Bruce K. Driver

Math 180C (Introduction to Probability) Notes

June 6, 2008 *File:180Notes.tex*

Contents

Part Math 180C

s

Part

Math 180C

0

Math 180C Homework Problems

The problems from Karlin and Taylor are referred to using the conventions. 1) II.1: E1 refers to Exercise 1 of section 1 of Chapter II. While II.3: P4 refers to Problem 4 of section 3 of Chapter II.

0.1 Homework #1 (Due Monday, April 7)

Exercise 0.1 (2nd order recurrence relations). Let a, b, c be real numbers with $a \neq 0 \neq c$ and suppose that $\{y_n\}_{n=-\infty}^{\infty}$ solves the second order homogeneous recurrence relation:

$$a y_{n+1} + b y_n + c y_{n-1} = 0. \tag{0.1}$$

Show:

1. for any $\lambda \in \mathbb{C}$,
$$a\lambda^{n+1} + b\lambda^n + c\lambda^{n-1} = \lambda^{n-1} p(\lambda) \tag{0.2}$$
where $p(\lambda) = a\lambda^2 + b\lambda + c$.
2. Let $\lambda_\pm = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ be the roots of p and suppose for the moment that $b^2 - 4ac \neq 0$. Show
$$y_n := A_+ \lambda_+^n + A_- \lambda_-^n$$
solves Eq. (0.1) for any choice of A_+ and A_-.
3. Now suppose that $b^2 = 4ac$ and $\lambda_0 := -b/(2a)$ is the double root of $p(\lambda)$. Show that
$$y_n := (A_0 + A_1 n) \lambda_0^n$$
solves Eq. (0.1) for any choice of A_0 and A_1. **Hint:** Differentiate Eq. (0.2) with respect to λ and then set $\lambda = \lambda_0$.
4. Show that every solution to Eq. (0.1) is of the form found in parts 2. and 3.

In the next couple of exercises you are going to use first step analysis to show that a simple unbiased random walk on $\mathbb{Z}$ is null recurrent. We let $\{X_n\}_{n=0}^{\infty}$ be the Markov chain with values in $\mathbb{Z}$ with transition probabilities given by

$$P(X_{n+1} = j \pm 1 | X_n = j) = 1/2 \text{ for all } n \in \mathbb{N}_0 \text{ and } j \in \mathbb{Z}.$$

Further let $a, b \in \mathbb{Z}$ with $a < 0 < b$ and

$$T_{a,b} := \min\{n : X_n \in \{a, b\}\} \text{ and } T_b := \inf\{n : X_n = b\}.$$

We know by Proposition 3.15 that $\mathbb{E}_0[T_{a,b}] < \infty$ from which it follows that $P(T_{a,b} < \infty) = 1$ for all $a < 0 < b$.

Exercise 0.2. Let $w_j := P_j\left(X_{T_{a,b}} = b\right) := P\left(X_{T_{a,b}} = b | X_0 = j\right)$.

1. Use first step analysis to show for $a < j < b$ that
$$w_j = \frac{1}{2}(w_{j+1} + w_{j-1}) \tag{0.3}$$
provided we define $w_a = 0$ and $w_b = 1$.
2. Use the results of Exercise 0.1 to show
$$P_j\left(X_{T_{a,b}} = b\right) = w_j = \frac{1}{b-a}(j-a). \tag{0.4}$$
3. Let
$$T_b := \begin{cases} \min\{n : X_n = b\} & \text{if} \quad \{X_n\} \text{ hits } b \\ \infty & \text{otherwise} \end{cases}$$
be the first time $\{X_n\}$ hits b. Explain why, $\left\{X_{T_{a,b}} = b\right\} \subset \{T_b < \infty\}$ and use this along with Eq. (0.4) to conclude that $P_j(T_b < \infty) = 1$ for all $j < b$. (By symmetry this result holds true for all $j \in \mathbb{Z}$.)

