is a probability of at least .1 of hitting 3 on the next step so $P_3(T_3 > n) \leq (.9)^n$. As $n \to \infty$, $(.9)^n \to 0$ so $P_3(T_3 < \infty) = 1$, i.e., we will return to 3 with probability 1. The last argument applies even more strongly to states 1 and 2, since the probability of jumping to them on the next step is always at least .2. Thus all three states are recurrent.

The last argument generalizes to the give the following useful fact.

Lemma 1.3. *Suppose $P_x(T_y \leq k) \geq \alpha > 0$ for all x in the state space S. Then*

$$P_x(T_y > nk) \leq (1 - \alpha)^n$$

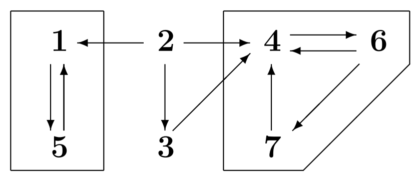
To be able to analyze any finite state Markov chain we need some theory. To motivate the developments consider

Example 1.14. A Seven-state chain. Consider the transition probability:

	1	**2**	**3**	**4**	**5**	**6**	**7**
1	.3	0	0	0	.7	0	0
2	.1	.2	.3	.4	0	0	0
3	0	0	.5	.5	0	0	0
4	0	0	0	.5	0	.5	0
5	.6	0	0	0	.4	0	0
6	0	0	0	0	0	.2	.8
7	0	0	0	1	0	0	0

To identify the states that are recurrent and those that are transient, we begin by drawing a graph that will contain an arc from i to j if $p(i, j) > 0$ and $i \neq j$. We do not worry about drawing the self-loops corresponding to states with $p(i, i) > 0$ since such transitions cannot help the chain get somewhere new.

In the case under consideration we draw arcs from $1 \to 5$, $2 \to 1$, $2 \to 3$, $2 \to 4$, $3 \to 4$, $4 \to 6$, $4 \to 7$, $5 \to 1$, $6 \to 4$, $6 \to 7$, $7 \to 4$.



The state 2 communicates with 1, which does not communicate with it, so Theorem 1.3 implies that 2 is transient. Likewise 3 communicates with 4, which doesn't communicate with it, so 3 is transient. To conclude that all the remaining states are recurrent we will introduce two definitions and a fact.

Definition 1.5. *A set A is* **closed** *if it is impossible to get out, i.e., if $i \in A$ and $j \notin A$ then $p(i,j) = 0$.*

In Example 1.14, $\{1,5\}$ and $\{4,6,7\}$ are closed sets. Their union, $\{1,4,5,6,7\}$ is also closed. One can add 3 to get another closed set $\{1,3,4,5,6,7\}$. Finally, the whole state space $\{1,2,3,4,5,6,7\}$ is always a closed set.

Among the closed sets in the last example, some are obviously too big. To rule them out, we need a definition.

Definition 1.6. *A set B is called* **irreducible** *if whenever $i, j \in B$, i communicates with j.*

The irreducible closed sets in the Example 1.14 are $\{1,5\}$ and $\{4,6,7\}$. The next result explains our interest in irreducible closed sets.

Theorem 1.4. *If C is a finite closed and irreducible set, then all states in C are recurrent.*

Before entering into an explanation of this result, we note that Theorem 1.4 tells us that 1, 5, 4, 6, and 7 are recurrent, completing our study of the Example 1.14 with the results we had claimed earlier.

In fact, the combination of Theorem 1.3 and 1.4 is sufficient to classify the states in any finite state Markov chain. An algorithm will be explained in the proof of the following result.

Theorem 1.5. *If the state space S is finite, then S can be written as a disjoint union $T \cup R_1 \cup \cdots \cup R_k$, where T is a set of transient states and the R_i, $1 \le i \le k$, are closed irreducible sets of recurrent states.*

Proof. Let T be the set of x for which there is a y so that $x \to y$ but $y \not\to x$. The states in T are transient by Theorem 1.3. Our next step is to show that all the remaining states, $S - T$, are recurrent.

Pick an $x \in S - T$ and let $C_x = \{y : x \to y\}$. Since $x \notin T$ it has the property if $x \to y$, then $y \to x$. To check that C_x is closed note that if $y \in C_x$ and $y \to z$, then Lemma 1.1 implies $x \to z$ so $z \in C_x$. To check irreducibility, note that if $y, z \in C_x$, then by our first observation $y \to x$ and we have $x \to z$ by definition, so Lemma 1.1 implies $y \to z$. C_x is closed and irreducible so all states in C_x are recurrent. Let $R_1 = C_x$. If $S - T - R_1 = \emptyset$, we are done. If not, pick a site $w \in S - T - R_1$ and repeat the procedure. □

* * * * * * *

The rest of this section is devoted to the proof of Theorem 1.4. To do this, it is enough to prove the following two results.

Lemma 1.4. *If x is recurrent and $x \to y$, then y is recurrent.*

Lemma 1.5. *In a finite closed set there has to be at least one recurrent state.*

To prove these results we need to introduce a little more theory. Recall the time of the kth visit to y defined by

$$T_y^k = \min\{n > T_y^{k-1} : X_n = y\}$$

and $\rho_{xy} = P_x(T_y < \infty)$ the probability we ever visit y at some time $n \ge 1$ when we start from x. Using the strong Markov property as in the proof of (1.4) gives

$$P_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}. \tag{1.5}$$

Let $N(y)$ be the number of visits to y at times $n \ge 1$. Using (1.5) we can compute $EN(y)$.

Lemma 1.6. $E_x N(y) = \rho_{xy}/(1 - \rho_{yy})$

Proof. Accept for the moment the fact that for any nonnegative integer valued random variable X, the expected value of X can be computed by

$$EX = \sum_{k=1}^{\infty} P(X \geq k) \tag{1.6}$$

We will prove this after we complete the proof of Lemma 1.6. Now the probability of returning at least k times, $\{N(y) \geq k\}$, is the same as the event that the kth return occurs, i.e., $\{T_y^k < \infty\}$, so using (1.5) we have

$$E_x N(y) = \sum_{k=1}^{\infty} P(N(y) \geq k) = \rho_{xy} \sum_{k=1}^{\infty} \rho_{yy}^{k-1} = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

since $\sum_{n=0}^{\infty} \theta^n = 1/(1-\theta)$ whenever $|\theta| < 1$. □

Proof of (1.6). Let $1_{\{X \geq k\}}$ denote the random variable that is 1 if $X \geq k$ and 0 otherwise. It is easy to see that

$$X = \sum_{k=1}^{\infty} 1_{\{X \geq k\}}.$$

Taking expected values and noticing $E1_{\{X \geq k\}} = P(X \geq k)$ gives

$$EX = \sum_{k=1}^{\infty} P(X \geq k)$$ □

Our next step is to compute the expected number of returns to y in a different way.

Lemma 1.7. *$E_x N(y) = \sum_{n=1}^{\infty} p^n(x,y)$.*

Proof. Let $1_{\{X_n = y\}}$ denote the random variable that is 1 if $X_n = y$, 0 otherwise. Clearly

$$N(y) = \sum_{n=1}^{\infty} 1_{\{X_n = y\}}.$$

Taking expected values now gives

$$E_x N(y) = \sum_{n=1}^{\infty} P_x(X_n = y)$$ □

With the two lemmas established we can now state our next main result.

Theorem 1.6. *y is recurrent if and only if*

$$\sum_{n=1}^{\infty} p^n(y,y) = E_y N(y) = \infty$$

Proof. The first equality is Lemma 1.7. From Lemma 1.6 we see that $E_y N(y) = \infty$ if and only if $\rho_{yy} = 1$, which is the definition of recurrence. $\square$

With this established we can easily complete the proofs of our two lemmas .

Proof of Lemma 1.4 . Suppose x is recurrent and $\rho_{xy} > 0$. By Lemma 1.2 we must have $\rho_{yx} > 0$. Pick j and ℓ so that $p^j(y,x) > 0$ and $p^\ell(x,y) > 0$. $p^{j+k+\ell}(y,y)$ is probability of going from y to y in $j+k+\ell$ steps while the product $p^j(y,x)p^k(x,x)p^\ell(x,y)$ is the probability of doing this and being at x at times j and $j+k$. Thus we must have

$$\sum_{k=0}^{\infty} p^{j+k+\ell}(y,y) \geq p^j(y,x)\left(\sum_{k=0}^{\infty} p^k(x,x)\right) p^\ell(x,y)$$

If x is recurrent then $\sum_k p^k(x,x) = \infty$, so $\sum_m p^m(y,y) = \infty$ and Theorem 1.6 implies that y is recurrent. $\square$

Proof of Lemma 1.5. If all the states in C are transient then Lemma 1.6 implies that $E_x N(y) < \infty$ for all x and y in C. Since C is finite, using Lemma 1.7

$$\begin{aligned}
\infty > \sum_{y \in C} E_x N(y) &= \sum_{y \in C} \sum_{n=1}^{\infty} p^n(x,y) \\
&= \sum_{n=1}^{\infty} \sum_{y \in C} p^n(x,y) = \sum_{n=1}^{\infty} 1 = \infty
\end{aligned}$$

where in the next to last equality we have used that C is closed. This contradiction proves the desired result. $\square$

1.4 Stationary Distributions

In the next section we will show that if we impose an additional assumption called aperiodicity an irreducible finite state Markov chain converges to a stationary distribution

$$p^n(x,y) \to \pi(y)$$

To prepare for that this section introduces stationary distributions and shows how to compute them. Our first step is to consider

What happens in a Markov chain when the initial state is random? Breaking things down according to the value of the initial state and using the definition of conditional probability

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_0 = i, X_n = j) \\ &= \sum_i P(X_0 = i) P(X_n = j | X_0 = i) \end{aligned}$$

If we introduce $q(i) = P(X_0 = i)$, then the last equation can be written as

$$P(X_n = j) = \sum_i q(i) p^n(i,j) \tag{1.7}$$

In words, we multiply the transition matrix on the left by the vector q of initial probabilities. If there are k states, then $p^n(x,y)$ is a $k \times k$ matrix. So to make the matrix multiplication work out right, we should take q as a $1 \times k$ matrix or a "row vector."

Example 1.15. Consider the weather chain (Example 1.3) and suppose that the initial distribution is $q(1) = 0.3$ and $q(2) = 0.7$. In this case

$$\begin{pmatrix} .3 & .7 \end{pmatrix} \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} .32 & .68 \end{pmatrix}$$

$$\begin{aligned} \text{since} \qquad .3(.6) + .7(.2) &= .32 \\ .3(.4) + .7(.8) &= .68 \end{aligned}$$

Example 1.16. Consider the social mobility chain (Example 1.4) and suppose that the initial distribution: $q(1) = .5$, $q(2) = .2$, and $q(3) = .3$. Multiplying the vector q by the transition probability gives the vector of probabilities at time 1.

$$\begin{pmatrix} .5 & .2 & .3 \end{pmatrix} \begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} = \begin{pmatrix} .47 & .32 & .21 \end{pmatrix}$$

To check the arithmetic note that the three entries on the right-hand side are

$$\begin{aligned} .5(.7) + .2(.3) + .3(.2) &= .35 + .06 + .06 = .47 \\ .5(.2) + .2(.5) + .3(.4) &= .10 + .10 + .12 = .32 \\ .5(.1) + .2(.2) + .3(.4) &= .05 + .04 + .12 = .21 \end{aligned}$$

If the distribution at time 0 is the same as the distribution at time 1, then by the Markov property it will be the distribution at all times $n \geq 1$.

Definition 1.7. *If $qp = q$ then q is called a* **stationary distribution**.

Stationary distributions have a special importance in the theory of Markov chains, so we will use a special letter π to denote solutions of the equation

$$\pi p = \pi.$$

To have a mental picture of what happens to the distribution of probability when one step of the Markov chain is taken, it is useful to think that we have $q(i)$ pounds of sand at state i, with the total amount of sand $\sum_i q(i)$ being one pound. When a step is taken in the Markov chain, a fraction $p(i,j)$ of the sand at i is moved to j. The distribution of sand when this has been done is

$$qp = \sum_i q(i)p(i,j)$$

If the distribution of sand is not changed by this procedure q is a stationary distribution.

Example 1.17. Weather chain. To compute the stationary distribution we want to solve

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix}$$

Multiplying gives two equations:

$$\begin{aligned} .6\pi_1 + .2\pi_2 &= \pi_1 \\ .4\pi_1 + .8\pi_2 &= \pi_2 \end{aligned}$$

Both equations reduce to $.4\pi_1 = .2\pi_2$. Since we want $\pi_1 + \pi_2 = 1$, we must have $.4\pi_1 = .2 - .2\pi_1$, and hence

$$\pi_1 = \frac{.2}{.2 + .4} = \frac{1}{3} \qquad \pi_2 = \frac{.4}{.2 + .4} = \frac{2}{3}$$

To check this we note that

$$\begin{pmatrix} 1/3 & 2/3 \end{pmatrix} \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} \frac{.6}{3} + \frac{.4}{3} & \frac{.4}{3} + \frac{1.6}{3} \end{pmatrix}$$

General two state transition probability.

	1	**2**
1	$1 - a$	a
2	b	$1 - b$

We have written the chain in this way so the stationary distribution has a simple formula

$$\pi_1 = \frac{b}{a + b} \qquad \pi_2 = \frac{a}{a + b} \tag{1.8}$$

As a first check on this formula we note that in the weather chain $a = 0.4$ and $b = 0.2$ which gives $(1/3, 2/3)$ as we found before. We can prove this works in general by drawing a picture:

$$\frac{b}{a+b} \overset{\mathbf{1}}{\bullet} \underset{b}{\overset{a}{\rightleftarrows}} \overset{\mathbf{2}}{\bullet} \frac{a}{a+b}$$

In words, the amount of sand that flows from 1 to 2 is the same as the amount that flows from 2 to 1 so the amount of sand at each site stays constant. To check algebraically that $\pi p = \pi$:

$$\begin{aligned} \frac{b}{a+b}(1-a) + \frac{a}{a+b}b &= \frac{b - ba + ab}{a+b} = \frac{b}{a+b} \\ \frac{b}{a+b}a + \frac{a}{a+b}(1-b) &= \frac{ba + a - ab}{a+b} = \frac{a}{a+b} \end{aligned} \tag{1.9}$$

Formula (1.8) gives the stationary distribution for any two state chain, so we progress now to the three state case and consider the

Example 1.18. Social Mobility (continuation of 1.4).

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

The equation $\pi p = \pi$ says

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix}$$

which translates into three equations

$$\begin{aligned} .7\pi_1 + .3\pi_2 + .2\pi_3 &= \pi_1 \\ .2\pi_1 + .5\pi_2 + .4\pi_3 &= \pi_2 \\ .1\pi_1 + .2\pi_2 + .4\pi_3 &= \pi_3 \end{aligned}$$

Note that the columns of the matrix give the numbers in the rows of the equations. The third equation is redundant since if we add up the three equations we get

$$\pi_1 + \pi_2 + \pi_3 = \pi_1 + \pi_2 + \pi_3$$

If we replace the third equation by $\pi_1 + \pi_2 + \pi_3 = 1$ and subtract π_1 from each side of the first equation and π_2 from each side of the second equation we get

$$\begin{aligned} -.3\pi_1 + .3\pi_2 + .2\pi_3 &= 0 \\ .2\pi_1 - .5\pi_2 + .4\pi_3 &= 0 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned} \tag{1.10}$$

At this point we can solve the equations by hand or using a calculator.

By hand. We note that the third equation implies $\pi_3 = 1 - \pi_1 - \pi_2$ and substituting this in the first two gives

$$\begin{aligned} .2 &= .5\pi_1 - .1\pi_2 \\ .4 &= .2\pi_1 + .9\pi_2 \end{aligned}$$

Multiplying the first equation by .9 and adding .1 times the second gives

$$2.2 = (0.45 + 0.02)\pi_1 \qquad \text{or} \qquad \pi_1 = 22/47$$

Multiplying the first equation by .2 and adding $-.5$ times the second gives

$$-0.16 = (-.02 - 0.45)\pi_2 \qquad \text{or} \qquad \pi_2 = 16/47$$

Since the three probabilities add up to 1, $\pi_3 = 9/47$.

Using the TI83 calculator is easier. To begin we write (1.10) in matrix form as

$$\begin{pmatrix}\pi_1 & \pi_2 & \pi_3\end{pmatrix}\begin{pmatrix}-.2 & .1 & 1\\ .2 & -.4 & 1\\ .3 & .3 & 1\end{pmatrix} = \begin{pmatrix}0 & 0 & 1\end{pmatrix}$$

If we let A be the 3×3 matrix in the middle this can be written as $\pi A = (0, 0, 1)$. Multiplying on each side by A^{-1} we see that

$$\pi = (0, 0, 1)A^{-1}$$

which is the third row of A^{-1}. To compute A^{-1}, we enter A into our calculator (using the MATRX menu and its EDIT submenu), use the MATRIX menu to put $[A]$ on the computation line, press x^{-1}, and then ENTER. Reading the third row we find that the stationary distribution is

$$(0.468085,\ 0.340425,\ 0.191489)$$

Converting the answer to fractions using the first entry in the MATH menu gives

$$(22/47,\ 16/47,\ 9/47)$$

Example 1.19. Brand Preference (continuation of 1.5).

	1	**2**	**3**
1	.8	.1	.1
2	.2	.6	.2
3	.3	.3	.4

Using the first two equations and the fact that the sum of the π's is 1

$$\begin{aligned} .8\pi_1 + .2\pi_2 + .3\pi_3 &= \pi_1 \\ .1\pi_1 + .6\pi_2 + .3\pi_3 &= \pi_2 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Subtracting π_1 from both sides of the first equation and π_2 from both sides of the second, this translates into $\pi A = (0, 0, 1)$ with

$$A = \begin{pmatrix} -.2 & .1 & 1 \\ .2 & -.4 & 1 \\ .3 & .3 & 1 \end{pmatrix}$$

Note that here and in the previous example the first two columns of A consist of the first two columns of the transition probability with 1 subtracted from the diagonal entries, and the final column is all 1's. Computing the inverse and reading the last row gives

$$(0.545454,\ 0.272727,\ 0.181818)$$

Converting the answer to fractions using the first entry in the MATH menu gives

$$(6/11,\ 3/11,\ 2/11)$$

To check this we note that

$$\begin{aligned} &\begin{pmatrix} 6/11 & 3/11 & 2/11 \end{pmatrix} \begin{pmatrix} .8 & .1 & .1 \\ .2 & .6 & .2 \\ .3 & .3 & .4 \end{pmatrix} \\ &= \left(\frac{4.8 + .6 + .6}{11} \quad \frac{.6 + 1.8 + .6}{11} \quad \frac{.6 + .6 + .8}{11} \right) \end{aligned}$$

Example 1.20. Basketball (continuation of 1.10). To find the stationary matrix in this case we can follow the same procedure. A consists of the first three columns of the transition matrix with 1 subtracted from the diagonal, and a final column of all 1's.

$$\begin{array}{cccc} -1/4 & 1/4 & 0 & 1 \\ 0 & -1 & 2/3 & 1 \\ 2/3 & 1/3 & -1 & 1 \\ 0 & 0 & 1/2 & 1 \end{array}$$

The answer is given by the fourth row of A^{-1}:

$$(0.5,\ 0.1875,\ 0.1875,\ 0.125) = (1/2,\ 3/16,\ 3/16,\ 1/8)$$

Thus the long run fraction of time the player hits a shot is

$$\pi(HH) + \pi(MH) = 0.6875 = 11/36.$$

1.5 Limit Behavior

If y is a transient state, then $\sum_{n=1}^{\infty} p^n(x,y) < \infty$ for any initial state x and hence

$$p^n(x,y) \to 0$$

In view of the decomposition theorem, Theorem 1.5 we can now restrict our attention to chains that consist of a single irreducible class of recurrent states. Our first example shows one problem that can prevent the convergence of $p^n(x,y)$.

Example 1.21. Ehrenfest chain (continuation of 1.2). For concreteness, suppose there are three balls. In this case the transition probability is

	0	**1**	**2**	**3**
0	0	3/3	0	0
1	1/3	0	2/3	0
2	0	2/3	0	1/3
3	0	0	3/3	0

In the second power of p the zero pattern is shifted:

	0	**1**	**2**	**3**
0	1/3	0	2/3	0
1	0	7/9	0	2/9
2	2/9	0	7/9	0
3	0	2/3	0	1/3

To see that the zeros will persist, note that if we have an odd number of balls in the left urn, then no matter whether we add or subtract one the result will be an even number. Likewise, if the number is even, then it will be odd on the next one step. This alternation between even and odd means that it is impossible to be back where we started after an odd number of steps. In symbols, if n is odd then $p^n(x,x) = 0$ for all x.

To see that the problem in the last example can occur for multiples of any number N consider:

Example 1.22. Renewal chain. We will explain the name later. For the moment we will use it to illustrate "pathologies." Let f_k be a distribution on the positive integers and let $p(0,k-1) = f_k$. For states $i > 0$ we let $p(i, i-1) = 1$. In words the chain jumps from 0 to $k-1$ with probability f_k and then walks back to 0 one step at a time. If $X_0 = 0$ and the jump is to $k-1$ then it returns to 0 at time k. If say $f_5 = f_{15} = 1/2$ then $p^n(0,0) = 0$ unless n is a multiple of 5.

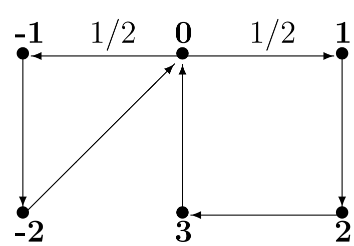
Definition 1.8. *The* **period** *of a state is the largest number that will divide all the $n \geq 1$ for which $p^n(x,x) > 0$. That is, it is the greatest common divisor of $I_x = \{n \geq 1 : p^n(x,x) > 0\}$.*

To check that this definition works correctly, we note that in Example 1.21, $\{n \geq 1 : p^n(x,x) > 0\} = \{2, 4, \ldots\}$, so the greatest common divisor is 2. Similarly, in Example 1.22, $\{n \geq 1 : p^n(x,x) > 0\} = \{5, 10, \ldots\}$, so the greatest common divisor is 5. As the next example shows, things aren't always so simple.

Example 4.4. Triangle and square. Consider the transition matrix:

	−2	**−1**	**0**	**1**	**2**	**3**
−2	0	0	1	0	0	0
−1	1	0	0	0	0	0
0	0	0.5	0	0.5	0	0
1	0	0	0	0	1	0
2	0	0	0	0	0	1

In words, from 0 we are equally likely to go to 1 or −1. From −1 we go with probability one to −2 and then back to 0, from 1 we go to 2 then to 3 and back to 0. The name refers to the fact that $0 \to -1 \to -2 \to 0$ is a triangle and $0 \to 1 \to 2 \to 3 \to 0$ is a square.



Clearly, $p^3(0,0) > 0$ and $p^4(0,0) > 0$ so $3, 4 \in I_0$. To compute I_0 the following is useful:

Lemma 1.8. *I_x is closed under addition. That is, if $i, j \in I_x$, then $i + j \in I_x$.*

Proof. If $i, j \in I_x$ then $p^i(x, x) > 0$ and $p^j(x, x) > 0$ so

$$p^{i+j}(x, x) \geq p^i(x, x)p^j(x, x) > 0$$

and hence $i + j \in I_x$. □

Using this we see that

$$I_0 = \{3, 4, 6, 7, 8, 9, 10, 11, \dots\}$$

Note that in this example once we have three consecutive numbers (e.g., 6,7,8) in I_0 then $6+3, 7+3, 8+3 \in I_0$ and hence it will contain all the larger integers.

For another unusual example consider the renewal chain (Example 1.22) with $f_5 = f_{12} = 1/2$. $5, 12 \in I_0$ so using Lemma 1.8

$$\begin{aligned} I_0 = \{&5, 10, 12, 15, 17, 20, 22, 24, 25, 27, 29, 30, 32, \\ &34, 35, 36, 37, 39, 40, 41, 42, 43, \dots\} \end{aligned}$$

To check this note that 5 gives rise to 10=5+5 and 17=5+12, 10 to 15 and 22, 12 to 17 and 24, etc. Once we have five consecutive numbers in I_0, here 39–43, we have all the rest. The last two examples motivate the following.

Lemma 1.9. *If x has period 1, i.e., the greatest common divisor I_x is 1, then there is a number n_0 so that if $n \geq n_0$, then $n \in I_x$. In words, I_x contains all of the integers after some value n_0.*

Proof. We begin by observing that it enough to show that I_x will contain two consecutive integers: k and $k+1$. For then it will contain $2k, 2k+1, 2k+2$, or in general $jk, jk+1, \dots jk+j$. For $j \geq k-1$ these blocks overlap and no integers are left out. In the last example $24, 25 \in I_0$ implies $48, 49, 50 \in I_0$ which implies $72, 73, 74, 75 \in I_0$ and $96, 97, 98, 99, 100 \in I_0$.

To show that there are two consecutive integers, we cheat and use a fact from number theory: if the greatest common divisor of a set I_x is 1 then there are integers $i_1, \dots i_m \in I_x$ and (positive or negative) integer coefficients c_i so that $c_1 i_1 + \cdots + c_m i_m = 1$. Let $a_i = c_i^+$ and $b_i = (-c_i)^+$. In words the a_i are the positive coefficients and the b_i are -1 times the negative coefficients. Rearranging the last equation gives

$$a_1 i_1 + \cdots + a_m i_m = (b_1 i_1 + \cdots + b_m i_m) + 1$$

and using Lemma 1.8 we have found our two consecutive integers in I_x. $\square$

While periodicity is a theoretical possibility, it rarely manifests itself in applications, except occasionally as an odd-even parity problem, e.g., the Ehrenfest chain. In most cases we will find (or design) our chain to be **aperiodic**, i.e., all states have period 1. To be able to verify this property for examples, we need to discuss some theory.

Lemma 1.10. *If $p(x,x) > 0$, then x has period 1.*

Proof. If $p(x,x) > 0$, then $1 \in I_x$, so the greatest common divisor is 1. $\square$

This is enough to show that all states in the weather chain (Example 1.3), social mobility (Example 1.4), and brand preference chain (Example 1.5) are aperiodic. For states with zeros on the diagonal the next result is useful.

Lemma 1.11. *If $\rho_{xy} > 0$ and $\rho_{yx} > 0$ then x and y have the same period.*

Why is this true? The short answer is that if the two states have different periods, then by going from x to y, from y to y in the various possible ways, and then from y to x, we will get a contradiction.

Proof. Suppose that the period of x is c, while the period of y is $d < c$. Let k be such that $p^k(x,y) > 0$ and let m be such that $p^m(y,x) > 0$. Since

$$p^{k+m}(x,x) \geq p^k(x,y)p^m(y,x) > 0$$

we have $k+m \in I_x$. Since x has period c, $k+m$ must be a multiple of c. Now let ℓ be any integer with $p^\ell(y,y) > 0$. Since

$$p^{k+\ell+m}(x,x) \geq p^k(x,y)p^\ell(y,y)p^m(y,x) > 0$$

$k+\ell+m \in I_x$, and $k+\ell+m$ must be a multiple of c. Since $k+m$ is itself a multiple of c, this means that ℓ is a multiple of c. Since $\ell \in I_y$ was arbitrary, we have shown that c is a divisor of every element of I_y, but $d < c$ is the greatest common divisor, so we have a contradiction. $\square$

Lemma 1.11 easily settles the question for the inventory chain (Example 1.6)

	0	**1**	**2**	**3**	**4**	**5**
0	0	0	.1	.2	.4	.3
1	0	0	.1	.2	.4	.3
2	.3	.4	.3	0	0	0
3	.1	.2	.4	.3	0	0
4	0	.1	.2	.4	.3	0
5	0	0	.1	.2	.4	.3

Since $p(x,x) > 0$ for $x = 2, 3, 4, 5$, Lemma 1.10 implies that these states are aperiodic. Since this chain is irreducible it follows from Lemma 1.11 that 0 and 1 are aperiodic.

Consider now the basketball chain (Example 1.10):

	HH	**HM**	**MH**	**MM**
HH	3/4	1/4	0	0
HM	0	0	2/3	1/3
MH	2/3	1/3	0	0
MM	0	0	1/2	1/2

Lemma 1.10 implies that **HH** and **MM** are aperiodic. Since this chain is irreducible it follows from Lemma 1.11 that **HM** and **MH** are aperiodic.

We now come to the main results of the chapter.

Theorem 1.7. Convergence theorem. *Suppose p is irreducible, aperiodic, and has a stationary distribution π. Then as $n \to \infty$, $p^n(x,y) \to \pi(y)$.*

Corollary. *If p is irreducible and has stationary distribution π, it is unique.*

Proof. First suppose that p is aperiodic. If there were two stationary distributions, π_1 and π_2, then by applying Theorem 1.7 we would conclude that

$$\pi_1(y) = \lim_{n\to\infty} p^n(x,y) = \pi_2(y)$$

To get rid of the aperiodicity assumption, let I be the transition probability for the chain that never moves, i.e., $I(x,x) = 1$ for all x, and define a new transition probability $q = (I + p)/2$, i.e., we either do nothing with probability 1/2 or take a step according to p. Since

$p(x,x) \geq 1/2$ for all x, $\hat{p}$ is aperiodic. The result now follows by noting that $\pi p = \pi$ if and only if $\pi q = \pi$. □

The next result considers the existence of stationary distributions:

Theorem 1.8. *If the state space S is finite and irreducible then there is a stationary distribution with $\pi(x) > 0$ for all x.*

Our final important fact about the stationary distribution is that it gives us the "limiting fraction for time we spend in each state."

Theorem 1.9. Asymptotic frequency. *Suppose p is irreducible and recurrent. If $N_n(y)$ be the number of visits to y up to time n, then*

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

Notice that this result and the next do not require aperiodicity. As a corollary we get the following.

Theorem 1.10. *If p is an irreducible and has stationary distribution π, then*

$$\pi(y) = 1/E_y T_y$$

In the next two examples we will be interested in the long run cost associated with a Markov chain. For this we will need the following extension of Theorem 1.9.

Theorem 1.11. *Suppose p is irreducible, has stationary distribution π, and $\sum_x |f(x)|\pi(x) < \infty$ then*

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to \sum_x f(x)\pi(x)$$

To illustrate the use of Theorem 1.11, we consider

Example 1.23. Repair chain (continuation of 3.2). A machine has three critical parts that are subject to failure, but can function as long as two of these parts are working. When two are broken, they are replaced and the machine is back to working order the next day. Declaring the state space to be the parts that are broken

$\{0, 1, 2, 3, 12, 13, 23\}$, we arrived at the following transition matrix:

	0	**1**	**2**	**3**	**12**	**13**	**23**
0	.93	.01	.02	.04	0	0	0
1	0	.94	0	0	.02	.04	0
2	0	0	.95	0	.01	0	.04
3	0	0	0	.97	0	.01	.02
12	1	0	0	0	0	0	0
13	1	0	0	0	0	0	0
23	1	0	0	0	0	0	0

and we asked: If we are going to operate the machine for 1800 days (about 5 years) then how many parts of types 1, 2, and 3 will we use?

To find the stationary distribution we compute the inverse of

$$\begin{pmatrix} -.07 & .01 & .02 & .04 & 0 & 0 & 1 \\ 0 & -.06 & 0 & 0 & .02 & .04 & 1 \\ 0 & 0 & -.05 & 0 & .01 & 0 & 1 \\ 0 & 0 & 0 & -.03 & 0 & .01 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Solving the equations and using FRAC:

$$\begin{array}{lll} & \pi(0) = 3000/8910 & \\ \pi(1) = 500/8910 & \pi(2) = 1200/8910 & \pi(3) = 4000/8910 \\ \pi(12) = 22/8910 & \pi(13) = 60/8910 & \pi(23) = 128/8910 \end{array}$$

We use up one part of type 1 on each visit to 12 or to 13, so on the average we use 82/8910 of a part per day. Over 1800 days we will use an average of $1800 \cdot 82/8910 = 16.56$ parts of type 1. Similarly type 2 and type 3 parts are used at the long run rates of 150/8910 and 188/8910 per day, so over 1800 days we will use an average of 30.30 parts of type 2 and 37.98 parts of type 3.

Example 1.24. Inventory chain (continuation of 1.6). We have an electronics store that sells a videogame system, selling 0, 1, 2, or 3 of these units each day with probabilities .3, .4, .2, and .1. Each night at the close of business new units can be ordered which will be available when the store opens in the morning. Suppose that sales produce a profit of \$12 but it costs \$2 a day to keep unsold units in the store overnight. Since it is impossible to sell 4 units in a day, and it costs us to have unsold inventory we should never have more than 3 units on hand.

Suppose we use a 2,3 inventory policy. That is, we order if there are ≤ 2 units and we order enough stock so that we have 3 units at the beginning of the next day. In this case we always start the day with 3 units, so the transition probability has constant rows

	0	**1**	**2**	**3**
0	.1	.2	.4	.3
1	.1	.2	.4	.3
2	.1	.2	.4	.3
3	.1	.2	.4	.3

In this case it is clear that the stationary distribution is $\pi(0) = .1$, $\pi(1) = .2$, $\pi(2) = .4$, and $\pi(3) = .3$. If we end the day with k units then we sold $3 - k$ and have to keep k over night. Thus our long run sales under this scheme are

$$.1(36) + .2(24) + .4(12) = 3.6 + 4.8 + 4.8 = 13.2 \quad \text{dollars per day}$$

while the inventory holding costs are

$$2(.2) + 4(.4) + 6(.3) = .4 + 1.6 + 1.8 = 3.8$$

for a net profit of 9.4 dollars per day.

Suppose we use a 1,3 inventory policy. In this case the transition probability is

	0	**1**	**2**	**3**
0	.1	.2	.4	.3
1	.1	.2	.4	.3
2	.3	.4	.3	0
3	.1	.2	.4	.3

Solving for the stationary distribution we get

$$\pi(0) = 19/110 \quad \pi(1) = 30/110 \quad \pi(2) = 40/110 \quad \pi(3) = 21/110$$

To compute the profit we make from sales note that if we always had enough stock then by the calculation in the first case, we would make 13.2 dollars per day. However, when $X_n = 2$ and the demand is 3, an event with probability $(4/11)\cdot .1 = .036$, we lose exactly one of our sales. From this it follows that in the long run we make a profit of

$$13.2 - (.036)12 = 12.768 \quad \text{dollars per day}$$

Our inventory holding cost under the new system is

$$2 \cdot \frac{30}{110} + 4 \cdot \frac{40}{110} + 6 \cdot \frac{21}{110} = \frac{60 + 160 + 126}{110} = 3.145$$

so now our profits are $12.768 - 3.145 = 9.623$.

Suppose we use a 0,3 inventory policy. In this case the transition probability is

	0	**1**	**2**	**3**
0	.1	.2	.4	.3
1	.7	.3	0	0
2	.3	.4	.3	0
3	.1	.2	.4	.3

From the equations for the stationary distribution we get

$$\pi(0) = 343/1070 \quad \pi(1) = 300/1070 \quad \pi(2) = 280/1070 \quad \pi(3) = 147/1070$$

To compute our profit we note, as in the previous calculation if we always had enough stock then we would make 13.2 dollars per day. Considering the various lost sales scenarios shows that in the long run we make sales of

$$13.2 - 12 \cdot \left(\frac{280}{1070} .1 + \frac{300}{1070} (.1 \cdot 2 + .2 \cdot 1) \right) = 11.54 \text{ dollars per day}$$

Our inventory holding cost until the new scheme is

$$2 \cdot \frac{300}{1070} + 4 \cdot \frac{280}{1070} + 6 \cdot \frac{147}{1070} \quad = \frac{600 + 1120 + 882}{1070} = \frac{4720}{1472} = 2.43$$

so the long run profit is $11.54 - 2.43 = 9.11$ dollars per day.

At this point we have computed

policy	0,3	1,3	2,3
profit per day	\$ 9.11	\$9.62	\$ \$9.40

so the 1,3 inventory policy is optimal.

1.6 Special Examples

1.6.1 Doubly stochastic chains

Definition 1.9. *A transition matrix p is said to be* **doubly stochastic** *if its COLUMNS sum to 1, or in symbols* $\sum_x p(x,y) = 1$.

The adjective "doubly" refers to the fact that by its definition a transition probability matrix has ROWS that sum to 1, i.e., $\sum_y p(x,y) = 1$. The stationary distribution is easy to guess in this case:

Theorem 1.12. *If p is a doubly stochastic transition probability for a Markov chain with N states, then the uniform distribution, $\pi(x) = 1/N$ for all x, is a stationary distribution.*

Proof. To check this claim we note that if $\pi(x) = 1/N$ then

$$\sum_x \pi(x)p(x,y) = \frac{1}{N}\sum_x p(x,y) = \frac{1}{N} = \pi(y)$$

Looking at the second equality we see that conversely, if the stationary distribution is uniform then p is doubly stochastic. □

Example 1.25. Symmetric reflecting random walk on the line. The state space is $\{0, 1, 2 \ldots, L\}$. The chain goes to the right or left at each step with probability 1/2, subject to the rules that if it tries to go to the left from 0 or to the right from L it stays put. For example, when $L = 4$ the transition probability is

	0	**1**	**2**	**3**	**4**
0	0.5	0.5	0	0	0
1	0.5	0	0.5	0	0
2	0	0.5	0	0.5	0
3	0	0	0.5	0	0.5
4	0	0	0	0.5	0.5

It is clear in the example $L = 4$ that each column adds up to 1. With a little thought one sees that this is true for any L, so the stationary distribution is uniform, $\pi(i) = 1/(L + 1)$.

Example 1.26. Tiny Board Game. Consider a circular board game with only six spaces $\{0, 1, 2, 3, 4, 5\}$. On each turn we roll a die with 1 on three sides, 2 on two sides, and 3 on one side to decide how far to move. Here we consider 5 to be adjacent to 0, so if we are there and we roll a 2 then the result is $5 + 2 \bmod 6 = 1$, where $i + k \bmod 6$ is the remainder when $i + k$ is divided by 6. In this case the transition probability is

	0	**1**	**2**	**3**	**4**	**5**
0	0	1/3	1/3	1/6	0	0
1	0	0	1/2	1/3	1/6	0
2	0	0	0	1/2	1/3	1/6
3	1/6	0	0	0	1/2	1/3
4	1/3	1/6	0	0	0	1/2
5	1/2	1/3	1/6	0	0	0

It is clear that the columns add to one, so the stationary distribution is uniform. To check the hypothesis of the convergence theorem, we note that after 3 turns we will have moved between 3 and 9 spaces so $p^3(i, j) > 0$ for all i and j.

Example 1.27. Mathematician's Monopoly. The game Monopoly is played on a game board that has 40 spaces arranged around the outside of a square. The squares have names like *Reading Railroad* and *Park Place* but we will number the squares 0 (*Go*), 1 (*Baltic Avenue*), ... 39 (*Boardwalk*). In Monopoly you roll two dice and move forward a number of spaces equal to the sum. For the moment, we will ignore things like *Go to Jail*, *Chance*, and other squares that make the transitions complicated and formulate the dynamics as following. Let r_k be the probability that the sum of two dice is k ($r_2 = 1/36$, $r_3 = 2/36$, ... $r_7 = 6/36$, ..., $r_{12} = 1/36$) and let

$$p(i, j) = r_k \qquad \text{if } j = i + k \bmod 40$$

where $i + k \bmod 40$ is the remainder when $i + k$ is divided by 40. To explain suppose that we are sitting on *Park Place* $i = 37$ and roll $k = 6$. $37 + 6 = 43$ but when we divide by 40 the remainder is 3, so $p(37, 3) = r_6 = 5/36$.

This example is larger but has the same structure as the previous example. Each row has the same entries but shift one unit to the right each time with the number that goes off the right edge emerging in the 0 column. This structure implies that each entry in the row appears once in each column and hence the sum of the entries in the column is 1, and the stationary distribution is uniform. To check the hypothesis of the convergence theorem note that in four

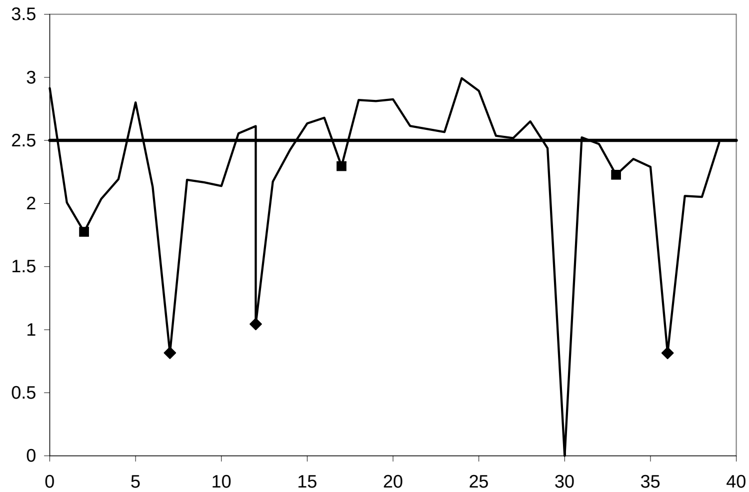


Figure 1.1: Stationary distribution for monopoly.

rolls you can move forward by 8 to 48 squares, so $p^4(i,j) > 0$ for all i and j.

Example 1.28. Real Monopoly has two complications:

- Square 30 is "Go to Jail," which sends you to square 10. You can buy your way out of jail but in the results we report below, we assume that you are cheap. If you roll a double then you get out for free. If you don't get doubles in three tries you have to pay.

- There are three *Chance* squares at 7, 12, and 36 (diamonds on the graph), and three *Community Chest* squares at 2, 17, 33 (squares on the graph), where you draw a card, which can send you to another square.

The graph gives the long run frequencies of being in different squares on the Monopoly board at the end of your turn, as computed by simulation. To make things easier to see we have removed the 9.46% chance of being *In Jail* to make the probabilities easier to see. The value reported for 10 is the 2.14% probability of *Just Visiting Jail*, i.e., being brought there by the roll of the dice. Square 30, *Go to Jail*, has probability 0 for the obvious reasons. The other three lowest values occur for *Chance* squares. Due to the transition from

30 to 10, frequencies for squares near 20 are increased relative to the average of 2.5% while those after 30 or before 10 are decreased. Squares 0 (*Go*) and 5 (*Reading Railroad*) are exceptions to this trend since there are Chance cards that instruct you to go there.

1.6.2 Detailed balance condition

π is said to satisfy the **detailed balance condition** if

$$\pi(x)p(x,y) = \pi(y)p(y,x) \tag{1.11}$$

To see that this is a stronger condition than $\pi p = \pi$, we sum over x on each side to get

$$\sum_x \pi(x)p(x,y) = \pi(y)\sum_x p(y,x) = \pi(y)$$

As in our earlier discussion of stationary distributions, we think of $\pi(x)$ as giving the amount of sand at x, and one transition of the chain as sending a fraction $p(x,y)$ of the sand at x to y. In this case the detailed balance condition says that the amount of sand going from x to y in one step is exactly balanced by the amount going back from y to x. In contrast the condition $\pi p = \pi$ says that after all the transfers are made, the amount of sand that ends up at each site is the same as the amount that starts there.

Many chains do not have stationary distributions that satisfy the detailed balance condition.

Example 1.29. Social Mobility

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

The stationary distribution computed in Example 1.18 is $(22/47, 16/47, 9/47)$ but

$$\pi(1)p(1,2) = \frac{22}{47}(.2) \neq \frac{16}{47}(0.3) = \pi(2)p(2,1)$$

To get an even simpler example modify the chain above so that

	1	**2**	**3**
1	.7	.3	0
2	.3	.5	.2
3	.2	.4	.4

$\pi(1)p(1,3) = 0$ but $p(3,1) > 0$ so if (1.11) holds then $\pi(3) = 0$ and using $\pi(3)p(3,i) = \pi(i)p(i,3)$ with $i = 1,2$ we conclude that all the $\pi(i) = 0$.

Example 1.30. Birth and death chains are defined by the property that the state space is some sequence of integers $\ell, \ell+1, \ldots r-1, r$ and it is impossible to jump by more than one:

$$p(x,y) = 0 \quad \text{when } |x-y| > 1$$

Suppose that the state space is $\{\ell, \ell+1, \ldots, r-1, r\}$ and the transition probability has

$$\begin{array}{ll} p(x,x+1) = p_x & \text{for } x < r \\ p(x,x-1) = q_x & \text{for } x > \ell \\ p(x,x) = r_x & \text{for } \ell \le x \le r \end{array}$$

while the other $p(x,y) = 0$. If $x < r$ detailed balance between x and $x+1$ implies $\pi(x)p_x = \pi(x+1)q_{x+1}$, so

$$\pi(x+1) = \frac{p_x}{q_{x+1}} \cdot \pi(x) \tag{1.12}$$

Using this with $x = \ell$ gives $\pi(\ell+1) = \pi(\ell)p_\ell/q_{\ell+1}$. Taking $x = \ell+1$

$$\pi(\ell+2) = \frac{p_{\ell+1}}{q_{\ell+2}} \cdot \pi(\ell+1) = \frac{p_{\ell+1} \cdot p_\ell}{q_{\ell+2} \cdot q_{\ell+1}} \cdot \pi(\ell)$$

Extrapolating from the first two results we see that in general

$$\pi(\ell+i) = \pi(\ell) \cdot \frac{p_{\ell+i-1} \cdot p_{\ell+i-2} \cdots p_{\ell+1} p_\ell}{q_{\ell+i} \cdot q_{\ell+i-1} \cdots q_{\ell+2} \cdots q_{\ell+1}} \tag{5.5}$$

To keep the indexing straight note that: (i) there are i terms in the numerator and in the denominator, (ii) the indices decrease by 1 each time, (iii) the answer will not depend on $p_{\ell+i}$ or q_ℓ.

For a concrete example to illustrate the use of this formula consider

Example 1.31. Ehrenfest chain. For concreteness, suppose there are three balls. In this case the transition probability is

	0	**1**	**2**	**3**
0	0	3/3	0	0
1	1/3	0	2/3	0
2	0	2/3	0	1/3
3	0	0	3/3	0

The detailed balance equations say:

$$\pi(0) = \pi(1)/3 \qquad 2\pi(1)/3 = 2\pi(2)/3 \qquad \pi(2)/3 = \pi(3)$$

Setting $\pi(0) = c$ we can solve to get $\pi(1) = 3c$, $\pi(2) = \pi(1) = 3c$, and $\pi(3) = c$. The sum of the π's should be one, so we pick $c = 1/8$ to get

$$\pi(0) = 1/8, \qquad \pi(1) = 3/8, \qquad \pi(2) = 3/8, \qquad \pi(3) = 1/8$$

Knowing the answer in general, one can look at the last equation and see that π represents the distribution of the number of Heads when we flip three coins, then guess in general that the binomial distribution with $p = 1/2$ is the stationary distribution:

$$\pi(i) = 2^{-n}\binom{n}{i}$$

Here $m! = 1 \cdot 2 \cdots (m-1) \cdot m$, with $0! = 1$, and

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

is the binomial coefficient which gives the number of ways of choosing x objects out of a set of n. To check that our guess satisfies the detailed balance condition, we note that

$$\begin{aligned}\pi(x)p(x, x+1) &= 2^{-n}\frac{n!}{x!(n-x)!} \cdot \frac{n-x}{n} \\ &= 2^{-n}\frac{n!}{(x+1)!(n-x-1)!} \cdot \frac{x+1}{n} = \pi(x+1)p(x+1, x)\end{aligned}$$

Example 1.32. Three machines, one repairman. Suppose that an office has three machines that each break with probability .1 each day, but when there is at least one broken, then with probability 0.5 the repairman can fix one of them for use the next day. If we ignore the possibility of two machines breaking on the same day, then the number of working machines can be modeled as a birth and death chain with the following transition matrix:

	0	**1**	**2**	**3**
0	.5	.5	0	0
1	.05	.5	.45	0
2	0	.1	.5	.4
3	0	0	.3	.7

Rows 0 and 3 are easy to see. To explain row 1, we note that the state will only decrease by 1 if one machine breaks and the repairman fails to repair the one he is working on, an event of probability (.1)(.5), while the state can only increase by 1 if he succeeds and there is no new failure, an event of probability .5(.9). Similar reasoning shows $p(2,1) = (.2)(.5)$ and $p(2,3) = .5(.8)$.

To find the stationary distribution we use the recursive formula (1.12) to conclude that if $\pi(0) = c$ then

$$\begin{aligned}
\pi(1) &= \pi(0) \cdot \frac{p_0}{q_1} = \quad c \cdot \frac{.5}{.05} = 10c \\
\pi(2) &= \pi(1) \cdot \frac{p_1}{q_2} = 10c \cdot \frac{.45}{.1} = 45c \\
\pi(3) &= \pi(2) \cdot \frac{p_2}{q_3} = 45c \cdot \frac{.4}{.3} = 60c
\end{aligned}$$

The sum of the π's is $116c$, so if we let $c = 1/116$ then we get

$$\pi(3) = \frac{60}{116}, \quad \pi(2) = \frac{45}{116}, \quad \pi(1) = \frac{10}{116}, \quad \pi(0) = \frac{1}{116}$$

There are many other Markov chains that are not birth and death chains but have stationary distributions that satisfy the detailed balance condition. A large number of possibilities are provided by

Example 1.33. Random walks on graphs. A graph is described by giving two things: (i) a set of vertices V (which, for the moment, we will suppose is a finite set) and (ii) an adjacency matrix $A(u,v)$, which is 1 if u and v are "neighbors" and 0 otherwise. By convention we set $A(v,v) = 0$ for all $v \in V$. The degree of a vertex u is equal to the number of neighbors it has. In symbols,

$$d(u) = \sum_v A(u,v)$$

since each neighbor of u contributes 1 to the sum. We write the degree this way to make it clear that

$$(*) \qquad p(u,v) = \frac{A(u,v)}{d(u)}$$

defines a transition probability. In words, if $X_n = u$, we jump to a randomly chosen neighbor of u at time $n+1$.

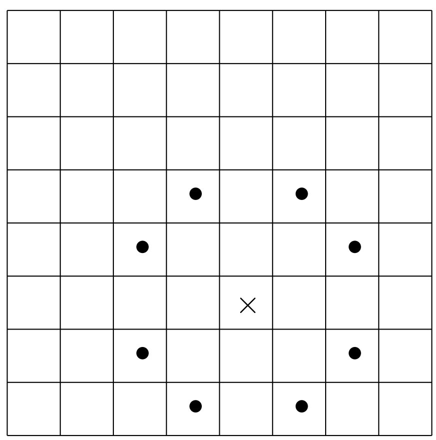
It is immediate from $(*)$ that if c is a positive constant then $\pi(u) = cd(u)$ satisfies the detailed balance condition:

$$\pi(u)p(u,v) = cA(u,v) = cA(v,u) = \pi(v)p(u,v)$$

Thus, if we take $c = 1/\sum_u d(u)$, we have a stationary probability distribution.

For a concrete example, consider

Example 1.34. Random walk of a knight on a chess board. A chess board is an 8 by 8 grid of squares. A knight moves by walking two steps in one direction and then one step in a perpendicular direction.



The degrees of the vertices are

2	3	4	4	4	4	3	2
3	4	6	6	6	6	4	3
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
4	6	8	8	8	8	6	4
3	4	6	6	6	6	4	3
2	3	4	4	4	4	3	2

The sum of the degrees is $4 \cdot 2 + 8 \cdot 3 + 20 \cdot 4 + 16 \cdot 6 + 16 \cdot 8 = 336$, so the stationary probabilities are the degrees divided by 336.

This problem is boring for a rook which has 14 possible moves from any square and hence a uniform stationary distribution. In the exercises, we will consider the other three interesting examples: king, bishop, and queen.

Example 1.35. The Metropolis-Hastings algorithm is a method for generating samples from a distribution $\pi(x)$. We begin with a Markov chain $q(x,y)$ that is the proposed jump distribution. A move is accepted with probability

$$r(x,y) = \min\left\{\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right\}$$

so the transition probability

$$p(x,y) = q(x,y)r(x,y)$$

To check that π satisfies the detailed balance condition we can suppose that $\pi(y)q(y,x) > \pi(x)q(x,y)$. In this case

$$\begin{aligned}
\pi(x)p(x,y) &= \pi(x)q(x,y)\cdot 1 \\
\pi(y)p(y,x) &= \pi(y)q(y,x)\frac{\pi(x)q(x,y)}{\pi(y)q(y,x)} = \pi(x)q(x,y)
\end{aligned}$$

To generate one sample from $\pi(x)$ we run the chain for a long time so that it reaches equilibrium. For many samples we output the state at widely separated times. Of course there is an art of knowing how long is long enough to wait between outputting the state.

This method is used when $\pi(x)$ has a complicated formula or involves a normalization that is difficult to compute. For a concrete example we consider the one dimensional Ising model. $x = (\eta_1, \eta_2, \dots \eta_N)$ where the spins $\eta_i = \pm 1$. Given an interaction parameter β which is inversely proportional to the temperature, the equilibirum state is

$$\pi(x) = \frac{1}{Z(\beta)}\exp\left(-\beta\sum_{j=1}^{N-1}(-\eta_i\eta_{i+1})\right)$$

There are two minus signs because the energy $H(x) = \sum_{j=1}^{N-1}(-\eta_i\eta_{i+1})$ is minimized when all the spins are equal and the probability of a state is proportional to $\exp(-\beta H(x))$. $Z(\beta)$ is a constant that makes the probabilities sum to one.

For the proposed jump distribution we let $q(x,y) = 1/N$ is the two configurations differ at exactly one spin. In this case the transition probability is

$$p(x,y) = q(x,y) \min\left\{\frac{\pi(y)}{\pi(x)}, 1\right\}$$

Note that the ratio $\pi(y)/\pi(x)$ is easy to compute because $Z(\beta)$ cancels out, as do all the terms in the sum that do not involve the site that flipped in going from x to y. In words $p(x,y)$ can be described by saying that we accept the proposed move with probability 1 if it lowers the energy and with probability $\pi(y)/\pi(x)$ if not.

1.6.3 Reversibility

Let $p(i,j)$ be a transition probability with stationary distribution $\pi(i)$. Let X_n be a realization of the Markov chain starting from the stationary distribution, i.e., $P(X_0 = i) = \pi(i)$. The next result says that if we watch the process X_m, $0 \le m \le n$, backwards, then it is a Markov chain.

Theorem 1.13. *Fix n and let $Y_m = X_{n-m}$ for $0 \le m \le n$. Then Y_m is a Markov chain with transition probability*

$$\hat{p}(i,j) = P(Y_{m+1} = j | Y_m = i) = \frac{\pi(j)p(j,i)}{\pi(i)} \tag{1.13}$$

Proof. We need to calculate the conditional probability.

$$\begin{aligned} &P(Y_{m+1} = i_{m+1} | Y_m = i_m, Y_{m-1} = i_{m-1} \dots Y_0 = i_0) \\ &\qquad = \frac{P(X_{n-(m+1)} = i_{m+1}, X_{n-m} = i_m, X_{n-m+1} = i_{m-1} \dots X_n = i_0)}{P(X_{n-m} = i_m, X_{n-m+1} = i_{m-1} \dots X_n = i_0)} \end{aligned}$$

Using the Markov property, we see the numerator is equal to

$$\pi(i_{m+1})p(i_{m+1}, i_m)P(X_{n-m+1} = i_{m-1}, \dots X_n = i_0 | X_{n-m} = i_m)$$

Similarly the denominator can be written as

$$\pi(i_m)P(X_{n-m+1} = i_{m-1}, \dots X_n = i_0 | X_{n-m} = i_m)$$

Dividing the last two formulas and noticing that the conditional probabilities cancel we have

$$P(Y_{m+1} = i_{m+1} | Y_m = i_m, \dots Y_0 = i_0) = \frac{\pi(i_{m+1}) p(i_{m+1}, i_m)}{\pi(i_m)}$$

This shows Y_m is a Markov chain with the indicated transition probability. $\square$

The formula for the transition probability in (1.13) may look a little strange, but it is easy to see that it works; i.e., the $r(i,j) \geq 0$, and have

$$\sum_j \hat{p}(i,j) = \sum_j \pi(j) p(j,i) \pi(i) = \frac{\pi(i)}{\pi(i)} = 1$$

since $\pi p = \pi$. When π satisfies the detailed balance conditions:

$$\pi(i) p(i,j) = \pi(j) p(j,i)$$

the transition probability for the reversed chain,

$$\hat{p}(i,j) = \frac{\pi(j) p(j,i)}{\pi(i)} = p(i,j)$$

is the same as the original chain. In words, if we make a movie of the Markov chain X_m, $0 \leq m \leq n$ starting from an initial distribution that satisfies the detailed balance condition and watch it backwards (i.e., consider $Y_m = X_{n-m}$ for $0 \leq m \leq n$), then we see a random process with the same distribution.

1.7 Proofs of the Theorems 1.7–1.11

To prepare for the proof of the convergence theorem, Theorem 1.7, we need the following:

Lemma 1.12. *If there is a stationary distribution, then all states y that have $\pi(y) > 0$ are recurrent.*

Proof. Lemma 1.7 tells us that $E_x N(y) = \sum_{n=1}^\infty p^n(x,y)$, so

$$\sum_x \pi(x) E_x N(y) = \sum_x \pi(x) \sum_{n=1}^\infty p^n(x,y)$$

Interchanging the order of summation and using $\pi p^n = \pi$, the above

$$= \sum_{n=1}^\infty \sum_x \pi(x) p^n(x,y) = \sum_{n=1}^\infty \pi(y) = \infty$$

since $\pi(y) > 0$. Using Lemma 1.6 now gives $E_x N(y) = \rho_{xy}/(1-\rho_{yy})$, so

$$\infty = \sum_x \pi(x) \frac{\rho_{xy}}{1-\rho_{yy}} \le \frac{1}{1-\rho_{yy}}$$

the second inequality following from the facts that $\rho_{xy} \le 1$ and π is a probability measure. This shows that $\rho_{yy} = 1$, i.e., y is recurrent. □

With Lemma 1.12 in hand we are ready to tackle the proof of:

Theorem 1.7. Convergence theorem. *Suppose p is irreducible, aperiodic, and has stationary distribution π. Then as $n \to \infty$, $p^n(x,y) \to \pi(y)$.*

Proof. Let S be the state space for p. Define a transition probability $\bar{p}$ on $S \times S$ by

$$\bar{p}((x_1,y_1),(x_2,y_2)) = p(x_1,x_2)p(y_1,y_2)$$

In words, each coordinate moves independently.

Step 1. We will first show that if p is aperiodic and irreducible then $\bar{p}$ is irreducible. Since p is irreducible, there are K, L, so that $p^K(x_1,x_2) > 0$ and $p^L(y_1,y_2) > 0$. Since x_2 and y_2 have period 1, it follows from Lemma 1.9 that if M is large, then $p^{L+M}(x_2, x_2) > 0$ and $p^{K+M}(y_2, y_2) > 0$, so

$$\bar{p}^{K+L+M}((x_1, y_1), (x_2, y_2)) > 0$$

Step 2. Since the two coordinates are independent $\bar{\pi}(a, b) = \pi(a)\pi(b)$ defines a stationary distribution for $\bar{p}$, and Lemma 1.12 implies that all states are recurrent for $\bar{p}$. Let (X_n, Y_n) denote the chain on $S \times S$, and let T be the first time that the two coordinates are equal, i.e., $T = \min\{n \geq 0 : X_n = Y_n\}$. Let $V_{(x,x)} = \min\{n \geq 0 : X_n = Y_n = x\}$ be the time of the first visit to (x, x). Since $\bar{p}$ is irreducible and recurrent, $V_{(x,x)} < \infty$ with probability one. Since $T \leq V_{(x,x)}$ for any x we must have

$$P(T < \infty) = 1. \tag{1.14}$$

Step 3. By considering the time and place of the first intersection and then using the Markov property we have

$$\begin{aligned}
P(X_n = y, T \leq n) &= \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x, X_n = y) \\
&= \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x) P(X_n = y | X_m = x) \\
&= \sum_{m=1}^{n} \sum_{x} P(T = m, Y_m = x) P(Y_n = y | Y_m = x) \\
&= P(Y_n = y, T \leq n)
\end{aligned}$$

Step 4. To finish up we observe that since the distributions of X_n and Y_n agree on $\{T \leq n\}$

$$|P(X_n = y) - P(Y_n = y)| \leq P(X_n = y, T > n) + P(Y_n = y, T > n)$$

and summing over y gives

$$\sum_{y} |P(X_n = y) - P(Y_n = y)| \leq 2P(T > n)$$

If we let $X_0 = x$ and let Y_0 have the stationary distribution π, then Y_n has distribution π, and Using (1.14) it follows that

$$\sum_{y} |p^n(x, y) - \pi(y)| \leq 2P(T > n) \to 0$$

proving the convergence theorem. □

Our next topic is the existence of stationary distributions. We first construct something less. $\mu(x) \geq 0$ is said to be a stationary measure if

$$\sum_x \mu(x)p(x,y) = \mu(y)$$

If $0 < \sum_x \mu(x) < \infty$ we can divide by the sum to get a stationary distribution.

Theorem 1.14. *Suppose p is irreducible and recurrent. Let $x \in S$ and let $T_x = \inf\{n \geq 1 : X_n = x\}$.*

$$\mu(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

defines a stationary measure with $0 < \mu(y) < \infty$ for all y.

Why is this true? This is called the "cycle trick." $\mu(y)$ is the expected number of visits to y in $\{0, \ldots, T_x - 1\}$. Multiplying by p moves us forward one unit in time so $\mu p(y)$ is the expected number of visits to y in $\{1, \ldots, T_x\}$. Since $X(T_x) = X_0 = x$ it follows that $\mu = \mu p$.

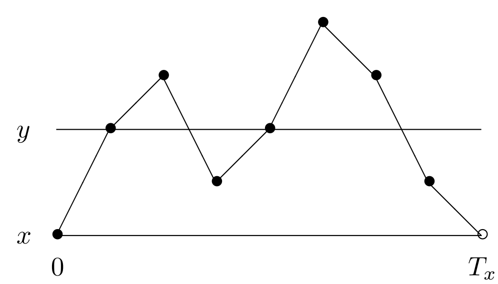


Figure 1.2: Picture of the cycle trick.

Proof. To formalize this intuition, let $\bar{p}_n(x,y) = P_x(X_n = y, T_x > n)$ and interchange sums to get

$$\sum_y \mu(y)p(y,z) = \sum_{n=0}^{\infty} \sum_y \bar{p}_n(x,y)p(y,z)$$

Case 1. Consider the generic case first: $z \neq x$.

$$\begin{aligned}\sum_y \bar{p}_n(x,y)p(y,z) &= \sum_y P_x(X_n = y, T_x > n, X_{n+1} = z) \\ &= P_x(T_x > n+1, X_{n+1} = z) = \bar{p}_{n+1}(x,z)\end{aligned}$$

Here the second equality holds since the chain must be somewhere at time n, and the third is just the definition of $\bar{p}_{n+1}$. Summing from $n = 0$ to ∞, we have

$$\sum_{n=0}^{\infty} \sum_y \bar{p}_n(x,y)p(y,z) = \sum_{n=0}^{\infty} \bar{p}_{n+1}(x,z) = \mu(z)$$

since $\bar{p}_0(x,z) = 0$.

Case 2. Now suppose that $z = x$. Reasoning as above we have

$$\sum_y \bar{p}_n(x,y)p(y,x) = \sum_y P_x(X_n = y, T_x > n, X_{n+1} = x) = P_x(T_x = n+1)$$

Summing from $n = 0$ to ∞ we have

$$\sum_{n=0}^{\infty} \sum_y \bar{p}_n(x,y)p(y,x) = \sum_{n=0}^{\infty} P_x(T_x = n+1) = 1 = \mu(x)$$

since $P_x(T = 0) = 0$.

To prove the final conclusion note that irreducibility implies $\mu(y) > 0$, since if K is the smallest k with $p^k(x,y) > 0$, then any k step from x to y cannot visit x at a positive time. To check $\mu(y) < \infty$ we note that $\mu(x) = 1$ and

$$1 = \mu(x) = \sum_y \mu(y)p^n(y,x) \geq \mu(y)$$

so if we pick n with $p^n(y,x) > 0$ then we conclude $\mu(y) < \infty$. $\square$

The next result now follows easily.

Theorem 1.8. *If the state space S is finite and is irreducible there is a stationary distribution with $\pi(x) > 0$ for all x.*

Proof. Theorem 1.14 implies that there is a stationary measure μ. Since S is finite then we can divide by the sum to get a stationary distribution. To prove that $\pi(y) > 0$ we note that this is trivial for $y = x$ the point used to define the measure. For $y \neq x$, we borrow an idea from Theorem 1.3. Let $K = \min\{k : p^k(x, y) > 0\}$. Since $p^K(x, y) > 0$ there must be a sequence $y_1, \ldots y_{K-1}$ so that

$$p(x, y_1)p(y_1, y_2) \cdots p(y_{K-1}, y) > 0$$

Since K is minimal all the $y_i \neq y$, so $P_x(X_K = y, T_x > K) > 0$ and hence $\pi(y) > 0$. □

We can now prove

Theorem 1.9. *Suppose p is irreducible and recurrent. Let $N_n(y)$ be the number of visits to y at times $\leq n$. As $n \to \infty$*

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

Why is this true? Suppose first that we start at y. The times between returns, $t_1, t_2, \ldots$ are independent and identically distributed so the strong law of large numbers for nonnegative random variables implies that the time of the kth return to y, $R(k) = \min\{n \geq 1 : N_n(y) = k\}$, has

$$\frac{R(k)}{k} \to E_y T_y \leq \infty \tag{1.15}$$

If we do not start at y then $t_1 < \infty$ and $t_2, t_3, \ldots$ are independent and identically distributed and we again have (1.15). Writing $a_k \sim b_k$ when $a_k/b_k \to 1$ we have $R(k) \sim kE_y T_y$. Taking $k = n/E_y T_y$ we see that there are about $n/E_y T_y$ returns by time n.

Proof. We have already shown (1.15). To turn this into the desired result, we note that from the definition of $R(k)$ it follows that $R(N_n(y)) \leq n < R(N_n(y) + 1)$. Dividing everything by $N_n(y)$ and then multiplying and dividing on the end by $N_n(y) + 1$, we have

$$\frac{R(N_n(y))}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{R(N_n(y) + 1)}{N_n(y) + 1} \cdot \frac{N_n(y) + 1}{N_n(y)}$$

Letting $n \to \infty$, we have $n/N_n(y)$ trapped between two things that converge to $E_y T_y$, so

$$\frac{n}{N_n(y)} \to E_y T_y$$

and we have proved the desired result. □

Theorem 1.10. *If p is an irreducible has stationary distribution π, then*

$$\pi(y) = 1/E_y T_y$$

Proof. Suppose X_0 has distribution π. From Theorem 1.9 it follows that

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

Taking expected value and using the fact that $N_n(y) \le n$, it can be shown that this implies

$$\frac{E_\pi N_n(y)}{n} \to \frac{1}{E_y T_y}$$

but since π is a stationary distribution $E_\pi N_n(y) = n\pi(y)$. □

Theorem 1.11. *Suppose p is irreducible, has stationary distribution π, and $\sum_x |f(x)|\pi(x) < \infty$ then*

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to \sum_x f(x)\pi(x)$$

The key idea here is that by breaking the path at the return times to x we get a seqeunce of random variables to which we can apply the law of large numbers.

Sketch of proof. Suppose that the chain starts at x. Let $T_0 = 0$ and $T_k = \min\{n > T_{k-1} : X_n = x\}$ be the time of the kth return to x. By the strong Markov property, the random variables

$$Y_k = \sum_{m=T_{k-1}+1}^{T_k} f(X_m)$$

are independent and identically distributed. By the cycle trick in the proof of Theorem 1.14

$$EY_k = \sum_x \mu_x(y) f(y)$$

Using the law of large numbers for i.i.d. variables

$$\frac{1}{L}\sum_{m=1}^{T_L} f(X_m) = \frac{1}{L}\sum_{k=1}^{L} Y_k \to \sum_x \mu_x(y) f(y)$$

By the proof of Theorem 1.9, $T_L/L \to E_x T_x$, so if we let $L = n/E_x T_x$ then

$$\frac{E_x T_x}{n} \sum_{m=1}^{n} f(X_m) \approx \frac{1}{L} \sum_{k=1}^{L} Y_k \to \sum_x \mu_x(y) f(y)$$

and it follows that

$$\frac{1}{n} \sum_{m=1}^{n} f(X_m) \to \sum_y \frac{\mu_x(y)}{E_x T_x} f(y) = \sum_y \pi(y) f(y) \qquad \square$$

1.8 Exit distributions

To motivate developments we begin with an example.

Example 1.36. Two year college. At a local two year college, 60% of freshmen become sophomores, 25% remain freshmen, and 15% drop out. 70% of sophomores graduate and transfer to a four year college, 20% remain sophomores and 10% drop out. What fraction of new students eventually graduate?

We use a Markov chain with state space 1 = freshman, 2 = sophomore, G = graduate, D = dropout. The transition probability is

	1	**2**	**G**	**D**
1	0.25	0.6	0	0.15
2	0	0.2	0.7	0.1
G	0	0	1	0
D	0	0	0	1

Let $h(x)$ be the probability that a student currently in state x eventually graduates. By considering what happens on one step

$$\begin{aligned} h(1) &= 0.25h(1) + 0.6h(2) \\ h(2) &= 0.2h(2) + 0.7 \end{aligned}$$

To solve we note that the second equation implies $h(2) = 7/8$ and then the first that

$$h(1) = \frac{0.6}{0.75} \cdot \frac{7}{8} = 0.7$$

Example 1.37. Tennis. In tennis the winner of a game is the first player to win four points, unless the score is $4-3$, in which case the game must continue until one player wins by two points. Suppose that the game has reached the point where one player is trying to get two points ahead to win and that the server will independently win the point with probability 0.6. What is the probability the server will win the game if the score is tied 3-3? if she is ahead by one point? Behind by one point?

We formulate the game as a Markov chain in which the state is the difference of the scores. The state space is $2, 1, 0, -1, -2$ with 2 (win for server) and -2 (win for opponent). The transition probability is

	2	**1**	**0**	**-1**	**-2**
2	1	0	0	0	0
1	.6	0	.4	0	0
0	0	.6	0	.4	0
-1	0	0	.6	0	.4
-2	0	0	0	0	1

If we let $h(x)$ be the probability of the server winning when the score is x then

$$h(x) = \sum_y p(x,y)h(y)$$

with $h(2) = 1$ and $h(-2) = 0$. This gives us three equations in three unknowns

$$\begin{aligned} h(1) &= .6 + .4h(0) \\ h(0) &= .6h(1) + .4h(-1) \\ h(-1) &= .6h(0) \end{aligned}$$

Rearranging we have

$$\begin{aligned} h(1) - .4h(0) + 0h(-1) &= .6 \\ -.6h(1) + h(0) - .4h(-1) &= 0 \\ 0h(1) - .6h(0) + h(-1) &= 0 \end{aligned}$$

which can be written in matrix form as

$$\begin{pmatrix} 1 & -.4 & 0 \\ -.6 & 1 & -.4 \\ 0 & -.6 & 1 \end{pmatrix} \begin{pmatrix} h(1) \\ h(0) \\ h(-1) \end{pmatrix} = \begin{pmatrix} .6 \\ 0 \\ 0 \end{pmatrix}$$

If we let $C = \{1, 0, -1\}$ be the nonabsorbing states, $r(x,y)$ the restriction of p to $x, y \in C$ (i.e., the 3×3 matrix inside the black lines in the transition probability) then the matrix is $I - r$. Solving gives

$$\begin{pmatrix} h(1) \\ h(0) \\ h(-1) \end{pmatrix} = (I-r)^{-1} \begin{pmatrix} .6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} .8769 \\ .6923 \\ .4154 \end{pmatrix}$$

The computations in this example become much simpler if we look at

$$p^2 = \begin{array}{cccccc} & \mathbf{2} & \mathbf{1} & \mathbf{0} & \mathbf{-1} & \mathbf{-2} \\ \mathbf{2} & 1 & 0 & 0 & 0 & 0 \\ \mathbf{1} & .6 & .24 & 0 & .16 & 0 \\ \mathbf{0} & .36 & 0 & .48 & 0 & .16 \\ \mathbf{-1} & 0 & .36 & 0 & .24 & .4 \\ \mathbf{-2} & 0 & 0 & 0 & 0 & 1 \end{array}$$

From p^2 we see that

$$h(0) = 0.36 + 0.48h(0)$$

so $h(0) = 0.36/0.52 = 0.6923$. By considering the outcome of the first point we see that $h(1) = 0.6 + 0.4h(0) = 0.8769$ and $h(-1) = 0.6h(0) = 0.4154$.

General solution. Suppose that the server wins each point with probability w. If the game is tied then after two points, the server will have won with probability w^2, lost with probability $(1-w)^2$, and returned to a tied game with probability $2w(1-w)$, so $h(0) = w^2 + 2w(1-w)h(0)$. Since $1 - 2w(1-w) = w^2 + (1-w)^2$, solving gives

$$h(0) = \frac{w^2}{w^2 + (1-w)^2}$$

Figure 1.3 graphs this function.

Having worked two examples, it is time to show that we have computed the right answer. In some cases we will want to guess and verify the answer. In those situations it is nice to know that the solution is unique. The next result proves this.

Theorem 1.15. *Consider a Markov chain with finite state space S. Let a and b be two points in S, and let $C = S - \{a, b\}$. Suppose $h(a) = 1$, $h(b) = 0$, and that for $x \in C$ we have*

$$h(x) = \sum_y p(x,y)h(y) \tag{1.16}$$

If $P_x(V_a \wedge V_b < \infty) > 0$ for all $x \in C$, then $h(x) = P_x(V_a < V_b)$.

Proof. Let $T = V_a \wedge V_b$. It follows from Lemma 1.3 that $P_x(T < \infty) = 1$ for all $x \in C$. (1.16) implies that $h(x) = E_x h(X_1)$ when $x \neq a, b$. The Markov property implies

$$h(x) = E_x h(X_{T \wedge n}).$$

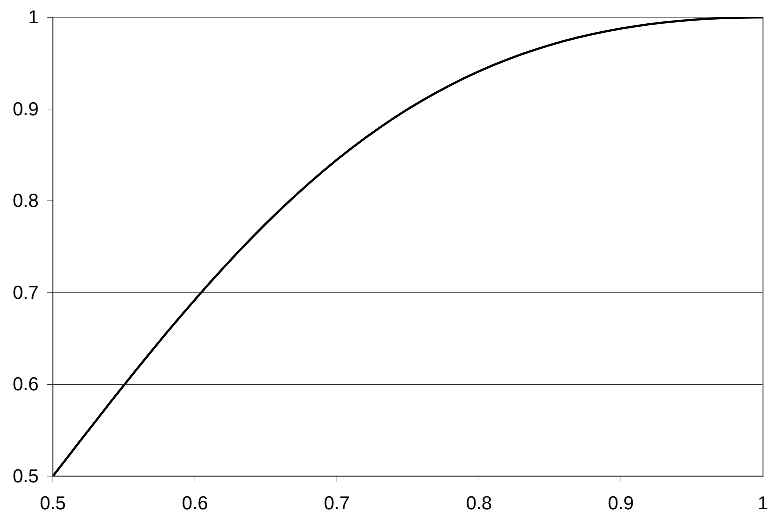


Figure 1.3: Probability the server winning a tied game as a function of the probability of winning a point.

We have to stop at time T because the equation is not assumed to be valid for $x = a, b$. Since S is finite, $P_x(T < \infty) = 1$ for all $x \in C$, $h(a) = 1$, and $h(b) = 0$, it is not hard to prove that $E_x h(X_{T\wedge n}) \to P_x(V_a < V_b)$ which gives the desired result. □

Example 1.38. Gambler's ruin. Consider a gambling game in which on any turn you win \$1 with probability p or lose \$1 with probability $1-p$. Suppose further that you will quit playing if your fortune reaches $\$N$. Of course, if your fortune reaches \$0, then the casino makes you stop. For reasons that will become clear in a moment, we depart from our usual definition and let

$$V_y = \min\{n \geq 0 : X_n = y\}$$

be the time of the first visit to y. Let

$$h(x) = P_x(V_N < V_0)$$

be the happy event that our gambler reaches the goal of $\$N$ before going bankrupt when starting with $\$x$. Thanks to our definition of V_x as the minimum of $n \geq 0$ we have $h(0) = 0$, and $h(N) = 1$. To calculate $h(x)$ for $0 < x < N$, we set $q = 1-p$ to simplify the formulas and consider what happens on the first step to arrive at

$$h(x) = ph(x+1) + qh(x-1) \qquad (1.17)$$

To solve this we rearrange to get $p(h(x+1)-h(x)) = q(h(x)-h(x-1))$ and conclude

$$h(x+1) - h(x) = \frac{q}{p} \cdot (h(x) - h(x-1)) \tag{1.18}$$

Symmetric case. If $p = 1/2$ this says that $h(x+1) - h(x) = h(x) - h(x-1)$. In words, this says that h has constant slope. Since $h(0) = 0$ and $h(N) = 1$ the slope must be $1/N$ and we must have $h(x) = x/N$. To argue this algebraically, we can instead observe that if $h(x) - h(x-1) = c$ for $1 \le i \le N$ then

$$1 = h(N) - h(0) = \sum_{i=1}^{N} h(x) - h(x-1) = Nc$$

so $c = 1/N$. Using the last identity again with the fact that $h(0) = 0$, we have

$$h(x) = h(x) - h(0) = \sum_{y=1}^{x} h(y) - h(y-1) = x/N$$

Recalling the definition of $h(x)$ this means

$$P_x(V_N < V_0) = x/N \quad \text{for } 0 \le x \le N \tag{1.19}$$

To see what the last formula says we will consider a concrete example.

Example 1.39. Matching pennies. Bob, who has 15 pennies, and Charlie, who has 10 pennies, decide to play a game. They each flip a coin. If the two coins match, Bob gets the two pennies (for a profit of 1). If the two coins are different, then Charlie gets the two pennies. They quit when someone has all of the pennies. What is the probability Bob will win the game?

Let X_n be the number of pennies Bob has after n plays. X_n is a gambler's ruin chain with $p = 1/2$, $N = 25$, and $X_0 = 15$, so by (1.19) the probability Bob wins is 15/25. Notice that the answer is simply Bob's fraction of the total supply of pennies.

Asymmetric case. When $p \neq 1/2$ the ideas are the same but the details are more difficult. If we set $c = h(x) - h(0)$ then (1.18) implies that for $i \geq 1$

$$h(x) - h(x-1) = c\left(\frac{q}{p}\right)^{x-1}$$

Summing from $x = 1$ to N, we have

$$1 = h(N) - h(0) = \sum_{x=1}^{N} h(x) - h(x-1) = c\sum_{x=1}^{N}\left(\frac{q}{p}\right)^{x-1}$$

Now for $\theta \neq 1$ the partial sum of the geometric series is

$$\sum_{j=0}^{N-1} \theta^j = \frac{1-\theta^N}{1-\theta} \tag{1.20}$$

To check this note that

$$\begin{aligned}(1-\theta)(1+\theta+\cdots\theta^{N-1}) = (1+\theta+\cdots\theta^{N-1})\\ -(\theta+\theta^2+\cdots\theta^N) = 1-\theta^N\end{aligned}$$

Using (1.20) we see that $c = (1-\theta)/(1-\theta^N)$ with $\theta = (1-p)/p$. Summing and using the fact that $h(0) = 0$, we have

$$h(x) = h(x) - h(0) = c\sum_{i=1}^{x} \theta^{i-1} = c \cdot \frac{1-\theta^x}{1-\theta} = \frac{1-\theta^x}{1-\theta^N}$$

Recalling the definition of $h(x)$ and rearranging the fraction we have

$$P_x(V_N < V_0) = \frac{\theta^x - 1}{\theta^N - 1} \qquad \text{where } \theta = \tfrac{1-p}{p} \tag{1.21}$$

To see what (1.21) says we consider:

Example 1.40. Roulette. If we bet \$1 on red on a roulette wheel with 18 red, 18 black, and 2 green (0 and 00) holes, we win \$1 with probability $18/38 = 0.4737$ and lose \$1 with probability $20/38$. Suppose we bring \$50 to the casino with the hope of reaching \$100 before going bankrupt. What is the probability we will succeed?

Here $\theta = q/p = 20/18$, so (1.21) implies

$$P_{50}(S_T = 100) = \frac{\left(\frac{20}{18}\right)^{50} - 1}{\left(\frac{20}{18}\right)^{100} - 1}$$

Using $(20/18)^{50} = 194$, we have

$$P_{50}(S_T = 100) = \frac{194 - 1}{(194)^2 - 1} = \frac{1}{194 + 1} = .005128$$

Now let's turn things around and look at the game from the viewpoint of the casino, i.e., $p = 20/38$. Suppose that the casino starts with the rather modest capital of $x = 100$. (1.21) implies that the probability they will reach N before going bankrupt is

$$\frac{(9/10)^{100} - 1}{(9/10)^N - 1}$$

If we let $N \to \infty$, $(9/10)^N \to 0$ so the answer converges to

$$1 - (9/10)^{100} = 1 - 2.656 \times 10^{-5}$$

If we increase the capital to \$200 then the failure probability is squared, since to become bankrupt we must first lose \$100 and then lose our second \$100. In this case the failure probability is incredibly small: $(2.656 \times 10^{-5})^2 = 7.055 \times 10^{-10}$.

From the last analysis we see that if $p > 1/2$, $q/p < 1$ and letting $N \to \infty$ in (1.21) gives

$$P_x(V_0 = \infty) = 1 - \left(\frac{q}{p}\right)^x \qquad P_x(V_0 < \infty) = \left(\frac{q}{p}\right)^x \tag{1.22}$$

to see that the form of the last answer makes sense note that to get from x to 0 we must go $x \to x - 1 \to x_2 \ldots \to 1 \to 0$, so

$$P_x(V_0 < \infty) = P_1(V_0 < \infty)^x.$$

Using this result, we can continue our investigation of Example 1.9.

Example 1.41. Wright–Fisher model with no mutation. The state space is $S = \{0, 1, \ldots N\}$ and the transition probability is

$$p(x, y) = \frac{N}{y} \left(\frac{x}{N}\right)^y \left(\frac{N - x}{N}\right)^{N-y}$$

The right-hand side is the binomial$(N, x/N)$ distribution, i.e., the number of successes in N trials when success has probability x/N, so the mean number of successes is x. From this it follows that if we define $h(x) = x/N$, then

$$h(x) = \sum_y p(x,y)h(y)$$

Taking $a = N$ and $b = 0$, we have $h(a) = 1$ and $h(b) = 0$. Since $P_x(V_a \wedge V_b < \infty) > 0$ for all $0 < x < N$, it follows from Lemma 1.15 that

$$P_x(V_N < V_0) = x/N$$

i.e., the probability of fixation to all A's is equal to the fraction of the genes that are A.

1.9 Exit times

To motivate developments we begin with an example.

Example 1.42. Two year college. In Example 1.36 we introduced a Markov chain with state space 1 = freshman, 2 = sophomore, G = graduate, D = dropout, and transition probability

	1	**2**	**G**	**D**
1	0.25	0.6	0	0.15
2	0	0.2	0.7	0.1
G	0	0	1	0
D	0	0	0	1

On the average how many years does a student take to graduate or drop out?

Let $g(x)$ be the expected time for a student to graduate or drop out. $g(G) = g(D) = 0$. By considering what happens on one step

$$\begin{aligned} g(1) &= 1 + 0.25g(1) + 0.6g(2) \\ g(2) &= 1 + 0.2g(2) \end{aligned}$$

where the 1+ is due to the fact that when the jump is made one year has elapsed. To solve we note that the second equation implies $g(2) = 1/0.8 = 1.25$ and then the first that

$$g(1) = \frac{1 + 0.6(1.25)}{0.75} = \frac{1.75}{0.75} = 2.3333$$

Example 1.43. Tennis. In Example 1.37 we formulated the game as a Markov chain in which the state is the difference of the scores. The state space is $S = \{2, 1, 0, -1, -2\}$ with 2 (win for server) and -2 (win for opponent). The transition probability is

	2	**1**	**0**	**-1**	**-2**
2	1	0	0	0	0
1	.6	0	.4	0	0
0	0	.6	0	.4	0
-1	0	0	.6	0	.4
-2	0	0	0	0	1

Let $g(x)$ be the expected time to complete the game when the current state is x. By considering what happens on one step

$$g(x) = 1 + \sum_y p(x,y)g(y)$$

Since $g(2) = g(-2) = 0$, if we let $r(x,y)$ be the restriction of the transition probability to $1, 0, -1$ we have

$$g(x) - \sum_y r(x,y)g(y) = 1$$

Writing $\mathbf{1}$ for a 3×1 matrix (i.e., column vector) with all 1's we can write this as

$$g(I-r) = \mathbf{1}$$

so $g = (I-r)^{-1}\mathbf{1}$.

There is another way to see this. If $N(y)$ is the number of visits to y at times $n \geq 0$, then from (1.7)

$$E_x N(y) = \sum_{n=0}^{\infty} r^n(x,y)$$

To see that this is $(I-r)^{-1}(x,y)$ note that $(I-r)(I+r+r^2+r^3+\cdots)$

$$= (I + r + r^2 + r^3 + \cdots) - (r + r^2 + r^3 + r^4 \cdots) = I$$

If T is the duration of the game then $T = \sum_y N(y)$ so

$$E_x T = (I-r)^{-1}\mathbf{1} \tag{1.23}$$

To solve the problem now we note that

$$I - r = \begin{pmatrix} 1 & -.4 & 0 \\ -.6 & 1 & -.4 \\ 0 & -.6 & 1 \end{pmatrix} \qquad (I-r)^{-1} = \begin{pmatrix} 19/13 & 10/13 & 4/13 \\ 15/13 & 25/13 & 10/13 \\ 9/13 & 15/13 & 19/13 \end{pmatrix}$$

so $E_0 T = (15 + 25 + 10)/13 = 50/13 = 3.846$ points.

Having worked two examples, it is time to show that we have computed the right answer. In some cases we will want to guess and verify the answer. In those situations it is nice to know that the solution is unique. The next result proves this.

Theorem 1.16. *Consider a Markov chain with finite state space S. Let $A \subset S$ and $V_A = \inf\{n \geq 0 : X_n \in A\}$. We suppose $C = S - A$ is finite, and that $P_x(V_A < \infty) > 0$ for any $x \in C$. Suppose $g(a) = 0$ for all $a \in A$, and that for $x \in C$ we have*

$$g(x) = 1 + \sum_y p(x,y)g(y) \tag{1.24}$$

Then $g(x) = E_x(V_A)$.

Proof. Let $T = V_a \wedge V_b$. It follows from Lemma 1.3 that $E_x T < \infty$ for all $x \in C$. (1.24) implies that $g(x) = 1 + E_x g(X_1)$ when $x \notin A$. The Markov property implies

$$g(x) = E_x(T \wedge n) + E_x g(X_{T\wedge n} n).$$

We have to stop at time t because the equation is not valid for $x \in A$. It follows from the definition of the expected value that $E_x(T \wedge n) \uparrow E_x T$. Since S is finite, $P_x(T < \infty) = 1$ for all $x \in C$, $g(a) = 0$ for $a \in A$, it is not hard to see that $E_x g(X_{T\wedge n}) \to 0$. □

Example 1.44. Waiting time for TT. Let T_{TT} be the (random) number of times we need to flip a coin before we have gotten Tails on two consecutive tosses. To compute the expected value of T_{TT} we will introduce a Markov chain with states 0, 1, 2 = the number of Tails we have in a row.

Since getting a Tails increases the number of Tails we have in a row by 1, but getting a Heads sets the number of Tails we have in a row to 0, the transition matrix is

	0	**1**	**2**
0	1/2	1/2	0
1	1/2	0	1/2
2	0	0	1

Since we are not interested in what happens after we reach 2 we have made 2 an absorbing state. If we let $V_2 = \min\{n \geq 0 : X_n = 2\}$ and $g(x) = E_x V_2$ then one step reasoning gives

$$g(0) = 1 + .5g(0) + .5g(1)$$
$$g(1) = 1 + .5g(0)$$

Plugging the second equation into the first gives $h(0) = 1.5 + .75h(0)$, so $.25h(0) = 1.5$ or $h(0) = 6$. To do this with the previous approach we note

$$I - r = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1 \end{pmatrix} \qquad (I-r)^{-1} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

so $E_0 V_2 = 6$.

Example 1.45. Waiting time for HT. Let T_{HT} be the (random) number of times we need to flip a coin before we have gotten a Heads followed by a Tails. Consider X_n is Markov chain with transition probability:

	HH	**HT**	**TH**	**TT**
HH	1/2	1/2	0	0
HT	0	0	1/2	1/2
TH	1/2	1/2	0	0
TT	0	0	1/2	1/2

If we eliminate the row and the column for HT then

$$I - r = \begin{pmatrix} 1/2 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -1/2 & 1/2 \end{pmatrix} \quad (I-r)^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix}$$

To compute the expected waiting time we note that after the first two tosses we have each of the four possibilities with probability 1/4 so

$$ET_{HT} = 2 + \frac{1}{4}(0 + 2 + 2 + 4) = 4$$

Why is $ET_{TT} = 6$ while $ET_{HT} = 4$? To explain we begin by noting that $E_y T_y = 1/\pi(y)$ and the stationary distribution assigns probability 1/4 to each state. One can verify this and check that convergence to equilibrium is rapid by noting that all the entries of p^2 are equal to 1/4.

$$E_{HT} T_{HT} = \frac{1}{\pi(HT)} = 4$$

To get from this to what we wanted to calculate, note that if we start with a H at time -1 and a T at time 0, then we have nothing that will help us in the future, so the expected waiting time for a HT when we start with nothing is the same.

If we apply this reasoning to TT we conclude

$$E_{TT}T_{TT} = \frac{1}{\pi(TT)} = 4$$

However this time if we start with a T at time -1 and a T at time 0, so a T at time 1 will give us a TT and a return at time 1, while if we get a H at time 1 we have wasted 1 turn and we have nothing that can help us later, so

$$4 = E_{TT}T_{TT} = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + ET_{TT})$$

Solving gives $ET_{TT} = 6$, so it takes longer to observe TT. The reason for this, which can be seen in the last equation, is that once we have one TT, we will get another one with probability 1/2, while occurrences of HT cannot overlap.

Example 1.46. Duration of fair games. Consider the gambler's ruin chain in which $p(i, i+1) = p(i, i-1) = 1/2$. Let $\tau = \min\{n : X_n \notin (0, N)\}$. We claim that

$$E_x\tau = x(N - x) \tag{1.25}$$

To see what formula (1.25) says, consider matching pennies. There $N = 25$ and $x = 15$, so the game will take $15 \cdot 10 = 150$ flips on the average. If there are twice as many coins, $N = 50$ and $x = 30$, then the game takes $30 \cdot 20 = 600$ flips on the average, or four times as long.

There are two ways to prove this.

Verify the guess. Let $g(x) = x(N - x)$. Clearly, $g(0) = g(N) = 0$. If $0 < x < N$ then by considering what happens on the first step we have

$$g(x) = 1 + \frac{1}{2}g(x+1) + \frac{1}{2}g(x-1)$$

If $g(x) = x(N - x)$ then the right-hand side is

$$\begin{aligned}
&= 1 + \frac{1}{2}(x+1)(N-x-1) + \frac{1}{2}(x-1)(N-x+1) \\
&= 1 + \frac{1}{2}[x(N-x) - x + N - x - 1] + \frac{1}{2}[x(N-x) + x - (N-x+1)] \\
&= 1 + x(N-x) - 1 = x(N-x)
\end{aligned}$$

Derive the answer. (1.24) implies that

$$g(x) = 1 + (1/2)g(x+1) + (1/2)g(x-1)$$

Rearranging gives

$$g(x+1) - g(x) = -2 + g(x) - g(x-1)$$

Setting $g(1)-g(0) = c$ we have $g(2)-g(1) = c-2$, $g(3)-g(2) = c-4$ and in general that

$$g(k) - g(k-1) = c - 2(k-1)$$

Using $g(0) = 0$ and summing we have

$$0 = g(N) = \sum_{k=1}^{N} c - 2(k-1) = cN - 2 \cdot \frac{N(N-1)}{2}$$

since, as one can easily check by induction, $\sum_{j=1}^{m} j = m(m+1)/2$. Solving gives $c = (N-1)$. Summing again, we see that

$$g(x) = \sum_{k=1}^{x} (N_1) - 2(k-1) = x(N-1) - x(x+1) = x(N-x)$$

Example 1.47. Duration of nonfair games. Consider the gambler's ruin chain in which $p(i, i+1)p$ and $p(i, i-1) = q$, where $p \neq q$. Let $\tau = \min\{n : X_n \notin (0, N)\}$. We claim that

$$E_x\tau = \frac{x}{q-p} - \frac{N}{q-p} \cdot \frac{1-(q/p)^x}{1-(q/p)^N} \tag{1.26}$$

This time the derivation is somewhat tedious so we will just verify the guess. We want to show that $g(x) = 1 + pg(x+1) + qg(x-1)$. Plugging the formula into the right-hand side:

$$\begin{aligned}
=&1 + p\frac{x+1}{q-p} + q\frac{x-1}{q-p} - \frac{N}{q-p}\left[p \cdot \frac{1-(q/p)^{x+1}}{1-(q/p)^N} + q\frac{1-(q/p)^{x-1}}{1-(q/p)^N}\right] \\
=&1 + \frac{x}{q-p} + \frac{p-q}{q-p} - \frac{N}{q-p}\left[\frac{p+q-(q/p)^x(q+p)}{1-(q/p)^N}\right]
\end{aligned}$$

which $= g(x)$ since $p + q = 1$.

To see what this says note that if $p < q$ then $q/p > 1$ so

$$\frac{N}{1-(q/p)^N} \to 0 \quad \text{and} \quad g(x) = \frac{x}{q-p} \tag{1.27}$$

To see this is reasonable note that our expected value on one play is $p - q$, so we lose an average of $q - p$ per play, and it should take an average of $x/(q - p)$ to lose x dollars.

When $p > q$, $(q/p)^N \to 0$, so doing some algebra

$$g(x) \approx \frac{N-x}{p-q}[1-(q/p)^x] + \frac{x}{p-q}(q/p)^x$$

Using (1.22) we see that the probability of not hitting 0 is $1-(q/p)^x$. In this case, since our expected winnings per play is $p-q$, it should take about $(N - x)/(p - q)$ plays to get to N. The second term represents the contribution to the expected value from paths that end at 0, but without a lot of calculation it is hard to explain why the term has exactly this form.

1.10 Infinite State Spaces*

In this section we delve into the complications that can arise when the state space for the chain is infinite. The new complication is that recurrence is not enough to guarantee the existence of a stationary distribution.

Example 1.48. Reflecting random walk. Imagine a particle that moves on $\{0, 1, 2, \ldots\}$ according to the following rules. It takes a step to the right with probability p. It attempts to take a step to the left with probability $1-p$, but if it is at 0 and tries to jump to the left, it stays at 0, since there is no -1 to jump to. In symbols,

$$\begin{aligned} p(i, i+1) &= p \quad \text{when } i \geq 0 \\ p(i, i-1) &= 1-p \quad \text{when } i \geq 1 \\ p(0,0) &= 1-p \end{aligned}$$

This is a birth and death chain, so we can solve for the stationary distribution using the detailed balance equations:

$$p\pi(i) = (1-p)\pi(i+1) \quad \text{when } i \geq 0$$

Rewriting this as $\pi(i+1) = \pi(i) \cdot p/(1-p)$ and setting $\pi(0) = c$, we have

$$\pi(i) = c \left(\frac{p}{1-p} \right)^i \tag{1.28}$$

There are now three cases to consider:

$p < 1/2$: $p/(1-p) < 1$. $\pi(i)$ decreases exponentially fast, so $\sum_i \pi(i) < \infty$, and we can pick c to make π a stationary distribution. To find the value of c to make π a probability distribution we recall

$$\sum_{i=0}^{\infty} \theta^i = 1/(1-\theta) \quad \text{when } \theta < 1.$$

Taking $\theta = p/(1-p)$ and hence $1 - \theta = (1-2p)/(1-p)$, we see that the sum of the $\pi(i)$ defined in $(*)$ is $c(1-p)/(1-2p)$, so

$$\pi(i) = \frac{1-2p}{1-p} \cdot \left(\frac{p}{1-p} \right)^i = (1-\theta)\theta^i \tag{1.29}$$

To confirm that we have succeeded in making the $\pi(i)$ add up to 1, note that if we are flipping a coin with a probability θ of Heads, then the probability of getting i Heads before we get our first Tails is given by $\pi(i)$.

The reflecting random walk is clearly irreducible. To check that it is aperiodic note that $p(0,0) > 0$ implies 0 has period 1, and then Lemma 1.11 implies that all states have period 1. Using the convergence theorem, Theorem 1.7, now we see that

I. *When $p < 1/2$, $P(X_n = j) \to \pi(j)$, the stationary distribution in (1.29).*

Using Theorem 1.10 now,

$$E_0 T_0 = \frac{1}{\pi(0)} = \frac{1}{1-\theta} = \frac{1-p}{1-2p} \tag{1.30}$$

It should not be surprising that the system stabilizes when $p < 1/2$. In this case movements to the left have a higher probability than to the right, so there is a drift back toward 0. On the other hand if steps to the right are more frequent than those to the left, then the chain will drift to the right and wander off to ∞.

II. *When $p > 1/2$ all states are transient.*

(1.22) implies that if $x > 0$, $P_x(T_0 < \infty) = ((1-p)/p)^x$.

To figure out what happens in the borderline case $p = 1/2$, we use results from Sections 1.8 and 1.9. Recall we have defined $V_y = \min\{n \geq 0 : X_n = y\}$ and (1.19) tells us that if $x > 0$

$$P_x(V_N < V_0) = x/N$$

If we keep x fixed and let $N \to \infty$, then $P_x(V_N < V_0) \to 0$ and hence

$$P_x(V_0 < \infty) = 1$$

In words, for any starting point x, the random walk will return to 0 with probability 1. To compute the mean return time we note that if $\tau_N = \min\{n : X_n \notin (0, N)\}$, then we have $\tau_N \leq V_0$ and by (1.25) we have $E_1 \tau_N = N - 1$. Letting $N \to \infty$ and combining the last two facts shows $E_1 V_0 = \infty$. Reintroducing our old hitting time $T_0 = \min\{n > 0 : X_n = 0\}$ and noting that on our first step we go to 0 or to 1 with probability 1/2 shows that

$$E_0 T_0 = (1/2) \cdot 1 + (1/2) E_1 V_0 = \infty$$

Summarizing the last two paragraphs, we have

III. *When* $p = 1/2$, $P_0(T_0 < \infty) = 1$ *but* $E_0 T_0 = \infty$.

Thus when $p = 1/2$, 0 is recurrent in the sense we will certainly return, but it is not recurrent in the following sense:

x is said to be **positive recurrent** if $E_x T_x < \infty$.

If a state is recurrent but not positive recurrent, i.e., $P_x(T_x < \infty) = 1$ but $E_x T_x = \infty$, then we say that x is **null recurrent**.

In our new terminology, our results for reflecting random walk say

If $p < 1/2$, 0 is positive recurrent
If $p = 1/2$, 0 is null recurrent
If $p > 1/2$, 0 is transient

In reflecting random walk, null recurrence thus represents the borderline between recurrence and transience. This is what we think in general when we hear the term. To see the reason we might be interested in positive recurrence recall that by Theorem 1.10

$$\pi(x) = \frac{1}{E_x T_x}$$

If $E_x T_x = \infty$, then this gives $\pi(x) = 0$. This observation motivates:

Theorem 1.17. *For an irreducible chain the following are equivalent:*
(i) Some state is positive recurrent.
(ii) There is a stationary distribution π*.*
(iii) All states are positive recurrent.

Proof. The stationary measure constructed in Theorem 1.8 has total mass

$$\begin{aligned}\sum_y \mu(y) &= \sum_{n=0}^{\infty} \sum_y P_x(X_n = y, T_x > n) \\ &= \sum_{n=0}^{\infty} P_x(T_x > n) = E_x T_x\end{aligned}$$

so (i) implies (ii). Noting that irreducibility implies $\pi(y) > 0$ for all y and then using $\pi(y) = 1/E_y T_y$ shows that (ii) implies (iii). It is trivial that (iii) implies (i). □

Our next example may at first seem to be quite different. In a branching process 0 is an absorbing state, so by Theorem 1.3 all the other states are transient. However, as the story unfolds we will see that branching processes have the same trichotomy as random walks do.

Example 1.49. Branching Processes. Consider a population in which each individual in the nth generation gives birth to an independent and identically distributed number of children. The number of individuals at time n, X_n is a Markov chain with transition probability given in Example 1.8. As announced there, we are interested in the question:

Q. What is the probability the species avoids extinction?

Here "extinction" means becoming absorbed state at 0. As we will now explain, whether this is possible or not can be determined by looking at the average number of offspring of one individual:

$$\mu = \sum_{k=0}^{\infty} kp_k$$

If there are m individuals at time $n-1$, then the mean number at time n is $m\mu$. More formally the conditional expectation given X_{n-1}

$$E(X_n|X_{n-1}) = \mu X_{n-1}$$

Taking expected values of both sides gives $EX_n = \mu EX_{n-1}$. Iterating gives

$$EX_n = \mu^n EX_0 \tag{1.31}$$

If $\mu < 1$, then $EX_n \to 0$ exponentially fast. Using the inequality

$$EX_n \geq P(X_n \geq 1)$$

it follows that $P(X_n \geq 1) \to 0$ and we have

I. *If $\mu < 1$ then extinction occurs with probability 1.*

To treat the cases $\mu \geq 1$ we will use a one-step calculation. Let ρ be the probability that this process dies out (i.e., reaches the absorbing state 0) starting from $X_0 = 1$. If there are k children in the first generation, then in order for extinction to occur, the family

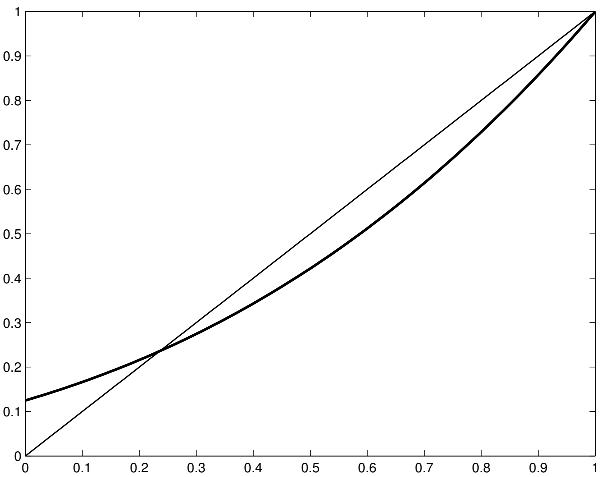


Figure 1.4: Generating function for Binomial(3,1/2).

line of each child must die out, an event of probability ρ^k, so we can reason that

$$\rho = \sum_{k=0}^{\infty} p_k \rho^k \tag{1.32}$$

If we let $\phi(\theta) = \sum_{k=0}^{\infty} p_k \theta^k$ be the generating function of the distribution p_k, then the last equation can be written simply as $\rho = \phi(\rho)$.

The equation in (1.32) has a trivial root at $\rho = 1$ since $\phi(\rho) = \sum_{k=0}^{\infty} p_k = 1$. The next result identifies the root that we want:

Lemma 1.13. *The extinction probability ρ is the smallest solution of the equation $\phi(x) = x$ with $0 \le x \le 1$.*

Proof. Extending the reasoning for (1.32) we see that in order for the process to hit 0 by time n, all of the processes started by first-generation individuals must hit 0 by time $n-1$, so

$$P(X_n = 0) = \sum_{k=0}^{\infty} p_k P(X_{n-1} = 0)^k$$

From this we see that if $\rho_n = P(X_n = 0)$ for $n \ge 0$, then $\rho_n = \phi(\rho_{n-1})$ for $n \ge 1$.

Since 0 is an absorbing state, $\rho_0 \le \rho_1 \le \rho_2 \le \ldots$ and the sequence converges to a limit ρ_∞. Letting $n \to \infty$ in $\rho_n = \phi(\rho_{n-1})$ implies that $\rho_\infty = \phi(\rho_\infty)$, i.e., ρ_∞ is a solution of $\phi(x) = x$. To complete the proof now let ρ be the smallest solution. Clearly $\rho_0 = 0 \le \rho$. Using the fact that ϕ is increasing, it follows that $\rho_1 = \phi(\rho_0) \le \phi(\rho) = \rho$. Repeating the argument we have $\rho_2 \le \rho$, $\rho_3 \le \rho$ and so on. Taking limits we have $\rho_\infty \le \rho$. However, ρ is the smallest solution, so we must have $\rho_\infty = \rho$. □

To see what this says, let us consider a concrete example.

Example 1.50. Binary branching. Suppose $p_2 = a$, $p_0 = 1 - a$, and the other $p_k = 0$. In this case $\phi(\theta) = a\theta^2 + 1 - a$, so $\phi(x) = x$ means

$$0 = ax^2 - x + 1 - a = (x - 1)(ax - (1 - a))$$

The roots are 1 and $(1 - a)/a$. If $a \le 1/2$, then the smallest root is 1, while if $a > 1/2$ the smallest root is $(1 - a)/a$.

Noting that $a \le 1/2$ corresponds to mean $\mu \le 1$ in binary branching motivates the following guess:

II. *If $\mu > 1$, then there is positive probability of avoiding extinction.*

Proof. In view of Lemma 1.13, we only have to show there is a root < 1. We begin by discarding a trivial case. If $p_0 = 0$, then $\phi(0) = 0$, 0 is the smallest root, and there is no probability of dying out. If $p_0 > 0$, then $\phi(0) = p_0 > 0$. Differentiating the definition of ϕ, we have

$$\phi'(x) = \sum_{k=1}^{\infty} p_k \cdot kx^{k-1} \quad \text{so} \quad \phi'(1) = \sum_{k=1}^{\infty} kp_k = \mu$$

If $\mu > 1$ then the slope of ϕ at $x = 1$ is larger than 1, so if ϵ is small, then $\phi(1 - \epsilon) < 1 - \epsilon$. Combining this with $\phi(0) > 0$ we see there must be a solution of $\phi(x) = x$ between 0 and $1 - \epsilon$. See the figure in the proof of (7.6). □

Turning to the borderline case:

III. *If $\mu = 1$ and we exclude the trivial case $p_1 = 1$, then extinction occurs with probability 1.*

Proof. By Lemma 1.13 we only have to show that there is no root < 1. To do this we note that if $p_1 < 1$, then for $y < 1$

$$\phi'(y) = \sum_{k=1}^{\infty} p_k \cdot kx^{k-1} < \sum_{k=1}^{\infty} p_k k = 1$$

so if $x < 1$ then $\phi(x) = \phi(1) - \int_x^1 \phi'(y)\,dy > 1 - (1-x) = x$. Thus $\phi(x) > x$ for all $x < 1$. $\square$

Note that in binary branching with $a = 1/2$, $\phi(x) = (1+x^2)/2$, so if we try to solve $\phi(x) = x$ we get

$$0 = 1 - 2x + x^2 = (1-x)^2$$

i.e., a double root at $x = 1$. In general when $\mu = 1$, the graph of ϕ is tangent to the diagonal (x,x) at $x = 1$. This slows down the convergence of ρ_n to 1 so that it no longer occurs exponentially fast.

In more advanced treatments, it is shown that if the offspring distribution has mean 1 and variance $\sigma^2 > 0$, then

$$P_1(X_n > 0) \sim \frac{2}{n\sigma^2}$$

This is not easy even for the case of binary branching, so we refer to reader to Section 1.9 of Athreya and Ney (1972) for a proof. We mention the result here because it allows us to see that the expected time for the process to die out $\sum_n P_1(T_0 > n) = \infty$. If we modify the branching process, so that $p(0,1) = 1$ then in the modified process

If $\mu < 1$, 0 is positive recurrent
If $\mu = 1$, 0 is null recurrent
If $\mu > 1$, 0 is transient

Our final example gives an application of branching processes to queueing theory.

Example 1.51. M/G/1 queue. We will not be able to explain the name of this example until we consider continuous-time Markov chains in Chapter 2. However, imagine a queue of people waiting to use an automated teller machine. Let X_n denote the number of people in line at the moment of the departure of the nth customer. To model this as a Markov chain we let a_k be the probability $k$customers arrive during one service time and write down the transition probability

$$p(0,k) = a_k \quad \text{and} \quad p(i, i-1+k) = a_k \quad \text{for } k \geq 0$$

with $p(i,j) = 0$ otherwise.

To explain this, note that if there is a queue, it is reduced by 1 by the departure of a customer, but k new customers will come with probability k. On the other hand if there is no queue, we must first wait for a customer to come and the queue that remains at her departure is the number of customers that arrived during her service time. The pattern becomes clear if we write out a few rows and columns of the matrix:

$$\begin{array}{c|cccccc c} & \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \dots \\ \mathbf{0} & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & \\ \mathbf{1} & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & \\ \mathbf{2} & 0 & a_0 & a_1 & a_2 & a_3 & a_4 & \\ \mathbf{3} & 0 & 0 & a_0 & a_1 & a_2 & a_3 & \\ \mathbf{4} & 0 & 0 & 0 & a_0 & a_1 & a_2 & \end{array}$$

If we regard the customers that arrive during a person's service time to be her children, then this queueing process gives rise to a branching process. From the results above for branching processes we see that if we denote the mean number of children by $\mu = \sum_k k a_k$, then

If $\mu < 1$, 0 is positive recurrent
If $\mu = 1$, 0 is null recurrent
If $\mu > 1$, 0 is transient

To bring out the parallels between the three examples, note that when $\mu > 1$ or $p > 1/2$ the process drifts away from 0 and is transient. When $\mu < 1$ or $p < 1/2$ the process drifts toward 0 and is positive recurrent. When $\mu = 1$ or $p = 1/2$, there is no drift. The process eventually hits 0 but not in finite expected time, so 0 is null recurrent.

1.11 Chapter Summary

A Markov chain with transition probability p is defined by the property that given the present state the rest of the past is irrelevant for predicting the future:

$$P(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = p(x, y)$$

The m step transition probability

$$p^m(i, j) = P(X_{n+m} = y | X_n = x)$$

is the mth power of the matrix p.

Recurrence and transience

The first thing we need to determine about a Markov chain is which states are recurrent and which are transient. To do this we let $T_y = \min\{n \geq 1 : X_n = y\}$ and let

$$\rho_{xy} = P_x(T_y < \infty)$$

When $x \neq y$ this is the probability X_n ever visits y starting at x. When $x = y$ this is the probability X_n returns to y when it starts at y. We restrict to times $n \geq 1$ in the definition of T_y so that we can say: y is recurrent if $\rho_{yy} = 1$ and transient if $\rho_{yy} < 1$.

Transient states in a finite state space can all be identified using

Theorem 1.3. *If $\rho_{xy} > 0$, but $\rho_{yx} < 1$, then x is transient.*

Once the transient states are removed we can use

Theorem 1.4. *If C is a finite closed and irreducible set, then all states in C are recurrent.*

Here A is closed if $x \in A$ and $y \notin A$ implies $p(x, y) = 0$, and B is irreducible if $x, y \in B$ implies $\rho_{xy} > 0$.

The keys to the proof of Theorem 1.4 are that: (i) If x is recurrent and $\rho_{xy} > 0$ then y is recurrent, and (ii) In a finite closed set there has to be at least one recurrent state. To prove these results, it was useful to know that if $N(y)$ is the number of visits to y at times $n \geq 1$ then

$$\sum_{n=1}^{\infty} p^n(x, y) = E_x N(y) = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

so y is recurrent if and only if $E_y N(y) = \infty$.

Theorems 1.3 and 1.4 two results allow us to decompose the state space and simplify the study of Markov chains.

Theorem 1.5. *If the state space S is finite, then S can be written as a disjoint union $T \cup R_1 \cup \cdots \cup R_k$, where T is a set of transient states and the R_i, $1 \le i \le k$, are closed irreducible sets of recurrent states.*

Stationary distributions

A stationary distribution is a solution of $\pi p = \pi$. Here if p is a $k \times k$ matrix, then π is a row vector, i.e., a $1 \times k$ matrix. If the state space S is finite and irreducible there is a unique stationary distribution. If there are k states then π can be computed by the following procedure. Form a matrix A by taking the first $k-1$ columns of $p - I$ and adding a final column of 1's. The equations $\pi p = \pi$ and $\pi_1 + \cdots \pi_k = 1$ are equivalent to

$$\pi A = \begin{pmatrix} 0 & \dots & 0 & 1 \end{pmatrix}$$

so we have

$$\pi = \begin{pmatrix} 0 & \dots & 0 & 1 \end{pmatrix} A^{-1}$$

or π is the bottom row of A^{-1}.

In two situations, the stationary distribution is easy to compute. (i) If the chain is doubly stochastic, i.e., $\sum_x p(x,y) = 1$, and has k states, then the stationary distribution is uniform $\pi(x) = 1/k$. (ii) π is a stationary distribution if the detailed balance condition holds

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

Birth and death chains, defined by the condition that $p(x,y) = 0$ if $|x - y| > 1$ always have stationary distributions with this property. If the state space is $\ell, \ell+1, \dots r$ π can be found by setting $\pi(\ell) = c$, solving for $\pi(x)$ for $\ell < x \le r$, and then choosing c to make the probabilities sum to 1.

Convergence theorems

Transient states y have $p^n(x,y) \to 0$, so to investigate the convergence of $p^n(x,y)$ it is enough by the decomposition theorem to suppose the chain is irreducible and all states are recurrent. The period of a state is the greatest common divisor of $I_x = \{n \ge 1 :$

$p^n(x,x) > 0\}$. If the period is 1, x is said to be aperiodic. A simple sufficient condition to be aperiodic is that $p(x,x) > 0$. To compute the period it is useful to note that if $\rho_{xy} > 0$ and $\rho_{yx} > 0$ then x and y have the same period. In particular all of the states in an irreducible set have the same period.

The main results about the asymptotic behavior of Markov chains are:

Theorem 1.7. *Suppose p is irreducible, aperiodic, and has a stationary distribution π. Then as $n \to \infty$, $p^n(x,y) \to \pi(y)$.*

Theorem 1.9. *Suppose p is irreducible and has stationary distribution π. If $N_n(y)$ be the number of visits to y up to time n, then*

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

As a corollary we get

Theorem 1.10. *If p is irreducible and has stationary distribution π then*

$$\pi(y) = E_y T_y$$

Chains with absorbing states

In this case there are two interesting questions. Where does the chain get absorbed? How long does it take? Let $V_y = \min\{n \geq 0 : X_n = y\}$ be the time of the first visit to y, i.e., now being there at time 0 counts.