Exercise 0.3. The goal of this exercise is to give a second proof of the fact that $P_j(T_b < \infty) = 1$. Here is the outline:

1. Let $w_j := P_j(T_b < \infty)$. Again use first step analysis to show that w_j satisfies Eq. (0.3) for all j with $w_b = 1$.
2. Use Exercise 0.1 to show that there is a constant, c, such that
$$w_j = c(j-b) + 1 \text{ for all } j \in \mathbb{Z}.$$
3. Explain why c must be zero to again show that $P_j(T_b < \infty) = 1$ for all $j \in \mathbb{Z}$.

Exercise 0.4. Let $T = T_{a,b}$ and $u_j := \mathbb{E}_j T := \mathbb{E}[T | X_0 = j]$.

1. Use first step analysis to show for $a < j < b$ that
$$u_j = \frac{1}{2}\left(u_{j+1} + u_{j-1}\right) + 1 \tag{0.5}$$
with the convention that $u_a = 0 = u_b$.
2. Show that
$$u_j = A_0 + A_1 j - j^2 \tag{0.6}$$
solves Eq. (0.5) for any choice of constants A_0 and A_1.
3. Choose A_0 and A_1 so that u_j satisfies the boundary conditions, $u_a = 0 = u_b$. Use this to conclude that
$$\mathbb{E}_j T_{a,b} = -ab + (b+a)\, j - j^2 = -a\,(b-j) + bj - j^2. \tag{0.7}$$

Remark 0.1. Notice that $T_{a,b} \uparrow T_b = \inf\{n : X_n = b\}$ as $a \downarrow -\infty$, and so passing to the limit as $a \downarrow -\infty$ in Eq. (0.7) shows
$$\mathbb{E}_j T_b = \infty \text{ for all } j < b.$$

Combining the last couple of exercises together shows that $\{X_n\}$ is null - recurrent.

Exercise 0.5. Let $T = T_b$. The goal of this exercise is to give a second proof of the fact and $u_j := \mathbb{E}_j T = \infty$ for all $j \neq b$. Here is the outline. Let $u_j := \mathbb{E}_j T \in [0,\infty] = [0,\infty) \cup \{\infty\}$.

1. Note that $u_b = 0$ and, by a first step analysis, that u_j satisfies Eq. (0.5) for all $j \neq b$ – allowing for the possibility that some of the u_j may be infinite.
2. Argue, using Eq. (0.5), that if $u_j < \infty$ for some $j < b$ then $u_i < \infty$ for all $i < b$. Similarly, if $u_j < \infty$ for some $j > b$ then $u_i < \infty$ for all $i > b$.
3. If $u_j < \infty$ for all $j > b$ then u_j must be of the form in Eq. (0.6) for some A_0 and A_1 in $\mathbb{R}$ such that $u_b = 0$. However, this would imply, $u_j = \mathbb{E}_j T \to -\infty$ as $j \to \infty$ which is impossible since $\mathbb{E}_j T \geq 0$ for all j. Thus we must conclude that $\mathbb{E}_j T = u_j = \infty$ for all $j > b$. (A similar argument works if we assume that $u_j < \infty$ for all $j < b$.)

0.2 Homework #2 (Due Monday, April 14)

- IV.1 (p. 208 –): E5, E8, P1, P5
- IV.3 (p. 243 –): E1, E2, E3,
- IV.4 (p.254 –): E2

0.3 Homework #3 (Due Monday, April 21)

Exercises 0.6 – 0.9 refer to the following Markov matrix:

$$P := \begin{array}{cc} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} & \\ \left[\begin{array}{cccccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 3/4 & 0 \end{array}\right] & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \end{array} \tag{0.8}$$

We will let $\{X_n\}_{n=0}^{\infty}$ denote the Markov chain associated to P.

Exercise 0.6. Make a jump diagram for this matrix and identify the recurrent and transient classes. Also find the invariant destitutions for the chain restricted to each of the recurrent classes.

Exercise 0.7. Find all of the invariant distributions for P.