Theorem 1.15. *Consider a Markov chain with finite state space S. Let a and b be two points in S, and let $C = S - \{a,b\}$. Suppose $h(a) = 1$, $h(b) = 0$, and that for $x \in C$ we have*

$$h(x) = \sum_y p(x,y)h(y)$$

If $\rho_{xa} + \rho_{xb} > 0$ for all $x \in C$, then $h(x) = P_x(V_a < V_b)$.

Let $r(x,y)$ be the part of the matrix $p(x,y)$ with $x, y \in C$. Since $h(a) = 1$ and $h(b) = 0$, the equation for h can be written for $x \in C$ as

$$h(x) = r(x,a) + \sum_y r(x,y)h(y)$$

so if we let v be the column vector with entries $r(x,a)$ then the last equation says $(I-r)h = v$ and

$$h = (I-r)^{-1}v.$$

Theorem 1.16. *Consider a Markov chain with finite state space S. Let $A \subset S$ and $V_A = \inf\{n \geq 0 : X_n \in A\}$. Suppose $g(a) = 0$ for all $a \in A$, and that for $x \in C = S - A$ we have*

$$g(x) = 1 + \sum_y p(x,y)g(y)$$

If $P_x(V_A < \infty) > 0$ for all $x \in C$, then $g(x) = E_x(V_A)$.

Since $g(x) = 0$ for $x \in A$ the equation for g can be written for $x \in C$ as

$$g(x) = 1 + \sum_y r(x,y)g(y)$$

so if we let $\vec{1}$ be a column vector consisting of all 1's then the last equation says $(I-r)g = \vec{1}$ and

$$g = (I-r)^{-1}\vec{1}.$$

1.12 Exercises

Understanding the definitions

1.1. A fair coin is tossed repeatedly with results $Y_0, Y_1, Y_2, \dots$ that are 0 or 1 with probability 1/2 each. For $n \geq 1$ let $X_n = Y_n + Y_{n-1}$ be the number of 1's in the $(n-1)$th and nth tosses. Is X_n a Markov chain?

1.2. Five white balls and five black balls are distributed in two urns in such a way that each urn contains five balls. At each step we draw one ball from each urn and exchange them. Let X_n be the number of white balls in the left urn at time n. Compute the transition probability for X_n.

1.3. We repeated roll two four sided dice with numbers 1, 2, 3, and 4 on them. Let Y_k be the sum on the kth roll, $S_n = Y_1 + \cdots + Y_n$ be the total of the first n rolls, and $X_n = S_n \pmod 6$. Find the transition probability for X_n.

1.4. The 1990 census showed that 36% of the households in the District of Columbia were homeowners while the remainder were renters. During the next decade 6% of the homeowners became renters and 12% of the renters became homeowners. What percentage were homeowners in 2000? in 2010?

1.5. Consider a gambler's ruin chain with $N = 4$. That is, if $1 \leq i \leq 3$, $p(i, i+1) = 0.4$, and $p(i, i-1) = 0.6$, but the endpoints are absorbing states: $p(0,0) = 1$ and $p(4,4) = 1$ Compute $p^3(1,4)$ and $p^3(1,0)$.

1.6. A taxicab driver moves between the airport A and two hotels B and C according to the following rules. If he is at the airport, he will be at one of the two hotels next with equal probability. If at a hotel then he returns to the airport with probability 3/4 and goes to the other hotel with probability 1/4. (a) Find the transition matrix for the chain. (b) Suppose the driver begins at the airport at time 0. Find the probability for each of his three possible locations at time 2 and the probability he is at hotel B at time 3.

1.7. Suppose that the probability it rains today is 0.3 if neither of the last two days was rainy, but 0.6 if at least one of the last two days was rainy. Let the weather on day n, W_n, be R for rain, or S for sun. W_n is not a Markov chain, but the weather for the last two days $X_n = (W_{n-1}, W_n)$ is a Markov chain with four states $\{RR, RS, SR, SS\}$. (a) Compute its transition probability. (b) Compute the two-step transition probability. (c) What is the probability it will rain on Wednesday given that it did not rain on Sunday or Monday.

1.8. Consider the following transition matrices. Identify the transient and recurrent states, and the irreducible closed sets in the Markov chains. Give reasons for your answers.

$$
\begin{array}{cccccc}
(a) & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\
\mathbf{1} & .4 & .3 & .3 & 0 & 0 \\
\mathbf{2} & 0 & .5 & 0 & .5 & 0 \\
\mathbf{3} & .5 & 0 & .5 & 0 & 0 \\
\mathbf{4} & 0 & .5 & 0 & .5 & 0 \\
\mathbf{5} & 0 & .3 & 0 & .3 & .4
\end{array}
\qquad
\begin{array}{ccccccc}
(b) & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\
\mathbf{1} & .1 & 0 & 0 & .4 & .5 & 0 \\
\mathbf{2} & .1 & .2 & .2 & 0 & .5 & 0 \\
\mathbf{3} & 0 & .1 & .3 & 0 & 0 & .6 \\
\mathbf{4} & .1 & 0 & 0 & .9 & 0 & 0 \\
\mathbf{5} & 0 & 0 & 0 & .4 & 0 & .6 \\
\mathbf{6} & 0 & 0 & 0 & 0 & .5 & .5
\end{array}
$$

$$
\begin{array}{cccccc}
(c) & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\
\mathbf{1} & 0 & 0 & 0 & 0 & 1 \\
\mathbf{2} & 0 & .2 & 0 & .8 & 0 \\
\mathbf{3} & .1 & .2 & .3 & .4 & 0 \\
\mathbf{4} & 0 & .6 & 0 & .4 & 0 \\
\mathbf{5} & .3 & 0 & 0 & 0 & .7
\end{array}
\qquad
\begin{array}{ccccccc}
(d) & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\
\mathbf{1} & .8 & 0 & 0 & .2 & 0 & 0 \\
\mathbf{2} & 0 & .5 & 0 & 0 & .5 & 0 \\
\mathbf{3} & 0 & 0 & .3 & .4 & .3 & 0 \\
\mathbf{4} & .1 & 0 & 0 & .9 & 0 & 0 \\
\mathbf{5} & 0 & .2 & 0 & 0 & .8 & 0 \\
\mathbf{6} & .7 & 0 & 0 & .3 & 0 & 0
\end{array}
$$

1.9. Find the stationary distributions for the Markov chains with transition matrices:

$$
\begin{array}{cccc}
(a) & \mathbf{1} & \mathbf{2} & \mathbf{3} \\
\mathbf{1} & .5 & .4 & .1 \\
\mathbf{2} & .2 & .5 & .3 \\
\mathbf{3} & .1 & .3 & .6
\end{array}
\qquad
\begin{array}{cccc}
(b) & \mathbf{1} & \mathbf{2} & \mathbf{3} \\
\mathbf{1} & .5 & .4 & .1 \\
\mathbf{2} & .3 & .4 & .3 \\
\mathbf{3} & .2 & .2 & .6
\end{array}
\qquad
\begin{array}{cccc}
(c) & \mathbf{1} & \mathbf{2} & \mathbf{3} \\
\mathbf{1} & .6 & .4 & 0 \\
\mathbf{2} & .2 & .4 & .2 \\
\mathbf{3} & 0 & .2 & .8
\end{array}
$$

1.10. Find the stationary distributions for the Markov chains on $\{1, 2, 3, 4\}$ with transition matrices:

$$
(a) \begin{pmatrix} .7 & 0 & .3 & 0 \\ .6 & 0 & .4 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & .4 & 0 & .6 \end{pmatrix}
\qquad
(b) \begin{pmatrix} .7 & .3 & 0 & 0 \\ .2 & .5 & .3 & 0 \\ .0 & .3 & .6 & .1 \\ 0 & 0 & .2 & .8 \end{pmatrix}
\qquad
(c) \begin{pmatrix} .7 & 0 & .3 & 0 \\ .2 & .5 & .3 & 0 \\ .1 & .2 & .4 & .3 \\ 0 & .4 & 0 & .6 \end{pmatrix}
$$

(c) The matrix is doubly stochastic so $\pi(i) = 1/4$, $i = 1, 2, 3, 4$.

1.11. Find the stationary distributions for the chains in exercises (a) 1.2, (b) 1.3, and (c) 1.7.

1.12. (a) Find the stationary distribution for the transition probability

$$\begin{array}{ccccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & 0 & 2/3 & 0 & 1/3 \\ \mathbf{2} & 1/3 & 0 & 2/3 & 0 \\ \mathbf{3} & 0 & 1/6 & 0 & 5/6 \\ \mathbf{4} & 2/5 & 0 & 3/5 & 0 \end{array}$$

and show that it does not satisfy the detailed balance condition (1.11).
(b) Consider

$$\begin{array}{ccccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & 0 & a & 0 & 1-a \\ \mathbf{2} & 1-b & 0 & b & 0 \\ \mathbf{3} & 0 & 1-c & 0 & c \\ \mathbf{4} & d & 0 & 1-d & 0 \end{array}$$

and show that there is a stationary distribution satisfying (1.11) if

$$0 < abcd = (1-a)(1-b)(1-c)(1-d).$$

1.13. Consider the Markov chain with transition matrix:

$$\begin{array}{ccccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & 0 & 0 & 0.1 & 0.9 \\ \mathbf{2} & 0 & 0 & 0.6 & 0.4 \\ \mathbf{3} & 0.8 & 0.2 & 0 & 0 \\ \mathbf{4} & 0.4 & 0.6 & 0 & 0 \end{array}$$

(a) Compute p^2. (b) Find the stationary distributions of p and all of the stationary distributions of p^2. (c) Find the limit of $p^{2n}(x, x)$ as $n \to \infty$.

1.14. Do the following Markov chains converge to equilibrium?

$$\begin{array}{ccccc} (a) & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & 0 & 0 & 1 & 0 \\ \mathbf{2} & 0 & 0 & .5 & .5 \\ \mathbf{3} & .3 & .7 & 0 & 0 \\ \mathbf{4} & 1 & 0 & 0 & 0 \end{array} \qquad \begin{array}{ccccc} (b) & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{1} & 0 & 1 & .0 & 0 \\ \mathbf{2} & 0 & 0 & 0 & 1 \\ \mathbf{3} & 1 & 0 & 0 & 0 \\ \mathbf{4} & 1/3 & 0 & 2/3 & 0 \end{array}$$

(c)	**1**	**2**	**3**	**4**	**5**	**6**
1	0	.5	.5	0	0	0
2	0	0	0	1	0	0
3	0	0	0	.4	0	.6
4	1	0	0	0	0	0
5	0	1	0	0	0	0
6	.2	0	0	0	.8	0

1.15. Find $\lim_{n\to\infty} p^n(i,j)$ for

$$p = \begin{array}{cccccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\ \mathbf{1} & 1 & 0 & 0 & 0 & 0 \\ \mathbf{2} & 0 & 2/3 & 0 & 1/3 & 0 \\ \mathbf{3} & 1/8 & 1/4 & 5/8 & 0 & 0 \\ \mathbf{4} & 0 & 1/6 & 0 & 5/6 & 0 \\ \mathbf{5} & 1/3 & 0 & 1/3 & 0 & 1/3 \end{array}$$

You are supposed to do this and the next problem by solving equations. However you can check your answers by using your calculator to find FRAC(p^{100}).

1.16. If we renumber the seven state chain in Example 1.14 we get

	1	**2**	**3**	**4**	**5**	**6**	**7**
1	.2	.3	.1	0	.4	0	0
2	0	.5	0	.2	.3	0	0
3	0	0	.7	.3	0	0	0
4	0	0	.6	.4	0	0	0
5	0	0	0	0	.5	.5	0
6	0	0	0	0	0	.2	.8
7	0	0	0	0	1	0	0

Find $\lim_{n\to\infty} p^n(i,j)$.

Two state Markov chains

1.17. Market research suggests that in a five year period 8% of people with cable television will get rid of it, and 26% of those without it will sign up for it. Compare the predictions of the Markov chain model with the following data on the fraction of people with cable TV: 56.4% in 1990, 63.4% in 1995, and 68.0% in 2000. What is the long run fraction of people with cable TV?

1.18. A sociology professor postulates that in each decade 8% of women in the work force leave it and 20% of the women not in it begin to work. Compare the predictions of his model with the following data on the percentage of women working: 43.3% in 1970, 51.5% in 1980, 57.5% in 1990, and 59.8% in 2000. In the long run what fraction of women will be working?

1.19. A rapid transit system has just started operating. In the first month of operation, it was found that 25% of commuters are using the system while 75% are travelling by automobile. Suppose that each month 10% of transit users go back to using their cars, while 30% of automobile users switch to the transit system. (a) Compute the three step transition probaiblity p^3. (b) What will be the fractions using rapid transit in the fourth month? (c) In the long run?

1.20. A regional health study indicates that from one year to the next, 75% percent of smokers will continue to smoke while 25% will quit. 8% of those who stopped smoking will resume smoking while 92% will not. If 70% of the population were smokers in 1995, what fraction will be smokers in 1998? in 2005? in the long run?

1.21. Three of every four trucks on the road are followed by a car, while only one of every five cars is followed by a truck. What fraction of vehicles on the road are trucks?

1.22. In a test paper the questions are arranged so that 3/4's of the time a True answer is followed by a True, while 2/3's of the time a False answer is followed by a False. You are confronted with a 100 question test paper. Approximately what fraction of the answers will be True.

1.23. In unprofitable times corporations sometimes suspend dividend payments. Suppose that after a dividend has been paid the next one will be paid with probability 0.9, while after a dividend is suspended the next one will be suspended with probability 0.6. In the long run what is the fraction of dividends that will be paid?

1.24. Census results reveal that in the United States 80% of the daughters of working women work and that 30% of the daughters of nonworking women work. (a) Write the transition probability for this model. (b) In the long run what fraction of women will be working?

1.25. When a basketball player makes a shot then he tries a harder shot the next time and hits (H) with probability 0.4, misses (M) with probability 0.6. When he misses he is more conservative the next time and hits (H) with probability 0.7, misses (M) with probability 0.3. (a) Write the transition probability for the two state Markov chain with state space $\{H, M\}$. (b) Find the long-run fraction of time he hits a shot.

1.26. Folk wisdom holds that in Ithaca in the summer it rains 1/3 of the time, but a rainy day is followed by a second one with probability 1/2. Suppose that Ithaca weather is a Markov chain. What is its transition probability?

Chains with three or more states

1.27. (a) Suppose brands A and B have consumer loyalties of .7 and .8, meaning that a customer who buys A one week will with probability .7 buy it again the next week, or try the other brand with .3. What is the limiting market share for each of these products? (b) Suppose now there is a third brand with loyalty .9, and that a consumer who changes brands picks one of the other two at random. What is the new limiting market share for these three products?

1.28. A midwestern university has three types of health plans: a health maintenance organization (HMO), a preferred provider organization (PPO), and a traditional fee for service plan (FFS). Experience dictates that people change plans according to the following transition matrix

	HMO	**PPO**	**FFS**
HMO	.85	.1	.05
PPO	.2	.7	.1
FFS	.1	.3	.6

In 2000, the percentages for the three plans were HMO:30%, PPO:25%, and FFS:45%. (a) What will be the percentages for the three plans in 2001? (b) What is the long run fraction choosing each of the three plans?

1.29. Bob eats lunch at the campus food court every week day. He either eats Chinese food, Quesadila, or Salad. His transition matrix is

	C	**Q**	**S**
C	.15	.6	.25
Q	.4	.1	.5
S	.1	.3	.6

He had Chinese food on Monday. (a) What are the probabilities for his three meal choices on Friday (four days later). (b) What are the long run frequencies for his three choices?

1.30. The liberal town of Ithaca has a "free bikes for the people program." You can pick up bikes at the library (L), the coffee shop (C) or the cooperative grocery store (G). The director of the program has determined that bikes move around accroding to the following Markov chain

	L	**C**	**G**
L	.5	.2	.3
C	.4	.5	.1
G	.25	.25	.5

On Sunday there are an equal number of bikes at each place. (a) What fraction of the bikes are at the three locations on Tuesday? (b) on the next Sunday? (c) In the long run what fraction are at the three locations?

1.31. A plant species has red, pink, or white flowers according to the genotypes RR, RW, and WW, respectively. If each of these genotypes is crossed with a pink (RW) plant then the offspring fractions are

	RR	**RW**	**WW**
RR	.5	.5	0
RW	.25	.5	.25
WW	0	.5	.5

What is the long run fraction of plants of the three types?

1.32. The weather in a certain town is classified as rainy, cloudy, or sunny and changes according to the following transition probability is

	R	**C**	**S**
R	1/2	1/4	1/4
C	1/4	1/2	1/4
S	1/2	1/2	0

In the long run what proportion of days in this town are rainy? cloudy? sunny?

1.33. A sociologist studying living patterns in a certain region determines that the pattern of movement between urban (U), suburban (S), and rural areas (R) is given by the following transition matrix.

	U	**S**	**R**
U	.86	.08	.06
S	.05	.88	.07
R	.03	.05	.92

In the long run what fraction of the population will live in the three areas?

1.34. In a large metropolitan area, commuters either drive alone (A), carpool (C), or take public transportation (T). A study showed that transportation changes according to the following matrix:

	A	**C**	**T**
A	.8	.15	.05
C	.05	.9	.05
S	.05	.1	.85

In the long run what fraction of commuters will use the three types of transportation?

1.35. (a) Three telephone companies A, B, and C compete for customers. Each year customers switch between companies according the followinng transition probability

	A	**B**	**C**
A	.75	.05	.20
B	.15	.65	.20
C	.05	.1	.85

What is the limiting market share for each of these companies?

1.36. A professor has two light bulbs in his garage. When both are burned out, they are replaced, and the next day starts with two working light bulbs. Suppose that when both are working, one of the two will go out with probability .02 (each has probability .01 and we ignore the possibility of losing two on the same day). However, when only one is there, it will burn out with probability .05. (i) What is the long-run fraction of time that there is exactly one bulb working? (ii) What is the expected time between light bulb replacements?

1.37. An individual has three umbrellas, some at her office, and some at home. If she is leaving home in the morning (or leaving work at night) and it is raining, she will take an umbrella, if one is there. Otherwise, she gets wet. Assume that independent of the past, it rains on each trip with probability 0.2. To formulate a Markov chain, let X_n be the number of umbrellas at her current location. (a) Find the transition probability for this Markov chain. (b) Calculate the limiting fraction of time she gets wet.

1.38. Let X_n be the number of days since David last shaved, calculated at 7:30AM when he is trying to decide if he wants to shave today. Suppose that X_n is a Markov chain with transition matrix

	1	**2**	**3**	**4**
1	1/2	1/2	0	0
2	2/3	0	1/3	0
3	3/4	0	0	1/4
4	1	0	0	0

In words, if he last shaved k days ago, he will not shave with probability 1/(k+1). However, when he has not shaved for 4 days his mother orders him to shave, and he does so with probability 1. (a) What is the long-run fraction of time David shaves? (b) Does the stationary distribution for this chain satisfy the detailed balance condition?

1.39. In a particular county voters declare themselves as members of the Republican, Democrat, or Green party. No voters change directly from the Republican to Green party or vice versa. Other transitions occur according to the following matrix:

	R	**D**	**G**
R	.85	.15	0
D	.05	.85	.10
G	0	.05	.95

In the long run what fraction of voters will belong to the three parties?

1.40. An auto insurance company classifies its customers in three categories: poor, satisfactory and excellent. No one moves from poor to excellent or from excellent to poor in one year.

	P	**S**	**E**
P	.6	.4	0
S	.1	.6	.3
E	0	.2	.8

What is the limiting fraction of drivers in each of these categories?

1.41. *Reflecting random walk on the line.* Consider the points $1, 2, 3, 4$ to be marked on a straight line. Let X_n be a Markov chain that moves to the right with probability 2/3 and to the left with probability 1/3, but subject this time to the rule that if X_n tries to go to the left from 1 or to the right from 4 it stays put. Find (a) the transition probability for the chain, and (b) the limiting amount of time the chain spends at each site.

1.42. At the end of a month, a large retail store classifies each of its customer's accounts according to current (0), 30–60 days overdue (1), 60–90 days overdue (2), more than 90 days (3). Their experience indicates that the accounts move from state to state according to a Markov chain with transition probability matrix:

	0	**1**	**2**	**3**
0	.9	.1	0	0
1	.8	0	.2	0
2	.5	0	0	.5
3	.1	0	0	.9

In the long run what fraction of the accounts are in each category?

1.43. At the beginning of each day, a piece of equipment is inspected to determine its working condition, which is classified as state 1 = new, 2, 3, or 4 = broken. We assume the state is a Markov chain with the following transition matrix:

	1	**2**	**3**	**4**
1	.95	.05	0	0
2	0	.9	.1	0
3	0	0	.875	.125

(a) Suppose that a broken machine requires three days to fix. To incorporate this into the Markov chain we add states 5 and 6 and suppose that $p(4,5) = 1$, $p(5,6) = 1$, and $p(6,1) = 1$. Find the fraction of time that the machine is working. (b) Suppose now that we have the option of performing preventative maintenance when the machine is in state 3, and that this maintenance takes one day and returns the machine to state 1. This changes the transition probability to

	1	**2**	**3**
1	.95	.05	0
2	0	.9	.1
3	1	0	0

Find the fraction of time the machine is working under this new policy.

1.44. *Landscape dynamics.* To make a crude model of a forest we might introduce states 0 = grass, 1 = bushes, 2 = small trees, 3 = large trees, and write down a transition matrix like the following:

	0	**1**	**2**	**3**
0	1/2	1/2	0	0
1	1/24	7/8	1/12	0
2	1/36	0	8/9	1/12
3	1/8	0	0	7/8

The idea behind this matrix is that if left undisturbed a grassy area will see bushes grow, then small trees, which of course grow into large trees. However, disturbances such as tree falls or fires can reset the system to state 0. Find the limiting fraction of land in each of the states.

More Theoretical Exercises

1.45. Consider a general chain with state space $S = \{1, 2\}$ and write the transition probability as

	1	**2**
1	$1-a$	a
2	b	$1-b$

Use the Markov property to show that

$$P(X_{n+1} = 1) - \frac{b}{a+b} = (1-a-b)\left\{P(X_n = 1) - \frac{b}{a+b}\right\}$$

and then conclude

$$P(X_n = 1) = \frac{b}{a+b} + (1-a-b)^n \left\{ P(X_0 = 1) - \frac{b}{a+b} \right\}$$

This shows that if $0 < a+b < 2$, then $P(X_n = 1)$ converges exponentially fast to its limiting value $b/(a+b)$.

1.46. *Bernoulli–Laplace model of diffusion.* Consider two urns each of which contains m balls; b of these $2m$ balls are black, and the remaining $2m-b$ are white. We say that the system is in state i if the first urn contains i black balls and $m-i$ white balls while the second contains $b-i$ black balls and $m-b+i$ white balls. Each trial consists of choosing a ball at random from each urn and exchanging the two. Let X_n be the state of the system after n exchanges have been made. X_n is a Markov chain. (a) Compute its transition probability. (b) Verify that the stationary distribution is given by

$$\pi(i) = \binom{b}{i}\binom{2m-b}{m-i} \Big/ \binom{2m}{m}$$

(c) Can you give a simple intuitive explanation why the formula in (b) gives the right answer?

1.47. *Library chain.* On each request the ith of n possible books is the one chosen with probability p_i. To make it quicker to find the book the next time, the librarian moves the book to the left end of the shelf. Define the state at any time to be the sequence of books we see as we examine the shelf from left to right. Since all the books are distinct this list is a permutation of the set $\{1, 2, \ldots n\}$, i.e., each number is listed exactly once. Show that

$$\pi(i_1, \ldots, i_n) = p_{i_1} \cdot \frac{p_{i_2}}{1 - p_{i_1}} \cdot \frac{p_{i_3}}{1 - p_{i_1} - p_{i_2}} \cdots \frac{p_{i_n}}{1 - p_{i_1} - \cdots p_{i_{n-1}}}$$

is a stationary distribution.

1.48. *Random walk on a clock.* Consider the numbers $1, 2, \ldots 12$ written around a ring as they usually are on a clock. Consider a Markov chain that at any point jumps with equal probability to the two adjacent numbers. (a) What is the expected number of steps that X_n will take to return to its starting position? (b) What is the probability X_n will visit all the other states before returning to its starting position?

The next three examples continue Example 1.34. Again we represent our chessboard as $\{(i,j) : 1 \le i,j \le 8\}$. How do you think that the pieces bishop, knight, king, queen, and rook rank in their answers to (b)?

1.49. *King's random walk.* A king can move one squares horizontally, vertically, or diagonally. Let X_n be the sequence of squares that results if we pick one of king's legal moves at random. Find (a) the stationary distribution and (b) the expected number of moves to return to corner (1,1) when we start there.

1.50. *Bishop's random walk.* A bishop can move any number of squares diagonally. Let X_n be the sequence of squares that results if we pick one of bishop's legal moves at random. Find (a) the stationary distribution and (b) the expected number of moves to return to corner (1,1) when we start there.

1.51. *Queen's random walk.* A queen can move any number of squares horizontally, vertically, or diagonally. Let X_n be the sequence of squares that results if we pick one of queen's legal moves at random. Find (a) the stationary distribution and (b) the expected number of moves to return to corner (1,1) when we start there.

1.52. *Wright–Fisher model.* Consider the chain described in Example 1.7.

$$p(x,y) = \binom{N}{y} (\rho_x)^y (1-\rho_x)^{N-y}$$

where $\rho_x = (1-u)x/N + v(N-x)/N$. (a) Show that if $u,v > 0$, then $\lim_{n\to\infty} p^n(x,y) = \pi(y)$, where π is the unique stationary distribution. There is no known formula for $\pi(y)$, but you can (b) compute the mean $\nu = \sum_y y\pi(y) = \lim_{n\to\infty} E_x X_n$.

1.53. *Ehrenfest chain.* Consider the Ehrenfest chain, Example 1.2, with transition probability $p(i,i+1) = (N-i)/N$, and $p(i,i-1) = i/N$ for $0 \le i \le N$. Let $\mu_n = E_x X_n$. (a) Show that $\mu_{n+1} = 1 + (1-2/N)\mu_n$. (b) Use this and induction to conclude that

$$\mu_n = \frac{N}{2} + \left(1 - \frac{2}{N}\right)^n (x - N/2)$$

From this we see that the mean μ_n converges exponentially rapidly to the equilibrium value of $N/2$ with the error at time n being $(1-2/N)^n(x-N/2)$.

1.54. Prove that if $p_{ij} > 0$ for all i and j then a necessary and sufficient condition for the existence of a reversible stationary distribution is

$$p_{ij}p_{jk}p_{ki} = p_{ik}p_{kj}p_{ji} \quad \text{for all } i, j, k$$

Hint: fix i and take $\pi_j = cp_{ij}/p_{ji}$.

Exit distributions and times

1.55. The Markov chain associated with a manufacturing process may be described as follows: A part to be manufactured will begin the process by entering step 1. After step 1, 20% of the parts must be reworked, i.e., returned to step 1, 10% of the parts are thrown away, and 70% proceed to step 2. After step 2, 5% of the parts must be returned to the step 1, 10% to step 2, 5% are scrapped, and 80% emerge to be sold for a profit. (a) Formulate a four-state Markov chain with states 1, 2, 3, and 4 where 3 = a part that was scrapped and 4 = a part that was sold for a profit. (b) Compute the probability a part is scrapped in the production process.

1.56. A bank classifies loans as paid in full (F), in good standing (G), in arrears (A), or as a bad debt (B). Loans move between the categories according to the following transition probability:

	F	**G**	**A**	**B**
F	1	0	0	0
G	.1	.8	.1	0
A	.1	.4	.4	.1
B	0	0	0	1

What fraction of loans in good standing are eventually paid in full? What is the answr for those in arrears?

1.57. A warehouse has a capacity to hold four items. If the warehouse is neither full nor empty, the number of items in the warehouse changes whenever a new item is produced or an item is sold. Suppose that (no matter when we look) the probability that the next event is "a new item is produced" is 2/3 and that the new event is a "sale" is 1/3. If there is currently one item in the warehouse, what is the probability that the warehouse will become full before it becomes empty.

1.58. Six children (Dick, Helen, Joni, Mark, Sam, and Tony) play catch. If Dick has the ball he is equally likely to throw it to Helen, Mark, Sam, and Tony. If Helen has the ball she is equally likely to throw it to Dick, Joni, Sam, and Tony. If Sam has the ball he is equally likely to throw it to Dick, Helen, Mark, and Tony. If either Joni or Tony gets the ball, they keep throwing it to each other. If Mark gets the ball he runs away with it. (a) Find the transition probability and classify the states of the chain. (b) Suppose Dick has the ball at the beginning of the game. What is the probability Mark will end up with it?

1.59. *Sucker bet.* Consider the following gambling game. Player 1 picks a three coin pattern (for example HTH) and player 2 picks another (say THH). A coin is flipped repeatedly and outcomes are recorded until one of the two patterns appears. Somewhat surprisingly player 2 has a considerable advantage in this game. No matter what player 1 picks, player 2 can win with probability $\geq 2/3$. Suppose without loss of generality that player 1 picks a pattern that begins with H:

case	Player 1	Player 2	Prob. 2 wins
1	HHH	THH	7/8
2	HHT	THH	3/4
3	HTH	HHT	2/3
4	HTT	HHT	2/3

Verify the results in the table. You can do this by solving six equations in six unknowns but this is not the easiest way.

1.60. At the New York State Fair in Syracuse, Larry encounters a carnival game where for one dollar he may buy a single coupon allowing him to play a guessing game. On each play, Larry has an even chance of winning or losing a coupon. When he runs out of coupons he loses the game. However, if he can collect three coupons, he wins a surprise. (a) What is the probability David will win the surprise? (b) What is the expected number of plays he needs to win or lose the game.

1.61. The Megasoft company gives each of its employees the title of programmer (P) or project manager (M). In any given year 70% of programmers remain in that position 20% are promoted to project manager and 10% are fired (state X). 95% of project managers remain in that position while 5% are fired. How long on the average does a programmer work before they are fired?

1.62. At a nationwide travel agency, newly hired employees are classified as beginners (B). Every six months the performance of each agent is reviewed. Past records indicate that transitions through the ranks to intermediate (I) and qualified (Q) are according to the following Markov chain, where F indicates workers that were fired:

	B	**I**	**Q**	**F**
B	.45	.4	0	.15
I	0	.6	.3	.1
Q	0	0	1	0
F	0	0	0	1

(a) What fraction are eventually promoted? (b) What is the expected time until a beginner is fired or becomes qualified?

1.63. At a manufacturing plant, employees are classified as trainee (R), technician (T) or supervisor (S). Writing Q for an employee who quits we model their progress through the ranks as a Markov chain with transition probability

	R	**T**	**S**	**Q**
R	.2	.6	0	.2
T	0	.55	.15	.3
S	0	0	1	0
Q	0	0	0	1

(a) What fraction of recruits eventually make supervisor? (b) What is the expected time until a trainee auits or becomes supervisor?

1.64. Customers shift between variable rate loans (V), thirty year fixed-rate loans (30), fifteen year fixed-rate loans (15), or enter the states paid in full (P), or foreclosed according to the following transition matrix:

	V	**30**	**15**	**P**	**f**
V	.55	.35	0	.05	.05
30	.15	.54	.25	.05	.01
15	.20	0	.75	.04	.01
P	0	0	0	1	0
F	0	0	0	0	1

(a) For each of the three loan types find (a) the expected time until paid or foreclosed. (b) the probability the loan is paid.

1.65. *Brother–sister mating.* In this genetics scheme two individuals (one male and one female) are retained from each generation and are mated to give the next. If the individuals involved are diploid and we are interested in a trait with two alleles, A and a, then each individual has three possible states AA, Aa, aa or more succinctly 2, 1, 0. If we keep track of the sexes of the two individuals the chain has nine states, but if we ignore the sex there are just six: 22, 21, 20, 11, 10, and 00. (a) Assuming that reproduction corresponds to picking one letter at random from each parent, compute the transition probability. (b) 22 and 00 are absorbing states for the chain. Show that the probability of absorption in 22 is equal to the fraction of A's in the state. (c) Let $T = \min\{n \geq 0 : X_n = 22 \text{ or } 00\}$ be the absorption time. Find $E_x T$ for all states x.

1.66. Use the second solution in Example 1.45 to compute the expected waiting times for the patterns HHH, HHT, HTT, and HTH. Which pattern has the longest waiting time? Which ones achieve the minimum value of 8?

1.67. Roll a fair die repeatedly and let $Y_1, Y_2, \ldots$ be the resulting numbers. Let $X_n = |\{Y_1, Y_2, \ldots, Y_n\}|$ be the number of values we have seen in the first n rolls for $n \geq 1$ and set $X_0 = 0$. X_n is a Markov chain. (a) Find its transition probability. (b) Let $T = \min\{n : X_n = 6\}$ be the number of trials we need to see all 6 numbers at least once. Find ET.

1.68. *Coupon collector's problem.* We are interested now in the time it takes to collect a set of N baseball cards. Let T_k be the number of cards we have to buy before we have k that are distinct. Clearly, $T_1 = 1$. A little more thought reveals that if each time we get a card chosen at random from all N possibilities, then for $k \geq 1$, $T_{k+1} - T_k$ has a geometric distribution with success probability $(N-k)/N$. Use this to show that the mean time to collect a set of N baseball cards is $\approx N \log N$, while the variance is $\approx N^2 \sum_{k=1}^{\infty} 1/k^2$.

1.69. *Algorthmic efficiency.* The simplex method minimizes linear functions by moving between extreme points of a polyhedral region so that each transition decreases the objective function. Suppose there are n extreme points and they are numbered in increasing

order of their values. Consider the Markov chain in which $p(1,1) = 1$ and $p(i,j) = 1/i - 1$ for $j < i$. In words, when we leave j we are equally likely to go to any of the extreme points with better value.
(a) Use (1.24) to show that for $i > 1$

$$E_i T_1 = 1 + 1/2 + \cdots + 1/(i-1)$$

(b) Let $I_j = 1$ if the chain visits j on the way from n to 1. Show that for $j < n$

$$P(I_j = 1 | I_{j+1}, \ldots I_n) = 1/j$$

to get another proof of the result and conclude that $I_1, \ldots I_{n-1}$ are independent.

Infinite State Space

1.70. *General birth and death chains.* The state space is $\{0, 1, 2, \ldots\}$ and the transition probability has

$$\begin{array}{ll} p(x, x+1) = p_x & \\ p(x, x-1) = q_x & \text{for } x > 0 \\ p(x, x) = r_x & \text{for } x \geq 0 \end{array}$$

while the other $p(x,y) = 0$. Let $V_y = \min\{n \geq 0 : X_n = y\}$ be the time of the first visit to y and let $h_N(x) = P_x(V_N < V_0)$. By considering what happens on the first step, we can write

$$h_N(x) = p_x h_N(x+1) + r_x h_N(x) + q_x h_N(x-1)$$

Set $h_N(1) = c_N$ and solve this equation to conclude that 0 is recurrent if and only if $\sum_{y=1}^{\infty} \prod_{x=1}^{y-1} q_x/p_x = \infty$ where by convention $\prod_{x=1}^{0} = 1$.

1.71. To see what the conditions in the last problem say we will now consider some concrete examples. Let $p_x = 1/2$, $q_x = e^{-cx^{-\alpha}}/2$, $r_x = 1/2 - q_x$ for $x \geq 1$ and $p_0 = 1$. For large x, $q_x \approx (1 - cx^{-\alpha})/2$, but the exponential formulation keeps the probabilities nonnegative and makes the problem easier to solve. Show that the chain is recurrent if $\alpha > 1$ or if $\alpha = 1$ and $c \leq 1$ but is transient otherwise.

1.72. Consider the Markov chain with state space $\{0, 1, 2, \ldots\}$ and transition probability

$$\begin{aligned} p(m, m+1) &= \frac{1}{2}\left(1 - \frac{1}{m+2}\right) \qquad \text{for } m \geq 0 \\ p(m, m-1) &= \frac{1}{2}\left(1 + \frac{1}{m+2}\right) \qquad \text{for } m \geq 1 \end{aligned}$$

and $p(0, 0) = 1 - p(0, 1) = 3/4$. Find the stationary distribution π.

1.73. Consider the Markov chain with state space $\{1, 2, \ldots\}$ and transition probability

$$\begin{aligned} p(m, m+1) &= m/(2m+2) \qquad \text{for } m \geq 1 \\ p(m, m-1) &= 1/2 \qquad \text{for } m \geq 2 \\ p(m, m) &= 1/(2m+2) \qquad \text{for } m \geq 2 \end{aligned}$$

and $p(1, 1) = 1 - p(1, 2) = 3/4$. Show that there is no stationary distribution.

1.74. Consider the aging chain on $\{0, 1, 2, \ldots\}$ in which for any $n \geq 0$ the individual gets one day older from n to $n+1$ with probability p_n but dies and returns to age 0 with probability $1-p_n$. Find conditions that guarantee that (a) 0 is recurrent, (b) positive recurrent. (c) Find the stationary distribution.

1.75. The opposite of the aging chain is the renewal chain with state space $\{0, 1, 2, \ldots\}$ in which $p(i, i-1) = 1$ when $i > 0$. The only nontrivial part of the transition probability is $p(0, i) = p_i$. Show that this chain is always recurrent but is positive recurrent if and only if $\sum_n np_n < \infty$.

1.76. Consider a branching process as defined in Example 7.2, in which each family has exactly three children, but invert Galton and Watson's original motivation and ignore male children. In this model a mother will have an average of 1.5 daughters. Compute the probability that a given woman's descendents will die out.

1.77. Consider a branching process as defined in Example 7.2, in which each family has a number of children that follows a shifted geometric distribution: $p_k = p(1-p)^k$ for $k \geq 0$, which counts the number of failures before the first success when success has probability p. Compute the probability that starting from one individual the chain will be absorbed at 0.

Chapter 2

Poisson Processes

2.1 Exponential Distribution

To prepare for our discussion of the Poisson process, we need to recall the definition and some of the basic properties of the exponential distribution. A random variable T is said to have **an exponential distribution with rate** λ, or $T = \text{exponential}(\lambda)$, if

$$P(T \le t) = 1 - e^{-\lambda t} \quad \text{for all } t \ge 0 \tag{2.1}$$

Here we have described the distribution by giving the **distribution function** $F(t) = P(T \le t)$. We can also write the definition in terms of the **density function** $f_T(t)$ which is the derivative of the distribution function.

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \ge 0 \\ 0 & \text{for } t < 0 \end{cases} \tag{2.2}$$

Integrating by parts with $f(t) = t$ and $g'(t) = \lambda e^{-\lambda t}$,

$$\begin{aligned} ET &= \int t\, f_T(t)\, dt = \int_0^\infty t \cdot \lambda e^{-\lambda t}\, dt \\ &= -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t}\, dt = 1/\lambda \end{aligned} \tag{2.3}$$

Integrating by parts with $f(t) = t^2$ and $g'(t) = \lambda e^{-\lambda t}$, we see that

$$\begin{aligned} ET^2 &= \int t^2\, f_T(t)\, dt = \int_0^\infty t^2 \cdot \lambda e^{-\lambda t}\, dt \\ &= -t^2 e^{-\lambda t}\Big|_0^\infty + \int_0^\infty 2te^{-\lambda t}\, dt = 2/\lambda^2 \end{aligned}$$

by the formula for ET. So the variance

$$\mathrm{var}\,(T) = ET^2 - (ET)^2 = 1/\lambda^2 \tag{2.4}$$

While calculus is required to know the exact values of the mean and variance, it is easy to see how they depend on λ. Let $T =$ exponential(λ), i.e., have an exponential distribution with rate λ, and let $S =$ exponential(1). To see that S/λ has the same distribution as T, we use (2.1) to conclude

$$P(S/\lambda \le t) = P(S \le \lambda t) = 1 - e^{-\lambda t} = P(T \le t)$$

Recalling that if c is any number then $E(cX) = cEX$ and $\mathrm{var}\,(cX) = c^2\,\mathrm{var}\,(X)$, we see that

$$ET = ES/\lambda \qquad \mathrm{var}\,(T) = \mathrm{var}\,(S)/\lambda^2$$

Lack of memory property. It is traditional to formulate this property in terms of waiting for an unreliable bus driver. In words, "if we've been waiting for t units of time then the probability we must wait s more units of time is the same as if we haven't waited at all." In symbols

$$P(T > t + s | T > t) = P(T > s) \tag{2.5}$$

To prove this we recall that if $B \subset A$, then $P(B|A) = P(B)/P(A)$, so

$$P(T > t + s | T > t) = \frac{P(T > t + s)}{P(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s)$$

where in the third step we have used the fact $e^{a+b} = e^a e^b$.

Exponential races. Let $S =$ exponential(λ) and $T =$ exponential(μ) be independent. In order for the minimum of S and T to be larger than t, each of S and T must be larger than t. Using this and independence we have

$$\begin{aligned} P(\min(S,T) > t) &= P(S > t, T > t) = P(S > t)P(T > t) \\ &= e^{-\lambda t} e^{-\mu t} = e^{-(\lambda+\mu)t} \end{aligned} \tag{2.6}$$

That is, $\min(S,T)$ has an exponential distribution with rate $\lambda+\mu$. The last calculation extends easily to a sequence of independent random variables $T_1, \ldots, T_n$ where $T_i = \text{exponential}(\lambda_i)$.

$$\begin{aligned} P(\min(T_1, \ldots, T_n) > t) &= P(T_1 > t, \ldots T_n > t) \\ &= \prod_{i=1}^n P(T_i > t) = \prod_{i=1}^n e^{-\lambda_i t} = e^{-(\lambda_1 + \cdots + \lambda_n)t} \end{aligned} \tag{2.7}$$

That is, the minimum, $\min(T_1, \ldots, T_n)$, of several independent exponentials has an exponential distribution with rate equal to the sum of the rates $\lambda_1 + \cdots \lambda_n$.

In the last paragraph we have computed the duration of a race between exponentially distributed random variables. We will now consider: "Who finishes first?" Going back to the case of two random variables, we break things down according to the value of S and then using independence with our formulas (2.1) and (2.2) for the distribution and density functions, to conclude

$$\begin{aligned} P(S < T) &= \int_0^\infty f_S(s) P(T > s)\, ds \\ &= \int_0^\infty \lambda e^{-\lambda s} e^{-\mu s}\, ds \\ &= \frac{\lambda}{\lambda+\mu} \int_0^\infty (\lambda+\mu) e^{-(\lambda+\mu)s}\, ds = \frac{\lambda}{\lambda+\mu} \end{aligned} \tag{2.8}$$

where on the last line we have used the fact that $(\lambda+\mu)e^{-(\lambda+\mu)s}$ is a density function and hence must integrate to 1. Of course, one can also use calculus to evaluate the integral.

From the calculation for two random variables, you should be able to guess that if $T_1, \ldots, T_n$ are independent exponentials, then

$$P(T_i = \min(T_1, \ldots, T_n)) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n} \tag{2.9}$$

That is, the probability of i finishing first is proportional to its rate.

Proof. Let $S = T_i$ and U be the minimum of T_j, $j \neq i$. (2.7) implies that U is exponential with parameter

$$\mu = (\lambda_1 + \cdots + \lambda_n) - \lambda_i$$

so using the result for two random variables

$$P(T_i = \min(T_1, \ldots, T_n)) = P(S < U) = \frac{\lambda_i}{\lambda_i + \mu} = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$$

proves the desired result. $\square$

Let I be the (random) index of the T_i that is smallest. In symbols,

$$P(I = i) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$$

You might think that the T_i's with larger rates might be more likely to win early. However,

$$I \textit{ and } V = \min\{T_1, \ldots T_n\} \textit{ are independent.} \tag{2.10}$$

Proof. Let $f_{i,V}(t)$ be the density function for V on the set $I = i$. In order for i to be first at time t, $T_i = t$ and the other $T_j > t$ so

$$\begin{aligned} f_{i,V}(t) &= \lambda_i e^{-\lambda_i t} \cdot \prod_{j \neq i} e^{-\lambda_j t} \\ &= \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n} \cdot (\lambda_1 + \cdots + \lambda_n) e^{-(\lambda_1 + \cdot + \lambda_n)t} \\ &= P(I = i) \cdot f_V(t) \end{aligned}$$

since V has an exponential$(\lambda_1 + \cdots + \lambda_n)$ distribution. $\square$

Our final fact in this section concerns sums of exponentials.

Theorem 2.1. *Let $\tau_1, \tau_2, \ldots$ be independent exponential(λ). The sum $T_n = \tau_1 + \cdots + \tau_n$ has a gamma(n, λ) distribution. That is, the density function of T_n is given by*

$$f_{T_n}(t) = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{n-1}}{(n-1)!} \quad \textit{for } t \geq 0 \tag{2.11}$$

and 0 otherwise.

Proof. The proof is by induction on n. When $n = 1$, T_1 has an exponential(λ) distribution. Recalling that the 0th power of any positive number is 1, and by convention we set 0!=1, the formula reduces to

$$f_{T_1}(t) = \lambda e^{-\lambda t}$$

and we have shown that our formula is correct for $n = 1$.

To do the induction step suppose that the formula is true for n. The sum $T_{n+1} = T_n + \tau_{n+1}$, so breaking things down according to the value of T_n, and using the independence of T_n and t_{n+1}, we have

$$f_{T_{n+1}}(t) = \int_0^t f_{T_n}(s) f_{t_{n+1}}(t-s)\, ds$$

Plugging the formula from (2.11) in for the first term and the exponential density in for the second and using the fact that $e^a e^b = e^{a+b}$ with $a = -\lambda s$ and $b = -\lambda(t-s)$ gives

$$\begin{aligned} \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} \cdot \lambda e^{-\lambda(t-s)}\, ds &= e^{-\lambda t} \lambda^n \int_0^t \frac{s^{n-1}}{(n-1)!}\, ds \\ &= \lambda e^{-\lambda t} \frac{\lambda^n t^n}{n!} \end{aligned}$$

which completes the proof. $\square$

2.2 Defining the Poisson Process

In this section we will give two definitions of the **Poisson process with rate** λ. The first, which will be our official definition, is nice because it allows us to construct the process easily.

Definition. Let $\tau_1, \tau_2, \ldots$ be independent exponential(λ) random variables. Let $T_n = \tau_1 + \cdots + \tau_n$ for $n \geq 1$, $T_0 = 0$, and define $N(s) = \max\{n : T_n \leq s\}$.

We think of the τ_n as times between arrivals of customers at a bank, so $T_n = \tau_1 + \cdots + \tau_n$ is the arrival time of the nth customer, and $N(s)$ is the number of arrivals by time s. To check the last interpretation, consider the following example:

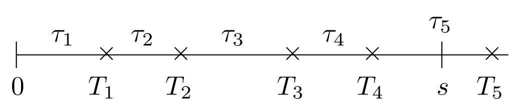


Figure 2.1: Poisson process definitions.

and note that $N(s) = 4$ when $T_4 \leq s < T_5$, that is, the 4th customer has arrived by time s but the 5th has not.

Recall that X has a **Poisson distribution** with mean λ, or $X =$ Poisson(λ), for short, if

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad \text{for } n = 0, 1, 2, \ldots$$

To explain why $N(s)$ is called the Poisson process rather than the exponential process, we will compute the distribution of $N(s)$.

Lemma 2.1. *$N(s)$ has a Poisson distribution with mean λs.*

Proof. Now $N(s) = n$ if and only if $T_n \leq s < T_{n+1}$; i.e., the nth customer arrives before time s but the $(n+1)$th after s. Breaking things down according to the value of $T_n = t$ and noting that for $T_{n+1} > s$, we must have $\tau_{n+1} > s - t$, and τ_{n+1} is independent of T_n, it follows that

$$P(N(s) = n) = \int_0^s f_{T_n}(t) P(t_{n+1} > s - t)\, dt$$

Plugging in (2.11) now, the last expression is

$$= \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \cdot e^{-\lambda(s-t)}\, dt$$
$$= \frac{\lambda^n}{(n-1)!} e^{-\lambda s} \int_0^s t^{n-1}\, dt = e^{-\lambda s} \frac{(\lambda s)^n}{n!}$$

which proves the desired result. □

Since this is our first mention of the Poisson distribution, we pause to derive some of its properties.

Theorem 2.2. *For any $k \geq 1$*

$$EX(X-1)\cdots(X-k+1) = \lambda^k \tag{2.12}$$

and hence $var(X) = \lambda$

Proof. $X(X-1)\cdots(X-k+1) = 0$ if $X \leq k-1$ so

$$EX(X-1)\cdots(X-k+1) = \sum_{j=k}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} j(j-1)\cdots(j-k+1)$$
$$= \lambda^k \sum_{j=k}^{\infty} e^{-\lambda} \frac{\lambda^{j-k}}{(j-k)!} = \lambda^k$$

since the sum gives the total mass of the Poisson distribution. Using $\text{var}(X) = E(X(X-1)) + EX - (EX)^2$ we conclude

$$\text{var}(X) = \lambda^2 + \lambda - (\lambda)^2 = \lambda$$ □

Theorem 2.3. *If X_i are independent Poissson(λ_i) then*

$$X_1 + \cdots + X_k = Poisson(\lambda_1 + \cdots + \lambda_n).$$

Proof. It suffices to prove the result for $k = 2$, for then the general result follows by induction. If $X_1 + X_2 = n$ then for some $0 \leq m \leq n$, $X_1 = m$ and $X_2 = n - m$

$$P(X_1 + X_2 = n) = \sum_{m=0}^{n} P(X_1 = m)P(X_2 = n-m)$$
$$= \sum_{m=0}^{n} e^{-\lambda_1 t} \frac{(\lambda_1 t)^m}{m!} \cdot e^{-\lambda_2 t} \frac{(\lambda_2 t)^{n-m}}{(n-m)!}$$

Knowing the answer we want, we can rewrite the last expression as

$$e^{-(\lambda_1+\lambda_2)t}\frac{(\lambda_1+\lambda_2)^n t)^n}{n!}\cdot\sum_{m=0}^{n}\binom{n}{m}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^m\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-m}$$

The sum is 1, since it is the sum of all the probabilities for a binomial(n,p) distribution with $p=\lambda_1/(\lambda_1+\lambda_2)$. The term outside the sum is the desired Poisson probability, so have proved the desired result. $\square$

The property of the Poisson process in Lemma 2.1 is the first part of our second definition. To start to develop the second part we show:

Lemma 2.2. *$N(t+s)-N(s)$, $t\geq 0$ is a rate λ Poisson process and independent of $N(r)$, $0\leq r\leq s$.*

Why is this true? Suppose for concreteness (and so that we can use Figure 2.2 at the beginning of this section again) that by time s there have been four arrivals T_1,T_2,T_3,T_4 that occurred at times t_1,t_2,t_3,t_4. We know that the waiting time for the fifth arrival must have $\tau_5>s-t_4$, but by the lack of memory property of the exponential distribution (2.5)

$$P(\tau_5>s-t_4+t|\tau_5>s-t_4)=P(\tau_5>t)=e^{-\lambda t}$$

This shows that the distribution of the first arrival after s is exponential(λ) and independent of T_1,T_2,T_3,T_4. It is clear that $\tau_6,\tau_7,\ldots$ are independent of T_1,T_2,T_3,T_4, and τ_5. This shows that the interarrival times after s are independent exponential(λ), and hence that $N(t+s)-N(s)$, $t\geq 0$ is a Poisson process. $\square$

From Lemma 2.2 we get easily the following:

Lemma 2.3. *$N(t)$ has* **independent increments***: if $t_0<t_1<\ldots<t_n$, then*

$$N(t_1)-N(t_0),N(t_2)-N(t_1),\ldots N(t_n)-N(t_{n-1})\quad \textit{are independent}$$

Why is this true? Lemma 2.2 implies that $N(t_n)-N(t_{n-1})$ is independent of $N(r)$, $r\leq t_{n-1}$ and hence of $N(t_{n-1})-N(t_{n-2}),\ldots N(t_1)-N(t_0)$. The desired result now follows by induction. $\square$

We are now ready for our second definition. It is in terms of the process $\{N(s):s\geq 0\}$ that counts the number of arrivals in $[0,s]$.

Theorem 2.4. *If $\{N(s), s \geq 0\}$ is a Poisson process, then*
(i) $N(0) = 0$,
(ii) $N(t+s) - N(s) = \text{Poisson}(\lambda t)$, and
(iii) $N(t)$ has independent increments.
Conversely, if (i), (ii), and (iii) hold, then $\{N(s), s \geq 0\}$ is a Poisson process.

Why is this true? As we remarked above, Lemmas 2.1 and 2.3 give the first statement. To start to prove the converse, let T_n be the time of the nth arrival. The first arrival occurs after time t if and only if there were no arrivals in $[0, t]$. So using the formula for the Poisson distribution

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

This shows that $\tau_1 = T_1$ is exponential(λ). For $\tau_2 = T_2 - T_1$ we note that

$$\begin{aligned} P(t_2 > t | t_1 = s) &= P(\text{ no arrival in } (s, s+t] \, | t_1 = s) \\ &= P(N(t+s) - N(s) = 0 | N(r) = 0 \text{ for } r < s, N(s) = 1) \\ &= P(N(t+s) - N(s) = 0) = e^{-\lambda t} \end{aligned}$$

by the independent increments property in (iii), so τ_2 is exponential(λ) and independent of τ_1. Repeating this argument we see that $\tau_1, \tau_2, \ldots$ are independent exponential(λ). □

Up to this point we have been concerned with the mechanics of defining the Poisson process, so the reader may be wondering:

Why is the Poisson process important for applications?

Our answer is based on the Poisson approximation to the binomial. Consider the restaurant Trillium on the Cornell campus. Suppose that each of the n students on campus independently decides to go to Trillium between 12:00 and 1:00 with probability λ/n, and suppose that a person who chooses to go, will do so at a time chosen at random between 12:00 and 1:00. The probability that exactly k students will go is given by the binomial($n, \lambda/n$) distribution

$$\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \tag{2.13}$$

Theorem 2.5. *If n is large the binomial($n, \lambda/n$) distribution is approximately Poisson(λ).*

Proof. Exchanging the numerators of the first two fractions and breaking the last term into two, (2.13) becomes

$$\frac{\lambda^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \tag{2.14}$$

Taking the four terms from left to right, we have

(i) $\lambda^k/k!$ does not depend on n.

(ii) There are k terms on the top and k terms on the bottom, so we can write this fraction as

$$\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}$$

The first term here is 1; the second is $1 - \frac{1}{n} \to 1$ as $n \to \infty$. This holds true for any fixed value of j, so the second term converges to 1 as $n \to \infty$.

(iii) It is one of the famous facts of calculus that

$$(1 - \lambda/n)^n \to e^{-\lambda} \quad \text{as } n \to \infty.$$

We have broken off the last term to be able to exactly apply this fact.

(iv) $\lambda/n \to 0$, so $1 - \lambda/n \to 1$. The power $-k$ is fixed so

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1^{-k} = 1$$

Combining (i)–(iv), we see that (2.14) converges to

$$\frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1$$

which is the Poisson distribution with mean λ. □

By extending the last argument we can also see why the number of individuals that arrive in two disjoint time intervals should be independent. Using the multinomial instead of the binomial, we see that the probability j people will go between 12:00 and 12:20 and k people will go between 12:20 and 1:00 is

$$\frac{n!}{j!k!(n-j-k)!} \left(\frac{\lambda}{3n}\right)^j \left(\frac{2\lambda}{3n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-(j+k)}$$

Rearranging gives

$$\frac{(\lambda/3)^j}{j!} \cdot \frac{(2\lambda/3)^k}{k!} \cdot \frac{n(n-1)\cdots(n-j-k+1)}{n^{j+k}} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-(j+k)}$$

Reasoning as before shows that when n is large, this is approximately

$$\frac{(\lambda/3)^j}{j!} \cdot \frac{(2\lambda/3)^k}{k!} \cdot 1 \cdot e^{-\lambda}$$

Writing $e^{-\lambda} = e^{-\lambda/3}e^{-2\lambda/3}$ and rearranging we can write the last expression as

$$e^{-\lambda/3}\frac{(\lambda/3)^j}{j!} \cdot e^{-2\lambda/3}\frac{(2\lambda/3)^k}{k!}$$

This shows that the number of arrivals in the two time intervals are independent Poissons with means $\lambda/3$ and $2\lambda/3$.

The last proof can be easily generalized to show that if we divide the hour between 12:00 and 1:00 into any number of intervals, then the arrivals are independent Poissons with the right means. However, the argument gets very messy to write down.

More realistic models.

Two of the weaknesses of the derivation above are:

(i) All students are assumed to have exactly the same probability of going to Trillium.

(ii) Students who choose to go, do so at a time chosen at random between 12:00 and 1:00, so the arrival rate of customers is constant during the hour.

(i) is a very strong assumption but can be weakened by using a more general Poisson approximation result like the following:

Theorem 2.6. *For each n let $X_{n,m}$ be independent random variables with $P(X_{n,m} = 1) = p_{n,m}$ and $P(X_{n,m} = 0) = 1 - p_{n,m}$. Let*

$$S_n = X_{n,1} + \cdots + X_{n,n}, \quad \lambda_n = ES_n = p_{n,1} + \cdots + p_{n,n},$$

and $Z_n = \text{Poisson}(\lambda_n)$. Then for any set A

$$|P(S_n \in A) - P(Z_n \in A)| \leq \sum_{m=1}^{n} p_{n,m}^2$$

Why is this true? If X and Y are integer valued random variables then for any set A

$$|P(X \in A) - P(Y \in A)| \leq \frac{1}{2} \sum_n |P(X = n) - P(Y = n)|$$

The right-hand side is called the **total variation distance** between the two distributions and is denoted $\|X - Y\|$. If $P(X = 1) = p$, $P(X = 0) = 1 - p$, and $Y = \text{Poisson}(p)$ then

$$\sum_n |P(X = n) - P(Y = n)| = |(1-p) - e^{-p}| + |p - pe^{-p}| + 1 - (1+p)e^{-p}$$

Since $1 \geq e^{-p} \geq 1 - p$ the right-hand side is

$$e^{-p} - 1 + p + p - pe^{-p} + 1 - e^{-p} - pe^{-p} = 2p(1 - e^{-p} \leq 2p^2$$

At this point we have shown $\|X - Y\| \leq p^2$. Let $Y_{n,m}$ be independent with a Poisson($p_{n,m}$). With some work one can show

$$\|(X_{n,1} + \cdots + X_{n,n}) - (Y_{n,1} + \cdots + Y_{n,n})\|$$
$$\|(X_{n,1}, \cdots, X_{n,n}) - (Y_{n,1}, \cdots, Y_{n,n})\| \leq \sum_{m=1}^{n} \|X_{n,m} - Y_{n,m}\|$$

and the desired result follows. □

Theorem 2.6 is useful because it gives a bound on the difference between the distribution of S_n and the Poisson distribution with mean $\lambda_n = ES_n$. From it, we immediately get the following convergence theorem.

Theorem 2.7. *If in addition to the assumptions in Theorem 2.6, we suppose that $\lambda_n \to \lambda < \infty$ and $\max_k p_{n,k} \to 0$, then*

$$\max_A |P(S_n \in A) - P(Z_n \in A)| \to 0$$

Proof. Since $p_{n,m}^2 \leq p_{n,m} \cdot \max_k p_{n,k}$, summing over m gives

$$\sum_{m=1}^{n} p_{n,m}^2 \leq \max_k p_{n,k} \left(\sum_m p_{n,m} \right)$$

The first term on the right goes to 0 by assumption. The second is $\lambda_n \to \lambda$. Since we have assumed $\lambda < \infty$, the product of the two terms converges to $0 \cdot \lambda = 0$. □

The last results handles problem (i). We do not need to assume that all students have the same probability, just that all of the probabilities are samll. However, those of us who go to Trillium know that since the fourth class period of the day is 11:15-12:05, the arrival rate is large between 12:05 and 12:15 and only returns to a low level at about 12:30. To address the problem of varying arrival rates, we generalize the definition.

Nonhomogeneous Poisson processes. *We say that* $\{N(s), s \geq 0\}$ *is a Poisson process with rate* $\lambda(r)$ *if*

(i) $N(0) = 0$,

(ii) $N(t)$ *has independent increments, and*

(iii) $N(t+s) - N(s)$ *is Poisson with mean* $\int_s^t \lambda(r)\, dr$.

The first definition does not work well in this setting since the interarrival times $\tau_1, \tau_2, \ldots$ are no longer exponentially distributed or independent. To demonstrate the first claim, we note that

$$P(\tau_1 > t) = P(N(t) = 0) = e^{-\int_0^t \lambda(s)\, ds}$$

since the last expression gives the probability a Poisson with mean $\mu(t) = \int_0^t \lambda(s)\, ds$ is equal to 0. Differentiating gives the density function

$$P(\tau_1 = t) = -\frac{d}{dt} P(t_1 > t) = \lambda(t) e^{-\int_0^t \lambda(s)\, ds} = \lambda(t) e^{-\mu(t)}$$

Generalizing the last computation shows that the joint distribution

$$f_{T_1,T_2}(u, v) = \lambda(u) e^{-\mu(u)} \cdot \lambda(v) e^{-(\mu(v) - \mu(u))}$$

Changing variables, $s = u$, $t = v - u$, the joint density

$$f_{\tau_1,\tau_2}(s, t) = \lambda(s) e^{-\mu(s)} \cdot \lambda(s+t) e^{-(\mu(s+t) - \mu(s))}$$

so τ_1 and τ_2 are not independent when $\lambda(s)$ is not constant.

2.3 Compound Poisson Processes

In this section we will embellish our Poisson process by associating an independent and identically distributed (i.i.d.) random variable Y_i with each arrival. By independent we mean that the Y_i are independent of each other and of the Poisson process of arrivals. To explain why we have chosen these assumptions, we begin with two examples for motivation.

Example 2.1. Consider the McDonald's restaurant on Route 13 in the southern part of Ithaca. By arguments in the last section, it is not unreasonable to assume that between 12:00 and 1:00 cars arrive according to a Poisson process with rate λ. Let Y_i be the number of people in the ith vehicle. There might be some correlation between the number of people in the car and the arrival time, e.g., more families come to eat there at night, but for a first approximation it seems reasonable to assume that the Y_i are i.i.d. and independent of the Poisson process of arrival times.

Example 2.2. Messages arrive at a central computer to be transmitted across the Internet. If we imagine a large number of users working at terminals connected to a central computer, then the arrival times of messages can be modeled by a Poisson process. If we let Y_i be the size of the ith message, then again it is reasonable to assume $Y_1, Y_2, \ldots$ are i.i.d. and independent of the Poisson process of arrival times.

Having introduced the Y_i's, it is natural to consider the sum of the Y_i's we have seen up to time t:

$$S(t) = Y_1 + \cdots + Y_{N(t)}$$

where we set $S(t) = 0$ if $N(t) = 0$. In Example 2.1, $S(t)$ gives the number of customers that have arrived up to time t. In Example 2.2, $S(t)$ represents the total number of bytes in all of the messages up to time t. In each case it is interesting to know the mean and variance of $S(t)$.

Theorem 2.8. *Let $Y_1, Y_2, \ldots$ be independent and identically distributed, let N be an independent nonnegative integer valued random variable, and let $S = Y_1 + \cdots + Y_N$ with $S = 0$ when $N = 0$.*

(i) If $E|Y_i|$, $EN < \infty$, then $ES = EN \cdot EY_i$.

(ii) If EY_i^2, $EN^2 < \infty$, *then* $var(S) = EN\ var(Y_i) + var(N)(EY_i)^2$.

(iii) If N *is Poisson*(λ), *then* $var(S) = \lambda EY_i^2$.

Why is this reasonable? The first of these is natural since if $N = n$ is nonrandom $ES = nEY_i$. (i) then results by setting $n = EN$. The formula in (ii) is more complicated but it clearly has two of the necessary properties:

If $N = n$ is nonrandom, $\text{var}(S) = n \,\text{var}(Y_i)$.

If $Y_i = c$ is nonrandom $\text{var}(S) = c^2 \,\text{var}(N)$.

Combining these two observations, we see that $EN \,\text{var}(Y_i)$ is the contribution to the variance from the variability of the Y_i, while $\text{var}(N)(EY_i)^2$ is the contribution from the variability of N.

Proof. When $N = n$, $S = X_1 + \cdots + X_n$ has $ES = nEY_i$. Breaking things down according to the value of N,

$$
\begin{aligned}
ES &= \sum_{n=0}^{\infty} E(S|N = n) \cdot P(N = n) \\
&= \sum_{n=0}^{\infty} nEY_i \cdot P(N = n) = EN \cdot EY_i
\end{aligned}
$$

For the second formula we note that when $N = n$, $S = X_1 + \cdots + X_n$ has $\text{var}(S) = n\ \text{var}(Y_i)$ and hence,

$$E(S^2|N = n) = n\ \text{var}(Y_i) + (nEY_i)^2$$

Computing as before we get

$$
\begin{aligned}
ES^2 &= \sum_{n=0}^{\infty} E(S^2|N = n) \cdot P(N = n) \\
&= \sum_{n=0}^{\infty} \{n \cdot\ \text{var}(Y_i) + n^2(EY_i)^2\} \cdot P(N = n) \\
&= (EN) \cdot\ \text{var}(Y_i) + EN^2 \cdot (EY_i)^2
\end{aligned}
$$

To compute the variance now, we observe that

$$
\begin{aligned}
\text{var}(S) &= ES^2 - (ES)^2 \\
&= (EN) \cdot\ \text{var}(Y_i) + EN^2 \cdot (EY_i)^2 - (EN \cdot EY_i)^2 \\
&= (EN) \cdot\ \text{var}(Y_i) + \ \text{var}(N) \cdot (EY_i)^2
\end{aligned}
$$

where in the last step we have used $\operatorname{var}(N) = EN^2 - (EN)^2$ to combine the second and third terms.

For part (iii), we note that in the special case of the Poisson, we have $EN = \lambda$ and $\operatorname{var}(N) = \lambda$, so the result follows from $\operatorname{var}(Y_i) + (EY_i)^2 = EY_i^2$. $\square$

For a concrete example of the use of Theorem 2.8 consider

Example 2.3. Suppose that the number of customers at a liquor store in a day has a Poisson distribution with mean 81 and that each customer spends an average of \$8 with a standard deviation of \$6. It follows from (i) in Theorem 2.8 that the mean revenue for the day is $81 \cdot \$8 = \$648$. Using (iii), we see that the variance of the total revenue is

$$81 \cdot \left\{(\$6)^2 + (\$8)^2\right\} = 8100$$

Taking square roots we see that the standard deviation of the revenue is \$90 compared with a mean of \$648.

2.4 Transformations

2.4.1 Thinning

In the previous section, we added up the Y_i's associated with the arrivals in our Poisson process to see how many customers, etc., we had accumulated by time t. In this section we will use the Y_i to split one Poisson process into several. Let $N_j(t)$ be the number of $i \le N(t)$ with $Y_i = j$. In Example 2.1, where Y_i is the number of people in the ith car, $N_j(t)$ will be the number of cars that have arrived by time t with exactly j people. The somewhat remarkable fact is:

Theorem 2.9. *$N_j(t)$ are independent Poisson processes with rate $\lambda P(Y_i = j)$.*

Why is this remarkable? There are two "surprises" here: the resulting processes are Poisson and they are independent. To drive the point home consider a Poisson process with rate 10 per hour, and then flip coins to determine whether the arriving customers are male or female. One might think that seeing 40 men arrive in one hour would be indicative of a large volume of business and hence a larger than normal number of women, but Theorem 2.9 tells us that the number of men and the number of women that arrive per hour are independent.

Proof. To begin we suppose that $P(Y_i = 1) = p$ and $P(Y_i = 2) = 1-p$, so there are only two Poisson processes to consider: $N_1(t)$ and $N_2(t)$. We will check the second definition given in Theorem 2.4. It should be clear that the independent increments property of the Poisson process implies that the pairs of increments

$$(N_1(t_i) - N_1(t_{i-1}), N_2(t_i) - N_2(t_{i-1})), \quad 1 \le i \le n$$

are independent of each other. Since $N_1(0) = N_2(0) = 0$ by definition, it only remains to check that the components $X_i = N_i(t+s) - N_i(s)$ are independent and have the right Poisson distributions. To do this, we note that if $X_1 = j$ and $X_2 = k$, then there must have been $j+k$ arrivals between s and $s+t$, j of which were assigned 1's and k of which were assigned 2's, so

$$\begin{aligned} P(X_1 = j, X_2 = k) &= e^{-\lambda t} \frac{(\lambda t)^{j+k}}{(j+k)!} \cdot \frac{(j+k)!}{j!k!} p^j (1-p)^k \\ &= e^{-\lambda p t} \frac{(\lambda p t)^j}{j!} e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^k}{k!} \end{aligned} \tag{2.15}$$

so $X_1 = \text{Poisson}(\lambda p)$ and $X_2 = \text{Poisson}(\lambda(1-p))$. For the general case, we use the multinomial to conclude that if $p_j = P(Y_i = j)$ for $1 \le j \le m$ then

$$\begin{aligned} P(X_1 = k_1, \dots X_m = k_m) &= e^{-\lambda t} \frac{(\lambda t)^{k_1 + \cdots k_m}}{(k_1 + \cdots k_m)!} \frac{(k_1 + \cdots k_m)!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m} \\ &= \prod_{j=1}^m e^{-\lambda p_j t} \frac{(\lambda p_j)^{k_j}}{k_j!} \end{aligned}$$

□

The thinning results generalizes easily to the nonhomogeneous case:

Theorem 2.10. *Suppose that in a Poisson process with rate λ, we keep a point that lands at s with probability $p(s)$. Then the result is a nonhomogeneous Poisson process with rate $\lambda p(s)$.*

For an application of this consider

Example 2.4. M/G/∞ queue. In modeling telephone traffic, we can, as a first approximation, suppose that the number of phone lines is infinite, i.e., everyone who tries to make a call finds a free line. This certainly is not always true but analyzing a model in which we pretend this is true can help us to discover how many phone lines we need to be able to provide service 99.99% of the time.

The argument for arrivals at Trillium implies that the beginnings of calls follow a Poisson process. As for the calls themselves there is no reason to suppose that their duration follows a special distribution like the exponential, so use a general distribution function G with $G(0) = 0$ and mean μ. Suppose that the system starts empty at time 0. The probability a call started at s has ended by time t is $G(t-s)$, so using Theorem 2.10 the number of calls still in progress at time t is Poisson with mean

$$\int_{s=0}^t \lambda(1 - G(t-s))\, ds = \lambda \int_{r=0}^r (1 - G(r))\, dr$$

Letting $t \to \infty$ we see that in the long run the number of calls in the system will be Poisson with mean

$$\lambda \int_{r=0}^{\infty} (1 - G(r))\, dr = \lambda\mu$$

That is, the mean number in the system is the rate at which calls enter times their average duration. In the argument above we supposed that the system starts empty. Since the number of initial calls in the system at time t decreases to 0 as $t \to \infty$, the limiting result is true for any initial number of calls X_0.

2.4.2 Superposition

Taking one Poisson process and splitting it into two or more by using an i.i.d. sequence Y_i is called **thinning**. Going in the other direction and adding up a lot of independent processes is called **superposition**. Since a Poisson process can be split into independent Poisson processes, it should not be too surprising that when the independent Poisson processes are put together, the sum is Poisson with a rate equal to the sum of the rates.

Theorem 2.11. *Suppose $N_1(t), \dots N_k(t)$ are independent Poisson processes with rates $\lambda_1, \dots, \lambda_k$, then $N_1(t) + \cdots + N_k(t)$ is a Poisson process with rate $\lambda_1 + \cdots + \lambda_k$.*

Proof. Again we consider only the case $k = 2$ and check the second definition given in Theorem 2.4. It is clear that the sum has independent increments and $N_1(0) + N_2(0) = 0$. The fact that the increments have the right Poisson distribution follows from Theorem 2.3. □

We will see in the next chapter that the ideas of compounding and thinning are very useful in computer simulations of continuous time Markov chains. For the moment we will illustrate their use in computing the outcome of races between Poisson processes.

Example 2.5. A Poisson race. Given a Poisson process of red arrivals with rate λ and an independent Poisson process of green arrivals with rate μ, what is the probability that we will get 6 red arrivals before a total of 4 green ones?

Solution. The first step is to note that the event in question is equivalent to having at least 6 red arrivals in the first 9. If this happens, then we have at most 3 green arrivals before the 6th red one. On the other hand if there are 5 or fewer red arrivals in the first 9, then we have had at least 4 red arrivals and at most 5 green.

Viewing the red and green Poisson processes as being constructed by starting with one rate $\lambda + \mu$ Poisson process and flipping coins with probability $p = \lambda/(\lambda + \mu)$ to decide the color, we see that the probability of interest is

$$\sum_{k=6}^{9} \binom{9}{k} p^k (1-p)^{9-k}$$

If we suppose for simplicity that $\lambda = \mu$ so $p = 1/2$, this expression becomes

$$\frac{1}{512} \cdot \sum_{k=6}^{9} \binom{9}{k} = \frac{1 + 9 + (9 \cdot 8)/2 + (9 \cdot 8 \cdot 7)/3!}{512} = \frac{140}{512} = 0.273$$

2.4.3 Conditioning

Let $T_1, T_2, T_3, \ldots$ be the arrival times of a Poisson process with rate λ, let $U_1, U_2, \ldots U_n$ be independent and uniformly distributed on $[0, t]$, and let $V_1 < \ldots V_n$ be the U_i rearranged into increasing order . This section is devoted to the proof of the following remarkable fact.

Theorem 2.12. *If we condition on $N(t) = n$, then the vector $(T_1, T_2, \ldots T_n)$ has the same distribution as $(V_1, V_2, \ldots V_n)$ and hence the set of arrival times $\{T_1, T_2, \ldots, T_n\}$ has the same distribution as $\{U_1, U_2, \ldots, U_n\}$.*

Why is this true? We begin by finding the joint density function of (T_1, T_2, T_3) given that there were 3 arrivals before time t. The probability is 0 unless $0 < v_1 < v_2 < v_3 < t$. To compute the answer in this case, we note that $P(N(t) = 4) = e^{-\lambda t}(\lambda t)^3/3!$, and that for $T_1 = t_1$, $T_2 = t_2$, $T_3 = t_3$, $N(t) = 4$ we must have $\tau_1 = t_1$, $\tau_2 = t_2 - t_1$, $\tau_3 = t_3 - t_2$, and $\tau > t - t_3$, so the desired conditional distribution is:

$$
\begin{aligned}
&= \frac{\lambda e^{-\lambda t_1} \cdot \lambda e^{-\lambda(t_2-t_1)} \cdot \lambda e^{-\lambda(t_3-t_2)} \cdot e^{-\lambda(t-t_3)}}{e^{-\lambda t}(\lambda t)^3/3!} \\
&= \frac{\lambda^3 e^{-\lambda t}}{e^{-\lambda t}(\lambda t)^3/3!} = \frac{3!}{t^3}
\end{aligned}
$$

Note that the answer does not depend on the values of v_1, v_2, v_3 (as long as $0 < v_1 < v_2 < v_3 < t$), so the resulting conditional distribution is uniform over

$$\{(v_1, v_2, v_3) : 0 < v_1 < v_2 < v_3 < t\}$$

This set has volume $t^3/3!$ since $\{(v_1, v_2, v_3) : 0 < v_1, v_2, v_3 < t\}$ has volume t^3 and $v_1 < v_2 < v_3$ is one of 3! possible orderings.

Generalizing from the concrete example it is easy to see that the joint density function of $(T_1, T_2, \ldots T_n)$ given that there were n arrivals before time t is $n!/t^n$ for all times $0 < t_1 < \ldots < t_n < t$. $\square$

Theorem 2.12 implies that if we condition on having n arrivals at time t, then the locations of the arrivals are the same as the location of n points thrown uniformly on $[0, t]$. From the last observation we immediately get:

Theorem 2.13. *If $s < t$ and $0 \le m \le n$, then*

$$P(N(s) = m | N(t) = n) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}$$

That is, the conditional distribution of $N(s)$ given $N(t) = n$ is binomial$(n, s/t)$.

Proof. The number of arrivals by time s is the same as the number of $U_i < s$. The events $\{U_i < s\}$ these events are independent and have probability s/t, so the number of $U_i < s$ will be binomial$(n, s/t)$. $\square$

2.5 Exercises

Exponential distribution

2.1. Suppose that the time to repair a machine is exponentially distributed random variable with mean 2. (a) What is the probability the repair takes more than 2 hours. (b) What is the probability that the repair takes more than 5 hours given that it takes more than 3 hours.

2.2. The lifetime of a radio is exponentially distributed with mean 5 years. If Ted buys a 7 year-old radio, what is the probability it will be working 3 years later?

2.3. A doctor has appointments at 9 and 9:30. The amount of time each appointment lasts is exponential with mean 30. What is the expected amount of time after 9:30 until the second patient has completed his appointment?

2.4. Copy machine 1 is in use now. Machine 2 will be turned on at time t. Suppose that the machines fail at rate λ_i. What is the probability that machine 2 is the first to fail?

2.5. Alice and Betty enter a beauty parlor simultaneously, Alice to get a manicure and Betty to get a haircut. Suppose the time for a manicure (haircut) is exponentially distributed with mean 20 (30) minutes. (a) What is the probability Alice gets done first? (b) What is the expected amount of time until Alice and Betty are both done?

2.6. Let S and T be exponentially distributed with rates λ and μ. Let $U = \min\{S, T\}$ and $V = \max\{S, T\}$. Find (a) EU. (b) $E(V - U)$, (c) EV. (d) Use the identity $V = S + T - U$ to get a different looking formula for EV and verify the two are equal.

2.7. Let S and T be exponentially distributed with rates λ and μ. Let $U = \min\{S, T\}$, $V = \max\{S, T\}$, and $W = V - U$. Find the variances of U, V, and W.

2.8. In a hardware store you must first go to server 1 to get your goods and then go to a server 2 to pay for them. Suppose that the times for the two activities are exponentially distributed with means 6 minutes and 3 minutes. Compute the average amount of time it take Bob to get his goods and pay if when he comes in there is one customer named Al with server 1 and no one at server 2.

2.9. Consider the set-up of the previous problem but suppose that the times for the two activities are exponentially distributed with rates λ and μ. Compute Bob's average waiting time.

2.10. Consider a bank with two tellers. Three people, Alice, Betty, and Carol enter the bank at almost the same time and in that order. Alice and Betty go directly into service while Carol waits for the first available teller. Suppose that the service times for each customer are exponentially distributed with mean 4 minutes. (a) What is the expected total amount of time for Carol to complete her businesses? (b) What is the expected total time until the last of the three customers leaves? (c) What is the probability Carol is the last one to leave? (d) Answer questions (a),(b), and (c) for a general exponential with rate λ.

2.11. Consider the set-up of the previous problem but now suppose that the two tellers have exponential service times with means 3 and 6 minutes. Answer questions (a), (b), and (c).

2.12. Consider the set-up of the previous problem but now suppose that the two tellers have exponential service times with rates $\lambda \leq \mu$. Again, answer questions (a), (b), and (c).

2.13. Three people are fishing and each catches fish at rate 2 per hour. How long do we have to wait until everyone has caught at least one fish?

2.14. A machine has two critically important parts and is subject to three different types of shocks. Shocks of type i occur at times of a Poisson process with rate λ_i. Shocks of types 1 break part 1, those of type 2 break part 2, while those of type 3 break both parts. Let U and V be the failure times of the two parts. (a) Find $P(U > s, V > t)$. (b) Find the distribution of U and the distribution of V. (c) Are U and V independent?

2.15. A submarine has three navigational devices but can remain at sea if at least two are working. Suppose that the failure times are exponential with means 1 year, 1.5 years, and 3 years. What is the average length of time the boat can remain at sea.

2.16. Ron, Sue, and Ted arrive at the beginning of a professor's office hours. The amount of time they will stay is exponentially distributed with means of 1, 1/2, and 1/3 hour. (a) What is the expected time until only one student remains? (b) For each student find the probability they are the last student left. (c) What is the expected time until all three students are gone?

2.17. Let T_i, $i = 1, 2, 3$ be independent exponentials with rate λ_i. Find

(a) $P(T_1 < T_2 < T_3)$
(b) $P(T_1 < T_2 | \max T_i = T_3)$
(c) $E(\max T_i | T_1 < T_2 < T_3)$
(d) Use (a) and (c) to find $E \max_i T_i$. The next exercise will give a much simpler formula.

2.18. Let T_i, $i = 1, 2, 3$ be independent exponentials with rate λ_i. (a) Show that for any numbers t_1, t_2, t_3

$$\begin{aligned}\max\{t_1, t_2, t_3\} = t_1 + t_2 + t_3 &- \min\{t_1, t_2\} - \min\{t_1, t_3\} \\ &- \min\{t_2, t_3\} + \min\{t_1, t_2, t_3\}\end{aligned}$$

(b) Use (a) to find $E \max\{T_1, T_2, T_3\}$. (c) Use the formula to give a simple solution of part (c) of Exercise 2.16.

2.19. A flashlight needs two batteries to be operational. You start with four batteries numbered 1 to 4. Whenever a battery fails it is replaced by the lowest-numbered working battery. Suppose that battery life is exponential with mean 100 hours. Let T be the time at which there is one working battery left and N be the number of the one battery that is still good. (a) Find ET. (b) Find the distribution of N. (c) Solve (a) and (b) for a general number of batteries.

Poisson approximation to binomial

2.20. Compare the Poisson approximation with the exact binomial probabilities of 1 success when $n = 20$, $p = 0.1$.

2.21. Compare the Poisson approximation with the exact binomial probabilities of no success when (a) $n = 10$, $p = 0.1$, (b) $n = 50$, $p = 0.02$.

2.22. The probability of a three of a kind in poker is approximately 1/50. Use the Poisson approximation to estimate the probability you will get at least one three of a kind if you play 20 hands of poker.

2.23. Suppose 1% of a certain brand of Christmas lights is defective. Use the Poisson approximation to compute the probability that in a box of 25 there will be at most one defective bulb.

Poisson processes: Basic properties

2.24. Suppose $N(t)$ is a Poisson process with rate 3. Let T_n denote the time of the nth arrival. Find (a) $E(T_{12})$, (b) $E(T_{12}|N(2) = 5)$, (c) $E(N(5)|N(2) = 5)$.

2.25. Customers arrive at a shipping office at times of a Poisson process with rate 3 per hour. (a) The office was supposed to open at 8AM but the clerk Oscar overslept and came in at 10AM. What is the probability that no customers came in the two-hour period? (b) What is the distribution of the amount of time Oscar has to wait until his first customer arrives?

2.26. Suppose that the number of calls per hour to an answering service follows a Poisson process with rate 4. (a) What is the probability that fewer (i.e., $<$) than 2 calls came in the first hour? (b) Suppose that 6 calls arrive in the first hour, what is the probability there will be < 2 in the second hour. (c) Suppose that the operator gets to take a break after she has answered 10 calls. How long are her average work periods?

2.27. Traffic on Rosedale Road in Princeton, NJ, follows a Poisson process with rate 6 cars per minute. A deer runs out of the woods and tries to cross the road. If there is a car passing in the next 5 seconds then there will be a collision. (a) Find the probability of a collision. (b) What is the chance of a collision if the deer only needs 2 seconds to cross the road.

.

2.28. Calls to the Dryden fire department arrive according to a Poisson process with rate 0.5 per hour. Suppose that the time required to respond to a call, return to the station, and get ready to respond to the next call is uniformly distributed between 1/2 and 1 hour. If a new call comes before the Dryden fire department is ready to respond, the Ithaca fire department is asked to respond. Suppose that the Dryden fire department is ready to respond now. (a) Find the probability distribution for the number of calls they will handle before they have to ask for help from the Ithaca fire department. (b) In the long run what fraction of calls are handled by the Ithaca fire department?

2.29. A math professor waits at the bus stop at the Mittag-Leffler Institute in the suburbs of Stockholm, Sweden. Since he has forgotten to find out about the bus schedule, his waiting time until the next bus is uniform on (0,1). Cars drive by the bus stop at rate 6 per hour. Each will take him into town with probability 1/3. What is the probability he will end up riding the bus?

2.30. The number of hours between successive trains is T which is uniformly distributed between 1 and 2. Passengers arrive at the station according to a Poisson process with rate 24 per hour. Let X denote the number of people who get on a train. Find (a) EX, (b) $\text{var}\,(X)$.

2.31. Consider a Poisson process with rate λ and let L be the time of the last arrival in the interval $[0, t]$, with $L = 0$ if there was no arrival. (a) Compute $E(t-L)$ (b) What happens when we let $t \to \infty$ in the answer to (a)?

2.32. Customers arrive according to a Poisson process of rate λ per hour. Joe does not want to stay until the store closes at $T = 10$PM, so he decides to close up when the first customer after time $T - s$ arrives. He wants to leave early but he does not want to lose any business so he is happy if he leaves before T and no one arrives after. (a) What is the probability he achieves his goal? (b) What is the optimal value of s and the corresponding success probability?

2.33. Let T be exponentially distributed with rate λ. (a) Use the definition of conditional expectation to compute $E(T|T < c)$. (b) Determine $E(T|T < c)$ from the identity

$$ET = P(T < c)E(T|T < c) + P(T > c)E(T|T > c)$$

2.34. *When did the chicken cross the road?* Suppose that traffic on a road follows a Poisson process with rate λ cars per minute. A chicken needs a gap of length at least c minutes in the traffic to cross the road. To compute the time the chicken will have to wait to cross the road, let $t_1, t_2, t_3, \ldots$ be the interarrival times for the cars and let $J = \min\{j : t_j > c\}$. If $T_n = t_1 + \cdots + t_n$, then the chicken will start to cross the road at time T_{J-1} and complete his journey at time $T_{J-1}+c$. Use the previous exercise to show $E(T_{J-1}+c) = (e^{\lambda c}-1)/\lambda$.

Random sums

2.35. Edwin catches trout at times of a Poisson process with rate 3 per hour. Suppose that the trout weigh an average of 4 pounds with a standard deviation of 2 pounds. Find the mean and standard deviation of the total weight of fish he catches in two hours.

2.36. An insurance company pays out claims at times of a Poisson process with rate 4 per week. Writing K as shorthand for "thousands of dollars," suppose that the mean payment is 10K and the standard deviation is 6K. Find the mean and standard deviation of the total payments for 4 weeks.

2.37. Customers arrive at an automated teller machine at the times of a Poisson process with rate of 10 per hour. Suppose that the amount of money withdrawn on each transaction has a mean of \$30 and a standard deviation of \$20. Find the mean and standard deviation of the total withdrawals in 8 hours.

2.38. Let S_t be the price of stock at time t and suppose that at times of a Poisson process with rate λ the price is multiplied by a random variable $X_i > 0$ with mean μ and variance σ^2. That is,

$$S_t = S_0 \prod_{i=1}^{N(t)} X_i$$

where the product is 1 if $N(t) = 0$. Find $ES(t)$ and $\operatorname{var} S(t)$.

2.39. Messages arrive to be transmitted across the internet at times of a Poisson process with rate λ. Let Y_i be the size of the ith message, measured in bytes, and let $g(z) = Ez^{Y_i}$ be the generating function of Y_i. Let $N(t)$ be the number of arrivals at time t and $S = Y_1 + \cdot + Y_{N(t)}$ be the total size of the messages up to time t. (a) Find the generating function $f(z) = E(z^S)$. (b) Differentiate and set $z = 1$ to find ES. (c) Differentiate again and set $z = 1$ to find $E\{S(S-1)\}$. (d) Compute $\operatorname{var}(S)$.

2.40. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. Let $T \geq 0$ be an independent with mean μ and variance σ^2. Find $\operatorname{cov}(T, N_T)$.

2.41. Let $t_1, t_2, \ldots$ be independent exponential(λ) random variables and let N be an independent random variable with $P(N = n) = (1-p)^{n-1}$. What is the distribution of the random sum $T = t_1 + \cdots + t_N$?

Thinning and conditioning

2.42. Traffic on Snyder Hill Road in Ithaca, NY, follows a Poisson process with rate 2/3's of a vehicle per minute. 10% of the vehicles are trucks, the other 90% are cars. (a) What is the probability at least one truck passes in a hour? (b) Given that 10 trucks have passed by in an hour, what is the expected number of vehicles that have passed by. (c) Given that 50 vehicles have passed by in a hour, what is the probability there were exactly 5 trucks and 45 cars.

2.43. Rock concert tickets are sold at a ticket counter. Females and males arrive at times of independent Poisson processes with rates 30 and 20. (a) What is the probability the first three customers are female? (b) If exactly 2 customers arrived in the first five minutes, what is the probability both arrived in the first three minutes. (c) Suppose that customers regardless of sex buy 1 ticket with probability 1/2, two tickets with probability 2/5, and three tickets with probability 1/10. Let N_i be the number of customers that buy i tickets in the first hour. Find the joint distribution of (N_1, N_2, N_3).

2.44. Ellen catches fish at times of a Poisson process with rate 2 per hour. 40% of the fish are salmon, while 60% of the fish are trout. What is the probability she will catch exactly 1 salmon and 2 trout if she fishes for 2.5 hours?

2.45. Signals are transmitted according to a Poisson process with rate λ. Each signal is successfully transmitted with probability p and lost with probability $1 - p$. The fates of different signals are independent. For $t \geq 0$ let $N_1(t)$ be the number of signals successfully transmitted and let $N_2(t)$ be the number that are lost up to time t. (a) Find the distribution of $(N_1(t), N_2(t))$. (b) What is the distribution of $L =$ the number of signals lost before the first one is successfully transmitted?

2.46. A copy editor reads a 200-page manuscript, finding 108 typos. Suppose that the author's typos follow a Poisson process with some unknown rate λ per page, while from long experience we know that the copyeditor finds 90% of the mistakes that are there. (a) Compute the expected number of typos found as a function of the arrival rate λ. (b) Use the answer to (a) to find an estimate of λ and of the number of undiscovered typos.

2.47. Two copy editors read a 300-page manuscript. The first found 100 typos, the second found 120, and their lists contain 80 errors in common. Suppose that the author's typos follow a Poisson process with some unknown rate λ per page, while the two copy editors catch errors with unknown probabilities of success p_1 and p_2. Let X_0 be the number of typos that neither found. Let X_1 and X_2 be the number of typos found only by 1 or only by 2, and let X_3 be the number of typos found by both. (a) Find the joint distribution of (X_0, X_1, X_2, X_3). (b) Use the answer to (a) to find an estimates of p_1, p_2 and then of the number of undiscovered typos.

2.48. A light bulb has a lifetime that is exponential with a mean of 200 days. When it burns out a janitor replaces it immediately. In addition there is a handyman who comes at times of a Poisson process at rate .01 and replaces the bulb as "preventive maintenance." (a) How often is the bulb replaced? (b) In the long run what fraction of the replacements are due to failure?

2.49. Starting at some fixed time, which we will call 0 for convenience, satellites are launched at times of a Poisson process with rate λ. After an independent amount of time having distribution function F and mean μ, the satellite stops working. Let $X(t)$ be the number of working satellites at time t. (a) Find the distribution of $X(t)$. (b) Let $t \to \infty$ in (a) to show that the limiting distribution is Poisson$(\lambda\mu)$.

2.50. Ignoring the fact that the bar exam is only given twice a year, let us suppose that new lawyers arrive in Los Angeles according to a Poisson process with mean 300 per year. Suppose that each lawyer independently practices for an amount of time T with a distribution function $F(t) = P(T \le t)$ that has $F(0) = 0$ and mean 25 years. Show that in the long run the number of lawyers in Los Angeles is Poisson with mean 7500.

2.51. Policy holders of an insurance company have accidents at times of a Poisson process with rate λ. The distribution of the time R until a claim is reported is random with $P(R \le r) = G(r)$ and $ER = \nu$. (a) Find the distriibution of the number of unreported claims. (b) Suppose each claim has mean μ and variance σ^2. Find the mean and variance of S the total size of the unreported claims.

2.52. Suppose $N(t)$ is a Poisson process with rate 2. Compute the conditional probabilities (a) $P(N(3) = 4|N(1) = 1)$, (b) $P(N(1) = 1|N(3) = 4)$.

2.53. Customers arrive at a bank according to a Poisson process with rate 10 per hour. Given that two customers arrived in the first 5 minutes, what is the probability that (a) both arrived in the first 2 minutes. (b) at least one arrived in the first 2 minutes.

2.54. Suppose that the number of calls per hour to an answering service follows a Poisson process with rate 4. Suppose that 3/4's of the calls are made by men, 1/4 by women, and the sex of the caller is independent of the time of the call. (a) What is the probability that in one hour exactly 2 men and 3 women will call the answering service? (b) What is the probability 3 men will make phone calls before 3 women do?

2.55. Hockey teams 1 and 2 score goals at times of Poisson processes with rates 1 and 2. Suppose that $N_1(0) = 3$ and $N_2(0) = 1$. (a) What is the probability that $N_1(t)$ will reach 5 before $N_2(t)$ does? (b) Answer part (a) for Poisson processes with rates λ_1 and λ_2.

2.56. Consider two independent Poisson processes $N_1(t)$ and $N_2(t)$ with rates λ_1 and λ_2. What is the probability that the two-dimensional process $(N_1(t), N_2(t))$ ever visits the point (i, j)?

Chapter 3

Renewal processes

3.1 Laws of large numbers

In the Poisson process the times between successive arrivals are independent and exponentially distributed. The lack of memory property of the exponential distribution is crucial for many of the special properties of the Poisson process derived in this chapter. However, in many situations the assumption of exponential interarrival times is not justified. In this section we will consider a generalization of Poisson processes called **renewal processes** in which the times $t_1, t_2, \ldots$ between events are independent and have distribution F.

In order to have a simple metaphor with which to discuss renewal processes, we will think of a single light bulb maintained by a very diligent janitor, who replaces the light bulb immediately after it burns out. Let t_i be the lifetime of the ith light bulb. We assume that the light bulbs are bought from one manufacturer, so we suppose

$$P(t_i \le t) = F(t)$$

where F is a distribution function with $F(0) = P(t_i \le 0) = 0$.

If we start with a new bulb (numbered 1) at time 0 and each light bulb is replaced when it burns out, then $T_n = t_1 + \cdots + t_n$ gives the time that the nth bulb burns out, and

$$N(t) = \max\{n : T_n \le t\}$$

is the number of light bulbs that have been replaced by time t. The picture is the same as the one for the Poisson process, see Figure 2.1.

If renewal theory were only about changing light bulbs, it would not be a very useful subject. The reason for our interest in this system is that it captures the essence of a number of different situations. On example that we have already seen is

Example 3.1. Markov chains. Let X_n be a Markov chain and suppose that $X_0 = x$. Let T_n be the nth time that the process returns to x. The strong Markov property implies that $t_n = T_n - T_{n-1}$ are independent, so T_n is a renewal process.

Example 3.2. Machine repair. Instead of a light bulb, think of a machine that works for an amount of time s_i before it fails, requiring an amount of time u_i to be repaired. Let $t_i = s_i + u_i$ be the length of the ith cycle of breakdown and repair. If we assume that the repair leaves the machine in a "like new" condition, then the t_i are independent and identically distributed (i.i.d.) and a renewal process results.

Example 3.3. Counter processes. The following situation arises, for example, in medical imaging applications. Particles arrive at a counter at times of a Poisson process with rate λ. Each arriving particle that finds the counter free gets registered and locks the counter for an amount of time τ. Particles arriving during the locked period have no effect. If we assume the counter starts in the unlocked state, then the times T_n at which it becomes unlocked for the nth time form a renewal process. This is a special case of the previous example: $u_i = \tau$, $s_i =$ exponential with rate λ.

In addition there will be several applications to queueing theory.

The main result about renewal processes that we will use is the following law of large numbers:

Theorem 3.1. *Let $\mu = Et_i$ be mean interarrival time. If $P(t_i > 0) > 0$ then with probability one,*

$$N(t)/t \to 1/\mu \quad \text{as } t \to \infty$$

In words, this says that if our light bulb lasts μ years on the average then in t years we will use up about t/μ light bulbs. Since the interarrival times in a Poisson process have mean $1/\lambda$ Theorem 3.1 implies that if $N(t)$ is the number of arrivals up to time t in a Poisson process, then

$$N(t)/t \to \lambda \quad \text{as } t \to \infty \tag{3.1}$$

Proof of Theorem 3.1. We use the

Theorem 3.2. Strong law of large numbers. *Let $x_1, x_2, x_3, \dots$ be i.i.d. with $Ex_i = \mu$, and let $S_n = x_1 + \cdots + x_n$. Then with probability one,*

$$S_n/n \to \mu \quad \text{as } n \to \infty$$

Taking $x_i = t_i$, we have $S_n = T_n$, so Theorem 3.2 implies that with probability one, $T_n/n \to \mu$ as $n \to \infty$. Now by definition,

$$T_{N(t)} \le t < T_{N(t)+1}$$

Dividing by $N(t)$, we have

$$\frac{T_{N(t)}}{N(t)} \le \frac{t}{N(t)} \le \frac{T_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}$$

By the strong law of large numbers, the left- and right-hand sides converge to μ. From this it follows that $t/N(t) \to \mu$ and hence $N(t)/t \to 1/\mu$. □

The next result looks like that it is weaker than Theorem 3.1 but is harder to prove.

Theorem 3.3. *Let $\mu = Et_i$ be mean interarrival time. If $P(t_i > 0) > 0$ then*

$$EN(t)/t \to 1/\mu \quad \text{as } t \to \infty$$

To show that there is something to prove let U be uniform on (0,1) and let $N(t) = t$ with if $U \le 1 - 1/t$ and $= t^2$ if $U > 1 - 1/t$. Then $N(t)/t \to 1$ but $EN(t)/t = 2 - 1/t \to 2$. It seem unlikely that something unusal like this will happen to $N(t)$ but this has to be ruled out. The proof is technical, so most readers will want to skip it.

Proof. Consider first the special case in which $\bar{t}_i = 0$ with probability $1 - p$ and $\bar{t}_i = \delta$ with probability p. This renewal process stays at $k\delta$ for a geometric number of times V_k with mean $1/\delta$. So if $m\delta \le t < (m+1)\delta$ we have

$$EN_\delta(t) = E\left(\sum_{k=0}^{m} V_k\right) = (m+1)/\delta.$$

To compute the second moment we note that

$$EN_\delta(t)^2 = E\left(\sum_{k=0}^m V_k\right)^2 = (m+1)EV_1^2 + (m+1)m(EV_1)^2$$

Thus if $t \geq \delta$

$$E\left(\frac{N_\delta(t)}{t}\right)^2 \leq \frac{(m+1)^2 EV_1^2}{(m\delta)^2} \leq 4\frac{EV_1^2}{\delta^2}$$

If $P(t_i > 0) > 0$ then we can pick a $\delta > 0$ so $p = P(t_i \geq \delta) > 0$. Since the t_i renewal process will have fewer renewals by time t than the one generated by $\bar{t}_i$ we have

$$\max_{t \geq \delta} E\left(\frac{N(t)}{t}\right)^2 \leq C < \infty$$

To show that this is enough to rule out trouble we note that if $M > 1/\mu$ then the bounded convergence theorem implies

$$E\left(\frac{N(t)}{t}\right) \geq E\left(\frac{N(t)}{t} \wedge M\right) \to \frac{1}{\mu}$$

To handle the missing part of the expected value

$$E\left(\frac{N(t)}{t}; \frac{N(t)}{t} > M\right) \leq \frac{1}{M} E\left(\frac{N(t)}{t}\right)^2 \leq \frac{C}{M}$$

Given $\epsilon > 0$, if M is large the last quantity is $< \epsilon$ for all $t \geq \delta$. Since

$$E\left(\frac{N(t)}{t}\right) \leq E\left(\frac{N(t)}{t} \wedge M\right) + E\left(\frac{N(t)}{t}; \frac{N(t)}{t} > M\right)$$

the desired result follows. $\square$

Our next topic is a simple extension of the notion of a renewal process that greatly extends the class of possible applications. We suppose that at the time of the ith renewal we earn a reward r_i. The reward r_i may depend on the ith interarrival time t_i, but we will assume that the pairs (r_i, t_i), $i = 1, 2, \ldots$ are independent and have the same distribution. Let

$$R(t) = \sum_{i=1}^{N(t)} r_i$$

be the total amount of rewards earned by time t. The main result about renewal reward processes is the following strong law of large numbers.

Theorem 3.4. *With probability one,*

$$\frac{R(t)}{t} \to \frac{Er_i}{Et_i} \tag{3.2}$$

Proof. Multiplying and dividing by $N(t)$, we have

$$\frac{R(t)}{t} = \left(\frac{1}{N(t)} \sum_{i=1}^{N(t)} r_i \right) \frac{N(t)}{t} \to Er_i \cdot \frac{1}{Et_i}$$

where in the last step we have used Theorem 3.1 and applied the strong law of large numbers to the sequence r_i. □

Intuitively, (3.2) can be written as

$$\text{reward/time} = \frac{\text{expected reward/cycle}}{\text{expected time/cycle}}$$

an equation that can be "proved" by pretending the words on the right-hand side are numbers and then canceling the "expected" and "1/cycle" that appear in numerator and denominator. The last calculation is not given to convince you that Theorem 3.4 is correct but to help you remember the result. A second approach to this is that if we earn a reward of ρ dollar every τ units of time then in the long run we earn ρ/τ dollars per unit time. To get from this to the answer given in 3.4, note that the answer there only depends on the means Er_i and Et_i, so the general answer must be

$$\rho/\tau = Er_i/Et_i$$

This device can be applied to remember many of the results in this chapter: when the answer only depends on the mean the limit must be the same as in the case when the times are not random.

To illustrate the use of Theorem 3.4 we consider

Example 3.4. Long run car costs. Suppose that the lifetime of a car is a random variable with density function h. Our methodical Mr. Brown buys a new car as soon as the old one breaks down or reaches T years. Suppose that a new car costs A dollars and that an additional cost of B dollars to repair the vehicle is incurred if it breaks down before time T. What is the long-run cost per unit time of Mr. Brown's policy?

Solution. The duration of the ith cycle, t_i, has

$$Et_i = \int_0^T th(t)\,dt + T\int_T^\infty h(t)\,dt$$

since the length of the cycle will be t_i if the car's life is $t_i < T$, but T if the car's life $t_i \geq T$. The reward (or cost) of the ith cycle has

$$Er_i = A + B\int_0^T h(t)\,dt$$

since Mr. Brown always has to pay A dollars for a new car but only owes the additional B dollars if the car breaks down before time T. Using Theorem 3.4 we see that the long run cost per unit time is

$$\frac{Er_i}{Et_i} = \frac{A + B\int_0^T h(t)\,dt}{\int_0^T th(t)\,dt + \int_T^\infty Th(t)\,dt}$$

Concrete example. Suppose that the lifetime of Mr. Brown's car is uniformlu distributed on $[0, 10]$. This is probably not a reasonable assumption, since when cars get older they have a greater tendency to break. However, having confessed to this weakness, we will proceed with this assumption since it makes calculations easier. Suppose that the cost of a new car is $A = 10$ (thousand dollars), while the breakdown cost is $B = 3$ (thousand dollars). If Mr. Brown replaces his car after T years then the expected values of interest are

$$\begin{aligned} Er_i &= 10 + 3\frac{T}{10} = 10 + 0.3T \\ Et_i &= \int_0^T \frac{t}{10}\,dt + T\left(1 - \frac{T}{10}\right) = \frac{T^2}{20} + T - \frac{T^2}{10} = T - 0.05T^2 \end{aligned}$$

Combining the expressions for the Er_i and Et_i we see that the long-run cost per unit time is

$$\frac{Er_i}{Et_i} = \frac{10 + 0.3T}{T - 0.05T^2}$$

To maximize we take the derivative

$$\frac{d}{dT}\frac{Er_i}{Et_i} = \frac{0.3(T - 0.05T^2) - (10 + 0.3T)(1 - 0.1T)}{(T - 0.1T^2)^2}$$
$$= \frac{0.3T - 0.015T^2 - 10 - 0.3T + T + 0.03T^2}{(T - 0.1T^2)^2}$$

The numerator is $0.015T^2 + T - 10$ which is 0 when

$$T = \frac{-1 \pm \sqrt{1 + 4(0.015)(10)}}{2(0.015)} = \frac{-1 \pm \sqrt{1.6}}{0.03}$$

We want the + root which is $T = 8.83$.

Using the idea of renewal reward processes, we can easily treat the following extension of renewal processes.

Example 2.5. Alternating renewal processes. Let $s_1, s_2, \dots$ be independent with a distribution F that has mean μ_F, and let $u_1, u_2, \dots$ be independent with distribution G that has mean μ_G. For a concrete example consider the machine in Example 1.1 that works for an amount of time s_i before needing a repair that takes u_i units of time. However, to talk about things in general we will say that the alternating renewal process spends an amount of time s_i in state 1, an amount of time u_i in state 2, and then repeats the cycle again.

Theorem 3.5. *In an alternating renewal process, the limiting fraction of time in state 1 is*

$$\frac{\mu_F}{\mu_F + \mu_G}$$

To see that this is reasonable and to help remember the formula, consider the nonrandom case. If the machine always works for exactly μ_F days and then needs repair for exactly μ_G days, then the limiting fraction of time spent working is $\mu_F/(\mu_F + \mu_G)$.

Why is Theorem 3.5 true? In order to compute the limiting fraction of time the machine is working we let $t_i = s_i + u_i$ be the duration of the ith cycle, and let the reward $r_i = s_i$, the amount of time the machine was working during the ith cycle. In this case, if we ignore the contribution from the cycle that is in progress at time t, then

$$R(t) = \sum_{i=1}^{N(t)} r_i$$

will be the amount of time the machine has been working up to time t. Thus, Theorem 3.4 implies that

$$\frac{R(t)}{t} \to \frac{Er_i}{Et_i} = \frac{\mu_F}{\mu_F + \mu_G} \qquad \square$$

Proof. In order to complete the proof we need to show that we can ignore the contribution from the cycle that is in progress at time t. Since $t \to P(r_i/\epsilon > t)$ is decreasing, it follows that for any $\epsilon > 0$

$$\sum_{n=1}^{\infty} P(r_n > n\epsilon) \leq \int_0^{\infty} P(r_n/\epsilon > t)\,dt \leq E(r_i/\epsilon) < \infty$$

Now for any M we have

$$\frac{1}{n}\left(\max_{1\leq m\leq n} r_m/n\right) \leq \frac{1}{n}\max_{1\leq m\leq M} r_m + \max_{m\geq M}\frac{r_m}{m}$$

For each fixed M the first term converges to 0 as $n \to \infty$. The second is a constant δ_M that does not depend on n so we have

$$\limsup_{n\to\infty}\frac{1}{n}\left(\max_{1\leq m\leq n} r_m\right) \leq \delta_M$$

Letting $M \to \infty$ now we have

$$\lim_{n\to\infty}\max_{1\leq m\leq n} r_m/n = 0 \qquad (3.3)$$

Let $C = 1 + (1/\mu)$. (3.3) implies that

$$\max_{m\leq Ct} r_m/t \to 0$$

Theorem 3.1 implies that $N(t)/t \to 1/\mu$, so for large t we will have $N(t) + 1 \leq ct$. When this holds we will have

$$r_{N(t)+1}/t \leq \max_{m\leq Ct} r_m/t \to 0$$

so the contribution from the incomplete cycle can be ignored. $\square$

For a concrete example of alternating renewal processes, consider

Example 3.5. Poisson janitor. A light bulb burns for an amount of time having distribution F with mean μ_F then burns out. A janitor comes at times of a rate λ Poisson process to check the bulb and will replace the bulb if it is burnt out. (a) At what rate are bulbs replaced? (b) What is the limiting fraction of time that the light bulb works? (c) What is the limiting fraction of visits on which the bulb is working?

Solution. Suppose that a new bulb is put in at time 0. It will last for an amount of time s_1. Using the lack of memory property of the exponential distribution, it follows that the amount of time until the next inspection, u_1, will have an exponential distribution with rate λ. The bulb is then replaced and the cycle starts again, so we have an alternating renewal process.

To answer (a), we note that the expected length of a cycle $Et_i = \mu_F + 1/\lambda$, so if $N(t)$ is the number of bulbs replaced by time t, then it follows from Theorem 3.1 that

$$\frac{N(t)}{t} \to \frac{1}{\mu_F + 1/\lambda}$$

In words, bulbs are replaced on the average every $\mu_F + 1/\lambda$ units of time.

To answer (b), we let $r_i = s_i$, so Theorem 3.5 implies that in the long run, the fraction of time the bulb has been working up to time t is

$$\frac{Er_i}{Et_i} = \frac{\mu_F}{\mu_F + 1/\lambda}$$

To answer (c), we note that if $V(t)$ is the number of visits the janitor has made by time t, then by the law of large numbers for the Poisson process we have

$$\frac{V(t)}{t} \to \lambda$$

Combining this with the result of (a), we see that the fraction of visits on which bulbs are replaced

$$\frac{N(t)}{V(t)} \to \frac{1/(\mu_F + 1/\lambda)}{\lambda} = \frac{1/\lambda}{\mu_F + 1/\lambda}$$

This answer is reasonable since it is also the limiting fraction of time the bulb is off.

3.2 Applications to Queueing Theory

In this section we will use the ideas of renewal theory to prove results for queueing systems with general service times. In the first part of this section we will consider general arrival times. In the second we will specialize to Poisson inputs.

3.2.1 GI/G/1 queue

Here the *GI* stands for general input. That is, we suppose that the times t_i between successive arrivals are independent and have a distribution F with mean $1/\lambda$. We make this somewhat unusual choice of notation for mean so that if $N(t)$ is the number of arrivals by time t, then Theorem 3.1 implies that the long-run arrival rate is

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{Et_i} = \lambda$$

The second G stands for general service times. That is, we assume that the ith customer requires an amount of service s_i, where the s_i are independent and have a distribution G with mean $1/\mu$. Again, the notation for the mean is chosen so that the service rate is μ. The final 1 indicates there is one server. Our first result states that the queue is stable if the arrival rate is smaller than the long-run service rate.

Theorem 3.6. *Suppose $\lambda < \mu$. If the queue starts with some finite number $k \geq 1$ customers who need service, then it will empty out with probability one.*

Why is this true? While the queue is not empty the server works at rate μ, which is larger than the rate λ at which arrivals come, so the number of the people in the system tends to decrease and long queues will not develop.

Proof. We will proceed by contradiction using the idea above. Specifically we will show that if the server is always busy then since she serves customers faster than they arrive, she is eventually done with the customers before they arrive, which is a contradiction.

Turning to the details, let $T_n = t_1 + \cdots + t_n$ be the time of the nth arrival. The strong law of large numbers, Theorem 3.2 implies that

$$\frac{T_n}{n} \to \frac{1}{\lambda}$$

Let Z_0 be the sum of the service times of the customers in the system at time 0 and let s_i be the service time of the ith customer to arrive after time 0. If the server stays busy until the nth customer arrives then that customer will depart the queue at time $Z_0 + S_n$, where $S_n = s_1 + \cdots + s_n$. The strong law of large numbers implies

$$\frac{Z_0 + S_n}{n} \to \frac{1}{\mu}$$

Since $1/\mu < 1/\lambda$, this means that if we assume that the server is always working, then when n is large enough the nth customer departs before he arrives. This contradiction implies that the probability that the server stays busy for all time must be 0. □

By looking at the last argument more carefully we can conclude:

Theorem 3.7. *Suppose $\lambda < \mu$. The limiting fraction of time the server is busy is λ/μ.*

Why is this true? Customers arrive at rate λ per unit time, say per hour. The server can serve μ customers per hour, but cannot serve more customers than arrive. Thus the server must be serving customers at rate μ per hour for a long-run fraction λ/μ of the time.

More details. Suppose for simplicity that the queue starts out empty. The nth customer arrives at time $T_n = t_1 + \cdots + t_n$. If A_n is the amount of time the server has been busy up to time T_n and S_n is the sum of the first n service times, then

$$A_n = S_n - Z_n$$

where Z_n is the amount of work in the system at time T_n, i.e., the amount of time needed to empty the system if there were no more arrivals.

The first term is easy to deal with

$$\frac{S_n}{T_n} = \frac{S_n/n}{T_n/n} \to \frac{Es_i}{Et_i} = \frac{\lambda}{\mu}$$

Since $Z_n \geq 0$ this implies

$$\limsup_{n\to\infty} \frac{A_n}{T_n} \leq \frac{\lambda}{\mu}$$

To argue that equality holds we need to show that $Z_n/n \to 0$. Intuitively, the condition $\lambda < \mu$ implies the queue reaches equilibrium, so EZ_n stays bounded, and hence $Z_n/n \to 0$. The details of completing this proof are too complicated to give here. However, in Example 3.8 we will give a simple proof of this. □

Cost equations. Let X_s be the number of customers in the system at time s. Let L be the long-run average number of customers in the system:

$$L = \lim_{t\to\infty} \frac{1}{t}\int_0^t X_s\, ds$$

Let W be the long-run average amount of time a customer spends in the system:

$$W = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^n W_m$$

where W_m is the amount of time the mth arriving customer spends in the system. Finally, let λ_a be the long-run average rate at which arriving customers join the system, that is,

$$\lambda_a = \lim_{t\to\infty} N_a(t)/t$$

where $N_a(t)$ is the number of customers who arrive before time t and enter the system. Ignoring the problem of proving the existence of these limits, we can assert that these quantities are related by

Theorem 3.8. Little's formula. *$L = \lambda_a W$.*

Why is this true? Suppose each customer pays \$1 for each minute of time she is in the system. When ℓ customers are in the system, we are earning \$$\ell$ per minute, so in the long run we earn an average of \$$L$ per minute. On the other hand, if we imagine that customers pay for their entire waiting time when they arrive then we earn at rate $\lambda_a W$ per minute, i.e., the rate at which customers enter the system multiplied by the average amount they pay. □

For a simple example of the use of this formula consider:

Example 3.6. M/M/1 queue. Here arrivals are a rate λ Poisson process, there is one server, and customers require an amount of service that is exponentially distributed with mean $1/\mu$. If we assume

$\lambda < \mu$, then it follows from Example 4.18 that the equilibrium probability of n people in the system is given by the shifted geometric distribution

$$\pi(n) = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n$$

and the mean queue length in equilibrium is $L = \lambda/(\mu - \lambda)$. Since all customers enter the system $\lambda_a = \lambda$, and it follows that from Theorem 3.8 that the average waiting time is

$$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}$$

For an example with $\lambda_a < \lambda$ we consider

Example 3.7. Barbershop chain. A barber can cut hair at rate 3, where the units are people per hour, i.e., each haircut requires an exponentially distributed amount of time with mean 20 minutes. Customers arrive at times of a rate 2 Poisson process, but will leave if both chairs in the waiting room are full. As we computed in Example 4.15 that the equilibrium distribution for the number of people in the system was given by

$$\pi(0) = 27/65, \quad \pi(1) = 18/65, \quad \pi(2) = 12/65, \quad \pi(3) = 8/65$$

From this it follows that

$$L = 1 \cdot \frac{18}{65} + 2 \cdot \frac{12}{65} + 3 \cdot \frac{8}{65} = \frac{66}{65}$$

Customers will only enter the system if there are < 3 people, so

$$\lambda_a = 2(1 - \pi(3)) = 114/65$$

Combining the last two results and using Little's formula, Theorem 3.8, we see that the average waiting time in the system is

$$W = \frac{L}{\lambda_a} = \frac{66/65}{114/65} = \frac{66}{114} = 0.579 \text{ hours}$$

Example 3.8. Waiting time in the queue. Consider the $GI/G/1$ queue and suppose that we are only interested in the customer's average waiting time in the queue, W_Q. If we know the average waiting time W in the system, this can be computed by simply subtracting out the amount of time the customer spends in service

$$W_Q = W - Es_i \tag{3.4}$$

For instance, in the previous example, subtracting off the 0.333 hours that his haircut takes we see that the customer's average time waiting in the queue $W_Q = 0.246$ hours or 14.76 minutes.