Exercise 0.8. Compute the hitting probabilities, $h_5 = P_5\left(X_n \text{ hits } \{3,4\}\right)$ and $h_6 = P_6\left(X_n \text{ hits } \{3,4\}\right)$.

Exercise 0.9. Find $\lim_{n\to\infty} P_6\left(X_n = j\right)$ for $j = 1,2,3,4,5,6$.

Exercise 0.10. Suppose that $\{T_k\}_{k=1}^{n}$ are independent exponential random variables with parameters $\{q_k\}_{k=1}^{n}$, i.e. $P\left(T_k > t\right) = e^{-q_k t}$ for all $t \geq 0$. Show that $T := \min\left(T_1, T_2, \dots, T_n\right)$ is again an exponential random variable with parameter $q = \sum_{k=1}^{n} q_k$.

Exercise 0.11. Let $\{T_k\}_{k=1}^{n}$ be as in Exercise 0.11. Since these are continuous random variables, $P\left(T_k = T_j\right) = 0$ for all $k \neq j$, i.e. there is no chance that any two of the $\{T_k\}_{k=1}^{n}$ are the same.

Find
$$P\left(T_1 < \min\left(T_2, \dots, T_n\right)\right).$$

Hints: 1. Let $S := \min\left(T_2, \dots, T_n\right)$, 2. write $P\left(T_1 < \min\left(T_2, \dots, T_n\right)\right) = \mathbb{E}\left[1_{T_1 < S}\right]$, 3. use Proposition 1.16 above.

Exercise 0.12. Consider the "pure birth" process with constant rates, $\lambda > 0$. In this case $S = \{0,1,2,\dots\}$ and if $\pi = \left(\pi_0, \pi_1, \pi_2, \dots\right)$ is a given initial distribution. In this case one may show that $\pi(t)$, satisfies the system of differential equations:

$$\begin{aligned}
\dot{\pi}_0(t) &= -\lambda\pi_0(t) \\
\dot{\pi}_1(t) &= \lambda\pi_0(t) - \lambda\pi_1(t) \\
\dot{\pi}_2(t) &= \lambda\pi_1(t) - \lambda\pi_2(t) \\
&\vdots \\
\dot{\pi}_n(t) &= \lambda\pi_{n-1}(t) - \lambda\pi_n(t) \\
&\vdots
\end{aligned}$$

Show that the solution to these equations are given by

$$\begin{aligned}
\pi_0(t) &= \pi_0 e^{-\lambda t} \\
\pi_1(t) &= e^{-\lambda t}(\pi_0\lambda t + \pi_1) \\
\pi_2(t) &= e^{-\lambda t}\left(\pi_0\frac{(\lambda t)^2}{2!} + \pi_1\lambda t + \pi_2\right) \\
&\vdots \\
\pi_n(t) &= e^{-\lambda t}\left(\sum_{k=0}^{n}\pi_{n-k}\frac{(\lambda t)^k}{k!}\right) \\
&\vdots
\end{aligned}$$

Note: There are two ways to do this problem. The first and more interesting way is to derive the solutions using Lemma 6.14. The second is to check that the given functions satisfy the differential equations.

0.4 Homework #4 (Due Monday, April 28)

- VI.1 (p. 342 –): E1, E2, E5, P3, P5*, P8**
- VI.2 (p. 353 –): E1, P2***

* Please show that W1 and W2 - W1 are independent exponentially distributed random variables by computing P(W1 > t and W2 - W1 > s) for all s,t>0.

**Hint: you can save some work using what we already have seen about two state Markov chains, see the notes or sections VI.3 or VI.6 of the book.