Let L_Q be the average queue length in equilibrium; i.e., we do not count the customer in service if there is one. If suppose that customers pay \$1 per minute in the queue and repeat the derivation of Little's formula, then

$$L_Q = \lambda W_Q \tag{3.5}$$

The length of the queue is 1 less than the number in the system, except when the number in the system is 0, so if $\pi(0)$ is the probability of no customers, then

$$L_Q = L - 1 + \pi(0)$$

Combining the last three equations with our first cost equation:

$$\pi(0) = L_Q - (L - 1) = 1 + \lambda(W_Q - W) = 1 - \lambda E s_i \tag{3.6}$$

Recalling that $Es_i = 1/\mu$, we have a simple proof of Theorem 3.7.

3.2.2 M/G/1 queue

Here the M stands for Markovian input and indicates we are considering the special case of the $GI/G/1$ queue in which the inputs are a rate λ Poisson process. The rest of the set-up is as before: there is a one server and the ith customer requires an amount of service s_i, where the s_i are independent and have a distribution G with mean $1/\mu$.

When the input process is Poisson, the system has special properties that allow us to go further. We learned in Theorem 3.6 that if $\lambda < \mu$ then a $GI/G/1$ queue will repeatedly return to the empty state. Thus the server experiences alternating busy periods with duration B_n and idle periods with duration I_n. In the case of Markovian inputs, the lack of memory property implies that I_n has an exponential distribution with rate λ. Combining this observation with our result for alternating renewal processes we see that the limiting fraction of time the server is idle is

$$\frac{1/\lambda}{1/\lambda + EB_n} = 1 - \frac{\lambda}{\mu}$$

by Theorem 3.7. Rearranging, we have

$$EB_n = \frac{1}{\lambda}\left(\frac{1}{1-(\lambda/\mu)} - 1\right) = \frac{1/\mu}{1-\lambda/\mu} = \frac{Es_i}{1-\lambda Es_i} \tag{3.7}$$

A second special property of Poisson arrivals is:

PASTA. These initials stand for "Poisson arrivals see time averages." To be precise, if $\pi(n)$ is the limiting fraction of time that there are n individuals in the queue and a_n is the limiting fraction of arriving customers that see a queue of size n, then

Theorem 3.9. $a_n = \pi(n)$.

Why is this true? If we condition on there being arrival at time t, then the times of the previous arrivals are a Poisson process with rate λ. Thus knowing that there is an arrival at time t does not affect the distribution of what happened before time t. □

Example 3.9. Workload in the M/G/1 queue. We define the workload in the system at time t, V_t, to be the sum of the remaining service times of all customers in the system. Suppose that each customer in the queue or in service pays at a rate of $\$y$ when his remaining *service* time is y; i.e., we do not count the remaining waiting time in the queue. If we let Y be the average total payment made by an arriving customer, then our cost equation reasoning implies that the average workload V satisfies

$$V = \lambda Y$$

(All customers enter the system, so $\lambda_a = \lambda$.) Since a customer with service time s_i pays s_i during the q_i units of time spent waiting in the queue and at rate $s_i - x$ after x units of time in service

$$Y = E(s_i q_i) + E\left(\int_0^{s_i} s_i - x\,dx\right)$$

Now a customer's waiting time in the queue can be determined by looking at the arrival process and at the service times of previous customers, so it is independent of her service time, i.e., $E(s_i q_i) = Es_i \cdot W_Q$ and we have

$$Y = (Es_i)W_Q + E(s_i^2/2)$$

PASTA implies that $V = W_Q$, so using $Y = V/\lambda$ and multiplying both sides by λ, we have

$$W_Q = \lambda(Es_i)W_Q + \lambda E(s_i^2/2)$$

Solving for W_Q now gives

$$W_Q = \frac{\lambda E(s_i^2)}{2(1 - \lambda Es_i)} \tag{3.8}$$

the so-called **Pollaczek-Khintchine formula**. Using formulas (3.5), (3.4), and (3.8) we can now compute

$$L_Q = \lambda W_Q \qquad W = W_Q + Es_i \qquad L = \lambda W$$

3.3 Age and Residual Life

Let $t_1, t_2, \dots$ be i.i.d. interarrival times, let $T_n = t_1 + \cdots + t_n$ be the time of the nth renewal, and let $N(t) = \max\{n : T_n \le t\}$ be the number of renewals by time t. Let

$$A(t) = t - T_{N(t)} \qquad \text{and} \qquad Z(t) = T_{N(t)+1} - t$$

$A(t)$ gives the age of the item in use at time t, while $Z(t)$ gives its residual lifetime.

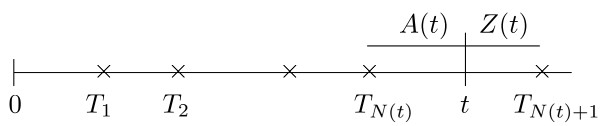


Figure 3.1: Age and residual life.

To explain the interest in $Z(t)$ note that the interrarival times after $T_{N(t)+1}$ will be independent of $Z(t)$ and i.i.d. with distribution F, so if we can show that $Z(t)$ converges in distribution, then the renewal process after time t will converge to an equilibrium.

3.3.1 Discrete case

The situation in which all the interarrival times are positive integers is very simple but also important because visits of a Markov chain to a fixed state, Example 3.1, are a special case. Let

$$V_m = \begin{cases} 1 & \text{if } m \in \{T_0, T_1, T_2, \dots\} \\ 0 & \text{otherwise} \end{cases}$$

$V_m = 1$ if a renewal occurs at time m, i.e., if T_n visits m. Let $Z_n = \min\{m - n : m \ge n, V_m = 1\}$ be the residual life. An example should help clarify the definitions:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
V_n	1	0	0	0	1	0	0	1	1	0	0	0	0	1
Z_n	0	3	2	1	0	2	1	0	0	4	3	2	1	0

It is clear that if $Z_n = i > 0$ then $Z_{n+1} = i - 1$. When $Z_n = 0$, a renewal has just occurred. If the time to the next renewal is k then $Z_{n+1} = k - 1$. To check this note that $Z_4 = 0$ and the time to the next renewal is 3 (it occurs at time 7) so $Z_5 = 2$.

To study Z_n we note that it is a Markov chain with state space $S = \{0, 1, 2, \ldots\}$ and transition probability

$$\begin{aligned} p(0,j) &= f_{j+1} && \text{for } j \geq 0 \\ p(i,i-1) &= 1 && \text{for } i \geq 1 \\ p(i,j) &= 0 && \text{otherwise} \end{aligned}$$

In this chain 0 is always recurrent. If there are infinitely many values of k with $f_k > 0$ then it is irreducible. If not and K is the largest value of k with $f_k > 0$ then $\{0, 1, \ldots K - 1\}$ is a closed irreducible set.

To define a stationary measure we will use the cycle trick, Theorem 1.14 with $x = 0$. Starting from 0 the chain will visit a site i at most 1 before it returns to 0, and this will happen if and only if the first jump is to a state $\geq i$, i.e., $t_1 \geq i + 1$. Thus the stationary measure is

$$\mu(i) = P(t_1 > i)$$

Using (A.26) we see that

$$\sum_{i=0}^{\infty} \mu(i) = Et_1$$

so the chain is positive recurrent if and only if $Et_1 < \infty$. In this case

$$\pi(i) = P(t_1 > i)/Et_1 \tag{3.9}$$

$I_0 \supset J_0 = \{k : f_k > 0\}$ so if the greatest common divisor of J_0 is 1 then 0 is aperiodic. To argue the converse note that I_0 consists of all finite sums of elements in J_0 so g.c.d. I_0 = g.c.d. J_0. Using the Markov chain convergence theorem now gives:

Theorem 3.10. *Suppose $Et_1 < \infty$ and the greatest common divisor of $\{k : f_k > 0\}$ is 1 then*

$$\lim_{n\to\infty} P(Z_n = i) = \frac{P(t_1 > i)}{Et_1}$$

In particular $P(Z_n = 0) \to 1/Et_1$.

Example 3.10. Visits to Go. In Monopoly one rolls two dice and then moves that number of squares. As in Example 1.27 we will ignore *Go to Jail*, *Chance*, and other squares that make the chain complicated. The average number of spaces moved in one roll is $Et_1 = 7$ so in the long run we land exactly on Go in 1/7 of the trips around the board.

3.3.2 Age

Let $A_n = \min\{n - m : m \le n,\, V_m = 1\}$ be the age of the item in use at time n. In the example considered earlier:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
V_n	1	0	0	0	1	0	0	1	1	0	0	0	0	1
A_n	0	1	2	3	0	1	2	0	0	1	2	3	4	0

Here when $A_n = 2$ we know that the associated renewal has $t_i > 2$. In order to move on to 3, we must have $t_i > 3$. So if we let $F_i = P(t_1 > i)$ then the transition probability q for A_n can be written as

$$q(j, j+1) = \frac{F_{j+1}}{F_j} \qquad q(j, 0) = 1 - \frac{F_{j+1}}{F_j} = \frac{f_{j+1}}{F_j} \quad \text{for } j \ge 0$$

Again this chain 0 is always recurrent. If there are infinitely many values of k with $f_k > 0$ then it is irreducible. If not and K is the largest value of k with $f_k > 0$ then $\{0, 1, \dots K - 1\}$ is a closed irreducible set.

To define a stationary measure we will use the cycle trick, Theorem 1.14 with $x = 0$. Starting from 0 the chain will visit a site i at most 1 before it returns to 0, and this will happen with probability

$$q(0,1)q(1,2)\cdots q(i-1,i) = \frac{F_1}{F_0} \cdot \frac{F_2}{F_1} \cdots \frac{F_i}{F_{i-1}} = F_i$$

so again the chain is positive recurrent if and only if $Et_1 < \infty$, and in this case

$$\pi(i) = P(t_1 > i)/Et_1$$

Using the Markov chain convergence theorem now gives:

Theorem 3.11. *Suppose $Et_1 < \infty$ and the greatest common divisor of $\{k : f_k > 0\}$ is 1 then*

$$\lim_{n\to\infty} P(A_n = i) = \frac{P(t_1 > i)}{Et_1}$$

Example 3.11. Shaving chain. The chain in Exercise 1.38 is a special case of the age chain. The transition probability is

	1	**2**	**3**	**4**
1	1/2	1/2	0	0
2	2/3	0	1/3	0
3	3/4	0	0	1/4
4	1	0	0	0

From this we see that

$$F_1/F_0 = 1/2 \quad F_2/F_1 = 1/3 \quad F_3/F_2 = 1/4 \quad F_4/F_3 = 0$$

so $F_1 = 1/2$, $F_2 = 1/6$, $F_3 = 1/24$ and it follows that

$$f_1 = 1/2 \quad f_2 = 1/3 \quad f_3 = 1/8 \quad f_4 = 1/24.$$

The mean $Et_1 = 41/24$ so the limit distribution has

$$\pi_0 = \frac{24}{41} \quad \pi_1 = \frac{12}{41} \quad \pi_2 = \frac{4}{41} \quad \pi_3 = \frac{1}{41}$$

It should be clear by comparing the numerical examples above that there is a close relationship between p and q. In fact, q is the dual chain of p, i.e., the chain p run backwards. To check this we need to verify that

$$q(i,j) = \frac{\pi(j)p(j,i)}{\pi(i)}$$

There are two cases to consider. If $i \geq 0$ and $j = i+1$

$$q(i,i+1) = \frac{F_{i+1}}{F_i} = \frac{\pi(i+1)p(i+1,i)}{\pi(i)}$$

since $\pi(j) = F_j/Et_1$. If $j = 0$ then

$$q(i,0) = \frac{f_{i+1}}{F_i} = \frac{\pi(0)p(0,i)}{\pi(i)}$$

3.3.3 General case

With the discrete case taken care of, we will proceed to the general case, which will be studied using renewal reward processes.

Example 3.12. Long run average residual life. To study the asymptotic behavior of $Z(t)$, we begin by noting that when $T_{i-1} < s < T_i$, we have $Z(s) = T_i - s$ and changing variables $r = T_i - s$ gives

$$\int_{T_{i-1}}^{T_i} Z(s)\,ds = \int_0^{t_i} r\,dr = t_i^2/2$$

So ignoring the contribution from the last incomplete cycle, $[T_{N(t)}, t]$ we have

$$\int_0^t Z(s)\,ds \approx \sum_{i=1}^{N(t)} t_i^2/2$$

The right-hand side is a renewal reward process with $r_i = t_i^2/2$, so it follows from Theorem 3.4 that

$$\frac{1}{t}\int_0^t Z(s)\,ds \to \frac{Et_i^2/2}{Et_i} \tag{3.10}$$

To see what this means we will consider two concrete examples:

A. Exponential. Suppose the t_i are exponential with rate λ. In this case $Et_i = 1/\lambda$, while integration by parts with $f(t) = t^2$ and $g'(t) = \lambda e^{-\lambda t}$ shows

$$\begin{aligned} Et_i^2 &= \int_0^\infty t^2\,\lambda e^{-\lambda t}\,dt = t^2(-e^{-\lambda t})\big|_0^\infty + \int_0^\infty 2te^{-\lambda t}\,dt \\ &= 0 + (2/\lambda)\int_0^\infty t\lambda e^{-\lambda t}\,dt = 2/\lambda^2 \end{aligned}$$

Using this in (3.10), it follows that

$$\frac{1}{t}\int_0^t Z(s)\,ds \to \frac{1/\lambda^2}{1/\lambda} = \frac{1}{\lambda}$$

This is not surprising since the lack of memory property of the exponential implies that for any s, $Z(s)$ has an exponential distribution with mean $1/\lambda$.

B. Uniform. Suppose the t_i are uniform on $(0, b)$, that is, the density function is $f(t) = 1/b$ for $0 < t < b$ and 0 otherwise. In this case the symmetry of the uniform distribution about $b/2$ implies $Et_i = b/2$, while a little calculus shows

$$Et_i^2 = \int_0^b t^2 \cdot \frac{1}{b}\, dt = \frac{t^3}{3b}\bigg|_0^b = \frac{b^2}{3}$$

Using this in (3.10) it follows that

$$\frac{1}{t}\int_0^t Z(s)\, ds \to \frac{b^2/6}{b/2} = \frac{b}{3}$$

Example 3.13. Long run average age. As in our analysis of $Z(t)$, we begin by noting that changing variables $s = T_{i-1} + r$ we have

$$\int_{T_{i-1}}^{T_i} A(s)\, ds = \int_0^{t_i} r\, dr = t_i^2/2$$

so ignoring the contribution from the last incomplete cycle $[T_{N(t)}, t]$ we have

$$\int_0^t A(s)\, ds \approx \sum_{i=1}^{N(t)} t_i^2/2$$

The right-hand side is the renewal reward process we encountered in Example 3.12, so it follows from Theorem 3.4 that

$$\frac{1}{t}\int_0^t A(s)\, ds \to Et_i^2/2Et_i \tag{3.11}$$

Combining the results from the last two examples leads to a surprise.

Example 3.14. Inspection paradox. Let $L(t) = A(t) + Z(t)$ be the lifetime of the item in use at time t. Adding (3.10) and (3.11), we see that the average lifetime of the items in use up to time t:

$$\frac{1}{t}\int_0^t L(s)\, ds \to \frac{Et_i^2}{Et_i} \tag{3.12}$$

To see that this is surprising, note that:

(i) If $\text{var}\,(t_i) = Et_i^2 - (Et_i)^2 > 0$, then $Et_i^2 > (Et_i)^2$, so the limit is $> Et_i$.

(ii) The average of the lifetimes of the first n items:

$$\frac{t_1 + \cdots + t_n}{n} \to Et_i$$

Thus, the average lifetime of items in use up to time t converges to a limit Et_i^2/Et_i, which is larger than Et_i, the limiting average lifetime of the first n items.

There is a simple explanation for this "paradox": taking the average age of the item in use up to time s is biased since items that last for time u are counted u times. For simplicity suppose that there are only a finite number of possible lifetimes $\ell_1 < \ell_2 \ldots < \ell_k$ with probabilities $p_1, \ldots, p_k > 0$ and $p_1 + \cdots + p_k = 1$.

By considering a renewal reward process in which $r_i = 1$ when $t_i = \ell_j$, we see that the number of items up to time t with lifetime ℓ_j is $\sim p_j N(t)$. The total length of these lifetimes is $\sim \ell_j p_j N(t)$ Thus the limiting fraction of time the lifetime is ℓ_j is by Theorem 3.4

$$\sim \frac{\ell_j p_j N(t)}{t} \to \frac{\ell_j p_j}{Et_i}$$

The expected value of this limiting distribution is

$$\sum_j \ell_j \frac{\ell_j p_j}{Et_i} = \frac{Et_i^2}{Et_i}$$

Example 3.15. Limiting distribution for the residual life. Let $Z(t) = T_{N(t)+1} - t$. Let $I_c(t) = 1$ if $Z(t) \le c$ and 0 otherwise. To study the asymptotic behavior of $I_c(t)$, we begin with the observation that

$$\int_{T_{i-1}}^{T_i} I_c(s)\, ds = \min\{t_i, c\}$$

To check this we consider two cases.

Case 1. $t_i \ge c$. $I_c(s) = 1$ for $T_i - c \le s \le T_i$, 0 otherwise, so the integral is c.

Case 2. $t_i \le c$. $I_c(s) = 1$ for all $s \in [T_{i-1}, T_i]$, so the integral is t_i.

Ignoring the contribution from the last incomplete cycle $[T_{N(t)}, t]$, we have

$$\int_0^t I_c(s)\, ds \approx \sum_{i=1}^{N(t)} \min\{t_i, c\}$$

The right-hand side is a renewal reward process with $r_i = \min\{t_i, c\}$, so it follows from Theorem 3.4 that

$$\frac{1}{t}\int_0^t I_c(s)\,ds \to \frac{E\min\{t_i, c\}}{Et_i} \tag{3.13}$$

Example 3.16. Limiting distribution for the age. Let $A(t) = t - T_{N(t)+1}$ and let $J_c(t) = 1$ if $A(t) \le c$ and 0 otherwise. Imitating the last argument, it is easy to show that

$$\frac{1}{t}\int_0^t J_c(s)\,ds \to \frac{E\min\{t_i, c\}}{Et_i} \tag{3.14}$$

Proof. To study the asymptotic behavior of $J_c(t)$ we begin with the observation that

$$\int_{T_{i-1}}^{T_i} J_c(s)\,ds = \min\{t_i, c\}$$

which can be checked by considering two cases as before.

Ignoring the contribution from the last incomplete cycle $[T_{N(t)}, t]$, we have

$$\int_0^t I_c(s)\,ds \approx \sum_{i=1}^{N(t)} \min\{t_i, c\}$$

The right-hand side is a renewal reward process with $r_i = \min\{t_i, c\}$, so (3.14) follows from Theorem 3.4. □

To evaluate the limits in (3.13) and (3.14), we note that (1.9) implies

$$E\min\{t_i, c\} = \int_0^\infty P(\min\{t_i, c\} > t)\,dt = \int_0^c P(t_i > t)\,dt$$

From this it follows that the limiting fraction of time that the age of the item in use at time t is $\le c$ is

$$G(c) = \frac{\int_0^c P(t_i > t)\,dt}{Et_i} \tag{3.15}$$

This is a distribution function of a nonnegative random variable since $G(0) = 0$, $t \to G(t)$ is nondecreasing, and $G(\infty) = 1$ by (1.9).

Differentiating we find that the density function of the limiting age is given by

$$g(c) = \frac{d}{dc}G(c) = \frac{P(t_i > c)}{Et_i} \tag{3.16}$$

which agrees with the results in Theorems 3.10 and 3.11.

Turning to our two concrete examples:

A. Exponential. In this case the limiting density given in (3.16) is

$$\frac{P(t_i > c)}{Et_i} = \frac{e^{-\lambda c}}{1/\lambda} = \lambda e^{-\lambda c}$$

For the residual life this is not surprising, since the distribution of $Z(s)$ is always exponential with rate λ.

B. Uniform. Plugging into (3.16) gives for $0 < c < b$:

$$\frac{P(t_i > c)}{Et_i} = \frac{(b-c)/b}{b/2} = \frac{2(b-c)}{b^2}$$

In words, the limiting density is a linear function that starts at $2/b$ at 0 and hits 0 at $c = b$.

In case A, the limit distribution $G = F$, while in case B, $G \neq F$. To show that case B is the rule to which case A is the only exception, we prove:

Theorem 3.12. *Suppose that $G = F$ in (3.15). Then $F(x) = 1 - e^{-\lambda x}$ for some $\lambda > 0$.*

Proof. Let $H(c) = 1 - G(c)$, and $\lambda = 1/Et_i$. (3.15) implies $H'(c) = -\lambda H(c)$. Combining this with the observation $H(0) = 1$ and using the uniqueness of the solution of the differential equation, we conclude that $H(c) = e^{-\lambda c}$. $\square$

3.4 Renewal equations

The results in the second half of the previous section concern the limiting behavior of the average distribution on $[0,t]$. In some cases due to a problem similar to periodicity, the distributions of $Z(t)$ and $A(t)$ do not converge. We say that t_i has an **arithmetic distribution** if there is a $\delta > 0$ so that $P(t_i \in \{\delta, 2\delta, \dots\}) = 1$. In the discrete case this holds with $\delta = 1$. If the distribution is arithmetic and $t = m\delta + \gamma$ where $\gamma \in [0, \delta)$ then $P(Z(t) \in \{\delta - \gamma, 2\delta - \gamma, \dots\}) = 1$ so the distribution of $Z(t)$ does not converge.

The arithmetic case can be reduced to the discrete case by dividing by δ so for the rest of this section we will suppose that the distribution of t_i is not arithmetic. Let

$$U(t) = 1 + EN(t) = \sum_{n=0}^{\infty} P(T_n \le t) \tag{3.17}$$

In words, $U(t)$ is the expected number of renewals including the one at time 0. The big result about the asymptotic behavior of $U(t)$ is

Theorem 3.13. Blackwell's renewal theorem. *Suppose the distribution of t_i is not arithmetic and let $\mu = Et_i \le \infty$. For any $h > 0$,*

$$U(t+h) - U(t) \to h/\mu \quad \text{as } t \to \infty.$$

Theorem 3.3 implies that $U(t)/t \to 1/\mu$, i.e., in the long run renewals occur at rate $1/t$. Theorem 3.13 implies this is true when we look at intervals of any length h.

Limit results for the age and residual life can be proved for nonarithmetic distributions. To avoid the need for Riemann-Steiltjes integrals, we will assume that the t_i have a density function f. In this case, by considering the value of t_1 we have

$$U(t) = 1 + \int_0^t U(t-s) f(s)\, ds \tag{3.18}$$

To check this, we note that if $t_1 > t$ then there is only one renewal before time t but if $t_1 = s$ then the expected number of renewals by time t is $1 + U(t-s)$.

To explain why $Z(t)$ should converge to the distribution in (3.15) we will consider

Example 3.17. Delayed renewal processes. Let $t_1, t_2, \ldots$ be independent and have density f. If $T_0 \geq 0$ is independent of $t_1, t_2, \ldots$ and has density function g, then $T_k = T_{k-1} + t_k$, $k \geq 1$ defines a **delayed renewal process**. If we let $R_t = \min\{k : T_k > t\}$ be the number of renewals in $[0, t]$ counting the one at T_0, and let $V(t) = ER_t$, then breaking things down according to the value of T_0 gives

$$V(t) = \int_0^t U(t-s)g(s)\,ds \tag{3.19}$$

To check this note that if $T_0 > t$ then there are no renewals, while if $T_0 = s$ it is as if we have an ordinary renewal process starting at s and run for time $t - s$.

We know $U(t) \sim t/\mu$. Our next step is to find a g so that $V(t) = t/\mu$. Using (3.18) with t replaced by $t-s$ and the integration done with respect to r, we want

$$\frac{t}{\mu} = \int_0^t g(s)\,ds + \int_0^t \int_0^{t-s} U(t-s-r)f(r)\,dr g(s)\,ds$$

Interchanging the order of integration, the double integral is

$$\int_0^t dr f(r) \int_0^{t-r} U(t-r-s)g(s)\,ds = \int_0^t dr f(r)\,\frac{t-r}{\mu}$$

Rearranging we want the distribution function $G(t) = \int_0^t g(s)\,ds$ to satisfy

$$G(t) = \frac{t}{\mu} - \int_0^t dr f(r)\frac{t-r}{\mu} \tag{3.20}$$

The integration by parts formula is

$$\int_0^t f_1'(s)f_2(s)\,ds = f_1(s)f_2(s)\,ds\big|_0^t - \int_0^t f_1(s)f_2'(s)\,ds$$

Using this with $f_1'(r) = -f(r)$, $f_2(r) = (t-r)/\mu$, $f_2'(r) = 1/\mu$ and choosing $f_1(r) = 1 - F(r)$, where $F(r) = \int_0^r f(s)\,ds$ is the distribution function, we have

$$G(t) = \frac{t}{\mu} - \frac{t}{\mu} + \frac{1}{\mu}\int_0^t 1 - F(r)\,dr$$

so we want $g(r) = (1 - F(r))/\mu$.

Intuitively, g is a stationary distribution for the residual life. To see this, note that the calculation above shows that if we start with this delay then the $V(t) = t/\mu$, and g is the only distribution with this property. Since the expected number of renewals in $[s,t]$ is $(t-s)/\mu$, it follows that the delay at time s must have distribution g.

To study the asymptotic behavior of the age and residual life we will study **renewal equations**.

$$H(t) = h(t) + \int_0^t H(t-s) f(s)\, ds \tag{3.21}$$

To explain our interest we will consider two examples:

Example 3.18. Residual life. Let $x > 0$ be fixed, and let $H(t) = P(Z(t) > x)$. In this case

$$H(t) = (1 - F(t+x)) + \int_0^t H(t-s) f(s)\, ds$$

To check this note that if $t_1 > t$ then the residual life will be $> x$ if $t > t + x$. If $t_1 = s$ then we have a renewal process starting at s and we are interested in its residual life after $t - s$

Example 3.19. Age. Let $x > 0$ be fixed, and let $H(t) = P(A(t) \leq x)$. In this case

$$H(t) = (1 - F(t)) 1_{[0,x]}(t) + \int_0^t H(t-s) f(s)\, ds$$

To check this note that if $t_1 > t$ then the age will be $\leq x$ if and only if $t \leq x$. If $t_1 = s$ then we have a renewal process starting at s and we are interested in its age at $t - s$.

To motivate the solution of the renewal equation we recall (3.18)

$$U(t) = 1 + \int_0^t U(t-s) f(s)\, ds$$

is a solution with $h \equiv 1$. Let $f^1 = f$ and for $n \geq 2$ let

$$f^n(t) = \int_0^t f^{n-1}(s) f(t-s)\, ds$$

be the distribution of T_n. Let

$$u(s) = \sum_{n=1}^{\infty} f^n(s) = U'(s)$$

Theorem 3.14. *If h is bounded then the function*

$$H(t) = h(t) + \int_0^t h(t-s)u(s)\,ds$$

is the unique solution of the renewal equation that is bounded on bounded intervals.

Proof. Replacing t by $t-s$ in (3.21) and changing variables in the integral we have

$$H(t-s) = h(t-s) + \int_0^{t-s} H(r)f(t-s-r)\,dr.$$

Using this in (3.21) gives

$$H(t) = h(t) + \int_0^t h(t-s)f(s)\,ds + \int_0^t \int_0^{t-s} H(r)f(t-s-r)\,dr f(s)\,ds$$

Interchanging the order of integration in the double integral

$$\int_0^t dr H(r) \int_0^{t-r} ds f(t-s-r)f(s) = \int_0^t dr H(r) f^2(t-r)$$

Changing variables in the last integral we have

$$H(t) = h(t) + \int_0^t h(t-s)f(s)\,ds + \int_0^t H(t-s)f^2(s)\,ds$$

Repeating the last calculation leads to

$$H(t) = h(t) + \sum_{i=1}^{n} \int_0^t h(t-s)f^m(s)\,ds + \int_0^t H(t-s)f^{n+1}(s)\,ds$$

If $|H(s)| \le K(t)$ for $s \le t$ then the last term is $\le K(t)P(T_{n+1} \le t)$ as $n \to \infty$ and we conclude that $H(t)$ must have the indicated form. Since our solution has $|H(t)| \le K(t)U(t)$ for $s \le t$, it is bounded on bounded intervals. □

The next result gives the asymptotic behavior of the solutions.

Theorem 3.15. *Suppose $h(t) \geq 0$ is nonincreasing and has $\int_0^\infty h(s)\,ds < \infty$ then*

$$H(t) \to \frac{1}{\mu} \int_0^\infty h(s)\,ds$$

Sketch of proof. If $h(t) = 1_{[0,h]}$ this follows from Blackwell's renewal theorem, 3.13. The result can now be proved by approximating h above and below by functions that are constant on $[mh, (m+1)h)$ and then letting $h \to 0$. □

Residual life. In Example 3.18, $h(t) = 1 - F(t+x)$ satisfies the assumptions of Theorem 3.15 so

$$P(Z(t) > x) \to \frac{1}{\mu} \int_0^\infty 1 - F(t+x)\,dt = \frac{1}{\mu} \int_x^\infty 1 - F(s)\,ds \quad (3.22)$$

Age. In Example 3.19, $h(t) = (1 - F(t))1_{[0,x]}(t)$ satisfies the assumptions of Theorem 3.15 so

$$P(Z(t) > x) \to \frac{1}{\mu} \int_0^x 1 - F(t)\,dt \quad (3.23)$$

Rate of convergence. (3.20) can be rewritten as

$$\frac{t}{\mu} = G(t) + \int_0^t \frac{t-s}{\mu} f(s)\,ds$$

Subtracting this from (3.18) shows that $H(t) = U(t) - t/\mu$ satisfies the renewal equation (3.21) with

$$h(t) = 1 - G(t) = (1/\mu) \int_t^\infty 1 - F(s)\,ds.$$

This function is ≥ 0 and nonincreasing.

$$\int_0^\infty h(t)\,dt = \frac{1}{\mu} \int_0^\infty \int_t^\infty 1 - F(s)\,ds\,dt$$

Interchanging the order of integration in the double integral we have

$$\int_0^\infty ds\,(1 - F(s)) \int_0^s dt = \int_0^\infty s(1 - F(s))\,ds = Et_1^2/2$$

by (A.23). Thus if $Et_i^2 < \infty$

$$U(t) - \frac{t}{\mu} \to \frac{Et_i^2}{2\mu^2} \quad (3.24)$$

3.5 Exercises

3.1. The weather in a certain locale consists of alternating wet and dry spells. Suppose that the number of days in each rainy spell is a Poisson distribution with mean 2, and that a dry spell follows a geometric distribution with mean 7. Assume that the successive durations of rainy and dry spells are independent. What is the long-run fraction of time that it rains?

3.2. Monica works on a temporary basis. The mean length of each job she gets is 11 months. If the amount of time she spends between jobs is exponential with mean 3 months, then in the long run what fraction of the time does she spend working?

3.3. Thousands of people are going to a Grateful dead concert in Pauley Pavillion at UCLA. They park their 10 foot cars on several of the long streets near the arena. There are no lines to tell the drivers where to park, so they park at random locations, and end up leaving spacings between the cars that are independent and uniform on $(0, 10)$. In the long run, what fraction of the street is covered with cars?

3.4. The times between the arrivals of customers at a taxi stand are independent and have a distribution F with mean μ_F. Assume an unlimited supply of cabs, such as might occur at an airport. Suppose that each customer pays a random fare with distribution G and mean μ_G. Let $W(t)$ be the total fares paid up to time t. Find $\lim_{t\to\infty} E\,W(t)/t$.

3.5. A policeman cruises (on average) approximately 10 minutes before stopping a car for speeding. 90% of the cars stopped are given speeding tickets with an \$80 fine. It takes the policeman an average of 5 minutes to write such a ticket. The other 10% of the stops are for more serious offenses, leading to an average fine of \$300. These more serious charges take an average of 30 minutes to process. In the long run, at what rate does he assign fines.

3.6. A group of n children continuously, and independently, climb up and then sled down a slope. Assume that each child's actions follow an alternating renewal process: climbing up for an amount of time with distribution F and mean μ_F, and then sliding down for an amount of time with distribution G and mean μ_G. Let

$U(t)$ be the number of children climbing the hill at time t. Find $\lim_{t\to\infty} P(U(t) = k)$.

3.7. *Counter processes.* As in Example 1.5, we suppose that arrivals at a counter come at times of a Poisson process with rate λ. An arriving particle that finds the counter free gets registered and then locks the counter for an amount of time τ. Particles that arrive while the counter is locked have no effect. (a) Find the limiting probability the counter is locked at time t. (b) Compute the limiting fraction of particles that get registered.

3.8. A cocaine dealer is standing on a street corner. Customers arrive at times of a Poisson process with rate λ. The customer and the dealer then disappear from the street for an amount of time with distribution G while the transaction is completed. Customers that arrive during this time go away never to return. (a) At what rate does the dealer make sales? (b) What fraction of customers are lost?

3.9. A worker has a number of machines to repair. Each time a repair is completed a new one is begun. Each repair independently takes an exponential amount of time with rate μ to complete. However, independent of this, mistakes occur according to a Poisson process with rate λ. Whenever a mistake occurs, the item is ruined and work is started on a new item. In the long run how often are jobs completed?

3.10. Three children take turns shooting a ball at a basket. The first shoots until she misses, then the second shoots until she misses, the third shoots until she misses, then the process starts again with the first child. Suppose that child i makes a basket with probability p_i and that successive trials are independent. Determine the proportion of time in the long run that each child shoots.

3.11. Solve the previous problem when $p_1 = 2/3$, $p_2 = 3/4$, $p_3 = 4/5$.

3.12. Bruno works for an amount of time that is uniformly distributed on $[6, 10]$ hours then he relaxes for an exponentially distributed number of hours; with mean 4. (a) Bruno never sleeps. Ignoring mundane details like night and day, what is the long-run fraction of time he spends working? (b) Suppose that when Bruno first begins relaxing he drinks a beer and then drinks one beer each hour after that. That, is if his relaxation period is from 12:13 to 3:27 he drinks 4 beers, one each at 12:13, 1:13, 2:13, and 3:13. Find the probability he drinks n beers in one relaxation period and the long-run average number of beers he drinks each 24 hours.

3.13. A young doctor is working at night in an emergency room. Emergencies come in at times of a Poisson process with rate 0.5 per hour. The doctor can only get to sleep when it has been 36 minutes (.6 hours) since the last emergency. For example, if there is an emergency at 1:00 and a second one at 1:17 then she will not be able to get to sleep until at least 1:53, and it will be even later if there is another emergency before that time.
(a) Compute the long-run fraction of time she spends sleeping, by formulating a renewal reward process in which the reward in the ith interval is the amount of time she gets to sleep in that interval.
(b) The doctor alternates between sleeping for an amount of time s_i and being awake for an amount of time u_i. Use the result from (a) to compute Eu_i.

3.14. A scientist has a machine for measuring ozone in the atmosphere that is located in the mountains just north of Los Angeles. At times of a Poisson process with rate 1, storms or animals disturb the equipment so that it can no longer collect data. The scientist comes every L units of time to check the equipment. If the equipment has been disturbed then she can usually fix it quickly so we will assume the the repairs take 0 time. (a) What is the limiting fraction of time the machine is working? (b) Suppose that the data that is being collected is worth a dollars per unit time, while each inspection costs $c < a$. Find the best value of the inspection time L.

3.15. The city of Ithaca, New York, allows for two-hour parking in all downtown spaces. Methodical parking officials patrol the downtown area, passing the same point every two hours. When an official encounters a car, he marks it with chalk. If the car is still there two hours later, a ticket is written. Suppose that you park your car for a random amount of time that is uniformly distributed on $(0, 4)$ hours. What is the probability you will get a ticket.

3.16. Each time the frozen yogurt machine at the mall breaks down, it is replaced by a new one of the same type. In the long run what percentage of time is the machine in use less than one year old if the the lifetime distribution of a machine is: (a) uniformly distributed on (0,2)? (b) exponentially distributed with mean 1?

3.17. While visiting Haifa, Sid Resnick discovered that people who wish to travel quickly from the port area up the mountain to The Carmel frequently take a taxi known as a sherut. The system operates as follows: Sherut-eem are lined up in a row at a taxi stand. The capacity of each car is 5 people. Potential customers arrive according to a Poisson process with rate λ. As soon as 5 people are in the car, it departs in a cloud of diesel emissions. The next car moves up, accepts passengers until it is full, then departs for The Carmel, and so on. A local resident (who has no need of a ride) wanders onto the scene. What is the distribution of the time he has to wait to see a cab depart?

3.18. In front of terminal C at the Chicago airport is an area where hotel shuttle vans park. Customers arrive at times of a Poisson process with rate 10 per hour looking for transportation to the Hilton hotel nearby. When 7 people are in the van it leaves for the 36-minute round trip to the hotel. Customers who arrive while the van is gone go to some other hotel instead. (a) What fraction of the customers actually go to the Hilton? (b) What is the average amount of time that a person who actually goes to the Hilton ends up waiting in the van?

3.19. Let $A(t)$ and $Z(t)$ be, respectively, the age and the residual life at time t in a renewal process in which the interarrival times have distribution F. Compute $P(Z(t) > x | A(t) = s)$.

3.20. Let $A(t)$ and $Z(t)$ be, respectively, the age and the residual life at time t in a renewal process in which the interarrival times have distribution F. Use the methods of Section 5.2 to compute the limiting behavior of the joint distribution $P(A(t) > x, Z(t) > y)$.

3.21. Let $t_1, t_2, \ldots$ be independent and identically distributed with $Et_i = \mu$ and $\text{var}\,(t_i) = \sigma^2$. Use (A.23) with $p = 2$ to compute the mean of the limiting distribution for the residual lifetime given in (3.16).

Chapter 4

Markov Chains in Continuous Time

4.1 Definitions and Examples

In Chapter 1 we considered Markov chains X_n with a discrete time index $n = 0, 1, 2, \dots$ In this chapter we will extend the notion to a continuous time parameter $t \geq 0$, a setting that is more convenient for some applications. In discrete time we formulated the Markov property as: for any possible values of $j, i, i_{n-1}, \dots i_0$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

In continuous time, it is technically difficult to define the conditional probability given all of the X_r for $r \leq s$, so we instead say that X_t, $t \geq 0$ is a Markov chain if for any $0 \leq s_0 < s_1 \cdots < s_n < s$ and possible states $i_0, \dots, i_n, i, j$ we have

$$P(X_{t+s} = j | X_s = i, X_{s_n} = i_n, \dots, X_{s_0} = i_0) = P(X_t = j | X_0 = i)$$

In words, given the present state, the rest of the past is irrelevant for predicting the future. Note that built into the definition is the fact that the probability of going from i at time s to j at time $s + t$ only depends on t the difference in the times.

Our first step is to construct a large collection of examples. In Example 4.6 we will see that this is almost the general case.

Example 4.1. Let $N(t)$, $t \geq 0$ be a Poisson process with rate λ and let Y_n be a discrete time Markov chain with transition probability

$u(i,j)$. Then $X_t = Y_{N(t)}$ is a continuous-time Markov chain. In words, X_t takes one jump according to $u(i,j)$ at each arrival of $N(t)$.

Why is this true? Intuitively, this follows from the lack of memory property of the exponential distribution. If $X_s = i$, then independent of what has happened in the past, the time to the next jump will be exponentially distributed with rate λ and will go to state j with probability $u(i,j)$. □

Discrete time Markov chains were described by giving their transition probabilities $p(i,j)$ = the probability of jumping from i to j in one step. In continuous time there is no first time $t > 0$ so we introduce for each $t > 0$ a **transition probability**

$$p_t(i,j) = P(X_t = j | X_0 = i)$$

To compute this for Example 4.1, we note that $N(t)$ has a Poisson number of jumps with mean λt, so

$$p_t(i,j) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} u^n(i,j)$$

where $u^n(i,j)$ is the nth power of the transition probability $u(i,j)$.

In continuous time, as in discrete time, the transition probability satisfies

Theorem 4.1. Chapman–Kolmogorov equation.

$$\sum_k p_s(i,k) p_t(k,j) = p_{s+t}(i,j)$$

Why is this true? In order for the chain to go from i to j in time $s+t$, it must be in some state k at time s, and the Markov property implies that the two parts of the journey are independent. □

Proof. Breaking things down according to the state at time s, we have

$$P(X_{s+t} = j | X_0 = i) = \sum_k P(X_{s+t} = j, X_s = k | X_0 = i)$$

Using the definition of conditional probability and the Markov property, the above is

$$= \sum_k P(X_{s+t} = j | X_s = k, X_0 = i) P(X_s = k | X_0 = i) = \sum_k p_t(k,j) p_s(i,k)$$

□

(4.1) shows that if we know the transition probability for $t < t_0$ for any $t_0 > 0$, we know it for all t. This observation and a large leap of faith (which we will justify later) suggests that the transition probabilities p_t can be determined from their derivatives at 0:

$$q(i,j) = \lim_{h \to 0} \frac{p_h(i,j)}{h} \qquad \text{for } j \neq i \tag{4.1}$$

If this limit exists (and it will in all the cases we consider) we will call $q(i,j)$ the **jump rate** from i to j. To explain this name we will compute the:

Jump rates for Example 4.1. The probability of at least two jumps by time h is 1 minus the probability of 0 or 1 jumps

$$\begin{aligned} 1 - \left(e^{-\lambda h} + \lambda h e^{-\lambda h}\right) &= 1 - (1 + \lambda h)\left(1 - \lambda h + \frac{(\lambda h)^2}{2!} + \dots\right) \\ &= (\lambda h)^2/2! + \dots = o(h) \end{aligned}$$

That is, when we divide it by h it tends to 0 as $h \to 0$. The probability of going from i to j in zero steps, $u^0(i,j) = 0$, when $j \neq i$, so

$$\frac{p_h(i,j)}{h} \approx \lambda e^{-\lambda h} u(i,j) \to \lambda u(i,j)$$

as $h \to 0$. Comparing the last equation with the definition of the jump rate in (4.1) we see that $q(i,j) = \lambda u(i,j)$. In words we leave i at rate λ and go to j with probability $u(i,j)$. □

Example 4.1 is atypical. There we started with the Markov chain and then computed its rates. In most cases it is much simpler to describe the system by writing down its transition rates $q(i,j)$ for $i \neq j$, which describe the rates at which jumps are made from i to j. The simplest possible example is:

Example 4.2. Poisson process. Let $X(t)$ be the number of arrivals up to time t in a Poisson process with rate λ. Since arrivals occur at rate λ in the Poisson process the number of arrivals, $X(t)$, increases from n to $n+1$ at rate λ, or in symbols

$$q(n, n+1) = \lambda \quad \text{for all } n \geq 0$$

This simplest example is a building block for other examples:

Example 4.3. M/M/s queue. Imagine a bank with s tellers that serve customers who queue in a single line if all of the servers are busy. We imagine that customers arrive at times of a Poisson process with rate λ, and that each service time is an independent exponential with rate μ. As in Example 4.2, $q(n, n+1) = \lambda$. To model the departures we let

$$q(n, n-1) = \begin{cases} n\mu & 0 \leq n \leq s \\ s\mu & n \geq s \end{cases}$$

To explain this, we note that when there are $n \leq s$ individuals in the system then they are all being served and departures occur at rate $n\mu$. When $n > s$, all s servers are busy and departures occur at $s\mu$.

This model is in turn a stepping stone to another, more realistic one:

Example 4.4. M/M/s queue with balking. Again customers arrive at the bank in Example 4.3 at rate λ, but this time they only join the queue with probability a_n if there are n customers in line. Since customers flip coins to determine if they join the queue, this thins the Poisson arrival rate so that

$$q(n, n+1) = \lambda a_n \quad \text{for } n \geq 0$$

Of course, the service rates $q(n, n-1)$ remain as they were previously.

Having seen several examples, it is natural to ask:

Given the rates, how do you construct the chain?

Let $\lambda_i = \sum_{j \neq i} q(i, j)$ be the rate at which X_t leaves i. If $\lambda_i = \infty$, it will want to leave immediately, so we will always suppose that each state i has $\lambda_i < \infty$. If $\lambda_i = 0$, then X_t will never leave i. So suppose $\lambda_i > 0$ and let

$$r(i,j) = q(i,j)/\lambda_i$$

Here r, short for "routing matrix," is the probability the chain goes to j when it leaves i.

Informal construction. If X_t is in a state i with $\lambda_i = 0$ then X_t stays there forever and the construction is done. If $\lambda_i > 0$, X_t stays at i for an exponentially distributed amount of time with rate λ_i, then goes to state j with probability $r(i,j)$.

Formal construction. Suppose, for simplicity, that $\lambda_i > 0$ for all i. Let Y_n be a Markov chain with transition probability $r(i,j)$. The discrete-time chain Y_n, gives the road map that the continuous-time process will follow. To determine how long the process should stay in each state let $\tau_0, \tau_1, \tau_2, \ldots$ be independent exponentials with rate 1.

At time 0 the process is in state X_0 and should stay there for an amount of time that is exponential with rate $\lambda(X_0)$, so we let the time the process stays in state X_0 be $t_1 = \tau_0/\lambda(X_0)$.

At time $T_1 = t_1$ the process jumps to X_1, where it should stay for an exponential amount of time with rate $\lambda(X_1)$, so we let the time the process stays in state X_1 be $t_2 = \tau_1/\lambda(X_1)$.

At time $T_2 = t_0 + t_2$ the process jumps to X_2, where it should stay for an exponential amount of time with rate $\lambda(X_2)$, so we let the time the process stays in state X_2 be $t_3 = \tau_2/\lambda(X_2)$.

Continuing in the obvious way, we can let the amount of time the process stays in X_{n-1} be $t_n = \tau_{n-1}/\lambda(X_{n-1})$, so that the process jumps to X_n at time

$$T_n = t_1 + \cdots + t_n$$

In symbols, if we let $T_0 = 0$, then for $n \geq 0$ we have

$$X(t) = Y_n \quad \text{for } T_n \leq t < T_{n+1} \tag{4.2}$$

Computer simulation. Before we turn to the dark side of the construction above, the reader should observe that it gives rise to a recipe for simulating a Markov chain. Generate independent standard exponentials τ_i, say, by looking at $\tau_i = -\ln U_i$ where U_i are uniform on $(0,1)$. Using another sequence of random numbers, generate the transitions of X_n, then define t_i, T_n, and X_t as above.

The good news about the formal construction above is that if $T_n \to \infty$ as $n \to \infty$, then we have succeeded in defining the process for all time and we are done. This will be the case in almost all the examples we consider. The bad news is that $\lim_{n\to\infty} T_n < \infty$ can happen.

Example 4.5. An exploding Markov chain. Think of a ball that is dropped and returns to half of its previous height on each bounce. Summing $1/2^n$ we conclude that all of its infinitely many bounces will be completed in a finite amount of time. To turn this idea into an example of a badly behaved Markov chain on the state space $S = \{1, 2, \ldots\}$, suppose $q(i, i+1) = 2^i$ for $i \geq 1$ with all the other $q(i,j) = 0$. The chain stays at i for an exponentially distributed amount of time with mean 2^{-i} before it goes to $i+1$. Let T_j be the first time the chain reaches j. By the formula for the rates

$$\sum_{i=1}^{\infty} E_1(T_{i+1} - T_i) = \sum_{i=1}^{\infty} 2^{-i} = 1$$

This implies $T_\infty = \lim_{n\to\infty} T_n$ has $E_1 T_\infty = 1$ and hence $P_1(T_\infty < \infty) = 1$.

In most models, it is senseless to have the process make an infinite amount of jumps in a finite amount of time so we introduce a "cemetery state" Δ to the state space and complete the definition by letting $T_\infty = \lim_{n\to\infty} T_n$ and setting

$$X(t) = \Delta \quad \text{for all } t \geq T_\infty$$

The simplest way to rule out explosions (i.e., $P_x(T_\infty < \infty) > 0$) is to consider

Example 4.6. Markov chains with bounded rates. When the maximum transition rate

$$\Lambda = \max_i \lambda_i < \infty$$

we can use a trick to reduce the process to one with constant transition rates $\lambda_i \equiv \Lambda$. Let

$$\begin{aligned} u(i,j) &= q(i,j)/\Lambda \quad \text{for } j \neq i \\ u(i,i) &= 1 - \sum_{j \neq i} q(i,j)/\Lambda \end{aligned}$$

In words, while the chain is at any state i, it attempts to make transitions at rate Λ. On each attempt it jumps from i to j with probability $u(i,j)$ and stays put with the remaining probability $1 - \sum_{j \neq i} u(i,j)$. Since the jump rate is independent of the state, the jump times are a Poisson process and $T_n \to \infty$.

Here, we have come full circle by writing a general Markov chain with bounded flip rates in the form given in Example 1.1. This observation is often useful in simulation. Since the holding times in each state are exponentially distributed with the same rate Λ, then, as we will see in Section 4, if we are interested in the long-run fraction of time the continuous time chain spends in each state, we can ignore the holding times and simulate the discrete time chain with transition probability u.

There are many interesting examples with unbounded rates. Perhaps the simplest and best known is

Example 4.7. The Yule process. In this simple model of the growth of a population (of bacteria for example), there are no deaths and each particle splits into birth at rate β, so $q(i, i+1) = \beta i$ and the other $q(i,j)$ are 0. If we start with one individual, then the jump to $n+1$ is made at time $T_n = t_1 + \cdots + t_n$, where t_n is exponential with rate βn. $Et_n = 1/\beta i$, so

$$ET_n = (1/\beta) \sum_{m=1}^{n} 1/m \sim (\log n)/\beta$$

as $n \to \infty$. This is by itself not enough to establish that $T_n \to \infty$, but it is not hard to fill in the missing details.

Proof. $\operatorname{var}(T_n) = \sum_{m=1}^{n} 1/m^2\beta^2 \leq C = \sum_{m=1}^{\infty} 1/m^2\beta^2$. Chebyshev's inequality implies

$$P(T_n \leq ET_n/2) \leq 4C/(ET_n)^2 \to 0$$

as $n \to \infty$. Since $n \to T_n$ is increasing, it follows that $T_n \to \infty$. $\square$

The next example shows that linear growth of the transition rates is at the borderline of explosive behavior.

Example 4.8. Pure birth processes with power law rates. Suppose $q(i, i+1) = i^p$ and all the other $q(i, j) = 0$. In this case the jump to $n + 1$ is made at time $T_n = t_1 + \cdots + t_n$, where t_n is exponential with rate n^p. $Et_n = 1/n^p$, so if $p > 1$

$$ET_n = \sum_{m=1}^{n} 1/m^p$$

stays bounded as $n \to \infty$. This implies $ET_\infty = \sum_{m=1}^{\infty} 1/m^p$, so $T_\infty < \infty$ with probability one.

4.2 Computing the Transition Probability

In the last section we saw that given jump rates $q(i,j)$ we can construct a Markov chain that has these jump rates. This chain, of course, has a transition probability

$$p_t(i,j) = P(X_t = j | X_0 = i)$$

Our next question is: How do you compute the transition probability p_t from the jump rates q?

Our road to the answer starts by using the Chapman–Kolmogorov equations, Theorem 4.1, and then taking the $k = i$ term out of the sum.

$$\begin{aligned} p_{t+h}(i,j) - p_t(i,j) &= \left(\sum_k p_h(i,k) p_t(k,j) \right) - p_t(i,j) \\ &= \left(\sum_{k \neq i} p_h(i,k) p_t(k,j) \right) + \left[p_h(i,i) - 1 \right] p_t(i,j) \end{aligned} \tag{4.3}$$

Our goal is to divide each side by h and let $h \to 0$ to compute

$$p'_t(i,j) = \lim_{h \to 0} \frac{p_{t+h}(i,j) - p_t(i,j)}{h}$$

By the definition of the jump rates

$$q(i,j) = \lim_{h \to 0} \frac{p_h(i,j)}{h} \quad \text{for } i \neq j$$

so ignoring the detail of interchanging the limit and the sum, we have

$$\lim_{h \to 0} \frac{1}{h} \sum_{k \neq i} p_h(i,k) p_t(k,j) = \sum_{k \neq i} q(i,k) p_t(k,j) \tag{4.4}$$

For the other term we note that $1 - p_h(i,i) = \sum_{k \neq i} p_h(i,k)$, so

$$\lim_{h \to 0} \frac{p_h(i,i) - 1}{h} = - \lim_{h \to 0} \sum_{k \neq i} \frac{p_h(i,k)}{h} = - \sum_{k \neq i} q(i,k) = -\lambda_i$$

and we have

$$\lim_{h \to 0} \frac{p_h(i,i) - 1}{h} \, p_t(i,j) = -\lambda_i p_t(i,j) \tag{4.5}$$

Combining (4.4) and (4.5) with (4.3) and the definition of the derivative we have

$$p_t'(i,j) = \sum_{k \neq i} q(i,k) p_t(k,j) - \lambda_i p_t(i,j) \tag{4.6}$$

To neaten up the last expression we introduce a new matrix

$$Q(i,j) = \begin{cases} q(i,j) & \text{if } j \neq i \\ -\lambda_i & \text{if } j = i \end{cases}$$

For future computations note that the off-diagonal elements $q(i,j)$ with $i \neq j$ are nonnegative, while the diagonal entry is a negative number chosen to make the row sum equal to 0.

Using matrix notation we can write (4.6) simply as

$$p_t' = Q p_t \tag{4.7}$$

This is **Kolmogorov's backward equation**. If Q were a number instead of a matrix, the last equation would be easy to solve. We would set $p_t = e^{Qt}$ and check by differentiating that the equation held. Inspired by this observation, we define the matrix

$$e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} = \sum_{n=0}^{\infty} Q^n \cdot \frac{t^n}{n!} \tag{4.8}$$

and check by differentiating that

$$\frac{d}{dt} e^{Qt} = \sum_{n=1}^{\infty} Q^n \frac{t^{n-1}}{(n-1)!} = \sum_{n=1}^{\infty} Q \cdot \frac{Q^{n-1} t^{n-1}}{(n-1)!} = Q e^{Qt}$$

Fine print. Here we have interchanged the operations of summation and differentiation, a step that is not valid in general. However, one can show for all of the examples we will consider this is valid, so we will take the physicists' approach and ignore this detail in our calculations.

Kolmogorov's forward equation. This time we split $t+h$ up in a different way when we use the Chapman–Kolmogorov equations:

$$\begin{aligned} p_{t+h}(i,j) - p_t(i,j) &= \left(\sum_k p_t(i,k) p_h(k,j) \right) - p_t(i,j) \\ &= \left(\sum_{k \neq j} p_t(i,k) p_h(k,j) \right) + \left[p_h(j,j) - 1 \right] p_t(i,j) \end{aligned}$$

Computing as before we arrive at

$$p_t'(i,j) = \sum_{k \neq j} p_t(i,k)q(k,j) - p_t(i,j)\lambda_j \tag{4.9}$$

Introducing matrix notation again, we can write

$$p_t' = p_t Q \tag{4.10}$$

Comparing (4.10) with (4.7) we see that $p_t Q = Q p_t$ and that the two forms of Kolmogorov's differential equations correspond to writing the rate matrix on the left or the right. While we are on the subject of the choices, we should remember that in general for matrices $AB \neq BA$, so it is somewhat remarkable that $p_t Q = Q p_t$. The key to the fact that these matrices commute is that $p_t = e^{Qt}$ is made up of powers of Q:

$$Q \cdot e^{Qt} = \sum_{n=0}^{\infty} Q \cdot \frac{(Qt)^n}{n!} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} \cdot Q = e^{Qt} \cdot Q$$

To illustrate the use of Kolmogorov's equations we will now consider some examples. The simplest possible is

Example 4.9. Poisson process. Let $X(t)$ be the number of arrivals up to time t in a Poisson process with rate λ. In order to go from i arrivals at time s to j arrivals at time $t+s$ we must have $j \geq i$ and have exactly $j-i$ arrivals in t units of time, so

$$p_t(i,j) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} \tag{4.11}$$

To check the differential equation we have to first figure out what it is. Using the more explicit form of the backwards equation, (4.6), and plugging in our rates, we have

$$p_t'(i,j) = \lambda p_t(i+1,j) - \lambda p_t(i,j)$$

To check this we have to differentiate the formula in (4.11).

When $j > i$ we have that the derivative of (4.11) is

$$-\lambda e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} + e^{-\lambda t} \frac{(\lambda t)^{j-i-1}}{(j-i-1)!} \lambda = -\lambda p_t(i,j) + \lambda p_t(i+1,j)$$

When $j = i$, $p_t(i,i) = e^{-\lambda t}$, so the derivative is

$$-\lambda e^{-\lambda t} = -\lambda p_t(i,i) = -\lambda p_t(i,i) + \lambda p_t(i+1,i)$$

since $p_t(i+1,i) = 0$.

There are not many examples in which one can write down solutions of the Kolmogorov's differential equation. A remarkable exception is:

Example 4.10. Yule process. In this model each particle splits into two at rate β, so $q(i,i+1) = \beta i$. To find the transition probability of the Yule process we will guess and verify that

$$p_t(1,j) = e^{-\beta t}(1-e^{-\beta t})^{j-1} \qquad \text{for } j \geq 1 \tag{4.12}$$

i.e., a geometric distribution with success probability $e^{-\beta t}$.

To check this we will use the forward equation (4.9) to conclude

$$p_t'(1,j) = -\beta j p_t(1,j) + \beta(j-1)p_t(1,j-1) \tag{4.13}$$

The use of the forward equation here is dictated by the fact that we are only writing down formulas for $p_t(1,j)$. To check (4.13) we differentiate the formula proposed in (4.12) to see that if $j > 1$

$$\begin{aligned} p_t'(1,j) = &- \beta e^{-\beta t}(1-e^{-\beta t})^{j-1} \\ &+ e^{-\beta t}(j-1)(1-e^{-\beta t})^{j-2}(\beta e^{-\beta t}) \end{aligned}$$

Recopying the first term on the right and using $\beta e^{-\beta t} = -(1 - e^{-\beta t})\beta + \beta$ in the second, we can rewrite the right-hand side of the above as

$$\begin{aligned} -\beta e^{-\beta t}(1-e^{-\beta t})^{j-1} - e^{-\beta t}(j-1)(1-e^{-\beta t})^{j-1}\beta \\ + e^{-\beta t}(1-e^{-\beta t})^{j-2}(j-1)\beta \end{aligned}$$

Adding the first two terms then comparing with (4.13) shows that the above is

$$= -\beta j p_t(1,j) + \beta(j-1)p_t(1,j-1)$$

Having worked to find $p_t(1,j)$, it is fortunately easy to find $p_t(i,j)$. The chain starting with i individuals is the sum of i copies of the chain starting from 1 individual. Using this one can easily compute that

$$p_t(i,j) = \binom{j-1}{i-1}(e^{-\beta t})^i(1-e^{-\beta t})^{j-i} \tag{4.14}$$

In words, the sum of i geometrics has a negative binomial distribution.

Proof. To begin we note that if $N_1, \dots N_i$ have the distribution given in (4.12) and $n_1 + \cdots + n_i = j$, then

$$P(N_1 = n_1, \dots, N_i = n_i) = \prod_{k=1}^{i} e^{-\beta t}(1-e^{-\beta t})^{n_k-1} = (e^{-\beta t})^i(1-e^{-\beta t})^{j-i}$$

To count the number of possible $(n_1, \dots, n_i)$ with $n_k \ge 1$ and sum j, we think of putting j balls in a row. To divide the j balls into i groups of size $n_1, \dots, n_i$, we will insert cards in the slots between the balls and let n_k be the number of balls in the kth group. Having made this transformation it is clear that the number of $(n_1, \dots, n_i)$ is the number of ways of picking $i-1$ of the $j-1$ slot to put the cards or $\binom{j-1}{i-1}$. Multiplying this times the probability for each $(n_1, \dots, n_i)$ gives the result. $\square$

It is usually not possible to explicitly solve Kolmogorov's equations to find the transition probability. To illustrate the difficulties involved we will now consider:

Example 4.11. Two-state chains. For concreteness, we can suppose that the state space is $\{1, 2\}$. In this case, there are only two flip rates $q(1,2) = \lambda$ and $q(2,1) = \mu$, so when we fill in the diagonal with minus the sum of the flip rates on that row we get

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

Writing out the backward equation in matrix form, (2.3), now we have

$$\begin{pmatrix} p_t'(1,1) & p_t'(1,2) \\ p_t'(2,1) & p_t'(2,2) \end{pmatrix} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} p_t(1,1) & p_t(1,2) \\ p_t(2,1) & p_t(2,2) \end{pmatrix}$$

Doing the first column of matrix multiplication on the right, we have

$$\begin{aligned} p_t'(1,1) &= -\lambda p_t(1,1) + \lambda p_t(2,1) = -\lambda(p_t(1,1) - p_t(2,1)) \\ p_t'(2,1) &= \mu p_t(1,1) - \mu p_t(2,1) = \mu(p_t(1,1) - p_t(2,1)) \end{aligned} \tag{4.15}$$

Taking the difference of the two equations gives

$$[p_t(1,1) - p_t(2,1)]' = -(\lambda+\mu)[p_t(1,1) - p_t(2,1)]$$

Since $p_0(1,1) = 1$ and $p_0(2,1) = 0$ we have

$$p_t(1,1) - p_t(2,1) = e^{-(\lambda+\mu)t}$$

Using this in (4.15) and integrating

$$\begin{aligned} p_t(1,1) &= p_0(1,1) + \frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)s}\Big|_0^t = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)t} \\ p_t(2,1) &= p_0(2,1) + \frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)s}\Big|_0^t = \frac{\mu}{\mu+\lambda} - \frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)t} \end{aligned}$$

As a check on the constants note that $p_0(1,1) = 1$ and $p_0(2,1) = 0$.

To prepare for the developments in the next section note that the probability of being in state 1 converges exponentially fast to the equilibrium value $\mu/(\mu+\lambda)$.

4.3 Limiting Behavior

The study of the limiting behavior of continuous time Markov chains is simpler than the theory for discrete time chains, since the randomness of the exponential holding times implies that we don't have to worry about aperiodicity. We will first state the main convergence result that is a combination of the discrete-time convergence theorem, (4.5), and strong law, (4.8), from Chapter 1 and then explain the terms it uses.

Theorem 4.2. *If a continuous-time Markov chain X_t is irreducible and has a stationary distribution π, then*

$$\lim_{t\to\infty} p_t(i,j) = \pi(j)$$

Furthermore if $r(j)$ is the reward we earn in state i and $\sum_j \pi(j)\,|r(j)| < \infty$, then as $t \to \infty$

$$\frac{1}{t}\int_0^t r(X_s)\,ds \to \sum_y \pi(y)r(y)$$

Here by X_t is **irreducible**, we mean that for any two states x and y it is possible to get from x to y in a finite number of jumps. To be precise, there is a sequence of states $x_0 = x, x_1, \dots x_n = y$ so that $q(x_{m-1}, x_m) > 0$ for $1 \le m \le n$.

In discrete time a stationary distribution is a solution of $\pi p = \pi$. Since there is no first $t > 0$, in continuous time we need the stronger notion: π is said to be a **stationary distribution** if $\pi p_t = \pi$ for all $t > 0$. The last condition is difficult to check since it involves all of the p_t, and as we have seen in the previous section, the p_t are not easy to compute. The next result solves these problems by giving a test for stationarity in terms of the basic data used to describe the chain, the matrix of transition rates

$$Q(i,j) = \begin{cases} q(i,j) & j \ne i \\ -\lambda_i & j = i \end{cases}$$

where $\lambda_i = \sum_{j\ne i} q(i,j)$ is the total rate of transitions out of i.

Theorem 4.3. *π is a stationary distribution if and only if $\pi Q = 0$.*

Why is this true? Filling in the definition of Q and rearranging, the condition $\pi Q = 0$ becomes

$$\sum_{k \neq j} \pi(k) q(k,j) = \pi(j) \lambda_j$$

If we think of $\pi(k)$ as the amount of sand at k, the right-hand side represents the rate at which sand leaves j, while the left gives the rate at which sand arrives at j. Thus, π will be a stationary distribution if for each j the flow of sand in to j is equal to the flow out of j. □

More details. If $\pi p_t = \pi$ then

$$\begin{aligned} 0 = \frac{d}{dt} \pi p_t &= \sum_i \pi(i) p_t'(i,j) = \sum_i \pi(i) \sum_k p_t(i,k) Q(k,j) \\ &= \sum_k \sum_i \pi(i) p_t(i,k) Q(k,j) = \sum_k \pi(k) Q(k,j) \end{aligned}$$

Conversely if $\pi Q = 0$

$$\begin{aligned} \frac{d}{dt} \left(\sum_i \pi(i) p_t(i,j) \right) &= \sum_i \pi(i) p_t'(i,j) = \sum_i \pi(i) \sum_k Q(i,k) p_t(k,j) \\ &= \sum_k \sum_i \pi(i) Q(i,k) p_t(k,j) = 0 \end{aligned}$$

Since the derivative is 0, πp_t is constant and must always be equal to π its value at 0. □

We will now consider some examples. The simplest one was already covered in the last section.

Example 4.12. Two state chains. Suppose that the state space is $\{1, 2\}$, $q(1, 2) = \lambda$, and $q(2, 1) = \mu$, where both rates are positive. The equations $\pi Q = 0$ can be written as

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

The first equation says $-\lambda \pi_1 + \mu \pi_2 = 0$. Taking into account that we must have $\pi_1 + \pi_2 = 1$, it follows that

$$\pi_1 = \frac{\mu}{\lambda + \mu} \quad \text{and} \quad \pi_2 = \frac{\lambda}{\lambda + \mu}$$

Example 4.13. L.A. weather chain. There are three states: 1 = sunny, 2 = smoggy, 3 = rainy. The weather stays sunny for an exponentially distributed number of days with mean 3, then becomes smoggy. It stays smoggy for an exponentially distributed number of days with mean 4, then rain comes. The rain lasts for an exponentially distributed number of days with mean 1, then sunshine returns. Remembering that for an exponential the rate is 1 over the mean, the verbal description translates into the following Q-matrix

	1	**2**	**3**
1	$-1/3$	$1/3$	0
2	0	$-1/4$	$1/4$
3	1	0	-1

The relation $\pi Q = 0$ leads to three equations:

$$\begin{array}{rrrl} -\frac{1}{3}\pi_1 & & +\pi_3 & = 0 \\ \frac{1}{3}\pi_1 & -\frac{1}{4}\pi_2 & & = 0 \\ & \frac{1}{4}\pi_2 & -\pi_3 & = 0 \end{array}$$

Adding the three equations gives 0=0 so we delete the third equation and add $\pi_1 + \pi_2 + \pi_3 = 1$ to get an equation that can be written in matrix form as

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} A = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \quad \text{where} \quad A = \begin{pmatrix} -1/3 & 1/3 & 1 \\ 0 & -1/4 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

This is similar to our recipe in discrete time. To find the stationary distribution of a k state chain, form A by taking the first $k-1$ columns of Q, add a column of 1's and then

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} A^{-1}$$

i.e., the last row of A^{-1}. In this case we have

$$\pi(1) = 3/8, \quad \pi(2) = 4/8, \quad \pi(3) = 1/8$$

To check our answer, note that the weather cycles between sunny, smoggy, and rainy spending independent exponentially distributed amounts of time with means 3, 4, and 1, so the limiting fraction of time spent in each state is just the mean time spent in that state over the mean cycle time, 8.

Detailed balance condition. Generalizing from discrete time we can formulate this condition as:

$$\pi(k)q(k,j) = \pi(j)q(j,k) \quad \text{for all } j \neq k \tag{4.16}$$

The reason for interest in this concept is

Theorem 4.4. *If (4.16) holds, then π is a stationary distribution.*

Why is this true? The detailed balance condition implies that the flows of sand between each pair of sites are balanced, which then implies that the net amount of sand flowing into each vertex is 0, i.e., $\pi Q = 0$.

Proof. Summing 4.16 over all $k \neq j$ and recalling the definition of λ_j gives

$$\sum_{k \neq j} \pi(k)q(k,j) = \pi(j) \sum_{k \neq j} q(j,k) = \pi(j)\lambda_j$$

Rearranging we have

$$(\pi Q)_j = \sum_{k \neq j} \pi(k)q(k,j) - \pi(j)\lambda_j = 0$$ □

As in discrete time, (4.16) is much easier to check but does not always hold. In Example 4.13

$$\pi(2)q(2,1) = 0 < \pi(1)q(1,2)$$

As in discrete time, detailed balance holds for

Example 4.14. Birth and death chains. Suppose that $S = \{0, 1, \ldots, N\}$ with

$$\begin{aligned} q(n, n+1) &= \lambda_n \qquad \text{for } 0 \leq n < N \\ q(n, n-1) &= \mu_n \qquad \text{for } 0 < n \leq N \end{aligned}$$

Here λ_n represents the birth rate when there are n individuals in the system, and μ_n denotes the death rate in that case.

If we suppose that all the λ_n and μ_n listed above are positive then the birth and death chain is irreducible, and we can divide to write the detailed balance condition as

$$\pi(n) = \frac{\lambda_{n-1}}{\mu_n} \pi(n-1) \tag{4.17}$$

Using this again we have $\pi(n-1) = (\lambda_{n-2}/\mu_{n-1})\pi(n-2)$ and it follows that

$$\pi(n) = \frac{\lambda_{n-1}}{\mu_n} \cdot \frac{\lambda_{n-2}}{\mu_{n-1}} \cdot \pi(n-2)$$

Repeating the last reasoning leads to

$$\pi(n) = \frac{\lambda_{n-1} \cdot \lambda_{n-2} \cdots \lambda_0}{\mu_n \cdot \mu_{n-1} \cdots \mu_1} \pi(0) \tag{4.18}$$

To check this formula and help remember it, note that (i) there are n terms in the numerator and in the denominator, and (ii) if the state space was $\{0, 1, \ldots, n\}$, then $\mu_0 = 0$ and $\lambda_n = 0$, so these terms cannot appear in the formula.

To illustrate the use of (4.18) we consider two concrete examples.

Example 4.15. Barbershop. A barber can cut hair at rate 3, where the units are people per hour, i.e., each haircut requires an exponentially distributed amount of time with mean 20 minutes. Suppose customers arrive at times of a rate 2 Poisson process, but will leave if both chairs in the waiting room are full. (a) What fraction of time will both chairs be full? (b) In the long run, how many customers does the barber serve per hour?

Solution. We define our state to be the number of customers in the system, so $S = \{0, 1, 2, 3\}$. From the problem description it is clear that

$$\begin{aligned} q(i, i-1) &= 3 \quad \text{for } i = 1, 2, 3 \\ q(i, i+1) &= 2 \quad \text{for } i = 0, 1, 2 \end{aligned}$$

The detailed balance conditions say

$$2\pi(0) = 3\pi(1), \quad 2\pi(1) = 3\pi(2), \quad 2\pi(2) = 3\pi(3)$$

Setting $\pi(0) = c$ and solving, we have

$$\pi(1) = \frac{2c}{3}, \quad \pi(2) = \frac{2}{3} \cdot \pi(1) = \frac{4c}{9}, \quad \pi(3) = \frac{2}{3} \cdot \pi(2) = \frac{8c}{27}$$

The sum of the π's is $(27 + 18 + 12 + 8)c/27 = 65c/27$, so $c = 27/65$ and

$$\pi(0) = 27/65, \quad \pi(1) = 18/65, \quad \pi(2) = 12/65, \quad \pi(3) = 8/65$$

From this we see that 8/65's of the time both chairs are full, so that fraction of the arrivals are lost and hence 57/65's or 87.7% of the customers enter service. Since the original arrival rate is 2, this means he serves an average of $114/65 = 1.754$ customers per hour.

Example 4.16. Machine repair model. A factory has three machines in use and one repairman. Suppose each machine works for an exponential amount of time with mean 60 days between breakdowns, but each breakdown requires an exponential repair time with mean 4 days. What is the long-run fraction of time all three machines are working?

Solution. Let X_t be the number of working machines. Since there is one repairman we have $q(i, i+1) = 1/4$ for $i = 0, 1, 2$. On the other hand, the failure rate is proportional to the number of machines working, so $q(i, i-1) = i/60$ for $i = 1, 2, 3$. Setting $\pi(0) = c$ and plugging into the recursion (4.17) gives

$$\begin{aligned}
\pi(1) &= \frac{\lambda_0}{\mu_1} \cdot \pi(0) = \frac{1/4}{1/60} \cdot c = 15c \\
\pi(2) &= \frac{\lambda_1}{\mu_2} \cdot \pi(1) = \frac{1/4}{2/60} \cdot 15c = \frac{225c}{2} \\
\pi(3) &= \frac{\lambda_2}{\mu_3} \cdot \pi(2) = \frac{1/4}{3/60} \cdot \frac{225c}{2} = \frac{1225c}{2}
\end{aligned}$$

Adding up the π's gives $(1225 + 225 + 30 + 2)c/2 = 1480c/2$ so $c = 2/1480$ and we have

$$\pi(3) = \frac{1225}{1480} \quad \pi(2) = \frac{225}{1480} \quad \pi(1) = \frac{30}{1480} \quad \pi(0) = \frac{2}{1480}$$

Thus in the long run all three machines are working $1225/1480 = 0.8277$ of the time.

For our final example of the use of detailed balance we consider

Example 4.17. M/M/s queue with balking. A bank has s tellers that serve customers who need an exponential amount of service with rate μ nd queue in a single line if all of the servers are busy. Customers arrive at times of a Poisson process with rate λ but only join the queue with probability a_n if there are n customers in line. As noted in Example 1.?, the birth rate $\lambda_n = \lambda a_n$ for $n \geq 0$, while the death rate is

$$\mu_n = \begin{cases} n\mu & 0 \le n \le s \\ s\mu & n \ge s \end{cases}$$

for $n \ge 1$. It is reasonable to assume that if the line is long the probability the customer will join the queue is small. The next result shows that this is always enough to prevent the queue length from growing out of control.

Theorem 4.5. *If $a_n \to 0$ as $n \to \infty$, then there is a stationary distribution.*

Proof. It follows from (4.17) that if $n \ge s$, then

$$\pi(n+1) = \frac{\lambda_n}{\mu_{n+1}} \cdot \pi(n) = a_n \cdot \frac{\lambda}{s\mu} \cdot \pi(n)$$

If N is large enough and $n \ge N$, then $a_n\lambda/(s\mu) \le 1/2$ and it follows that

$$\pi(n+1) \le \frac{1}{2}\pi(n) \ldots \le \left(\frac{1}{2}\right)^{n-N} \pi(N)$$

This implies that $\sum_n \pi(n) < \infty$, so we can pick $\pi(0)$ to make the sum $= 1$. □

Concrete example. Suppose $s = 1$ and $a_n = 1/(n+1)$. In this case

$$\frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1} = \frac{\lambda^n}{\mu^n} \cdot \frac{1}{1 \cdot 2 \cdots n} = \frac{(\lambda/\mu)^n}{n!}$$

To find the stationary distribution we want to take $\pi(0) = c$ so that

$$c \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} = 1$$

Recalling the formula for the Poisson distribution with mean λ/μ, we see that $c = e^{-\lambda/\mu}$ and the stationary distribution is Poisson.

4.4 Markovian Queues

In this section we will take a systematic look at the basic models of queueing theory that have Poisson arrivals and exponential service times. The arguments concerning Wendy's in Section 3.2 explain why we can be happy assuming that the arrival process is Poisson. However, the assumption of exponential services times is hard to justify. Here, it is a necessary evil. The lack of memory property of the exponential is needed for the queue length to be a continuous Markov chain. We begin with the simplest example:

Example 4.18. M/M/1 queue. In this system customers arrive to a single server facility at the times of a Poisson process with rate λ, and each requires an independent amount of service that has an exponential distribution with rate μ. From the description it should be clear that the transition rates are

$$\begin{aligned} q(n, n+1) &= \lambda \qquad \text{if } n \geq 0 \\ q(n, n-1) &= \mu \qquad \text{if } n \geq 1 \end{aligned}$$

so we have a birth and death chain with birth rates $\lambda_n = \lambda$ and death rates $\mu_n = \mu$. Plugging into our formula for the stationary distribution, (4.18), we have

$$\pi(n) = \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1} \cdot \pi(0) = \left(\frac{\lambda}{\mu}\right)^n \pi(0) \tag{4.19}$$

To find the value of $\pi(0)$, we recall that for $|\theta| < 1$, $\sum_{n=0}^{\infty} \theta^n = 1/(1-\theta)$. From this we see that if $\lambda < \mu$, then

$$\sum_{n=0}^{\infty} \pi(n) = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n \pi(0) = \frac{\pi(0)}{1 - (\lambda/\mu)}$$

So to have the sum 1, we pick $\pi(0) = 1 - (\lambda/\mu)$, and the resulting stationary distribution is the shifted geometric distribution

$$\pi(n) = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n \quad \text{for } n \geq 0 \tag{4.20}$$

Having determined the stationary distribution we can now compute various quantities of interest concerning the queue. We might be interested, for example, in the distribution of the waiting time

W of a customer who arrives to find the queue in equilibrium. To do this we begin by noting that the only way to wait 0 is for the number of people waiting in the queue Q to be 0 so

$$P(W = 0) = P(Q = 0) = 1 - \frac{\lambda}{\mu}$$

When there is at least one person in the system, the arriving customer will spend a positive amount of time in the queue. Writing $f_W(x)$ for the density function of W on $(0, \infty)$, we note that if there are n people in the system when the customer arrives, then the amount of time he needs to enter service has a gamma(n, μ) density, so using (1.11) in Chapter 3

$$f_W(x) = \sum_{n=1}^{\infty} \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n e^{-\mu x} \frac{\mu^n x^{n-1}}{(n-1)!}$$

Changing variables $m = n - 1$ and rearranging, the above becomes

$$= \left(1 - \frac{\lambda}{\mu}\right) e^{-\mu x} \lambda \sum_{m=0}^{\infty} \frac{\lambda^m x^m}{m!} = \frac{\lambda}{\mu}(\mu - \lambda) e^{-(\mu-\lambda)x}$$

Recalling that $P(W > 0) = \lambda/\mu$, we can see that the last result says that the conditional distribution of W given that $W > 0$ is exponential with rate $\mu - \lambda$.

Example 4.19. M/M/1 queue with a finite waiting room. In this system customers arrive at the times of a Poisson process with rate λ. Customers enter service if there are $< N$ individuals in the system, but when there are $\geq N$ customers in the system, the new arrival leaves never to return. Once in the system, each customer requires an independent amount of service that has an exponential distribution with rate μ.

Taking the state to be the number of customers in the system, the state space is now $S = \{0, 1, \dots N\}$. The birth and death rates are changed a little

$$\begin{aligned} q(n, n+1) &= \lambda \qquad \text{if } 0 \leq n < N \\ q(n, n-1) &= \mu \qquad \text{if } 0 < n \leq N \end{aligned}$$

but our formula for the stationary distribution, (4.18), still gives

$$\pi(n) = \frac{\lambda_{n-1} \cdots \lambda_0}{\mu_n \cdots \mu_1} \cdot \pi(0) = \left(\frac{\lambda}{\mu}\right)^n \pi(0) \qquad \text{for } 1 \leq n \leq N$$

The first thing that changes in the analysis is the normalizing constant. To isolate the arithmetic from the rest of the problem we recall that if $\theta \neq 1$, then

$$\sum_{n=0}^{N} \theta^n = \frac{1-\theta^{N+1}}{1-\theta} \tag{4.21}$$

Suppose now that $\lambda \neq \mu$. Using (4.21), we see that if

$$c = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}}$$

then the sum is 1, so the stationary distribution is given by

$$\pi(n) = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}} \left(\frac{\lambda}{\mu}\right)^n \quad \text{for } 0 \leq n \leq N \tag{4.22}$$

The new formula is similar to the old one in (4.20) and when $\lambda < \mu$ reduces to it as $N \to \infty$. Of course, when the waiting room is finite, the state space is finite and we always have a stationary distribution, even when $\lambda > \mu$. The analysis above has been restricted to $\lambda \neq \mu$. However, it is easy to see that when $\lambda = \mu$ the stationary distribution is $\pi(n) = 1/(N+1)$ for $0 \leq n \leq N$.

To check formula (4.22), we note that the barbershop chain, Example 3.3, has this form with $N = 3$, $\lambda = 2$, and $\mu = 3$, so plugging into (4.22) and multiplying numerator and denominator by $3^4 = 81$, we have

$$\begin{aligned}
\pi(0) &= \frac{1-2/3}{1-(2/3)^4} = \frac{81-54}{81-16} = 27/65 \\
\pi(1) &= \frac{2}{3}\pi(0) = 18/65 \\
\pi(2) &= \frac{2}{3}\pi(1) = 12/65 \\
\pi(3) &= \frac{2}{3}\pi(2) = 8/65
\end{aligned}$$

From a single server with a finite waiting room we move now to s servers with an unlimited waiting room, a system described more fully in Example 4.3.

Example 4.20. M/M/s queue. Imagine a bank with $s \geq 1$ tellers that serve customers who queue in a single line if all servers are busy.

We imagine that customers arrive at the times of a Poisson process with rate λ, and each requires an independent amount of service that has an exponential distribution with rate μ. As explained in Example 1.3, the flip rates are $q(n, n+1) = \lambda$ and

$$q(n, n-1) = \begin{cases} \mu n & \text{if } n \leq s \\ \mu s & \text{if } n \geq s \end{cases}$$

The conditions that result from using the detailed balance condition are

$$\lambda\pi(0) = \mu\pi(1) \quad \lambda\pi(1) = 2\mu\pi(2) \quad \dots \quad \lambda\pi(s-1) = s\mu\pi(s) \quad (4.23)$$

Then for $k \geq 0$ we have

$$\pi(s+k-1)\lambda = s\mu\pi(s+k) \quad \text{or} \quad \pi(s+k) = \frac{\lambda}{s\mu}\pi(s+k-1)$$

Iterating the last equation, we have that for $k \geq 0$

$$\pi(s+k) = \left(\frac{\lambda}{s\mu}\right)^2 \pi(s+k-2) \ldots = \left(\frac{\lambda}{s\mu}\right)^{k+1} \pi(s-1) \quad (4.24)$$

From the last formula we see that if $\lambda < s\mu$ then $\sum_{k=0}^{\infty} \pi(s+k) < \infty$ so $\sum_{j=0}^{\infty} \pi(j) < \infty$ and it is possible to pick $\pi(0)$ to make the sum equal to 1. From this it follows that

If $\lambda < s\mu$, then the $M/M/s$ queue has as stationary distribution.

The condition $\lambda < s\mu$ for the existence of a stationary distribution is natural since it says that the service rate of the fully loaded system is larger than the arrival rate, so the queue will not grow out of control. Conversely,

If $\lambda > s\mu$, the $M/M/s$ queue is transient.

Why is this true? The conclusion comes from combining two ideas:

(i) An $M/M/s$ queue with s rate μ servers is less efficient than an $M/M/1$ queue with 1 rate $s\mu$ server, since the single server queue always has departures at rate $s\mu$, while the s server queue sometimes has departures at rate $n\mu$ with $n < s$.

(ii) An $M/M/1$ queue is transient if its arrival rate is larger than its service rate.

Formulas for the stationary distribution $\pi(n)$ for the $M/M/s$ queue are unpleasant to write down for a general number of servers s, but it is not hard to use (4.23) and (4.24) to find the stationary distribution in concrete cases:

Example 4.21. M/M/3 queue. When $s = 3$, $\lambda = 2$ and $\mu = 1$, the first two equations in (4.23) say

$$2\pi(0) = \pi(1) \qquad 2\pi(1) = 2\pi(2) \tag{4.25}$$

while (4.24) tells us that for $k \geq 0$

$$\pi(3+k) = \left(\frac{2}{3}\right)^{k+1} \pi(2)$$

Summing the last result from $k = 0$ to ∞, adding $\pi(2)$, and changing variables $j = k + 1$, we have

$$\sum_{m=2}^{\infty} \pi(m) = \pi(2) \sum_{j=0}^{\infty} (2/3)^j = \frac{\pi(2)}{1 - \frac{2}{3}} = 3\pi(2)$$

by the formula for the sum of the geometric series. Setting $\pi(2) = c$ and using (4.25), we see that

$$\pi(1) = c \qquad \pi(0) = \frac{1}{2}\pi(1) = c/2$$

Taking the contributions in order of increasing j, the sum of all the $\pi(j)$ is $(0.5+1+3)\pi(2)$. From this we conclude that $\pi(2) = 2/9$, so

$$\pi(0) = 1/9, \quad \pi(1) = 2/9, \quad \pi(k) = (2/9)(2/3)^{k-2} \text{ for } k \geq 2$$

Our next result is a remarkable property of the M/M/s queue.

Theorem 4.6. *If $\lambda < \mu s$, then the output process of the $M/M/s$ queue in equilibrium is a rate λ Poisson process.*

Your first reaction to this should be that it is crazy. Customers depart at rate 0, μ, 2μ, $\ldots$, $s\mu$, depending on the number of servers that are busy and it is usually the case that none of these numbers $= \lambda$. To further emphasize the surprising nature of Theorem 4.6, suppose for concreteness that there is one server, $\lambda = 1$, and $\mu = 10$. If, in this situation, we have just seen 30 departures in the last 2 hours, then it seems reasonable to guess that the server is busy and the next departure will be exponential(10). However, if the output process is Poisson, then the number of departures in disjoint intervals are independent.

Proof for $s = 1$. Our first step in making the result in Theorem 4.6 seem reasonable is to check by hand that if there is one server and the queue is in equilibrium, then the time of the first departure, D, has an exponential distribution with rate λ. There are two cases to consider.

Case 1. If there are $n \geq 1$ customers in the queue, then the time to the next departure has an exponential distribution with rate μ, i.e.,

$$f_D(t) = \mu e^{-\mu t}$$

Case 2. If there are $n = 0$ customers in the queue, then we have to wait an exponential(λ) amount of time until the first arrival, and then an independent exponential(μ) for that customer to depart. If we let T_1 and T_2 be the waiting times for the arrival and for the departure, then breaking things down according to the value of $T_1 = s$, the density of $D = T_1 + T_2$ in this case is

$$\begin{aligned} f_D(t) &= \int_0^t \lambda e^{-\lambda s} \cdot \mu e^{-\mu(t-s)}\, ds = \lambda\mu e^{-\mu t} \int_0^t e^{-(\lambda-\mu)s}\, ds \\ &= \frac{\lambda\mu e^{-\mu t}}{\lambda - \mu}\left(1 - e^{-(\lambda-\mu)t}\right) = \frac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu t} - e^{-\lambda t}\right) \end{aligned}$$

The probability of 0 customers in equilibrium is $1 - (\lambda/\mu)$ by (4.20). This implies the probability of ≥ 1 customer is λ/μ, so combining the two cases:

$$f_D(t) = \frac{\mu - \lambda}{\mu} \cdot \frac{\lambda\mu}{\lambda - \mu}\left(e^{-\mu t} - e^{-\lambda t}\right) + \frac{\lambda}{\mu} \cdot \mu e^{-\mu t}$$

At this point cancellations occur to produce the answer we claimed:

$$-\lambda\left(e^{-\mu t} - e^{-\lambda t}\right) + \lambda e^{-\mu t} = \lambda e^{-\lambda t}$$

We leave it to the adventurous reader to try to repeat the last calculation for the M/M/s queue with $s > 1$ where there is not a neat formula for the stationary distribution.

Proof of Theorem 4.6. By repeating the proof of (1.13) one can show

Lemma 4.1. *Fix t and let $Y_s = X_{t-s}$ for $0 \le s \le t$. Then Y_s is a Markov chain with transition probability*

$$\hat{p}_t(i,j) = \frac{\pi(j)p_t(j,i)}{\pi(i)}$$

If π satisfies the detailed balance condition $\pi(i)q(i,j) = \pi(j)q(j,i)$, then the reversed chain has transition probability $\hat{p}_t(i,j) = p_t(i,j)$.

As we learned in Example 4.20, when $\lambda < \mu s$ the $M/M/s$ queue is a birth and death chain with a stationary distribution π that satisfies the detailed balance condition. Lemma 4.1 implies that if we take the movie of the Markov chain in equilibrium then we see something that has the same distribution as the $M/M/s$ queue. Reversing time turns arrivals into departures, so the departures must be a Poisson process with rate λ. □

It should be clear from the proof just given that we also have:

Theorem 4.7. *Consider a queue in which arrivals occur according to a Poisson process with rate λ and customers are served at rate μ_n when there are n in the system. Then as along as there is a stationary distribution the output process will be a rate λ Poisson process.*

A second refinement that will be useful in the next section is

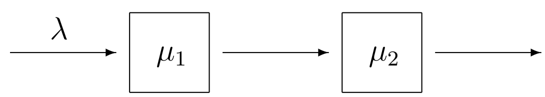
Theorem 4.8. *Let $N(t)$ be the number of departures between time 0 and time n for the $M/M/1$ queue $X(t)$ started from its equilibrium distribution. Then $\{N(s) : 0 \le s \le t\}$ and $X(t)$ are independent.*

Why is this true? At first it may sound deranged to claim that the output process up to time t is independent of the queue length. However, if we reverse time, then the departures before time t turn into arrivals after t, and these are obviously independent of the queue length at time t, $X(t)$.

4.5 Queueing Networks

In many situations we are confronted with more than one queue. For example, in California when you go to the Department of Motor Vehicles to renew your driver's license you must (i) take a test on the driving laws, (ii) have your test graded, (iii) pay your fees, and (iv) get your picture taken. A simple model of this type of situation with only two steps is:

Example 4.22. Two-station tandem queue. In this system customers at times of a Poisson process with rate λ arrive at service facility 1 where they each require an independent exponential amount of service with rate μ_1. When they complete service at the first site, they join a second queue to wait for an exponential amount of service with rate μ_2.



Our main problem is to find conditions that guarantee that the queue stabilizes, i.e., has a stationary distribution. This is simple in the tandem queue. The first queue is not affected by the second, so if $\lambda < \mu_1$, then (4.20) tells us that the equilibrium probability of the number of customers in the first queue, X_t^1, is given by the shifted geometric distribution

$$P(X_t^1 = m) = \left(\frac{\lambda}{\mu_1}\right)^m \left(1 - \frac{\lambda}{\mu_1}\right)$$

In the previous section we learned that the output process of an $M/M/1$ queue in equilibrium is a rate λ Poisson process. This means that if the first queue is in equilibrium, then the number of customers in the queue, X_t^2, is itself an $M/M/1$ queue with arrivals at rate λ (the output rate for 1) and service rate μ_2. Using the results in (4.20) again, the number of individuals in the second queue has stationary distribution

$$P(X_t^2 = n) = \left(\frac{\lambda}{\mu_2}\right)^n \left(1 - \frac{\lambda}{\mu_2}\right)$$

To specify the stationary distribution of the system, we need to know the joint distribution of X_t^1 and X_t^2. The answer is somewhat remarkable: in equilibrium the two queue lengths are independent.

$$P(X_t^1 = m, X_t^2 = n) = \left(\frac{\lambda}{\mu_1}\right)^m \left(1 - \frac{\lambda}{\mu_1}\right) \cdot \left(\frac{\lambda}{\mu_2}\right)^n \left(1 - \frac{\lambda}{\mu_2}\right) \tag{4.26}$$

Why is this true? Theorem 4.8 implies that the queue length and the departure process are independent.

Since there is more than a little hand-waving going on in the proof of Theorem 4.8 and its application here, it is comforting to note that one can simply verify from the definitions that

Lemma 4.2. *If $\pi(m,n) = c\lambda^{m+n}/(\mu_1^m \mu_2^n)$, where $c = (1-\lambda/\mu_1)(1-\lambda/\mu_2)$ is a constant chosen to make the probabilities sum to 1, then π is a stationary distribution.*

Proof. The first step in checking $\pi Q = 0$ is to compute the rate matrix Q. To do this it is useful to draw a picture which assumes $m, n > 0$

$$\begin{tikzcd}
(m-1,\, n+1) \arrow[d, dash, "(a)"'] & (m,\, n+1) \arrow[d, "\mu_2"] \arrow[dr, dash, "(c)"] & \\
(m-1,\, n) \arrow[r, "\lambda"] & (m,\, n) \arrow[ul, "\mu_1"'] \arrow[r, "\lambda"] \arrow[d, "\mu_2"] & (m+1,\, n) \\
& (m,\, n-1) \arrow[r, dash, "(b)"'] & (m+1,\, n-1) \arrow[ul, "\mu_1"']
\end{tikzcd}$$

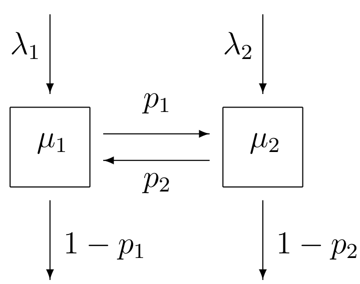
The rate arrows plus the ordinary lines on the picture, make three triangles. We will now check that the flows out of and into (m,n)

in each triangle balance. In symbols we note that

$$
\begin{aligned}
&(a) \qquad \mu_1\pi(m,n) = \frac{c\lambda^{m+n}}{\mu_1^{m-1}\mu_2^n} = \lambda\pi(m-1,n)\\
&(b) \qquad \mu_2\pi(m,n) = \frac{c\lambda^{m+n}}{\mu_1^{m}\mu_2^{n-1}} = \mu_1\pi(m+1,n-1)\\
&(c) \qquad \lambda\pi(m,n) = \frac{c\lambda^{m+n+1}}{\mu_1^{m}\mu_2^{n}} = \mu_2\pi(m,n+1)
\end{aligned}
$$

This shows that $\pi Q = 0$ when $m, n > 0$. There are three other cases to consider: (i) $m = 0$, $n > 0$, (ii) $m > 0$, $n = 0$, and (iii) $m = 0$, $n = 0$. In these cases some of the rates are missing. (i) those in (a), (ii) those in (b), and (iii) those in (a) and (b). However, since the rates in each group balance we have $\pi Q = 0$. □

Example 4.23. General two-station queue. Suppose that at station i: arrivals from outside the system occur at rate λ_i, service occurs at rate μ_i, and departures go to the other queue with probability p_i and leave the system with probability $1 - p_i$.



Our question is: When is the system stable? That is, when is there a stationary distribution? To get started on this question suppose that both servers are busy. In this case work arrives at station 1 at rate $\lambda_1 + p_2\mu_2$, and work arrives at station 2 at rate $\lambda_2 + p_1\mu_1$. It should be intuitively clear that:

(i) if $\lambda_1 + p_2\mu_2 < \mu_1$ and $\lambda_2 + p_1\mu_1 < \mu_2$, then each server can handle their maximum arrival rate and the system will have a stationary distribution.

(ii) if $\lambda_1 + p_2\mu_2 > \mu_1$ and $\lambda_2 + p_1\mu_1 > \mu_2$, then there is positive probability that both servers will stay busy for all time and the queue lengths will tend to infinity.

Not covered by (i) or (ii) is the situation in which server 1 can handle her worst case scenario but server 2 cannot cope with his:

$$\lambda_1 + p_2\mu_2 < \mu_1 \quad \text{and} \quad \lambda_2 + p_1\mu_1 > \mu_2$$

In some situations in this case, queue 1 will be empty often enough to reduce the arrivals at station 2 so that server 2 can cope with his workload. As we will see, a concrete example of this phenomenon occurs when

$$\lambda_1 = 1, \quad \mu_1 = 4, \quad p_1 = 1/2 \qquad \lambda_2 = 2, \quad \mu_2 = 3.5, \quad p_2 = 1/4$$

To check that for these rates server 1 can handle the maximum arrival rate but server 2 cannot, we note that

$$\lambda_1 + p_2\mu_2 = 1 + \frac{1}{4}\cdot 3.5 = 1.875 < 4 = \mu_1$$
$$\lambda_2 + p_1\mu_1 = 2 + \frac{1}{2}\cdot 4 = 4 > 3.5 = \mu_2$$

To derive general conditions that will allow us to determine when a two-station network is stable, let r_i be the long-run average rate that customers arrive at station i. If there is a stationary distribution, then r_i must also be the long run average rate at which customers leave station i or the queue would grow linearly in time. If we want the flow in and out of each of the stations to balance, then we need

$$r_1 = \lambda_1 + p_2 r_2 \quad \text{and} \quad r_2 = \lambda_2 + p_1 r_1 \tag{4.27}$$

Plugging in the values for this example and solving gives

$$r_1 = 1 + \frac{1}{4}r_2 \quad \text{and} \quad r_2 = 2 + \frac{1}{2}r_1 = 2 + \frac{1}{2}\left(1 + \frac{1}{4}r_2\right)$$

So $(7/8)r_2 = 5/2$ or $r_2 = 20/7$, and $r_1 = 1 + 20/28 = 11/7$. Since

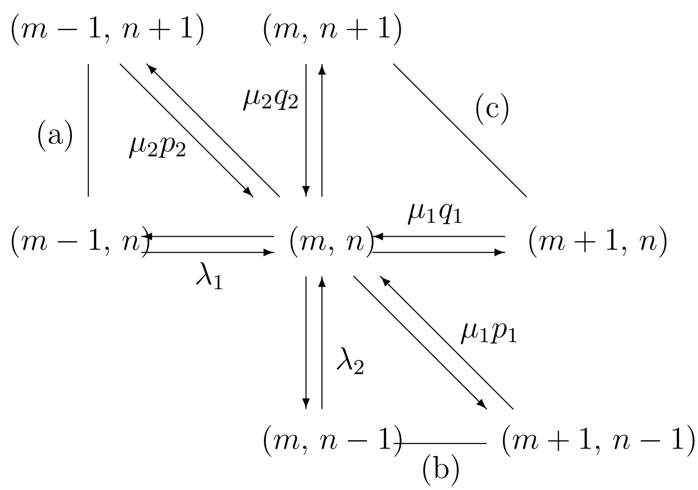
$$r_1 = 11/7 < 3 = \mu_1 \quad \text{and} \quad r_2 = 20/7 < 3.5$$

this analysis suggests that there will be a stationary distribution.

To prove that there is one, we return to the general situation and suppose that the r_i we find from solving (4.27) satisfy $r_i < \mu_i$. Thinking of two independent $M/M/1$ queues with arrival rates r_i, we let $\alpha_i = r_i/\mu_i$ and guess:

Theorem 4.9. *If $\pi(m,n) = c\alpha_1^m \alpha_2^n$ where $c = (1-\alpha_1)(1-\alpha_2)$ then π is a stationary distribution.*

Proof. The first step in checking $\pi Q = 0$ is to compute the rate matrix Q. To do this it is useful to draw a picture. Here, we have assumed that m and n are both positive. To make the picture slightly less cluttered, we have only labeled half of the arrows and have used $q_i = 1 - p_i$.



The rate arrows plus the dotted lines in the picture make three triangles. We will now check that the flows out of and into (m,n) in each triangle balance. In symbols we need to show that

$$
\begin{aligned}
(a)\quad & \mu_1\pi(m,n) = \mu_2 p_2 \pi(m-1,n+1) + \lambda_1\pi(m-1,n)\\
(b)\quad & \mu_2\pi(m,n) = \mu_1 p_1 \pi(m+1,n-1) + \lambda_2\pi(m,n-1)\\
(c)\quad & (\lambda_1+\lambda_2)\pi(m,n) = \mu_2(1-p_2)\pi(m,n+1) + \mu_1(1-p_1)\pi(m+1,n)
\end{aligned}
$$

Filling in $\pi(m,n) = c\alpha_1^m\alpha_2^n$ and canceling out c, we have

$$\begin{aligned}
\mu_1\alpha_1^m\alpha_2^n &= \mu_2 p_2\alpha_1^{m-1}\alpha_2^{n+1} + \lambda_1\alpha_1^{m-1}\alpha_2^n \\
\mu_2\alpha_1^m\alpha_2^n &= \mu_1 p_1\alpha_1^{m+1}\alpha_2^{n-1} + \lambda_2\alpha_1^m\alpha_2^{n-1} \\
(\lambda_1+\lambda_2)\alpha_1^m\alpha_2^n &= \mu_2(1-p_2)\alpha_1^m\alpha_2^{n+1} + \mu_1(1-p_1)\alpha_1^{m+1}\alpha_2^n
\end{aligned}$$

Canceling out the highest powers of α_1 and α_2 common to all terms in each equation gives

$$\begin{aligned}
\mu_1\alpha_1 &= \mu_2 p_2\alpha_2 + \lambda_1 \\
\mu_2\alpha_2 &= \mu_1 p_1\alpha_1 + \lambda_2 \\
(\lambda_1+\lambda_2) &= \mu_2(1-p_2)\alpha_2 + \mu_1(1-p_1)\alpha_1
\end{aligned}$$

Filling in $\mu_i\alpha_i = r_i$, the three equations become

$$\begin{aligned}
r_1 &= p_2 r_2 + \lambda_1 \\
r_2 &= p_1 r_1 + \lambda_2 \\
(\lambda_1+\lambda_2) &= r_2(1-p_2) + r_1(1-p_1)
\end{aligned}$$

The first two equations hold by (4.27). The third is the sum of the first two, so it holds as well.

This shows that $\pi Q = 0$ when $m, n > 0$. As in the proof for the tandem queue, there are three other cases to consider: (i) $m = 0$, $n > 0$, (ii) $m > 0$, $n = 0$, and (iii) $m = 0$, $n = 0$. In these cases some of the rates are missing. However, since the rates in each group balance we have $\pi Q = 0$. □

Example 4.24. Network of M/M/1 queues. Assume now that there are stations $1 \le i \le K$. Arrivals from outside the system occur to station i at rate λ_i and service occurs there at rate μ_i. Departures go to station j with probability $p(i,j)$ and leave the system with probability

$$q(i) = 1 - \sum_j p(i,j) \tag{4.28}$$

To have a chance of stability we must suppose

(A) *For each i it is possible for a customer entering at i to leave the system. That is, for each i there is a sequence of states $i = j_0, j_1, \ldots j_n$ with $p(j_{m-1}, j_m) > 0$ for $1 \le m \le n$ and $q(j_n) > 0$.*

Generalizing (4.27), we investigate stability by solving the system of equations for the r_j that represent the arrival rate at station j. As remarked earlier, the departure rate from station j must equal the arrival rate, or a linearly growing queue would develop. Thinking about the arrival rate at j in two different ways, it follows that

$$r_j = \lambda_j + \sum_{i=1}^{K} r_i p(i,j) \tag{4.29}$$

This equation can be rewritten in matrix form as $r = \lambda + rp$. In this form we can guess the solution

$$r = \sum_{n=0}^{\infty} \lambda p^n = \sum_{n=0}^{\infty} \sum_{i=1}^{K} \lambda_i p^n(i,j) \tag{4.30}$$

where p^n denotes the nth power of the matrix, and check that it works

$$r = \lambda + \sum_{m=0}^{\infty} \lambda p^m \cdot p = \lambda + rp$$

The last calculation is informal but it can be shown that under assumption (A) the series defining r converges and our manipulations are justified. Putting aside these somewhat tedious details, it is easy to see that the answer in (4.30) is reasonable: $p^n(i,j)$ is the probability a customer entering at i is at j after he has completed n services. The sum then adds the rates for all the ways of arriving at j.

Having found the arrival rates at each station, we can again be brave and guess that if $r_j < \mu_j$, then the stationary distribution is given by

$$\pi(n_1, \ldots, n_K) = \prod_{j=1}^{K} \left(\frac{r_j}{\mu_j}\right)^{n_j} \left(1 - \frac{r_j}{\mu_j}\right) \tag{4.31}$$

To prove this we will consider a more general collection of examples:

Example 4.25. Migration processes. As in the previous example, there are stations $1 \le i \le K$ and arrivals from outside the network occur to station i at rate λ_i. However, now when station i has n occupants, individuals depart at rate $\phi_i(n)$ where $\phi_i(n) \ge 0$

and $\phi_i(0) = 0$. Finally, a customer leaving i goes to station j with probability $p(i,j)$ independent of past events. Our main motivation for considering this more general set-up is that by taking $1 \le s_i \le \infty$ and letting

$$\phi_i(n) = \mu_i \min\{n, s_i\}$$

we can suppose that the ith station is an $M/M/s_i$ queue.

To find the stationary distribution for the migration process we first solve (4.29) to find the arrival and departure rates, r_i, for station i in equilibrium. Having done this we can let $\psi_i(n) = \prod_{m=1}^n \phi_i(m)$ and introduce our second assumption:

(B) *For* $1 \le j \le K$, *we have* $\sum_{n=0}^\infty r_j^n/\psi_j(n) < \infty$

This condition guarantees that there is a constant $c_j > 0$ so that

$$\sum_{n=0}^\infty c_j r_j^n/\psi_j(n) = 1$$

The next result says that $\pi_j(n) = c_j r_j^n/\psi_j(n)$ gives the equilibrium probability that queue j has n individuals and that in equilibrium the queue lengths are independent.

Theorem 4.10. *Suppose that conditions (A) and (B) hold. Then the migration process has stationary distribution*

$$\pi(n_1, \ldots, n_K) = \prod_{j=1}^K \frac{c_j\, r_j^{n_j}}{\psi_j(n_j)}$$

To make the connection with the results for Example 4.24, note that if all the queues are single server, then $\psi_j(n) = \mu_j^n$, so (B) reduces to $r_j < \mu_j$ and when this holds the queue lengths are independent shifted geometrics.

Proof. Write n as shorthand for $(n_1, n_2, \ldots n_K)$. Let A_j (for arrival) be the operator that adds one customer to queue j, let D_j (for departure) be the operator that removes one from queue j, and let T_{jk} be the operator that transfers one customer from j to k. That is,

$(A_j n)_j = n_j + 1$ with $(A_j n)_i = n_j$ otherwise
$(D_j n)_j = n_j - 1$ with $(D_j n)_i = n_j$ otherwise
$(T_{jk} n)_j = n_j - 1,\ (T_{jk} n)_k = n_k = n_k + 1,$ and $(T_{j,k} n)_i = n_j$ otherwise

Note that if $n_j = 0$ then $D_j n$ and $T_{jk} n$ have -1 in the jth coordinate, so in this case $q(n, D_j n)$ and $q(n, T_{jk} n) = 0$.

In equilibrium the rate at which probability mass leaves n is the same as the rate at which it enters n. Taking into account the various ways the chain can leave or enter state n it follows that the condition for a stationary distribution $\pi Q = 0$ is equivalent to $\pi_i \lambda_i = \sum_{j \neq i} \pi_j q(j, i)$ which in this case is

$$\begin{aligned}
\pi(n) & \left(\sum_{k=1}^{K} q(n, A_k n) + \sum_{j=1}^{K} q(n, D_j n) + \sum_{j=1}^{K} \sum_{k=1}^{K} q(n, T_{jk} n) \right) \\
& = \sum_{k=1}^{K} \pi(A_k n) q(A_k n, n) \\
& + \sum_{j=1}^{K} \pi(D_j n) q(D_j n, n) + \sum_{j=1}^{K} \sum_{k=1}^{K} \pi(T_{jk} n) q(T_{jk} n, n)
\end{aligned}$$

This will obviously be satisfied if we have

$$\pi(n) \sum_{k=1}^{K} q(n, A_k n) = \sum_{k=1}^{K} \pi(A_k n) q(A_k n, n) \tag{4.32}$$

and for each j we have

$$\begin{aligned}
\pi(n) & \left(q(n, D_j n) + \sum_{k=1}^{K} q(n, T_{jk} n) \right) \\
& = \pi(D_j n) q(D_j n, n) + \sum_{k=1}^{K} \pi(T_{jk} n) q(T_{jk} n, n)
\end{aligned} \tag{4.33}$$

Taking the second equation first, if $n_j = 0$, then both sides are 0, since $D_j n$ and $T_{jk} n$ are not in the state space. Supposing that $n_j > 0$ and filling in the values of our rates (4.33) becomes

$$\begin{aligned}
\pi(n) \phi_j(n_j) & \left(q(j) + \sum_{k=1}^{K} p(j, k) \right) \\
& = \pi(D_j n) \lambda_j + \sum_{k=1}^{K} \pi(T_{jk} n) \phi_k(n_k + 1) p(k, j)
\end{aligned} \tag{4.34}$$

The definition of q implies $q(j) + \sum_{k=1}^K p(j,k) = 1$, so filling in the proposed formula for $\pi(n)$ the left-hand side of (4.33) is

$$\pi(n)\phi_j(n_j) = \prod_{i=1}^K \frac{c_i\, r_i^{n_i}}{\psi_i(n_i)} \cdot \phi_j(n_j) = \prod_{i=1}^K \frac{c_i\, r_i^{\hat{n}_i}}{\psi_j(\hat{n}_i)} \cdot r_j = \pi(\hat{n}) \cdot r_j$$

where $\hat{n} = D_j n$ has $\hat{n}_j = n_j - 1$ and $\hat{n}_i = n_i$ for $i \neq j$. To compute the right-hand side of (4.33) we note that $(T_{jk}n)_i = \hat{n}_i$ for $i \neq k$ and $(T_{jk}n)_k = \hat{n}_k + 1 = n_k + 1$ so

$$\pi(T_{jk}n) = \pi(\hat{n}) \cdot \frac{r_k}{\phi_k(n_k+1)}$$

Since $D_j n = \hat{n}$, we can rewrite the right-hand side of (4.33) as

$$= \pi(\hat{n})\lambda_j + \pi(\hat{n}) \sum_{k=1}^K r_k p(k,j) = \pi(\hat{n}) \cdot r_j$$

where the last equality follows from (4.27): $\lambda_j + \sum_k r_k p(k,j) = r_j$.

At this point we have verified (4.33). Filling our rates into (4.32) and noting that $\pi(A_k n) = \pi(n) r_k / \phi_k(n_k + 1)$ we want to show

$$\pi(n) \sum_{k=1}^K \lambda_k = \pi(n) \sum_{k=1}^K \frac{r_k}{\phi_k(n_k+1)} \cdot \phi_k(n_k+1) q(k) \tag{4.35}$$

To derive this, we note that summing (4.27) from $j = 1$ to K and interchanging the order of summation in the double sum on the right gives

$$\begin{aligned} \sum_{j=1}^K r_j &= \sum_{j=1}^K \lambda_j + \sum_{k=1}^K r_k \sum_{j=1}^K p(k,j) \\ &= \sum_{j=1}^K \lambda_j + \sum_{k=1}^K r_k - \sum_{k=1}^K r_k q(i) \end{aligned}$$

since $\sum_{j=1}^K p(i,j) = 1 - q(i)$. Rearranging now gives

$$\sum_{k=1}^K r_k q(k) = \sum_{j=1}^K \lambda_j$$

This establishes (4.35), which implies (4.32), and completes the proof. □

4.6 Closed Queueing Networks

At first, the notion of N customers destined to move forever between K servers may sound like a queueing hell that might be the subject of a "Far Side" cartoon. However, as the next two examples show, this concept is useful for applications.

Example 4.26. Manufacturing system. The production of a part at a factory requires two operations. The first operation is always done at machine 1. The second is done at machine 2 or machine 3 with probabilities p and $1 - p$ after which the part leaves the system. Suppose that the factory has only a limited number of palettes, each of which holds one part. When a part leaves the system, the palette on which it rides is immediately used to bring a new part to queue at machine 1. If we ignore the parts, then the palettes are a closed queueing system. By computing the stationary distribution for this system we can compute the rate at which palettes leave machines 2 and 3, and hence compute the rate at which parts are made.

Example 4.27. Machine repair. For a concrete example consider trucks that can need engine repairs or tire repairs. To construct a closed queueing network model of this situation, we introduce three queues: 1 = the trucks that are working, 2 = those in engine repair, and 3 = those in tire repair. To simulate the breakdown mechanism, we will use an $M/M/\infty$ queue in which service times are exponential with rate λ if all trucks are always in use or an $M/M/s_1$ queue if there are never more than s_1 trucks in use. At repair station i we will suppose that there are s_i mechanics and hence an $M/M/s_i$ queue.

To begin to discuss these examples we need to introduce some notation and assumptions. Let $p(i, j)$ be the probability of going to station j after completing service at station i. We will suppose that

(A) *p is irreducible and has finitely many states*

so that there is a unique stationary distribution $\pi_i > 0$ for the routing matrix $p(i, j)$. If we let r_i denote the rate at which customers arrive at i in equilibrium then since there are no arrivals from outside, (4.29) becomes

$$r_j = \sum_{i=1}^{K} r_i\, p(i, j) \tag{4.36}$$

If we divide r_j by $R = \sum_j r_j$, the result is a stationary distribution for p. However, the stationary distribution for p is unique so we must have $r_j = R\pi_j$. Note that the r_j are departure (or arrival) rates, so their sum, R, is not the number of customers in the system, but instead represents the average rate at which services are completed in the system. This can be viewed as a measure of the system's throughput rate. In Example 4.26, each part exits from exactly two queues, so $R/2$ is the rate at which new parts are made.

To find the stationary distribution for the queueing system we take a clue from Section 4.6 and guess that it will have a product form. This is somewhat of a crazy guess since there are a fixed total number of customers, and hence the queue lengths cannot possibly be independent. However, we will see that it works. As in the previous section we will again consider something more general than our queueing system.

Example 4.28. Closed migration processes are defined by three rules.

(a) No customers enter or leave the system.

(b) When there are n_i customers at station i departures occur at rate $\phi_i(n_i)$ where $\phi_i(0) = 0$.

(c) An individual leaving i goes to j with probability $p(i,j)$.

Theorem 4.11. *Suppose (A) holds. The equilibrium distribution for the closed migration process with N individuals is*

$$\pi(n_1, \dots, n_K) = c_N \prod_{i=1}^{K} \frac{\pi_i^{n_i}}{\psi_i(n_i)}$$

if $n_1 + \cdots + n_K = N$ and 0 otherwise. Here $\psi_i(n) = \prod_{m=1}^{n} \phi_i(m)$ with $\psi_i(0) = 1$ and c_N is the constant needed to make the sum equal to 1.

Note. Here we used π_j instead of $r_j = R\pi_j$. Since $\sum_{j=1}^{K} n_j = N$, we have

$$\prod_{i=1}^{K} \frac{r_i^{n_i}}{\psi_i(n_i)} = R^N \prod_{i=1}^{K} \frac{\pi_i^{n_i}}{\psi_i(n_i)}$$

so this change only affects the value of the norming constant.

Proof. Write n as shorthand for $(n_1, n_2, \dots n_K)$ and let T_{jk} be the operator that transfers one customer from j to k. That is, $T_{jk}n = \bar{n}$, where $\bar{n}_j = n_j - 1$, $\bar{n}_k = n_k + 1$, and $\bar{n}_i = n_i$ otherwise. The condition for a stationary distribution $\pi Q = 0$ is equivalent to

$$\pi(n) \sum_{j=1}^{K} \sum_{k=1}^{K} q(n, T_{jk}n) = \sum_{j=1}^{K} \sum_{k=1}^{K} \pi(T_{jk}n) q(T_{jk}n, n) \tag{4.37}$$

This will obviously be satisfied if for each j we have

$$\pi(n) \sum_{k=1}^{K} q(n, T_{jk}n) = \sum_{k=1}^{K} \pi(T_{jk}n) q(T_{jk}n, n)$$

Filling in the values of our rates we want to show that

$$\pi(n)\phi_j(n_j) \sum_{k=1}^{K} p(j,k) = \sum_{k=1}^{K} \pi(T_{jk}n)\phi_k(n_k+1)p(k,j) \tag{4.38}$$

If $n_j = 0$, then both sides are 0, since $\phi_j(0) = 0$ and $T_{jk}n$ is not in the state space. Thus we can suppose that $n_j > 0$. $\sum_{k=1}^{K} p(j,k) = 1$, so filling in the proposed value of $\pi(n)$, the left-hand side of (4.38) is

$$\pi(n)\phi_j(n_j) = c_N \prod_{i=1}^{K} \frac{\pi_i^{n_i}}{\psi_i(n_i)} \cdot \phi_i(n_i) = \pi(\hat{n}) \cdot \pi_j \tag{4.39}$$

where $\hat{n}_j = n_j - 1$ and $\hat{n}_i = n_i$ otherwise. To compute the right-hand side of (4.38) we note that $(T_{jk}n)_i = \hat{n}_i$ for $i \neq k$ and $(T_{jk}n)_k = \hat{n}_k + 1 = n_k + 1$, so from the formula in Theorem 4.11,

$$\pi(T_{jk}n) = \pi(\hat{n}) \cdot \frac{\pi_k}{\phi_k(n_k+1)}$$

Using this we can rewrite:

$$\sum_{k=1}^{K} \pi(T_{jk}n)\phi_k(n_k+1)p(k,j) = \pi(\hat{n}) \sum_{k=1}^{K} \pi_k p(k,j) \tag{4.40}$$

Since $\sum_k \pi_k p(k,j) = \pi_j$, the expressions in (4.39) and (4.40) are equal. This verifies (4.38) and we have proved the result. $\square$

To see what Theorem 4.11 says, we will now reconsider our two previous examples.

Manufacturing system. Consider the special case of Example 4.26 in which the service rates are $\mu_1 = 1/8$, $\mu_2 = 1/9$, $\mu_3 = 1/12$ (per minute), and the routing matrix is

$$p(i,j) = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

To find the stationary distribution we set $\pi_1 = c$, then compute

$$\pi_2 = (2/3)c \quad \pi_3 = (1/3)c$$

The sum of the π's is $2c$, so $c = 1/2$, and we have

$$\pi_1 = 1/2 \quad \pi_2 = 1/3 \quad \pi_3 = 1/6$$

These queues are single servers so

$$\psi_i(n) = \prod_{m=1}^{n} \phi_i(m) = \mu_i^n$$

Plugging into the formula in Theorem 4.11, we have

$$\pi(n_1, n_2, n_3) = c_N \prod_{i=1}^{K} \left(\frac{\pi_i}{\mu_i}\right)^{n_i} = c_N \, 4^{n_1} \, 3^{n_2} \, 2^{n_3}$$

if $n_1 + n_2 + n_3 = N$ and 0 otherwise.

At this point we have to pick c_N to make the sum of the probabilities equal to 1. This is not as easy as it sounds. To illustrate the problems involved consider the simple-sounding case $N = 4$. Our task in this case is to enumerate the triples (n_1, n_2, n_3) with $n_i \geq 0$ and $n_1 + n_2 + n_3 = 4$, compute $4^{n_1} 3^{n_2} 2^{n_3}$ and then sum up the weights to determine the normalizing constant.

4,0,0	256	2,0,2	64	0,4,0	81
3,1,0	192	1,3,0	108	0,3,1	54
3,0,1	128	1,2,1	72	0,2,2	36
2,2,0	144	1,1,2	48	0,1,3	24
2,1,1	96	1,0,3	32	0,0,4	16

Summing up the weights we get 1351, so multiplying the table above by $c_4 = 1/1351$ gives the stationary distribution.

One can get a somewhat better idea of the nature of the stationary distribution by looking at the distribution of the lengths of the individual queues. To compute the answers note for example that the probability of 3 customers in the second queue is sum of the probabilities for 1,3,0 and 0,3,1 or (108 + 54)/1351.

	4	3	2	1	0
queue 1	0.189	0.237	0.225	0.192	0.156
queue 2	0.060	0.120	0.187	0.266	0.367
queue 3	0.012	0.041	0.110	0.259	0.578

Note that the second and third queues are quite often empty while queue 1 holds most of the palettes.

For $N = 10$ this gets to be considerably more complicated. To count the number of possible states, we observe that the number of (n_1, n_2, n_3) with integers $n_i \geq 0$ is the number of ways of arranging 10 o's and 2 x's in a row. To make the correspondence let n_1 be the number of o's before the first x, let n_2 be the number of o's between the first and the second x, and n_3 the number of o's after the second x. For example

$$o\,o\,o\,o\,o\,x\,o\,o\,x\,o\,o\,o$$

becomes $n_1 = 5$, $n_2 = 2$, and $n_3 = 3$. Having made this transformation it is clear that the number of possible states is the number of ways of picking 2 locations to put the x's or

$$\binom{12}{2} = \frac{12 \cdot 11}{2} = 66 \text{ states}$$

Each of these states has a weight between $2^{10} = 1,024$ and $4^{10} = 1,048,576$, $/1c_{10}$ will be quite large.

Example 4.29. Machine repair. Consider the special case of Example 4.27 in which the breakdown rate is $\mu_1 = 1$ trucks per week, the service rates are $\mu_2 = 2$, $\mu_3 = 4$, and the routing matrix is

$$p(i,j) = \begin{pmatrix} 0 & 1/4 & 3/4 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

That is, 1/4 of the breakdowns are engine repairs, while 3/4 are tire repairs. To find the stationary distribution, we set $\pi_1 = c$, then compute

$$\pi_2 = (1/4)c \quad \pi_3 = (3/4)c$$

The sum of the π's is $2c$ so $c = 1/2$.

The first queue is an $M/M/\infty$ queue with service rate 1 per customer so

$$\psi_1(n) = \prod_{m=1}^{n} \phi_i(m) = n!$$

The second and third queues are single servers with rates 2 and 4, so

$$\psi_2(n) = 2^n \qquad \psi_3(n) = 4^n$$

Plugging into the formula in (7.2) we have that if $n_1 + n_2 + n_3 = N$

$$\begin{aligned}\pi(n_1, n_2, n_3) &= c_N \, \frac{(1/2)^{n_1}}{n_1!} \, \frac{(1/8)^{n_2}}{2^{n_2}} \, \frac{(3/8)^{n_1}}{4^{n_3}} \\ &= c_N' \cdot \frac{N!}{n_1!} \, 16^{n_1} \, 2^{n_2} \, 3^{n_3}\end{aligned}$$

where in the last step we have multiplied the probabilities by $N!\,32^N$ and changed the normalizing constant.

At this point we have to pick c_N' to make the sum of the probabilities equal to 1. For simplicity, we take $N = 3$. Enumerating the triples (n_1, n_2, n_3) with $n_i \geq 0$ and $n_1 + n_2 + n_3 = 3$ and then computing $16^{n_1} 2^{n_2} 3^{n_3}/n_1!$ gives the following result:

3,0,0	4096	1,2,0	384	0,3,0	48
2,1,0	1536	1,1,1	576	0,2,1	72
2,0,1	2304	1,0,2	864	0,1,2	128
				0,0,3	162

Summing up the weights we get 10,170, so multiplying the table above by $1/10,170$ gives the stationary distribution. From this we see that the number of broken trucks is 0, 1, 2, 3 with probabilities 0.403, 0.378, 0.179, 0.040.

4.7 Exercises

4.1. A salesman flies around between Atlanta, Boston, and Chicago as follows. She stays in each city for an exponential amount of time with mean 1/4 month if the city is A or B, but with mean 1/5 month if the city is C. From A she goes to B or C with probability 1/2 each; from B she goes to A with probability 3/4 and to C with probability 1/4; from C she always goes back to A. (a) Find the limiting fraction of time she spends in each city. (b) What is her average number of trips each year from Boston to Atlanta?

4.2. A small computer store has room to display up to 3 computers for sale. Customers come at times of a Poisson process with rate 2 per week to buy a computer and will buy one if at least 1 is available. When the store has only 1 computer left it places an order for 2 more computers. The order takes an exponentially distributed amount of time with mean 1 week to arrive. Of course, while the store is waiting for delivery, sales may reduce the inventory to 1 and then to 0. (a) Write down the matrix of transition rates Q_{ij} and solve $\pi Q = 0$ to find the stationary distribution. (b) At what rate does the store make sales?

4.3. Consider two machines that are maintained by a single repairman. Machine i functions for an exponentially distributed amount of time with rate λ_i before it fails. The repair times for each unit are exponential with rate μ_i. They are repaired in the order in which they fail. (a) Formulate a Markov chain model for this situation with state space $\{0, 1, 2, 12, 21\}$. (b) Suppose that $\lambda_1 = 1$, $\mu_1 = 2$, $\lambda_2 = 3$, $\mu_2 = 4$. Find the stationary distribution.

4.4. Consider the set-up of the previous problem but now suppose machine 1 is much more important than 2, so the repairman will always service 1 if it is broken. (a) Formulate a Markov chain model for the this system with state space $\{0, 1, 2, 12\}$ where the numbers indicate the machines that are broken at the time. (b) Suppose that $\lambda_1 = 1$, $\mu_1 = 2$, $\lambda_2 = 3$, $\mu_2 = 4$. Find the stationary distribution.

4.5. Two people are working in a small office selling shares in a mutual fund. Each is either on the phone or not. Suppose that salesman i is on the phone for an exponential amount of time with rate μ_i and then off the phone for an exponential amount of time with rate λ_i. (a) Formulate a Markov chain model for this system with state space $\{0, 1, 2, 12\}$ where the state indicates who is on the phone. (b) Find the stationary distribution.

4.6. (a) Consider the special case of the previous problem in which $\lambda_1 = \lambda_2 = 1$, and $\mu_1 = \mu_2 = 3$, and find the stationary probabilities. (b) Suppose they upgrade their telephone system so that a call to one line that is busy is forwarded to the other phone and lost if that phone is busy. Find the new stationary probabilities.

4.7. Two people who prepare tax forms are working in a store at a local mall. Each has a chair next to his desk where customers can sit and be served. In addition there is one chair where customers can sit and wait. Customers arrive at rate λ but will go away if there is already someone sitting in the chair waiting. Suppose that server i requires an exponential amount of time with rate μ_i and that when both servers are free an arriving customer is equally likely to choose either one. (a) Formulate a Markov chain model for this system with state space $\{0, 1, 2, 12, 3\}$ where the first four states indicate the servers that are busy while the last indicates that there is a total of three customers in the system: one at each server and one waiting. (b) Consider the special case in which $\lambda = 2$, $\mu_1 = 3$ and $\mu_2 = 3$. Find the stationary distribution.

4.8. *Two queues in series.* Consider a two station queueing network in which arrivals only occur at the first server and do so at rate 2. If a customer finds server 1 free he enters the system; otherwise he goes away. When a customer is done at the first server he moves on to the second server if it is free and leaves the system if it is not. Suppose that server 1 serves at rate 4 while server 2 serves at rate 2. Formulate a Markov chain model for this system with state space $\{0, 1, 2, 12\}$ where the state indicates the servers who are busy. In the long run (a) what proportion of customers enter the system? (b) What proportion of the customers visit server 2?

Detailed balance

4.9. A hemoglobin molecule can carry one oxygen or one carbon monoxide molecule. Suppose that the two types of gases arrive at rates 1 and 2 and attach for an exponential amount of time with rates 3 and 4, respectively. Formulate a Markov chain model with state space $\{+, 0, -\}$ where $+$ denotes an attached oxygen molecule, $-$ an attached carbon monoxide molecule, and 0 a free hemoglobin molecule and find the long-run fraction of time the hemoglobin molecule is in each of its three states.

4.10. A machine is subject to failures of types $i = 1, 2, 3$ at rates λ_i and a failure of type i takes an exponential amount of time with rate μ_i to repair. Formulate a Markov chain model with state space $\{0, 1, 2, 3\}$ and find its stationary distribution.

4.11. Solve the previous problem in the concrete case $\lambda_1 = 1/24$, $\lambda_2 = 1/30$, $\lambda_3 = 1/84$, $\mu_1 = 1/3$, $\mu_2 = 1/5$, and $\mu_3 = 1/7$.

4.12. Customers arrive at a full-service one-pump gas station at rate of 20 cars per hour. However, customers will go to another station if there are at least two cars in the station, i.e., one being served and one waiting. Suppose that the service time for customers is exponential with mean 6 minutes. (a) Formulate a Markov chain model for the number of cars at the gas station and find its stationary distribution. (b) On the average how many customers are served per hour?

4.13. Solve the previous problem for a two-pump self-serve station under the assumption that customers will go to another station if there are at least four cars in the station, i.e., two being served and two waiting.

4.14. Three frogs are playing near a pond. When they are in the sun they get too hot and jump in the lake at rate 1. When they are in the lake they get too cold and jump onto the land at rate 2. Let X_t be the number of frogs in the sun at time t. (a) Find the stationary distribution for X_t. (b) Check the answer to (a) by noting that the three frogs are independent two-state Markov chains.

4.15. A computer lab has three laser printers, two that are hooked to the network and one that is used as a spare. A working printer will function for an exponential amount of time with mean 20 days. Upon failure it is immediately sent to the repair facility and replaced by another machine if there is one in working order. At the repair facility machines are worked on by a single repairman who needs an exponentially distributed amount of time with mean 2 days to fix one printer. In the long run how often are there two working printers?

4.16. A computer lab has three laser printers that are hooked to the network. A working printer will function for an exponential amount of time with mean 20 days. Upon failure it is immediately sent to the repair facility. There machines are worked on by two repairman who can each repair one printer in an exponential amount of time with mean 2 days. However, it is not possible for two people to work on one printer at once. (a) Formulate a Markov chain model for the number of working printers and find the stationary distribution. (b) How often are both repairmen busy? (c) What is the average number of machines in use?

4.17. Consider a barbershop with two barbers and two waiting chairs. Customers arrive at a rate of 5 per hour. Customers arriving to a fully occupied shop leave without being served. Find the stationary distribution for the number of customers in the shop, assuming that the service rate for each barber is 2 customers per hour.

4.18. Consider a barbershop with one barber who can cut hair at rate 4 and three waiting chairs. Customers arrive at a rate of 5 per hour. (a) Argue that this new set-up will result in fewer lost customers than the previous scheme. (b) Compute the increase in the number of customers served per hour.

4.19. There are two tennis courts. Pairs of players arrive at rate 3 per hour and play for an exponentially distributed amount of time with mean 1 hour. If there are already two pairs of players waiting new arrivals will leave. Find the stationary distribution for the number of courts occupied.

4.20. A taxi company has three cabs. Calls come in to the dispatcher at times of a Poisson process with rate 2 per hour. Suppose that each requires an exponential amount of time with mean 20 minutes, and that callers will hang up if they hear there are no cabs available. (a) What is the probability all three cabs are busy when a call comes in? (b) In the long run, on the average how many customers are served per hour?

4.21. There are 15 lily pads and 6 frogs. Each frog at rate 1 gets the urge to jump and when it does, it moves to one of the 9 vacant pads chosen at random. Find the stationary distribution for the set of occupied lily pads.

4.22. *Detailed balance for three state chains.* Consider a chain with state space $\{1, 2, 3\}$ in which $q(i, j) > 0$ if $i \neq j$ and suppose that there is a stationary distribution that satisfies the detailed balance condition. (a) Let $\pi(1) = c$. Use the detailed balance condition between 1 and 2 to find $\pi(2)$ and between 2 and 3 to find $\pi(3)$. (b) What conditions on the rates must be satisfied for there to be detailed balance between 1 and 3?

4.23. *Kolmogorov cycle condition.* Consider an irreducible Markov chain with state space S. We say that the cycle condition is satisfied if given a cycle of states $x_0, x_1, \ldots, x_n = x_0$ with $q(x_{i-1}, x_i) > 0$ for $1 \leq i \leq n$, we have

$$\prod_{i=1}^{n} q(x_{i-1}, x_i) = \prod_{i=1}^{n} q(x_i, x_{i-1})$$

(a) Show that if q has a stationary distribution that satisfies the detailed balance condition, then the cycle condition holds. (b) To prove the converse, suppose that the cycle condition holds. Let $a \in S$ and set $\pi(a) = c$. For $b \neq a$ in S let $x_0 = a, x_1 \ldots x_k = b$ be a path from a to b with $q(x_{i-1}, x_i) > 0$ for $1 \leq i \leq k$ let

$$\pi(b) = \prod_{j=1}^{k} \frac{q(x_{i-1}, x_i)}{q(x_i, x_{i-1})}$$

Show that $\pi(b)$ is well defined, i.e., is independent of the path chosen. Then conclude that π satisfies the detailed balance condition.

Markovian queues

4.24. Consider a taxi station at an airport where taxis and (groups of) customers arrive at times of Poisson processes with rates 2 and 3 per minute. Suppose that a taxi will wait no matter how many other taxis are present. However, if an arriving person does not find a taxi waiting he leaves to find alternative transportation. (a) Find the proportion of arriving customers that get taxis. (b) Find the average number of taxis waiting.

4.25. *Queue with impatient customers.* Customers arrive at a single server at rate λ and require an exponential amount of service with rate μ. Customers waiting in line are impatient and if they are not in service they will leave at rate δ independent of their position in the queue. (a) Show that for any $\delta > 0$ the system has a stationary distribution. (b) Find the stationary distribution in the very special case in which $\delta = \mu$.

4.26. Customers arrive at the Shortstop convenience store at a rate of 20 per hour. When two or fewer customers are present in the checkout line, a single clerk works and the service time is 3 minutes. However, when there are three or more customers are present, an assistant comes over to bag up the groceries and reduces the service time to 2 minutes. Assuming the service times are exponentially distributed, find the stationary distribution.

4.27. Customers arrive at a carnival ride at rate λ. The ride takes an exponential amount of time with rate μ, but when it is in use, the ride is subject to breakdowns at rate α. When a breakdown occurs all of the people leave since they know that the time to fix a breakdown is exponentially distributed with rate β. (i) Formulate a Markov chain model with state space $\{-1, 0, 1, 2, \ldots\}$ where -1 is broken and the states $0, 1, 2, \ldots$ indicate the number of people waiting or in service. (ii) Show that the chain has a stationary distribution of the form $\pi(-1) = a$, $\pi(n) = b\theta^n$ for $n \geq 0$.

4.28. Customers arrive at a two-server station according to a Poisson process with rate λ. Upon arriving they join a single queue to wait for the next available server. Suppose that the service times of the two servers are exponential with rates μ_a and μ_b and that a customer who arrives to find the system empty will go to each of the servers with probability $1/2$. Formulate a Markov chain model for this system with state space $\{0, a, b, 2, 3, \ldots\}$ where the states give the number of customers in the system, with a or b indicating there is one customer at a or b respectively. Show that this system is time reversible. Set $\pi(2) = c$ and solve to find the limiting probabilities in terms of c.

4.29. Let X_t be a Markov chain with a stationary distribution π that satisfies the detailed balance condition. Let Y_t be the chain constrained to stay in a subset A of the state space. That is, jumps which take the chain out of A are not allowed, but allowed jumps occur at the original rates. In symbols, $\bar{q}(x, y) = q(x, y)$ if $x, y \in A$ and 0 otherwise. Let $C = \sum_{y \in A} \pi(y)$. Show that $\nu(x) = \pi(x)/C$ is a stationary distribution for Y_t.

4.30. Two barbers share a common waiting room that has N chairs. Barber i gives service at rate μ_i and has customers that arrive at rate $\lambda_i < \mu_i$. Assume that customers always wait for the barber they came to see even when the other is free, but go away if the waiting room is full. Let N_t^i be the number of customers for barber i that are waiting or being served. Find the stationary distribution for (N_t^1, N_t^2).

4.31. Solve the previous problem when $\lambda_1 = 1$, $\mu_1 = 3$, $\lambda_2 = 2$, $\mu_2 = 4$, and $N = 2$.

4.32. Consider an $M/M/s$ queue with no waiting room. In words, requests for a phone line occur at a rate λ. If one of the s lines is free, the customer takes it and talks for an exponential amount of time with rate μ. If no lines are free, the customer goes away never to come back. Find the stationary distribution. You do not have to evaluate the normalizing constant.

Queueing networks

4.33. Consider a production system consisting of a machine center followed by an inspection station. Arrivals from outside the system occur only at the machine center and follow a Poisson process with rate λ. The machine center and inspection station are each single-server operations with rates μ_1 and μ_2. Suppose that each item independently passes inspection with probability p. When an object fails inspection it is sent to the machine center for reworking. Find the conditions on the parameters that are necessary for the system to have a stationary distribution.

4.34. Consider a three station queueing network in which arrivals to servers $i = 1, 2, 3$ occur at rates $3, 2, 1$, while service at stations $i = 1, 2, 3$ occurs at rates $4, 5, 6$. Suppose that the probability of going to j when exiting i, $p(i, j)$ is given by $p(1, 2) = 1/3$, $p(1, 3) = 1/3$, $p(2, 3) = 2/3$, and $p(i, j) = 0$ otherwise. Find the stationary distribution.

4.35. *Feed-forward queues.* Consider a k station queueing network in which arrivals to server i occur at rate λ_i and service at station i occurs at rate μ_i. We say that the queueing network is feed-forward if the probability of going from i to $j < i$ has $p(i, j) = 0$. Consider a general three station feed-forward queue. What conditions on the rates must be satisfied for a stationary distribution to exist?

4.36. *Queues in series.* Consider a k station queueing network in which arrivals to server i occur at rate λ_i and service at station i occurs at rate μ_i. In this problem we examine the special case of the feed-forward system in which $p(i, i+1) = p_i$ for $1 \leq i < k$. In words the customer goes to the next station or leaves the system. What conditions on the rates must be satisfied for a stationary distribution to exist?

4.37. At registration at a very small college, students arrive at the English table at rate 10 and at the Math table at rate 5. A student who completes service at the English table goes to the Math table with probability 1/4 and to the cashier with probability 3/4. A student who completes service at the Math table goes to the English table with probability 2/5 and to the cashier with probability 3/5. Students who reach the cashier leave the system after they pay. Suppose that the service times for the English table, Math table, and cashier are 25, 30, and 20, respectively. Find the stationary distribution.

4.38. Three vendors have vegetable stands in a row. Customers arrive at the stands 1, 2, and 3 at rates 10, 8, and 6. A customer visiting stand 1 buys something and leaves with probability 1/2 or visits stand 2 with probability 1/2. A customer visiting stand 3 buys something and leaves with probability 7/10 or visits stand 2 with probability 3/10. A customer visiting stand 2 buys something and leaves with probability 4/10 or visits stands 1 or 3 with probability 3/10 each. Suppose that the service rates at the three stands are large enough so that a stationary distribution exists. At what rate do the three stands make sales. To check your answer note that since each entering customers buys exactly once the three rates must add up to 10+8+6=24.

4.39. Four children are playing two video games. The first game, which takes an average of 4 minutes to play, is not very exciting, so when a child completes a turn on it they always stand in line to play the other one. The second one, which takes an average of 8 minutes, is more interesting so when they are done they will get back in line to play it with probability 1/2 or go to the other machine with probability 1/2. Assuming that the turns take an exponentially distributed amount of time, find the stationary distribution of the number of children playing or in line at each of the two machines.

4.40. A computer lab has 3 laser printers and 5 toner cartridges. Each machine requires one toner cartridges which lasts for an exponentially distributed amount of time with mean 6 days. When a toner cartridge is empty it is sent to a repairman who takes an exponential amount of time with mean 1 day to refill it. This system can be modeled as a closed queueing network with 5 customers (the toner cartridges), one $M/M/3$ queue, and one $M/M/1$ queue. Use this observation to compute the stationary distribution. How often are all three printers working?

4.41. The exercise room at the Wonderland motel has three pieces of equipment. Five businessmen who are trapped there by one of Ithaca's snowstorms use machines 1,2,3 for an exponential amount of time with means 15,10,5 minutes. When a person is done with one piece of equipment, he picks one of the other two at random. If it is occupied he stands in line to use it. Let (n_1, n_2, n_3) be the number of people using or in line to use each of the three pieces of equipment. (a) Find the stationary distribution. (b) Evaluate the norming constant.

4.42. Generalize the previous problem so that there are N machines and M businessmean, but simplify it by supposing that all machines have the same service time. Find the stationary distribution.

Chapter 5

Martingales

In this chapter we will introduce a class of process that can be thought of as the fortune of a gambler betting on a fair game. These results will be important when we consider applications to finance in the next chapter. In addition, they will allow us to give more transparent proofs of some rpoofs from Sections 1.8 and 1.9 concerning exit distributions and exit times for Markov chains.

5.1 Conditional Expectation

Our study of martingales will rely heavily on the notion of conditional expectation and involve some formulas that may not be familiar, so we will review them here. We begin with several definitions. Given an event A we define its **indicator function**

$$1_A = \begin{cases} 1 & x \in A \\ 0 & x \in A^c \end{cases}$$

In words, 1_A is "1 on A" (and 0 otherwise). Given a random variable Y, we define the **integral of Y over A** to be

$$E(Y; A) = E(Y 1_A)$$

Note that on the right multiplying by 1_A sets the product $= 0$ on A^c and leaves the values on A unchanged. Finally, we define the **conditional expectation of Y given A** to be

$$E(Y|A) = E(Y; A)/P(A)$$

This is, of course, the expected value for the probability defined by

$$P(\cdot|A) = P(\cdot \cap A)/P(A)$$

It is easy to see from the definition that the integral over A is linear:

$$E(Y + Z; A) = E(Y; A) + E(Z; A) \tag{5.1}$$

so dividing by $P(A)$, conditional expectation also has this property

$$E(Y + Z|A) = E(Y|A) + E(Z|A) \tag{5.2}$$

(Provided of course that all of the expected values exist.) In addition, as in ordinary integration one can take constants outside of the integral.

Lemma 5.1. *If X is a constant c on A, then $E(XY|A) = cE(Y|A)$.*

Proof. Since $X = c$ on A, $XY1_A = cY1_A$. Taking expected values and pulling the constant out front, $E(XY1_A) = E(cY1_A) = cE(Y1_A)$. Dividing by $P(A)$ now gives the result. □

Our last two properties concern the behavior of $E(Y; A)$ and $(Y|A)$ as a function of the set A.

Lemma 5.2. *If B is the disjoint union of $A_1, \ldots, A_k$, then*

$$E(Y; B) = \sum_{j=1}^{k} E(Y; A_j)$$

Proof. Our assumption implies $Y1_B = \sum_{j=1}^{k} Y1_{A_j}$, so taking expected values, we have

$$E(Y; B) = E(Y1_B) = E\left(\sum_{j=1}^{k} Y1_{A_j}\right) = \sum_{j=1}^{k} E(Y1_{A_j}) = \sum_{j=1}^{k} E(Y; A_j)$$

□

Lemma 5.3. *If B is the disjoint union of $A_1, \ldots, A_k$, then*

$$E(Y|B) = \sum_{j=1}^{k} E(Y|A_j) \cdot \frac{P(A_j)}{P(B)}$$

Proof. Using the definition of conditional expectation, formula (5.2), then doing some arithmetic and using the definition again, we have

$$\begin{aligned} E(Y|B) = E(Y;B)/P(B) &= \sum_{j=1}^{k} E(Y;A_j)/P(B) \\ &= \sum_{j=1}^{k} \frac{E(Y;A_j)}{P(A_j)} \cdot \frac{P(A_j)}{P(B)} = \sum_{j=1}^{k} E(Y|A_j) \cdot \frac{P(A_j)}{P(B)} \end{aligned}$$

which proves the desired result. $\square$

5.2 Examples of Martingales

We begin by giving the definition of a martingale. Thinking of M_n as the amount of money at time n for a gambler betting on a fair game, and X_n as the outcomes of the gambling game we say that $M_0, M_1, \ldots$ is a **martingale** with respect to $X_0, X_1, \ldots$ if for any $n \geq 0$ we have $E|M_n| < \infty$ and for any possible values $x_n, \ldots, x_0$

$$E(M_{n+1} - M_n | X_n = x_n, X_{n-1} = x_{n-1}, \ldots X_0 = x_0, M_0 = m_0) = 0 \tag{5.3}$$

The first condition, $E|M_n| < \infty$, is needed to guarantee that the conditional expectation makes sense. The second, defining property says that conditional on the past up to time n the average profit from the bet on the nth game is 0. It will take several examples to explain why this is a natural definition. It is motivated by the fact that in passing from the random variables X_n that are driving the process to our winnings M_n there may be a loss of information so it is convenient to condition on X_n rather than on M_n.

To explain the reason for our interest in martingales, and to help explain the definition we will now give a number of examples. In what follows we will often be forced to write the conditioning event so we introduce the short hand

$$A_{x,m} = \{X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0, M_0 = m\}$$

where x is short for $(x_n, \ldots, x_0)$

Example 5.1. Random walks. Let $X_1, X_2, \ldots$ be i.i.d. with $EX_i = \mu$. Let $S_n = S_0 + X_1 + \cdots + X_n$ be a random walk. $M_n = S_n - n\mu$ is a martingale with respect to X_n.

Proof. To check this, note that $M_{n+1} - M_n = X_{n+1} - \mu$ is independent of $X_n, \ldots, X_0, M_0$, so the conditional mean of the difference is just the mean:

$$E(M_{n+1} - M_n | A_{x,m}) = EX_{n+1} - \mu = 0 \qquad \square$$

In most cases, casino games are not fair but biased against the player. We say that M_n is a **supermartingale** with respect to X_n if a gambler's expected winnings on one play are negative:

$$E(M_{n+1} - M_n | A_{x,m}) \leq 0$$

To help remember the direction of the inequality, note that there is nothing "super" about a supermartingale. The definition traces its roots to the notion of superharmonic functions whose values at a point exceed the average value on balls centered around the point. If we reverse the sign and suppose

$$E(M_{n+1} - M_n | A_{x,m}) \geq 0$$

then M_n is called a **submartingale** with respect to X_n. A simple modification of the proof for Example 5.1 shows that if $\mu \leq 0$, then S_n defines a supermartingale, while if $\mu \geq 0$, then S_n is a submartingale.

The next result will lead to a number of examples.

Theorem 5.1. *Let X_n be a Markov chain with transition probability p and let $f(x, n)$ be a function of the state x and the time n so that*

$$f(x, n) = \sum_y p(x, y) f(y, n + 1)$$

Then $M_n = f(X_n, n)$ is a martingale with respect to X_n. In particular if $h(x) = \sum_y p(x, y)h(y)$ then $h(X_n)$ is a martingale.

Proof. By the Markov property and our assumption on f

$$E(f(X_{n+1}, n + 1)|A_{x,m}) = \sum_y p(x_n, y) f(y, n + 1) = f(x_n, n)$$

which proves the desired result. □

The next two examples begin to explain our interest in Theorem 5.1.

Example 5.2. Gambler's ruin. Let $X_1, X_2, \ldots$ be independent with

$$P(X_i = 1) = p \quad \text{and} \quad P(X_i = -1) = 1 - p$$

where $p \in (0, 1)$ and $p \neq 1/2$. Let $S_n = S_0 + X_1 + \cdots + X_n$. $M_n = \left(\frac{1-p}{p}\right)^{S_n}$ is a martingale with respect to X_n.

Proof. Using Theorem $h(x) = ((1 - p)/p)^x$, we need only check that

$h(x) = \sum_y p(x,y)h(y)$. To do this we note that

$$\begin{aligned}\sum_y p(x,y)h(y) &= p \cdot \left(\frac{1-p}{p}\right)^{x+1} + (1-p)\cdot\left(\frac{1-p}{p}\right)^{x-1} \\ &= (1-p)\cdot\left(\frac{1-p}{p}\right)^{x} + p\cdot\left(\frac{1-p}{p}\right)^{x} = \left(\frac{1-p}{p}\right)^{x}\end{aligned}$$

which proves the desired result. □

Example 5.3. Symmetric simple random walk. Let $Y_1, Y_2, \ldots$ be independent with

$$P(Y_i = 1) = P(Y_i = -1) = 1/2$$

and let $X_n = X_0 + Y_1 + \cdots + Y_n$. Then $M_n = X_n^2 - n$ is a martingale with respect to X_n. By Theorem 5.1 with $f(x,n) = x^2 - n$ it is enough to show that

$$\frac{1}{2}(x+1)^2 + \frac{1}{2}(x-1)^2 - 1 = x^2$$

To do this we work out the squares to conclude the left-hand side is

$$\frac{1}{2}[x^2 + 2x + 1 + x^2 - 2x + 1] = 1$$

Example 5.4. Products of independent random variables. To build a discrete time model of the stock market we let $X_1, X_2, \ldots$ be independent ≥ 0 with $EX_i = 1$. Then $M_n = M_0 X_1 \cdots X_n$ is a martingale with respect to X_m. To prove this we note that

$$E(M_{n+1} - M_n | A_{x,n}) = M_n E(X_{n+1} - 1 | A_{x,n}) = 0$$

The last example generalizes easily to give:

Example 5.5. Exponential martingale. Let $Y_1, Y_2, \ldots$ be independent and identically distributed with $\phi(\theta) = E\exp(\theta X_1) < \infty$. Let $S_n = S_0 + Y_1 + \cdots + Y_n$. Then $M_n = \exp(\theta S_n)/\phi(\theta)^n$ is a martingale with respect to X_n.

Proof. If we let $X_i = \exp(\theta Y_i)/\phi(\theta)$ then $M_n = M_) X_1 \cdots X_n$ with $EX_i = 1$ and this reduces to the previous example. □

5.3 Properties of Martingales

The first result should be intutitive if we think of supermartingale as betting on an unfavorable game: the expected value of our fortune will decline over time.

Theorem 5.2. *If M_m is a supermartingale and $m \le n$ then $EM_m \ge EM_n$.*

Proof. It is enough to show that the expected value decreases with each time step, i.e., $EM_k \ge EM_{k+1}$. To do this, we write x as shorthand for the vector $(x_n, x_{n-1}, \ldots x_0)$, let

$$A_{x,m} = \{X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0, M_0 = m\}$$

and note that (5.2) and the definition of conditional expectation imply

$$\begin{aligned} E(Y_{k+1} - Y_k) &= \sum_i E(Y_{k+1} - Y_k; A_i) \\ &= \sum_i P(A_i) E(Y_{k+1} - Y_k | A_i) \le 0 \end{aligned}$$

since each term $E(Y_{k+1} - Y_k | A_i) \le 0$. □

The result in Theorem 5.2 generalizes immediately to our other two types of processes. Multiplying by -1 we see:

Theorem 5.3. *If M_m is a submartingale and $0 \le m < n$, then $EM_m \le EM_n$.*

Since a process is a martingale if and only if it is both a supermartingale and submartingale, we can conclude that:

Theorem 5.4. *If M_m is a martingale and $0 \le m < n$ then $EM_m = EM_n$.*

The most famous result of martingale theory is that "you can't make money playing a fair game" and hence "you can't beat an unfavorable game." In this section we will prove two results that make these statements precise. Our first step is analyze a famous gambling system and show why it doesn't work.

Example 5.6. Doubling strategy. Suppose you are playing a game in which you will win or lose \$1 on each play. If you win you bet \$1 on the next play but if you lose then you bet twice the previous amount. The idea behind the system can be seen by looking at what happens if we lose four times in a row then win:

outcome	L	L	L	L	W
bet	1	2	4	8	16
net profit	-1	-3	-7	-15	1

In this example our net profit when we win is \$1. Since $1+2+\cdots+2^k = 2^{k+1}-1$, this is true if we lose k times in a row. Thus every time we win our net profit is up by \$1 from the previous time we won.

This system will succeed in making us rich as long as the probability of winning is positive, so where's the catch? The problem is that theorem will imply that if we use the doubling system on a supermartingale up to a fixed time n and we let W_n be our net winnings. Then $EW_n \le 0$.

To prove this we will introduce a family of betting strategies that generalize the doubling strategy. The amount of money we bet on the nth game, H_n, clearly, cannot depend on the outcome of that game nor is it sensible to allow it to depend on the outcomes of games that will be played later. We say that H_n is an admissible gambling strategy or **predictable process** if for each n the value of H_n can be determined from $X_{n-1}, X_{n-2}, \ldots, X_0, M_0$.

To motivate the next definition, think of H_m as the amount of stock we hold between time $m-1$ and m. Then our wealth at time n is

$$W_n = W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}) \tag{5.4}$$

since the change in our wealth from time $m-1$ to m is the amount we hold times the change in the price of the stock: $H_m(M_m - M_{m-1})$. To formulate the doubling strategy in this setting, let $X_m = 1$ if the mth coin flip is heads and -1 if the mth flip is tails, and let $M_n = X_1 + \cdots + X_n$ as the net profit of a gambler who bets 1 unit every time.

Theorem 5.5. *Suppose that M_n is a supermartingale with respect to X_n, H_n is predictable, and $0 \le H_n \le c_n$ where c_n is a constant that may depend on n. Then*

$$W_n = W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}) \quad \textit{is a supermartingale}$$

We need the condition $H_n \ge 0$ to prevent the bettor from becoming the house by betting a negative amount of money. The upper bound $H_n \le c_n$ is a technical condition that is needed to have expected values make sense. In the gambling context this assumption is harmless: even if the bettor wins every time there is an upper bound to the amount of money he can have at time n.

Proof. The gain at time $n+1$ is

$$W_{n+1} - W_n = H_{n+1}(Y_{n+1} - Y_n)$$

As in the proof of Theorem 5.2 let

$$A_{x,n} = \{X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0, M_0 = m_0\}.$$

H_{n+1} is constant on the event $A_{x,m}$, so Lemma 5.1 implies

$$E(H_{n+1}(M_{n+1} - M_n)|A_{x,m}) = H_{n+1}E(M_{n+1} - M_n|A_{x,m}) \le 0$$

verifying that W_n is a supermartingale. □

Arguing as in the discussion after Theorem 5.2 the same result holds for submartingales and for martingales with only the assumption that $|H_n| \le c_n$.

Though Theorem 5.5 may be depressing for gamblers, a simple special case gives us an important computational tool. To introduce this tool, we need one more notion.

We say that T is a **stopping time with respect to** X_n if the occurrence (or nonoccurrence) of the event $\{T = n\}$ can be determined from $X_n, X_{n-1} \ldots X_0, M_0$.

Example 5.7. Constant betting up to a stopping time. One possible gambling strategy is to bet \$1 each time until you stop playing at time T. In symbols, we let $H_m = 1$ if $T \ge m$ and 0 otherwise. To check that this is an admissible gambling strategy we note that the set on which H_m is 0 is

$$\{T \geq m\}^c = \{T \leq m-1\} = \cup_{k=1}^{m-1}\{T = k\}$$

By the definition of a stopping time, the event $\{T = k\}$ can be determined from the values of $M_0, X_0, \ldots, X_k$. Since the union is over $k \leq m-1$, H_m can be determined from the values of $M_0, X_0, X_1, \ldots, X_{m-1}$.

Having introduced the gambling strategy "Bet \$1 on each play up to time T" our next step is to compute the payoff we receive when $W_0 = M_0$. Letting $T \wedge n$ denote the minimum of T and n, i.e., it is T if $T < n$ and n if $T \geq n$, we can give the answer as:

$$W_n = M_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}) = M_{T\wedge n} \tag{5.5}$$

To check the last equality, consider two cases:

(i) if $T \geq n$ then $H_m = 1$ for all $m \leq n$, so

$$W_n = M_0 + (M_n - M_0) = M_n$$

(ii) if $T \leq n$ then $H_m = 0$ for $m > T$ and the sum stops at T. In this case,

$$W_n = Y_0 + (Y_T - Y_0) = Y_T$$

Combining (5.5) with Theorem 5.5 shows

Theorem 5.6. *If M_n is a supermartingale with respect to X_n and T is a stopping time then the stopped process $M_{T\wedge n}$ is a supermartingale with respect to X_n.*

As in the discussion after Theorem 5.2 the same conclusion is true for submartingales and martingales.

5.4 Applications

In this section we will apply the results from the previous section to rederive some of the results from Chapter 1 about hitting probabilities and exit times. In all the examples the method is the same. We have a martingale M_n and a stopping time. We use Theorems 5.6 and 5.4 to conclude $EM_0 = EM_{T\wedge n}$ then we let $n \to \infty$.

Example 5.8. Gambler's ruin. Continuing let $X_1, X_2, \dots \xi_n$ be independent with

$$P(X_i = 1) = p \quad \text{and} \quad P(\xi_i = -1) = 1 - p$$

Let $S_n = S_0 + X_1 + \cdots + X_n$ where $X_1, X_2, \dots$ are independent with $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q$ where $q = 1 - p$. Suppose $0 < p < 1$ with $p \neq 1/2$ and let $h(x) = (q/p)^x$. Example 5.2 implies that $M_n = h(S_n)$ is a martingale. Let $T = \min\{n : S_n \notin (a, b)\}$. It is easy to see that T is a stopping time. Lemma ?? implies that $P(T < \infty) = 1$. Using Theorems 5.6 and 5.4, we have

$$\begin{aligned}(q/p)^x = E_x M_{T\wedge n} &= (q/p)^a P(T \leq n, S_T = a) + (q/p)^b P(T \leq n, S_T = b) \\ &\quad + E((q/p)^{S_n}; T > n)\end{aligned}$$

Since $P(T < \infty) = 1$ and for $a < x < b$, $(q/p)^x \leq \max\{(q/p)^a, (q/p)^b\}$ the third term tends to 0 and we have

$$(q/p)^x = (q/p)^a P(S_T = a) + (q/p)^b [1 - P(S_T = a)]$$

Solving gives

$$P_x(S_T = a) = \frac{(q/p)^b - (q/p)^x}{(q/p)^b - (q/p)^a}$$

generalizing (1.21).

Example 5.9. Duration of fair games. Let $S_n = S_0 + X_1 + \cdots + X_n$ where $X_1, X_2, \dots$ are independent with $P(X_i = 1) = P(X_i = -1) = 1/2$. Let $T = \min\{n : S_n \notin (a, b)\}$ where $a < 0 < b$. S_n is a martingale so repeating the last proof shows that

$$P_0(S_T = a) = \frac{b}{b - a} \qquad P_(S_T = b) = \frac{-a}{b - a}$$

Our goal here is to prove a close relative of (1.25):

$$E_0 T = -ab$$

Example 5.3 implies that $S_n^2 - n$ is a martingale. Let $T = \min\{n : S_n \notin (a,b)\}$. From the previous example we have that T is a stopping time with $P(T < \infty) = 1$. Using Theorems 5.6 and 5.4 we have

$$0 = E_0(S_{T\wedge n}^2 - T \wedge n) = a^2 P(S_T = a, T \le n) + b^2 P(S_T = b, T \le n) \\ + E(S_n^2; T > n) - E_0(T \wedge n)$$

$P(T < \infty) = 1$ and on $\{T > n\}$ we have $S_{T\wedge n}^2 \le \max\{a^2, b^2\}$ so the third term tends to 0. As $n \uparrow \infty$ $E_0(T \wedge n) \uparrow E_0 T$ so using the result for the exist distribution we have

$$E_0 T = a^2 \frac{b}{b-a} + b^2 \frac{-a}{b-a} = \frac{-ab^2 + ba^2}{b-a} = -ab$$

Consider now a random walk $S_n = S_0 + X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are i.i.d. with mean μ. From Example 5.1, $M_n = S_n - n\mu$ is a martingale with respect to X_n.

Theorem 5.7. Wald's equation. *If T is a stopping time with $ET < \infty$, then*

$$E(S_T - S_0) = \mu ET$$

Why is this true? Theorems 5.6 and 5.4 give

$$ES_0 = E(S_{T\wedge n}) - \mu E(T \wedge n)$$

As $n \uparrow \infty$, $E_0(T \wedge n) \uparrow E_0 T$. To pass to the limit in the other term, we note that

$$E|S_T - S_{T\wedge n}| \le E\left(\sum_{m=n}^{T} |X_m|; T > n\right)$$

Using the assumptions $ET < \infty$ and $E|X| < \infty$ one can prove that the right-hand side tends to 0 and complete the proof. □

With Wald's equation in hand we can now better understand:

Example 5.10. Mean time to gambler's ruin. Let $S_n = S_0 + X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are independent with $P(X_i = 1) = p < 1/2$ and $P(X_i = -1) = 1 - p$. The mean movement on one step is $\mu = 2p - 1$ so $S_n - (2p-1)n$ is a martingale. Let $V_a = \min\{n \ge 0 : S_n = a\}$. Theorems 5.6 and 5.4 give

$$x = E_x S_{V_0 \wedge n} - (2p-1) E_x (V_0 \wedge n)$$

Rearranging we have

$$(1-2p)E_x(V_0 \wedge n) = x - E_x S_{V_0 \wedge n} \le x$$

This shows that $E_x V_0 < \infty$ so we can use Wald's equation to conclude that $-x = (2p-1)E_x V_0$ and we have a new derivation of (1.27)

$$E_x V_0 = x/(1-2p)$$

Example 5.11. Left-continuous random walk. Suppose that $X_1, X_2, \ldots$ are independent integer-valued random variables with $EX_i > 0$, $P(X_i \ge -1) = 1$, and $P(X_i = -1) > 0$. These walks are called left-continuous since they cannot jump over any integers when they are decreasing, which is going to the left as the number line is usually drawn. Let $\phi(\theta) = \exp(\theta X_i)$ and define $\alpha < 0$ by the requirement that $\phi(\alpha) = 1$. To see that such an α exists, note that (i) $\phi(0) = 1$ and

$$\phi'(\theta) = \frac{d}{d\theta} E e^{\theta x_i} = E(x_i e^{\theta x_i}) \quad \text{so} \quad \phi'(0) = Ex_i > 0$$

and $\phi(\theta) < 1$ for small negative θ. (ii) If $\theta < 0$, then $\phi(\theta) \ge e^{-\theta} P(x_i = -1) \to \infty$ as $\theta \to -\infty$. Our choice of α makes $\exp(\alpha S_n)$ a martingale. Having found the martingale it is easy now to conclude:

Theorem 5.8. *Consider a left continuous random walk. Let $a < 0$ and $V_a = \min\{n : S_n = a\}$.*

$$P_0(V_a < \infty) = e^{-\alpha a}$$

Proof. Theorems 5.6 and 5.4 give

$$1 = E_0 \exp(\alpha S_{V_a \wedge n}) = e^{\alpha a} P_0(V_a \le n) + E_0(\exp(\alpha S_n); T > n)$$

The strong law of large numbers implies that on $T = \infty$, $S_n/n \to \mu > 0$, so the second term $\to 0$ as $n \to \infty$ and it follows that $1 = e^{\alpha a} P_0(V_a < \infty)$. □

When the random walk is not left continuous we cannot get exact results on hitting probabilities but we can still get a bound.

Example 5.12. Cramér's estimate of ruin. Let S_n be the total assets of an insurance company at the end of year n. During year

n, premiums totaling c dollars are received, while claims totaling Y_n dollars are paid, so

$$S_n = S_{n-1} + c - Y_n$$

Let $X_n = c - Y_n$ and suppose that $X_1, X_2, \dots$ are independent random variables that are normal with mean $\mu > 0$ and variance σ^2. That is the density function of η_i is

$$(2\pi\sigma^2)^{-1/2} \exp(-(x-\mu)^2/2\sigma^2)$$

Let B for bankrupt be the event that the wealth of the insurance company is negative at some time n. We will show

$$P(B) \le \exp(-2\mu S_0/\sigma^2) \tag{5.6}$$

In words, in order to be successful with high probability, $\mu S_0/\sigma^2$ must be large, but the failure probability decreases exponentially fast as this quantity increases.

Proof. We begin by computing $\phi(\theta) = E\exp(\theta X_i)$. To do this we need a little algebra

$$-\frac{(x-\mu)^2}{2\sigma^2} + \theta(x-\mu) + \theta\mu = -\frac{(x-\mu-\sigma^2\theta)^2}{2\sigma^2} + \frac{\sigma^2\theta^2}{2} + \theta\mu$$

and a little calculus

$$\begin{aligned}\phi(\theta) &= \int e^{\theta x}(2\pi\sigma^2)^{-1/2}\exp(-(x-\mu)^2/2\sigma^2)\,dx \\ &= \exp(\sigma^2\theta^2/2 + \theta\mu)\int (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(x-\mu-\sigma^2\theta)^2}{2\sigma^2}\right)\,dx\end{aligned}$$

Since the integrand is the density of a normal with mean $\mu + \sigma^2\theta$ and variance σ^2 it follows that

$$\phi(\theta) = \exp(\sigma^2\theta^2/2 + \theta\mu) \tag{5.7}$$

If we pick $\theta = -2\mu/\sigma^2$, then

$$\sigma^2\theta^2/2 + \theta\mu = 2\mu^2/\sigma^2 - 2\mu^2/\sigma^2 = 0$$

So Example 5.5 implies $\exp(-2\mu S_n/\sigma^2)$ is a martingale. Let $T = \min\{n : S_n \le 0\}$. Theorems 5.6 and 5.4 gives

$$\exp(-2\mu S_0/\sigma^2) = E\exp(-2\mu S_{T\wedge n}) \ge P(T \le n)$$

since $\exp(-2\mu S_T/\sigma^2) \ge 1$ and the contribution to the expected value from $\{T > n\}$ is ≥ 0. Letting $n \to \infty$ now and noticing $P(T \le n) \to P(B)$ gives the desired result. $\square$

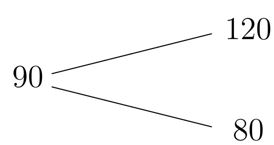
Chapter 6

Finance

6.1 Two simple examples

To warm up for the developments in the next section we will look at two simple concrete examples under the unrealistic assumption that the interest rate is 0.

One period case. In our first scenario the stock is at 90 at time 0 and may be 80 or 120 at time 1.



Suppose now that you are offered a **European call option** with **strike price** 100 and **expiry** 1. This means that after you see what happened to the stock, you have an option to buy the stock (but not an obligation to do so) for 100 at time 1. If the stock price is 80, you will not exercise the option to purchase the stock and your profit will be 0. If the stock price is 120 you will choose to buy the stock at 100 and then immediately sell it at 120 to get a profit of 20. Combining the two cases we can write the payoff in general as $(X_1 - 100)^+$, where $z^+ = \max\{z, 0\}$ denotes the positive part of z.

Our problem is to figure out the right price for this option. At first glance this may seem impossible since we have not assigned probabilities to the various events. However, it is a miracle of "pricing by the absence of arbitrage" that in this case we do not have to assign probabilities to the events to compute the price. To explain this we start by noting that X_1 will be 120 ("up") or 80 ("down") for a profit of 30 or a loss of 10, respectively. If we pay c for the option, then when X_1 is up we make a profit of $20 - c$, but when it is down we make $-c$. The last two sentences are summarized in the following table

	stock	option
up	30	$20 - c$
down	-10	$-c$

Suppose we buy x units of the stock and y units of the option, where negative numbers indicate that we sold instead of bought. One possible strategy is to choose x and y so that the outcome is the same if the stock goes up or down:

$$30x + (20 - c)y = -10x + (-c)y$$

Solving, we have $40x + 20y = 0$ or $y = -2x$. Plugging this choice of y into the last equation shows that our profit will be $(-10 + 2c)x$. If $c > 5$, then we can make a large profit with no risk by buying large amounts of the stock and selling twice as many options. Of course, if $c < 5$, we can make a large profit by doing the reverse. Thus, in this case the only sensible price for the option is 5.

A scheme that makes money without any possibility of a loss is called an **arbitrage opportunity**. It is reasonable to think that these will not exist in financial markets (or at least be short-lived) since if and when they exist people take advantage of them and the opportunity goes away. Using our new terminology we can say that the only price for the option which is consistent with absence of arbitrage is $c = 5$, so that must be the price of the option.

To find prices in general, it is useful to look at things in a different way. Let $a_{i,j}$ be the profit for the ith security when the jth outcome occurs.

Theorem 6.1. *Exactly one of the following holds:*

(i) There is an investment allocation x_i so that $\sum_{i=1}^m x_i a_{i,j} \geq 0$ for each j and $\sum_{i=1}^m x_i a_{i,k} > 0$ for some k.

(ii) There is a probability vector $p_j > 0$ so that $\sum_{j=1}^n a_{i,j} p_j = 0$ for all i.

Here an x satisfying (i) is an arbitrage opportunity. We never lose any money but for at least one outcome we gain a positive amount. Turning to (ii), the vector p_j is called a martingale measure since if the probability of the jth outcome is p_j, then the expected change in the price of the ith stock is equal to 0. Combining the two interpretations we can restate Theorem 6.1) as:

There is no arbitrage if and only if there is a strictly positive probability vector so that all the stock prices are martingale.

Proof of Theorem 6.1. One direction is easy. If (i) is true, then for any strictly positive probability vector $\sum_{i=1}^m \sum_{j=1}^n x_i a_{i,j} p_j > 0$, so (ii) is false.

Suppose now that (i) is false. The linear combinations $\sum_{i=1}^m x_i a_{i,j}$ when viewed as vectors indexed by j form a linear subspace of n-dimensional Euclidean space. Call it $\mathcal{L}$. If (i) is false, this subspace intersects the positive orthant $\mathcal{O} = \{y : y_j \geq 0 \text{ for all } j\}$ only at the origin. By linear algebra we know that $\mathcal{L}$ can be extended to an $n-1$ dimensional subspace $\mathcal{H}$ that only intersects $\mathcal{O}$ at the origin.

Since $\mathcal{H}$ has dimension $n-1$, it can be written as $\mathcal{H} = \{y : \sum_{j=1}^n y_j p_j = 0\}$. Since for each fixed i the vector $a_{i,j}$ is in $\mathcal{L} \subset \mathcal{H}$, (ii) holds. To see that all the $p_j > 0$ we leave it to the reader to check that if not, there would be a non-zero vector in $\mathcal{O}$ that would be in $\mathcal{H}$. $\square$

To apply Theorem 6.1 to our simplified example, we begin by noting that in this case $a_{i,j}$ is given by

		$j=1$	$j=2$
stock	$i=1$	30	-10
option	$i=2$	$20-c$	$-c$

By Theorem 6.1 if there is no arbitrage, then there must be an assignment of probabilities p_j so that

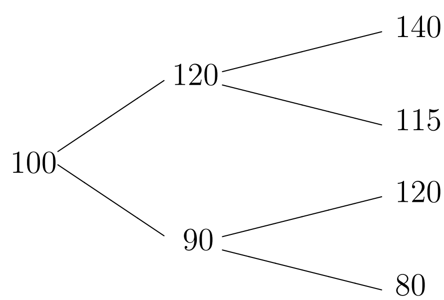
$$30p_1 - 10p_2 = 0 \qquad (20-c)p_1 + (-c)p_2 = 0$$

From the first equation we conclude that $p_1 = 1/4$ and $p_2 = 3/4$. Rewriting the second we have

$$c = 20p_1 = 20 \cdot (1/4) = 5$$

To prepare for the general case note that the equation $30p_1 - 10p_2 = 0$ says that under p_j the stock price is a martingale (i.e., the average value of the change in price is 0), while $c = 20p_1 + 0p_2$ says that the price of the option is then the expected value under the martingale probabilities.

Two-period binary tree. Suppose that a stock price starts at 100 at time 0. At time 1 (one day or one month or one year later) it will either be worth 120 or 90. If the stock is worth 120 at time 1, then it might be worth 140 or 115 at time 2. If the price is 90 at time 1, then the possibilities at time 2 are 120 and 80. The last three sentences can be simply summarized by the following tree.



Using these ideas we can quickly complete the computations in our example. When $X_1 = 120$ the two possible scenarios lead to a change of $+20$ or -5, so the relative probabilities of these two events should be $1/5$ and $4/5$. When $X_0 = 0$ the possible price changes on the first step are $+20$ and -10, so their relative probabilities are $1/3$ and $2/3$. Making a table of the possibilities, we have

X_1	X_2	probability	$(X_2 - 100)^+$
120	140	$(1/3) \cdot (1/5)$	40
120	115	$(1/3) \cdot (4/5)$	15
90	120	$(2/3) \cdot (1/4)$	20
90	80	$(2/3) \cdot (3/4)$	0

so the value of the option is

$$\frac{1}{15} \cdot 40 + \frac{4}{15} \cdot 15 + \frac{1}{6} \cdot 20 = \frac{80 + 120 + 100}{30} = 10$$

The last derivation may seem a little devious, so we will now give a second derivation of the price of the option. In the scenario described above, our investor has four possible actions:

A_0. Put \$1 in the bank and end up with \$1 in all possible scenarios.

A_1. Buy one share of stock at time 0 and sell it at time 1.

A_2. Buy one share at time 1 if the stock is at 120, and sell it at time 2.

A_3. Buy one share at time 1 if the stock is at 90, and sell it at time 2.

These actions produce the following payoffs in the indicated outcomes

X_1	X_2	A_0	A_1	A_2	A_3	option
120	140	1	20	20	0	40
120	115	1	20	-5	0	15
90	120	1	-10	0	30	20
90	80	1	-10	0	-10	0

Noting that the payoffs from the four actions are themselves vectors in four-dimensional space, it is natural to think that by using a linear combination of these actions we can reproduce the option exactly. To find the coefficients $Z - i$ for the actions A_i we write four equations in four unknowns,

$$\begin{aligned} z_0 + 20z_1 + 20z_2 &= 40 \\ z_0 + 20z_1 - 5z_2 &= 15 \\ z_0 - 10z_1 + 30z_3 &= 20 \\ z_0 - 10z_1 - 10z_3 &= 0 \end{aligned}$$

Subtracting the second equation from the first and the fourth from the third gives $25z_2 = 25$ and $40z_3 = 20$ so $z_2 = 1$ and $z_3 = 1/2$. Plugging in these values, we have two equations in two unknowns:

$$z_0 + 20z_1 = 20 \qquad z_0 - 10z_1 = 5$$

Taking differences, we conclude $30z_1 = 15$, so $z_1 = 1/2$ and $z_0 = 10$.

The reader may have already noticed that $z_0 = 10$ is the option price. This is no accident. What we have shown is that with \$10 cash we can buy and sell shares of stock to produce the outcome of the option in all cases. In the terminology of Wall Street, $z_1 = 1/2$, $z_2 = 1$, $z_3 = 1/2$ is a **hedging strategy** that allows us to **replicate the option**. Once we can do this it follows that the fair price must be \$10. To do this note that if we could sell it for \$12 then we can take \$10 of the cash to replicate the option and have a sure profit of \$2.

6.2 Binomial model

In this section we consider the general n period model. As in the previous section we begin with the

One period case. There are two possible outcomes for the stock which we call heads (H) and tails (T). When H occurs the value of the stock is multiplied by u (for 'up'). When T occurs the value of the stock is multiplied by d (for 'down').

$$
\begin{tikzcd}
 & S_1(H) = S_0 u \\
S_0 \arrow[ur, dash] \arrow[dr, dash] & \\
 & S_1(T) = S_0 d
\end{tikzcd}
$$

We assume that there is an interest rate r, which means that \$1 at time 0 is the same as \$$1+r$ at time 1. For the model to be sensible, we need

$$0 < d < 1 + r < u. \tag{6.1}$$

Consider now an option that pays off $V_1(H)$ or $V_1(T)$ at time 1. This could be a call option $(S_1-K)^+$, a put $(K-S_1)^+$, or something more exotic, so we will consider the general case. To find the "no arbitrage price" of this option we suppose we have V_0 in cash and Δ_0 shares of the stock at time 0, and want to pick these to match the option price exactly:

$$V_0 + \Delta_0 \left(\frac{1}{1+r} S_1(H) - S_0 \right) = \frac{1}{1+r} V_1(H) \tag{6.2}$$

$$V_0 + \Delta_0 \left(\frac{1}{1+r} S_1(T) - S_0 \right) = \frac{1}{1+r} V_1(T) \tag{6.3}$$

To find the values of V_0 and Δ_0 we define the risk neutral probability p^* so that

$$\frac{1}{1+r} \left(p^* S_0 u + (1-p^*) S_0 d \right) = S_0 \tag{6.4}$$

We divide by $1+r$ on the right because those amounts are at time 1. Solving we have

$$p^* = \frac{1+r-d}{u-d} \qquad 1 - p^* = \frac{u-(1+r)}{u-d} \tag{6.5}$$

The condition (6.1) implies $0 < p^* < 1$.

Taking $p^*(6.2) + (1 - p^*)(6.3)$ and using (6.4) we have

$$V_0 = \frac{1}{1+r}\left(p^* V_1(H) + (1-p^*)V_1(T)\right) \tag{6.6}$$

i.e., the value is the discounted expected value under the risk neutral probabilities. Taking the difference (6.2) − (6.3) we have

$$\Delta_0 \left(\frac{1}{1+r}(S_1(H) - S_1(T)) \right) = \frac{1}{1+r}(V_1(H) - S_1(T))$$

which implies that

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} \tag{6.7}$$

n period model. To solve the problem in general we work backwards from the end, repeatedly applying the solution of the one period problems. Let a be a string of H's and T's of length $n-1$ which represents the outcome of the first $n-1$ events. The value of the option at time n after the events in a have occurred, $V_n(a)$, and the amount of stock we need to hold in this situation, $\Delta_n(a)$, in order to replicate the option payoff satisfy:

$$V_n(a) + \Delta_n(a)\left(\frac{1}{1+r} S_{n+1}(aH) - S_n(a) \right) = \frac{1}{1+r} V_{n+1}(aH) \tag{6.8}$$

$$V_n(a) + \Delta_n(a)\left(\frac{1}{1+r} S_{n+1}(aT) - S_n(a) \right) = \frac{1}{1+r} V_{n+1}(aT) \tag{6.9}$$

Define the risk neutral probability $p_n^*(a)$ so that

$$S_n(a) = \frac{1}{1+r}[p_n^*(a) S_{n+1}(aH) + (1-p^*) S_{n+1}(aT)] \tag{6.10}$$

Here we allow the stock prices to general, subject only to the restriction that $0 < p_n^*(a) < 1$. Notice that these probabilities depend on the time n and the history a. In the binommial model one has $p_n^*(a) = (1 + r - d)/(u - d)$.

Taking $p_n^*(a)(6.8) + (1 - p_n^*(a))(6.9)$ and using (6.10) we have

$$V_n(a) = \frac{1}{1+r}[p_n^*(a) V_{n+1}(aH) + (1 - p_n^*(a)) V_{n+1}(aT)] \tag{6.11}$$

i.e., the value is the discounted expected value under the risk neutral probabilities. Taking the difference (6.8) − (6.9) we have

$$\Delta_n(a)\left(\frac{1}{1+r}(S_1(H)-S_1(T))\right)=\frac{1}{1+r}(V_1(H)-S_1(T))$$

which implies that

$$\Delta_n(a)=\frac{V_{n+1}(aH)-V_{n+1}(aT)}{S_{n+1}(aH)-S_{n+1}(aT)} \tag{6.12}$$

Turning to examples, we will often use the following binomial model because it leads to easy arithmetic

$$u=2, \qquad d=1/2, \qquad r=1/4 \tag{6.13}$$

The risk neutral probabilities

$$p^*2+(1-p^*)(1/2)=1+\frac{1}{4}$$

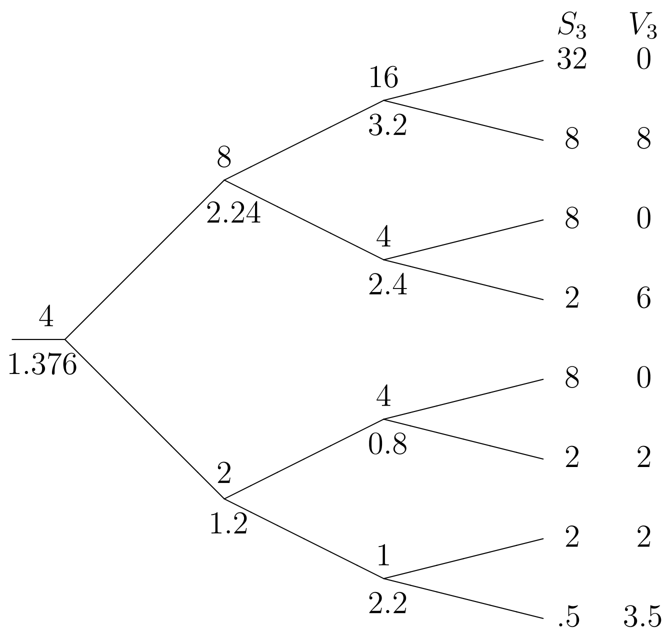
so $p^*=1-p^*=1/2$ and by (6.11) the option prices follow the recursion:

$$V_{n-1}(a)=.4[V_n(aH)+V_n(aT)] \tag{6.14}$$

Example 6.1. Lookback options. In this option you can buy the stock at time 3 at its current price and then sell it at the highest price seen in the past for a profit of

$$V_3=\max_{0\le n\le 3} S_m - S_3$$

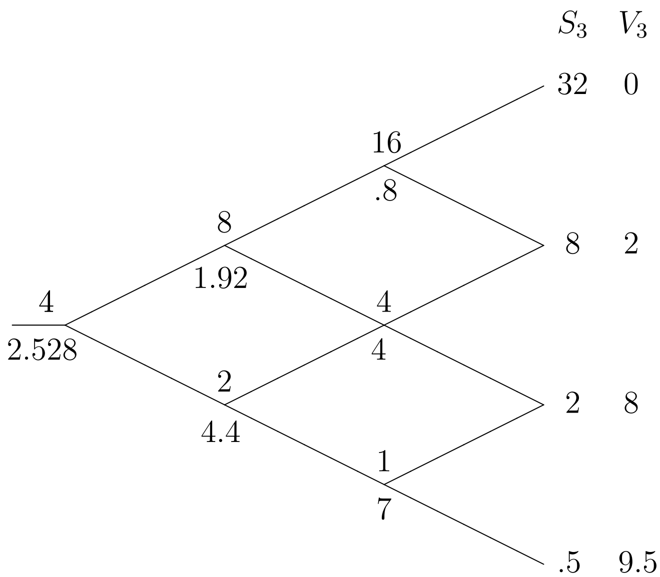
Our goal is to compute the value for this option in the binomial model given in (6.13) with $S_0=4$. For example $S_3(HTT)=2$ but the maximum in the past is $8=S_1(H)$ so $V_3(HTT)=8-2=6$. At the right edge of the picture the stock and option prices are given for time 3.



On the tree, stock prices are above the nodes and option prices below. To explain the computation of the option price note that by (6.14).

$$
\begin{aligned}
V_2(HH) &= 0.4(V_3(HHH) + V_3(HHT)) = 0.4(0 + 8) = 3.2 \\
V_2(HT) &= 0.4(V_3(HTH) + V_3(HTT)) = 0.4(0 + 6) = 2.4 \\
V_1(H) &= 0.4(V_2(HH) + V_2(HT)) = 0.4(3.2 + 2.4) = 2.24
\end{aligned}
$$

Example 6.2. Put option. We will use the binomial model in (6.13) with $S_0 = 4$, but now consider the put option with value $V_3 = (10 - S_3)^+$. The value of this option depends only on the price so we can reduce the tree considered above to:



On the tree itself stock prices are above the nodes and option prices below. To explain the computation of the option price note that by (6.14).

$$\begin{aligned}
V_2(1) &= 0.4(V_3(2) + V_3(0.5) = 0.4(8 + 9.5) = 7 \\
V_2(4) &= 0.4(V_3(8) + V_3(2)) = 0.4(2 + 8) = 4 \\
V_1(2) &= 0.4(V_2(4) + V_2(1)) = 0.4(4 + 7) = 4.4
\end{aligned}$$

The computation of the option price in the last case can be speeded up by observing that

Theorem 6.2. *In the binomial model, under the risk neutral probability measure $M_n = S_n/(1+r)^n$ is a martingale with respect to S_n.*

Proof. Let p^* and $1 - p^*$ be defined by (6.5). Given a string a of heads and tails of length n

$$P^*(a) = (p^*)^{H(a)}(1 - p^*)^{T(a)}$$

where $H(a)$ and $T(a)$ are the number of heads and tails in a. To check the martingale property we need to show that

$$E^*\left(\frac{S_{n+1}}{(1+r)^{n+1}} \,\middle|\, S_0 = s_0, \dots S_n = s_n \right) = \frac{S_n}{(1+r)^n}$$

where E^* indicates expected value with respect to P^*. Letting $X_{n+1} = S_{n+1}/S_n$ which is independent of S_n and is u with probability p^* and d with probability $1 - p^*$ we have

$$\begin{aligned} &E\left(\frac{S_{n+1}}{(1+r)^{n+1}}\middle| S_0 = s_0, \dots S_n = s_n\right) \\ &= \frac{S_n}{(1+r)^n} E\left(\frac{X_{n+1}}{1+r}\middle| S_0 = s_0, \dots S_n = s_n\right) = \frac{S_n}{(1+r)^n} \end{aligned}$$

since $EX_{n+1} = 1 + r$ by (6.10). □

Notation. To make it easier to write computations like the last one we will let

$$E_n(Y) = E(Y|S_0 = s_0, \dots S_n = s_n)$$

or in words, the conditional expectation of Y given the information at time n.

From Theorem 6.2 and the recursion for the option price (6.11) we immediately get:

Theorem 6.3. *In the binomial model we have* $V_0 = E^*(V_n/(1+r)^n)$.

Using this on Example 6.2 gives

$$V_0 = (4/5)^3 \cdot \left[2 \cdot \frac{3}{8} + 8 \cdot \frac{3}{8} + 9.5 \cdot \frac{1}{8}\right] = 2.528$$

Using this on Example 6.1 gives

$$V_0 = (4/5)^3 \cdot \frac{1}{8} \cdot [0 + 8 + 0 + 6 + 0 + 2 + 2 + 2 + 3.5] = 1.376$$

The option prices we have defined were motivated by the idea that by trading in the stock we could replicate the option exactly and hence they are the only price consistent with the absence of arbitrage. We will now go through the algebra needed to demonstrate this for the general n period model. Suppose we start with W_0 dollars and hold $\Delta_n(a)$ shares of stock between time n and $n + 1$. If we invest the money not in the stock in the money market account which pays interest r per period our wealth satisfies the recursion:

$$W_{n+1} = \Delta_n S_{n+1} + (1 + r)(W_n - \Delta_n S_n) \tag{6.15}$$

Theorem 6.4. *If $W_0 = V_0$ and we use the investment strategy in (6.12) then we have $W_n = V_n$.*

Proof. We proceed by induction. By assumption the result is true when $n = 0$. Let a be a string of H and T of length n. (6.15) implies

$$\begin{aligned} W_{n+1}(aH) &= \Delta_n(a)S_{n+1}(aH) + (1+r)(W_n(a) - \Delta_n(a)S_n(a)) \\ &= (1+r)W_n(a) + \Delta_n(a)[S_{n+1} - (1+r)S_n(a)] \end{aligned}$$

By induction the first term $= (1+r)V_n(a)$. Letting $q_n^*(a) = 1 - p_n^*(a)$, (6.10) implies

$$(1+r)S_n(a) = p_n^*(a)S_{n+1}(aH) + q_n^*(a)S_{n+1}(aT)$$

Subtracting this equation from $S_{n+1}(aH) = S_{n+1}(aH)$ we have

$$S_{n+1}(aH) - (1+r)S_n(a) = q_n^*(a)[S_{n+1}(aH) - S_{n+1}(aT)]$$

Using (6.12) now, we have

$$\Delta_n(a)[S_{n+1} - (1+r)S_n(a)] = q_n^*(a)[V_n(aH) - V_{n+1}(aT)]$$

Combining our results then using (6.11)

$$\begin{aligned} W_{n+1}(aH) &= (1+r)V_n(a) + q_n^*(a)[V_n(aH) - V_{n+1}(aT)] \\ &= p_n^*(a)V_n(aH) + q_n^*V_n(aT) + q_n^*(a)[V_n(aH) - V_{n+1}(aT)] = V_{n+1}(aH) \end{aligned}$$

The proof that $W_{n+1}(aT) = V_{n+1}(aT)$ is almost identical. $\square$

6.3 Black-Scholes formula

Many options take place over a time period of one or more months, so it is natural consider S_t to be the stock price after t years. We could use a binomial model in which prices change at the end of each day but it would also be natural to update prices several times during the day. Let h be the amount of time measured in years between updates of the stock price. This h will be very small e.g., $1/365$ for daily updates so it is natural to let $h \to 0$. Knowing what will happen when we take the limit we will let

$$S_{nh} = S_{(n-1)h} \exp(\mu h + \sigma\sqrt{h} X_n)$$

where $P(X_n = 1) = P(X_n = -1) = 1/2$. This is binomial model with a carefully chosen u and d. Iterating we see that

$$S_{nh} = S_0 \exp\left(\mu nh + \sigma\sqrt{h}\sum_{m=1}^{n} X_m\right) \tag{6.16}$$

If we let $t = nh$ the first term is just μt. Writing $h = t/n$ the second term becomes

$$\sigma\sqrt{t} \cdot \frac{1}{\sqrt{n}}\sum_{m=1}^{n} X_m$$

To take the limit as $n \to \infty$, we use the

Theorem 6.5. Central Limit Theorem. *Let $X_1, X_2, \ldots$ be i.i.d. with $EX_i = 0$ and $var(X_i) = 1$ Then for all x we have*

$$P\left(\frac{1}{\sqrt{n}}\sum_{m=1}^{n} X_m \le x\right) \to P(\chi \le x) \tag{6.17}$$

where χ has a standard normal distribution. That is,

$$P(\chi \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$

The conclusion in (6.17) is often written as

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n} X_m \Rightarrow \chi$$

where $\Rightarrow$ is read "converges in distribution to." Recalling that if we multiply a standard normal χ by a constant c then the result has a normal distribution with mean 0 and variance σ^2, we see that

$$\sqrt{t} \cdot \frac{1}{\sqrt{n}} \sum_{m=1}^{n} X_m \Rightarrow \sqrt{t}\chi$$

and the limit is a normal with mean 0 and variance t.

This motivates the following definition:

Definition. $B(t)$ is a standard Brownian motion if $B(0) = 0$ and it satisfies the following conditions:

(a) *Independent increments.* Whenever $0 = t_0 < t_1 < \ldots < t_k$

$$B(t_1) - B(t_0), \ldots, B(t_k) - B(t_{k-1}) \quad \text{are independent.}$$

(b) *Stationary increments.* The distribution of $B_t - B_s$ is normal $(0, t-s)$.

(c) $t \to B_t$ is continuous.

To explain (a) note that if $n_i = t_i/h$ then the sums

$$\sum_{n_{i-1} < m \le n_i} X_m \quad i = 1, \ldots k$$

are independent. For (b) we note that that the distribution of the sum only depends on the number of terms and use the previous calculation. Condition (c) is a natural assumption. From a technical point of view it is need because there are uncountably many values of t, so the finite dimensional distributions described in (a) and (b) do not completely describe the distribution of the process.

Using the new definition, our stock price model can be written as

$$S_t = S_0 \cdot \exp(\mu t + \sigma B_t)$$

where B_t is a standard Brownian motion Here μ is the **exponential growth** rate of the stock, and σ is its **volatility**. If we also assume that the per period interest rate in the approximating model is rh, and recall that

$$\left(\frac{1}{1+rh}\right)^{t/h} = \frac{1}{(1+rh)^{t/h}} \to \frac{1}{e^{rt}} = e^{-rt}$$

then the discounted stock price is

$$e^{-rt}S_t = S_0 \cdot \exp((\mu - r)t + \sigma B_t)$$

Extrapolating wildly from discrete time, we can guess that the option price is its expected value after changing the probabilities to make the stock price a martingale. By the formula for the moment generating function for the normal with mean 0 and variance $\sigma^2 t$

$$E \exp(-(\sigma^2/2)t + \sigma B_t) = 1$$

Since B_t has independent increments, if we let

$$\mu = r - \sigma^2/2 \tag{6.18}$$

then reasoning as for the exponential martingale, Example 5.5, the discounted stock price, $e^{-rt}S_t$ is a martingale.

A more satisfactory approach to this answer comes from using the discrete approximation. Note that

$$u = \exp(\mu h + \sigma\sqrt{h}) \qquad d = \exp(\mu h - \sigma\sqrt{h})$$

The risk neutral probabilities, p_h^*, which depend on the step size h, satisfy

$$p_h^* u + (1 - p_h^*)d = 1 + rh$$

Solving gives

$$p_h^* = \frac{1 + rh - d}{u - d}. \tag{6.19}$$

Recalling that $e^x = 1 + x + x^2/2 + \cdots$,

$$\begin{aligned} u &= 1 + \mu h + \sigma\sqrt{h} + \frac{1}{2}(\mu h + \sigma\sqrt{h})^2 + \ldots \\ &= 1 + \sigma\sqrt{h} + (\sigma^2/2 + \mu)h + \ldots \qquad (6.20) \\ d &= 1 - \sigma\sqrt{h} + (\sigma^2/2 + \mu)h + \ldots \end{aligned}$$

so from (6.19) we have

$$p_h^* \approx \frac{\sigma\sqrt{h} + (r - \mu - \sigma^2/2)h}{2\sigma\sqrt{h}} = \frac{1}{2} + \frac{r - \mu - \sigma^2/2}{2\sigma}\sqrt{h}$$

If $X_1^h, X_2^h, \ldots$ are i.i.d. with

$$P(X_1^h = 1) = p_h^* \qquad P(X_1^h = -1) = 1 - p_h^*$$

then the mean and variance are

$$\begin{aligned} EX_i^h &= \frac{(r-\mu-\sigma^2/2)}{\sigma}\sqrt{h} \\ \mathrm{var}\,(X_i^h) &= 1-(EX_i^h)^2 \to 1 \end{aligned}$$

To apply the central limit theorem we note that

$$\begin{aligned} \sigma\sqrt{h}\sum_{m=1}^{t/h} X_m^h &= \sigma\sqrt{h}\sum_{m=1}^{t/h}(X_m^h - EX_m^h) + \sigma\sqrt{h}\sum_{m=1}^{t/h} EX_m^h \\ &\to \sigma B_t + (r-\mu+\sigma^2/2)t \end{aligned}$$

so under the risk neutral measure, P^*,

$$e^{-rt}S_t = S_0\cdot\exp(-(\sigma^2/2)t+\sigma B_t) \tag{6.21}$$

Using the fact that $\log(S_t/S_0)$ has a normal$(\mu t, \sigma^2 t)$ distribution with $\mu = -\sigma^2/2$, we see that

$$E^*(e^{-rt}(S_t-K)^+) = e^{-rt}\int_{\log(K/S_0)}^{\infty}(S_0e^y-K)\frac{1}{\sqrt{2\pi\sigma^2 t}}e^{-(y-\mu t)^2/2\sigma^2 t}\,dy$$

Splitting the integral into two and then changing variables $y = \mu t + w\sigma\sqrt{t}$, $dy = \sigma\sqrt{t}\,dw$ the integral is equal to

$$= e^{-rt}S_0e^{\mu t}\frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{w\sigma\sqrt{t}}e^{-w^2/2}\,dw - e^{-rt}K\frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{-w^2/2}\,dw \tag{6.22}$$

where $\alpha = (\log(K/S_0)-\mu t)/\sigma\sqrt{t}$. The handle the first term, we note that

$$\begin{aligned} \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{w\sigma\sqrt{t}}e^{-w^2/2}\,dw &= e^{t\sigma^2/2}\int_\alpha^\infty \frac{1}{\sqrt{2\pi}}e^{-(w-\sigma\sqrt{t})^2/2}\,dw \\ &= e^{t\sigma^2/2}\ P(\mathrm{normal}(\sigma\sqrt{t},1)>\alpha) \end{aligned}$$

The last probability can be written in terms of the distribution function Φ of a normal(0,1) χ, i.e., $\Phi(t) = P(\chi \le t)$, by noting

$$\begin{aligned} P(\mathrm{normal}(\sigma\sqrt{t},1)>\alpha) &= P(\chi > \alpha-\sigma\sqrt{t}) \\ &= P(\chi \le \sigma\sqrt{t}-\alpha) = \Phi(\sigma\sqrt{t}-\alpha) \end{aligned}$$

where in the middle equality we have used the fact that χ and $-\chi$ have the same distribution. Using the last two computations in (6.22) converts it to

$$e^{-rt}S_0e^{\mu t}e^{\sigma^2 t/2}\Phi(\sigma\sqrt{t}-\alpha) - e^{-rt}K\Phi(-\alpha)$$

Using $\mu = -\sigma^2/2$ now the expression simplifies to the

Theorem 6.6. Black–Scholes formula. *The price of the European call option $(S_T - K)^+$ is given by*

$$S_0\Phi(\sigma\sqrt{T}-\alpha) - e^{-rT}K\Phi(-\alpha)$$

where $\alpha = \{\log(K/S_0e^{\mu T})\}/\sigma\sqrt{t}$ *and* $\mu = r - \sigma^2/2$.

The Black-Scholes differential equation

The last derivation used special properties of the call option. Suppose now that the payoff at time T is $g(S_T)$. Let $V(t,s)$ be the value of the option at time $t < T$ when the stock price is s. Reasoning with the discrete time approximation and ignoring the fact that the value in this case depends on h,

$$V(t-h,s) = \frac{1}{1+rh}\left[p^*V(t,su) + (1-p^*)V(t,sd)\right]$$

Doing some algebra we have

$$V(t,s)-(1+rh)V(t-s,h) = p^*[V(t,s)-V(t,su)]+(1-p^*)[V(t,s)-V(t,sd)]$$

Dividing by h we have

$$\begin{aligned} &\frac{V(t,s)-V(t-h,s)}{h} - rV(t-h,s) \\ &\quad = p^*\left[\frac{V(t,s)-V(t,su)}{h}\right] + (1-p^*)\left[\frac{V(t,s)-V(t,sd)}{h}\right] \end{aligned} \tag{6.23}$$

Letting $h \to 0$ the left-hand side of (6.23) converges to

$$\frac{\partial V}{\partial t}(t,s) - rV(t,s) \tag{6.24}$$

Expanding $V(t,s)$ in a power series in s

$$\begin{aligned} V(t,su) - V(t,s) &\approx \frac{\partial V}{\partial x}(t,s)(su-s) + \frac{\partial^2 V}{\partial x^2}(t,s)\frac{(su-s)^2}{2} \\ V(t,sd) - V(t,s) &\approx \frac{\partial V}{\partial x}(t,s)(sd-s) + \frac{\partial^2 V}{\partial x^2}(t,s)\frac{(sd-s)^2}{2} \end{aligned}$$

Using the last two equations, the right-hand side of (6.23) is

$$
\begin{aligned}
&= \frac{\partial V}{\partial x}(t,s)s[(1-u)p^* + (1-d)(1-p^*)] \\
&\quad - \frac{1}{2}\frac{\partial^2 V}{\partial x^2}(t,s)s^2[p^*(1-u)^2 + (1-p^*)(1-d)^2]
\end{aligned}
$$

From (6.19) and (6.20) the above is asymptotically

$$
\frac{\partial V}{\partial x}(t,s)s[-rh] + \frac{1}{2}\frac{\partial^2 V}{\partial x^2}(t,s)s^2\sigma^2 h
$$

Combining the last equation with (6.24) and (6.23) we have that the value function satisfies

$$
\frac{\partial V}{\partial t} - rV(t,s) + rs\frac{\partial V}{\partial x}(t,s) + \frac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial x^2}(t,s) = 0 \tag{6.25}
$$

for $0 \leq t < T$ with boundary condition $V(T,s) = g(s)$.

Appendix A

Review of Probability

Here we will review some of the basic facts usually taught in a first course in probability, concentrating on the ones that are used more than once in one of the six other chapters. This chapter may be read for review or skipped and referred to later if the need arises.

A.1 Probabilities, Independence

We begin with a vague but useful definition. (Here and in what follows, **boldface** indicates a word or phrase that is being defined or explained.) The term **experiment** is used to refer to any process whose outcome is not known in advance. Two simple experiments are flip a coin, and roll a die. The **sample space** associated with an experiment is the set of all possible outcomes. The sample space is usually denoted by Ω, the capital Greek letter Omega.

Example A.1. Flip three coins. The flip of one coin has two possible outcomes, called "Heads" and "Tails," and denoted by H and T. Flipping three coins leads to $2^3 = 8$ outcomes:

$$\begin{array}{cccc} & HHT & HTT & \\ HHH & HTH & THT & TTT \\ & THH & TTH & \end{array}$$

Example A.2. Roll two dice. The roll of one die has six possible outcomes: 1, 2, 3, 4, 5, and 6. Rolling two dice leads to $6^2 = 36$ outcomes $\{(m,n) : 1 \le m,n \le 6\}$.

The goal of probability theory is to compute the probability of various events of interest. Intuitively, an event is a statement about the outcome of an experiment. Formally, an **event** is a subset of the sample space. An example for flipping three coins is "two coins show Heads," or

$$A = \{HHT, HTH, THH\}$$

An example for rolling two dice is "the sum is 9," or

$$B = \{(6,3), (5,4), (4,5), (3,6)\}$$

Events are just sets, so we can perform the usual operations of set theory on them. For example, if $\Omega = \{1,2,3,4,5,6\}$, $A = \{1,2,3\}$, and $B = \{2,3,4,5\}$, then the **union** $A \cup B = \{1,2,3,4,5\}$, the **intersection** $A \cap B = \{2,3\}$, and the **complement of** A, $A^c = \{4,5,6\}$. To introduce our next definition, we need one more notion: two events are **disjoint** if their intersection is the empty set, $\emptyset$. A and B are not disjoint, but if $C = \{5,6\}$, then A and C are disjoint.

A **probability** is a way of assigning numbers to events that satisfies:

(i) For any event A, $0 \le P(A) \le 1$.

(ii) If Ω is the sample space, then $P(\Omega) = 1$.

(iii) For a finite or infinite sequence of disjoint events $P(\cup_i A_i) = \sum_i P(A_i)$.

In words, the probability of a union of disjoint events is the sum of the probabilities of the sets. We leave the index set unspecified since it might be finite,

$$P(\cup_{i=1}^k A_i) = \sum_{i=1}^k P(A_i)$$

or it might be infinite, $P(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i)$.

In Examples A.1 and A.2, all outcomes have the same probability, so

$$P(A) = |A|/|\Omega|$$

where $|B|$ is short for the number of points in B. For a very general example, let $\Omega = \{1,2,\ldots,n\}$; let $p_i \ge 0$ with $\sum_i p_i = 1$; and define

$P(A) = \sum_{i \in A} p_i$. Two basic properties that follow immediately from the definition of a probability are

$$P(A) = 1 - P(A^c) \tag{A.1}$$

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) \tag{A.2}$$

To illustrate their use consider the following:

Example A.3. Roll two dice and suppose for simplicity that they are red and green. Let A = "at least one 4 appears," B = "a 4 appears on the red die," and C = "a 4 appears on the green die," so $A = B \cup C$.

Solution 1. A^c = "neither die shows a 4," which contains $5 \cdot 5 = 25$ outcomes so (1.1) implies $P(A) = 1 - 25/36 = 11/36$.

Solution 2. $P(B) = P(C) = 1/6$ while $P(B \cap C) = P(\{4, 4\}) = 1/36$, so (1.2) implies $P(A) = 1/6 + 1/6 - 1/36 = 11/36$.

Conditional probability.

Suppose we are told that the event A with $P(A) > 0$ occurs. Then the sample space is reduced from Ω to A and the probability that B will occur given that A has occurred is

$$P(B|A) = P(B \cap A)/P(A) \tag{A.3}$$

To explain this formula, note that (i) only the part of B that lies in A can possibly occur, and (ii) since the sample space is now A, we have to divide by $P(A)$ to make $P(A|A) = 1$. Multiplying on each side of (A.3) by $P(A)$ gives us the **multiplication rule**:

$$P(A \cap B) = P(A)P(B|A) \tag{A.4}$$

Intuitively, we think of things occurring in two stages. First we see if A occurs, then we see what the probability B occurs given that A did. In many cases these two stages are visible in the problem.

Example A.4. Suppose we draw without replacement from an urn with 6 blue balls and 4 red balls. What is the probability we will get two blue balls? Let A = blue on the first draw, and B = blue on the second draw. Clearly, $P(A) = 6/10$. After A occurs, the urn has 5 blue balls and 4 red balls, so $P(B|A) = 5/9$ and it follows from (1.4) that

$$P(A \cap B) = P(A)P(B|A) = \frac{6}{10} \cdot \frac{5}{9}$$

To see that this is the right answer notice that if we draw two balls without replacement and keep track of the order of the draws, then there are $10 \cdot 9$ outcomes, while $6 \cdot 5$ of these result in two blue balls being drawn.

The multiplication rule is useful in solving a variety of problems. To illustrate its use we consider:

Example A.5. Suppose we roll a four-sided die then flip that number of coins. What is the probability we will get exactly one Heads? Let B = we get exactly one Heads, and A_i = an i appears on the first roll. Clearly, $P(A_i) = 1/4$ for $1 \le i \le 4$. A little more thought gives

$$P(B|A_1) = 1/2, \quad P(B|A_2) = 2/4, \quad P(B|A_3) = 3/8, \quad P(B|A_4) = 4/16$$

so breaking things down according to which A_i occurs,

$$\begin{aligned} P(B) &= \sum_{i=1}^{4} P(B \cap A_i) = \sum_{i=1}^{4} P(A_i)P(B|A_i) \\ &= \frac{1}{4}\left(\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16}\right) = \frac{13}{32} \end{aligned}$$

One can also ask the reverse question: if B occurs, what is the most likely cause? By the definition of conditional probability and the multiplication rule,

$$P(A_i|B) = \frac{P(A_i \cap B)}{\sum_{j=1}^4 P(A_j \cap B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^4 P(A_j)P(B|A_j)} \tag{A.5}$$

This little monster is called **Bayes' formula** but it will not see much action here.

Last but not least, two events A and B are said to be **independent** if $P(B|A) = P(B)$. In words, knowing that A occurs does not change the probability that B occurs. Using the multiplication rule this definition can be written in a more symmetric way as

$$P(A \cap B) = P(A) \cdot P(B) \tag{A.6}$$

Example A.6. Roll two dice and let A = "the first die is 4."

Let B_1 = "the second die is 2." This satisfies our intuitive notion of independence since the outcome of the first dice roll has nothing to do with that of the second. To check independence from (A.6), we note that $P(B_1) = 1/6$ while the intersection $A \cap B_1 = \{(4,2)\}$ has probability 1/36.

$$P(A \cap B_1) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{4}{36} = P(A)P(B_1)$$

Let B_2 = "the sum of the two dice is 3." The events A and B_2 are disjoint, so they cannot be independent:

$$P(A \cap B_2) = 0 < P(A)P(B_2)$$

Let B_3 = "the sum of the two dice is 9." This time the occurrence of A enhances the probability of B_3, i.e., $P(B_3|A) = 1/6 > 4/36 = P(B_3)$, so the two events are not independent. To check that this claim using (A.6), we note that (A.4) implies

$$P(A \cap B_3) = P(A)P(B_3|A) > P(A)P(B_3)$$

Let B_4 = "the sum of the two dice is 7." Somewhat surprisingly, A and B_4 are independent. To check this from (A.6), we note that $P(B_4) = 6/36$ and $A \cap B_4 = \{(4,3)\}$ has probability 1/36, so

$$P(A \cap B_3) = \frac{1}{36} = \frac{1}{6} \cdot \frac{6}{36} = P(A)P(B_3)$$

There are two ways of extending the definition of independence to more than two events.

$A_1, \ldots, A_n$ are said to be **pairwise independent** if for each $i \neq j$, $P(A_i \cap A_j) = P(A_i)P(A_j)$, that is, each pair is independent.

$A_1, \ldots, A_n$ are said to be **independent** if for any $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ we have

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

If we flip n coins and let A_i = "the ith coin shows Heads," then the A_i are independent since $P(A_i) = 1/2$ and for any choice of indices $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ we have $P(A_{i_1} \cap \ldots \cap A_{i_k}) = 1/2^k$. Our next example shows that events can be pairwise independent but not independent.

Example A.7. Flip three coins. Let $A =$ "the first and second coins are the same," $B =$ "the second and third coins are the same," and $C =$ "the third and first coins are the same." Clearly $P(A) = P(B) = P(C) = 1/2$. The intersection of any two of these events is

$$A \cap B = B \cap C = C \cap A = \{HHH, TTT\}$$

an event of probability $1/4$. From this it follows that

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$$

i.e., A and B are independent. Similarly, B and C are independent and C and A are independent; so A, B, and C are pairwise independent. The three events A, B, and C are not independent, however, since $A \cap B \cap C = \{HHH, TTT\}$ and hence

$$P(A \cap B \cap C) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3 = P(A)P(B)P(C)$$

The last example is somewhat unusual. However, the moral of the story is that to show several events are independent, you have to check more than just that each pair is independent.

A.2 Random Variables, Distributions

Formally, a **random variable** is a real-valued function defined on the sample space. However, in most cases the sample space is usually not visible, so we describe the random variables by giving their distributions. In the **discrete case** where the random variable can take on a finite or countably infinite set of values this is usually done using the **probability function**. That is, we give $P(X = x)$ for each value of x for which $P(X = x) > 0$.

Example A.8. Binomial distribution. If we perform an experiment n times and on each trial there is a probability p of success, then the number of successes S_n has

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, \dots, n$$

In words, S_n has a binomial distribution with parameters n and p, a phrase we will abbreviate as $S_n = \text{binomial}(n, p)$.

Example A.9. Geometric distribution. If we repeat an experiment with probability p of success until a success occurs, then the number of trials required, N, has

$$P(N = n) = (1-p)^{n-1} p \qquad \text{for } n = 1, 2, \dots$$

In words, N has a geometric distribution with parameter p, a phrase we will abbreviate as $N = \text{geometric}(p)$.

Example A.10. Poisson distribution. X is said to have a Poisson distribution with parameter $\lambda > 0$, or $X = \text{Poisson}(\lambda)$ if

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \dots$$

To see that this is a probability function we recall

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{A.7}$$

so the proposed probabilities are nonnegative and sum to 1.

In many situations random variables can take any value on the real line or in a certain subset of the real line. For concrete examples, consider the height or weight of a person chosen at random or the time it takes a person to drive from Los Angeles to San Francisco. A random variable X is said to have a **continuous distribution** with **density function** f if for all $a \le b$ we have

$$P(a \le X \le b) = \int_a^b f(x)\,dx \tag{A.8}$$

Geometrically, $P(a \le X \le b)$ is the area under the curve f between a and b.

In order for $P(a \le X \le b)$ to be nonnegative for all a and b and for $P(-\infty < X < \infty) = 1$ we must have

$$f(x) \ge 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{A.9}$$

Any function f that satisfies (A.9) is said to be a **density function**. We will now define three of the most important density functions.

Example A.11. The uniform distribution on (a,b).

$$f(x) = \begin{cases} 1/(b-a) & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

The idea here is that we are picking a value "at random" from (a, b). That is, values outside the interval are impossible, and all those inside have the same probability density. Note that the last property implies $f(x) = c$ for $a < x < b$. In this case the integral is $c(b - a)$, so we must pick $c = 1/(b - a)$.

Example A.12. The exponential distribution.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

Here $\lambda > 0$ is a parameter. To check that this is a density function, we note that

$$\int_0^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{\infty} = 0 - (-1) = 1$$

In a first course in probability, the next example is the star of the show. However, it will have only a minor role here.

Example A.13. The standard normal distribution.

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

Since there is no closed form expression for the antiderivative of f, it takes some ingenuity to check that this is a probability density. Those details are not important here so we will ignore them.

Any random variable (discrete, continuous, or in between) has a **distribution function** defined by $F(x) = P(X \le x)$. If X has a density function $f(x)$ then

$$F(x) = P(-\infty < X \le x) = \int_{-\infty}^{x} f(y)\, dy$$

That is, F is an antiderivative of f.

One of the reasons for computing the distribution function is explained by the next formula. If $a < b$, then $\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$ with the two sets on the right-hand side disjoint so

$$P(X \le b) = P(X \le a) + P(a < X \le b)$$

or, rearranging,

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) \qquad \text{(A.10)}$$

The last formula is valid for any random variable. When X has density function f, it says that

$$\int_a^b f(x)\, dx = F(b) - F(a)$$

i.e., the integral can be evaluated by taking the difference of the antiderivative at the two endpoints.

To see what distribution functions look like, and to explain the use of (A.10), we return to our examples.

Example A.14. The uniform distribution. $f(x) = 1/(b-a)$ for $a < x < b$.

$$F(x) = \begin{cases} 0 & x \le a \\ (x-a)/(b-a) & a \le x \le b \\ 1 & x \ge b \end{cases}$$

To check this, note that $P(a < X < b) = 1$ so $P(X \leq x) = 1$ when $x \geq b$ and $P(X \leq x) = 0$ when $x \leq a$. For $a \leq x \leq b$ we compute

$$P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy = \int_{a}^{x} \frac{1}{b-a}\, dy = \frac{x-a}{b-a}$$

In the most important special case $a = 0$, $b = 1$ we have $F(x) = x$ for $0 \leq x \leq 1$.

Example A.15. The exponential distribution. $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

The first line of the answer is easy to see. Since $P(X > 0) = 1$, we have $P(X \leq x) = 0$ for $x \leq 0$. For $x \geq 0$ we compute

$$P(X \leq x) = \int_{0}^{x} \lambda e^{-\lambda y}\, dy = -e^{-\lambda y}\Big|_{0}^{x} = 1 - e^{-\lambda x}$$

In many situations we need to know the relationship between several random variables $X_1, \ldots, X_n$. If the X_i are discrete random variables then this is easy, we simply give the probability function that specifies the value of

$$P(X_1 = x_1, \ldots, X_n = x_n)$$

whenever this is positive. When the individual random variables have continuous distributions this is described by giving the **joint density function** which has the interpretation that

$$P((X_1, \ldots, X_n) \in A) = \int \cdots \int_A f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

By analogy with (A.9) we must require that $f(x_1, \ldots, x_n) \geq 0$ and

$$\int \cdots \int f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = 1$$

Having introduced the joint distribution of n random variables, we will for simplicity restrict our attention for the rest of the section to $n = 2$, where will typically write $X_1 = X$ and $X_2 = Y$. The first question we will confront is: "Given the joint distribution of (X, Y), how do we recover the distributions of X and Y?" In the discrete case this is easy. The **marginal distributions** of X and Y are given by

$$\begin{aligned} P(X = x) &= \sum_y P(X = x, Y = y) \\ P(Y = y) &= \sum_x P(X = x, Y = y) \end{aligned} \tag{A.11}$$

To explain the first formula in words, if $X = x$, then Y will take on some value y, so to find $P(X = x)$ we sum the probabilities of the disjoint events $\{X = x, Y = y\}$ over all the values of y.

Formula (A.11) generalizes in a straightforward way to continuous distributions: we replace the sum by an integral and the probability functions by density functions. To make the analogy more apparent we will introduce the following:

If X and Y have joint density $f_{X,Y}(x, y)$ then the **marginal densities** of X and Y are given by

$$\begin{aligned} f_X(x) &= \int f_{X,Y}(x, y)\, dy \\ f_Y(y) &= \int f_{X,Y}(x, y)\, dx \end{aligned} \tag{A.12}$$

The verbal explanation of the first formula is similar to that of the discrete case: if $X = x$, then Y will take on some value y, so to find $f_X(x)$ we integrate the joint density $f_{X,Y}(x, y)$ over all possible values of y.

Two random variables are said to be **independent** if for any two sets A and B we have

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \tag{A.13}$$

In the discrete case, (A.13) is equivalent to

$$P(X = x, Y = y) = P(X = x)P(Y = y) \tag{A.14}$$

for all x and y. With our notation the condition for independence is exactly the same in the continuous case, though in that situation we must remember that the formula says that the joint distribution is the product of the marginal densities. In the traditional notation, the condition for continuous random variables is

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \tag{A.15}$$

The notions of independence extend in a straightforward way to n random variables. Using our notation that combines the discrete and the continuous case $X_1, \dots X_n$ are independent if

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n) \quad \text{(A.16)}$$

That is, if the joint probability or probability density is the product of the marginals. Two important consequences of independence are

Theorem A.1. *If $X_1, \dots X_n$ are independent, then*

$$E(X_1 \cdots X_n) = EX_1 \cdots EX_n$$

Theorem A.2. *If $X_1, \dots X_n$ are independent and $n_1 < \dots < n_k \le n$, then*

$$h_1(X_1, \dots X_{n_1}), h_2(X_{n_1+1}, \dots X_{n_2}), \dots h_k(X_{n_{k-1}+1}, \dots X_{n_k})$$

are independent.

In words, the second result says that functions of disjoint sets of independent random variables are independent.

Our last topic in this section is the distribution of $X + Y$ when X and Y are independent. In the discrete case this is easy:

$$P(X + Y = z) = \sum_x P(X = x)P(Y = z - x) \quad \text{(A.17)}$$

To see the first equality, note that if the sum is z then X must take on some value x and Y must be $z - x$. The first equality is valid for any random variables. The second holds since we have supposed X and Y are independent.

Example A.16. If $X = \text{binomial}(n, p)$ and $Y = \text{binomial}(m, p)$ are independent, then $X + Y = \text{binomial}(n + m, p)$.

Proof by direct computation.

$$\begin{aligned}
P(X + Y = i) &= \sum_{j=0}^{i} \binom{n}{j} p^j (1-p)^{n-j} \cdot \binom{m}{i-j} p^{i-j} (1-p)^{m-i+j} \\
&= p^i (1-p)^{n+m-i} \sum_{j=0}^{i} \binom{n}{j} \cdot \binom{m}{i-j} \\
&= \binom{n+m}{i} p^i (1-p)^{n+m-i}
\end{aligned}$$

The last equality follows from the fact that if we pick i individuals from a group of n boys and m girls, which can be done in $\binom{n+m}{i}$ ways, then we must have j boys and $i-j$ girls for some j with $0 \le j \le i$. □

Much easier proof. Consider a sequence of $n+m$ independent trials. Let X be the number of successes in the first n trials and Y be the number of successes in the last m. By (2.13), X and Y independent. Clearly their sum is binomial(n,p). □

Formula (A.17) generalizes in the usual way to continuous distributions: regard the probabilities as density functions and replace the sum by an integral.

$$P(X+Y=z) = \int P(X=x)P(Y=z-x)\,dx \tag{A.18}$$

Example A.17. Let U and V be independent and uniform on $(0,1)$. Compute the density function for $U+V$.

Solution. If $U+V=x$ with $0 \le x \le 1$, then we must have $U \le x$ so that $V \ge 0$. Recalling that we must also have $U \ge 0$

$$f_{U+V}(x) = \int_0^x 1 \cdot 1\,du = x \quad \text{when } 0 \le x \le 1$$

If $U+V=x$ with $1 \le x \le 2$, then we must have $U \ge x-1$ so that $V \le 1$. Recalling that we must also have $U \le 1$,

$$f_{U+V}(x) = \int_{x-1}^1 1 \cdot 1\,du = 2-x \quad \text{when } 1 \le x \le 2$$

Combining the two formulas we see that the density function for the sum is triangular. It starts at 0 at 0, increases linearly with rate 1 until it reaches the value of 1 at $x=1$, then it decreases linearly back to 0 at $x=2$. □

A.3 Expected Value, Moments

If X has a discrete distribution, then the **expected value** of $h(X)$ is

$$Eh(X) = \sum_x h(x)P(X = x) \tag{A.19}$$

When $h(x) = x$ this reduces to EX, the expected value, or **mean of** X, a quantity that is often denoted by μ or sometimes μ_X to emphasize the random variable being considered. When $h(x) = x^k$, $Eh(X) = EX^k$ is the k**th moment**. When $h(x) = (x - EX)^2$,

$$Eh(X) = E(X - EX)^2 = EX^2 - (EX)^2$$

is called the **variance** of X. It is often denoted by $\text{var}\,(X)$ or σ_X^2. The variance is a measure of how spread out the distribution is. However, if X has the units of feet then the variance has units of feet^2, so the **standard deviation** $\sigma(X) = \sqrt{\text{var}\,(X)}$, which has again the units of feet, gives a better idea of the "typical" deviation from the mean than the variance does.

Example A.18. Roll one die. $P(X = x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$ so

$$EX = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6} = \frac{21}{6} = 3\frac{1}{2}$$

In this case the expected value is just the average of the six possible values.

$$EX^2 = (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) \cdot \frac{1}{6} = \frac{91}{6}$$

so the variance is $91/6 - 49/4 = 70/24$. Taking the square root we see that the standard deviation is 1.71. The three possible deviations, in the sense of $|X - EX|$, are 0.5, 1.5, and 2.5 with probability 1/3 each, so 1.71 is indeed a reasonable approximation for the typical deviation from the mean.

Example A.19. Geometric distribution. Suppose

$$P(N = k) = p(1 - p)^{k-1} \quad \text{for } k = 1, 2, \dots$$

Starting with the sum of the geometric series

$$(1 - \theta)^{-1} = \sum_{n=0}^{\infty} \theta^n$$

and then differentiating twice and discarding terms that are 0, gives

$$(1-\theta)^{-2} = \sum_{n=1}^{\infty} n\theta^{n-1} \quad \text{and} \quad 2(1-\theta)^{-3} = \sum_{n=2}^{\infty} n(n-1)\theta^{n-2}$$

Using these with $\theta = 1 - p$, we see that

$$EN = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = p/p^2 = \frac{1}{p}$$

$$EN(N-1) = \sum_{n=2}^{\infty} n(n-1)(1-p)^{n-1}p = 2p^{-3}(1-p)p = \frac{2(1-p)}{p^2}$$

and hence

$$\begin{aligned} \text{var}\,(N) &= EN(N-1) + EN - (EN)^2 \\ &= \frac{2(1-p)}{p^2} + \frac{p}{p^2} - \frac{1}{p^2} = \frac{(1-p)}{p^2} \end{aligned}$$

The definition of expected value generalizes in the usual way to continuous random variables. We replace the probability function by the density function and the sum by an integral

$$Eh(X) = \int h(x) f_X(x)\,dx \tag{A.20}$$

Example A.20. Uniform distribution on (a,b). Suppose X has density function $f_X(x) = 1/(b-a)$ for $a < x < b$ and 0 otherwise. In this case

$$EX = \int_a^b \frac{x}{b-a}\,dx = \frac{b^2 - a^2}{2(b-a)} = \frac{(b+a)}{2}$$

since $b^2 - a^2 = (b-a)(b+a)$. Notice that $(b+a)/2$ is the midpoint of the interval and hence the natural choice for the average value of X. A little more calculus gives

$$EX^2 = \int_a^b \frac{x^2}{b-a}\,dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ba + a^2}{3}$$

since $b^3 - a^3 = (b-a)(b^2 + ba + a^2)$. Squaring our formula for EX gives $(EX)^2 = (b^2 + 2ab + a^2)/4$, so

$$\text{var}\,(X) = (b^2 - 2ab + a^2)/12 = (b-a)^2/12$$

Example A.21. Exponential distribution. Suppose X has density function $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise. To compute the expected value, we use the integration by parts formula:

$$\int_a^b g(x)h'(x)\,dx = g(x)h(x)\big|_a^b - \int_a^b g'(x)h(x)\,dx \tag{A.21}$$

with $g(x) = x$ and $h'(x) = \lambda e^{-\lambda x}$. Since $g'(x) = 1$, $h(x) = -e^{-\lambda x}$ (A.21) implies

$$\begin{aligned} EX &= \int_0^\infty x\,\lambda e^{-\lambda x}\,dx \\ &= -xe^{-\lambda x}\big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = 0 + 1/\lambda \end{aligned}$$

where to evaluate the last integral we used $\int_0^\infty \lambda e^{-\lambda x}\,dx = 1$. The second moment is computed similarly

$$\begin{aligned} EX^2 &= \int_0^\infty x^2\,\lambda e^{-\lambda x}\,dx \\ &= -x^2e^{-\lambda x}\big|_0^\infty + \int_0^\infty 2xe^{-\lambda x}\,dx = 0 + 2/\lambda^2 \end{aligned}$$

where we have used the result for the mean to evaluate the last integral. The variance is thus

$$\text{var}\,(X) = EX^2 - (EX)^2 = 1/\lambda^2$$

To help explain the answers we have found in the last two examples we use

Theorem A.3. *If c is a real number, then*

$$\begin{array}{llll} (a) & E(X+c) = EX + c & (b) & var\,(X+c) = \,var\,(X) \\ (c) & E(cX) = cEX & (d) & var\,(cX) = c^2\,var\,(X) \end{array}$$

Uniform distribution on (a,b). If X is uniform on $[(a-b)/2, (b-a)/2]$ then $EX = 0$ by symmetry. If $c = (a+b)/2$, then $Y = X + c$ is uniform on $[a, b]$, so it follows from (a) and (b) of Theorem A.3 that

$$EY = EX + c = (a+b)/2 \qquad \text{var}\,(Y) = \text{var}\,(X)$$

From the second formula we see that the variance of the uniform distribution will only depend on the length of the interval. To see that it will be a multiple of $(b-a)^2$ note that $Z = X/(b-a)$ is uniform on $[-1/2, 1/2]$ and then use part (d) of Theorem A.3 to conclude $\text{var}(X) = (b-a)^2 \text{var}(Z)$. Of course one needs calculus to conclude that $\text{var}(Z) = 1/12$.

Exponential distribution. Changing variables $x = y/\lambda$, $\lambda\, dx = dy$,

$$\int_0^\infty x^k \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda^k} \int_0^\infty y^k e^{-y}\, dy$$

What underlies this relationship between the moments is the fact that if Y has an exponential(1) distribution and $X = Y/\lambda$ then X has an exponential(λ) distribution. Using Theorem A.3 now, it follows that

$$EX = EY/\lambda \quad \text{and} \quad \text{var}(X) = \text{var}(Y)/\lambda^2$$

Again, we have to resort to calculus to show that $EY = 1$ and $EY^2 = 1$ but the scaling relationship tells us the dependence of the answer on λ.

The next two results give important properties of expected value and variance.

Theorem A.4. *If $X_1, \ldots, X_n$ are any random variables, then*

$$E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n$$

Theorem A.5. *If $X_1, \ldots, X_n$ are independent, then*

$$var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n)$$

To illustrate the use of these properties we consider the

Example A.22. Binomial distribution. If we perform an experiment n times and on each trial there is a probability p of success, then the number of successes S_n has

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, \ldots, n$$

To compute the mean and variance we begin with the case $n = 1$. Writing X instead of S_1 to simplify notation, we have $P(X = 1) = p$

and $P(X = 0) = 1 - p$, so

$$\begin{aligned} EX &= p \cdot 1 + (1-p) \cdot 0 = p \\ EX^2 &= p \cdot 1^2 + (1-p) \cdot 0^2 = p \\ \text{var}\,(X) &= EX^2 - (EX)^2 = p - p^2 = p(1-p) \end{aligned}$$

To compute the mean and variance of S_n, we observe that if $X_1, \ldots, X_n$ are independent and have the same distribution as X, then $X_1 + \cdots + X_n$ has the same distribution as S_n. Intuitively, this holds since $X_i = 1$ means one success on the ith trial so the sum counts the total number of success. Using Theorems A.4 and A.5, we have

$$ES_n = nEX = np \qquad \text{var}\,(S_n) = n\,\text{var}\,(X) = np(1-p)$$

In some cases an alternate approach to computing the expected value of X is useful. In the discrete case the formula is

Theorem A.6. *If $X \geq 0$ is integer valued then*

$$EX = \sum_{k=1}^{\infty} P(X \geq k) \tag{A.22}$$

Proof. Let $1_{\{X \geq k\}}$ denote the random variable that is 1 if $X \geq k$ and 0 otherwise. It is easy to see that

$$X = \sum_{k=1}^{\infty} 1_{\{X \geq k\}}.$$

Taking expected values and noticing $E1_{\{X \geq k\}} = P(X \geq k)$ gives

$$EX = \sum_{k=1}^{\infty} P(X \geq k)$$

which proves the desired result. □

The analogous result for the continuous case is:

Theorem A.7. *Let $X \geq 0$. Let H be a differentiable nondecreasing function with $H(0) = 0$. Then*

$$EH(X) = \int_0^{\infty} H'(t) P(X > t)\, dt$$

Proof. We assume H is nondecreasing only to make sure that the integral exists. (It may be ∞.) Introducing the indicator $1_{\{X>t\}}$ that is 1 if $X > t$ and 0 otherwise, we have

$$\int_0^\infty H'(t) 1_{\{X>t\}} = \int_0^X H'(t)\,dt = H(X)$$

and taking expected value gives the desired result. □

Taking $H(x) = x^p$ with $p > 0$ we have

$$EX^p = \int_0^\infty p t^{p-1} P(X > t)\,dt \tag{A.23}$$

When $p = 1$ this becomes

$$EX = \int_0^\infty P(X > t)\,dt \tag{A.24}$$

the analogue to (A.23) in the discrete case is

$$EX^p = \sum_{k=1}^\infty (k^p - (k-1)^p) P(X \geq k) \tag{A.25}$$

When $p = 1, 2$ this becomes

$$EX = \sum_{k=1}^\infty P(X \geq k) \tag{A.26}$$

$$EX^2 = \sum_{k=1}^\infty (2k-1) P(X \geq k) \tag{A.27}$$