*** Depending on how you choose to do this problem you may find Lemma 2.7 in the lecture notes useful.

0.5 Homework #5 (Due Monday, May 5)

- VI.2 (p. 353 –): P2.3 (**Hint:** look at the picture on page 345 to find an expression for the area in terms of the $\{S_k\}_{k=1}^N$.)
- VI.3 (p. 365 –): E3.1, E3.3, P3.3, P3.4
- VI.4 (p. 377 –):E4.2, P4.1

Test #1 is on Friday May 9

0.6 Homework #6 (Due Monday, May 12)

- VI.4 (p. 377 –): P4.3
- VI.5 (p. 392 –): P5.2
- VI.6 (p. 405-): P6.2

0.7 Homework #7 (Due Monday, May 19)

- VII.1 (p. 424-426): Ex. 1.2, 1.3; Pr. 1.1, 1.3
- VII.2 (p. 431-432): Ex. 2.1
- VII.3 (p. 435-437): Ex. 3.1*, 3.3*; Pr. 3.2

* **Hint.** Write the event $\{N(t) = n \text{ and } W_{N(t)+1} > t + s\}$ purely in terms of the Poisson process, N. Then use your knowledge of N in order to do the computations. Use facts you know about Poisson processes and make use of Ex. 3.1.

0.8 Homework #8 (Due Monday, June 2)

- VII.3 (p. 435-437): Pr. 3.4
- VII.4 (p. 445-447): Ex. 4.2, 4.3, 4.5; Pr. 4.1, 4.5
- VII.5 (p. 455-457): Ex. 5.1; Pr. 5.1, 5.4
- [VIII.1 (p. 487-491): Ex. 1.1, 1.4, 1.5; Pr. 1.1: **These have been removed from the assignment.**]

1

Independence and Conditioning

Definition 1.1. *We say that an event, A, is independent of an event, B, iff $P(A|B) = P(A)$ or equivalently that*

$$P(A \cap B) = P(A)\, P(B).$$

We further say a collection of events $\{A_j\}_{j\in J}$ are independent iff

$$P(\cap_{j\in J_0} A_j) = \prod_{j\in J_0} P(A_j)$$

for any finite subset, J_0, of J.

Lemma 1.2. *If $\{A_j\}_{j\in J}$ is an independent collection of events then so is $\left\{A_j, A_j^c\right\}_{j\in J}$.*

Proof. First consider the case of two independent events, A and B. By assumption, $P(A \cap B) = P(A)\, P(B)$. Since

$$A \cap B^c = A \setminus B = A \setminus (B \cap A),$$

it follows that

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(B \cap A) = P(A) - P(A)\, P(B) \\ &= P(A)\,(1 - P(B)) = P(A)\, P(B^c). \end{aligned}$$

Thus if $\{A, B\}$ are independent then so is $\{A, B^c\}$. Similarly we may show $\{A^c, B\}$ are independent and then that $\{A^c, B^c\}$ are independent. That is $P\left(A^\varepsilon \cap B^\delta\right) = P(A^\varepsilon)\, P\left(B^\delta\right)$ where ε, δ is either "nothing" or "c."

The general case now easily follows similarly. Indeed, if $\{A_1, \dots, A_n\} \subset \{A_j\}_{j\in J}$ we must show that

$$P\left(A_1^{\varepsilon_1} \cap \cdots \cap A_n^{\varepsilon_n}\right) = P\left(A_1^{\varepsilon_1}\right) \dots P\left(A_n^{\varepsilon_n}\right)$$

where $\varepsilon_j = c$ or $\varepsilon_j =$ " ". But this follows from above. For example, $\{A_1 \cap \cdots \cap A_{n-1}, A_n\}$ are independent implies that $\{A_1 \cap \cdots \cap A_{n-1}, A_n^c\}$ are independent and hence

$$\begin{aligned} P(A_1 \cap \cdots \cap A_{n-1} \cap A_n^c) &= P(A_1 \cap \cdots \cap A_{n-1})\, P(A_n^c) \\ &= P(A_1) \dots P(A_{n-1})\, P(A_n^c). \end{aligned}$$

Thus we have shown it is permissible to add A_j^c to the list for any $j \in J$. ■

Lemma 1.3. *If $\{A_n\}_{n=1}^\infty$ is a sequence of independent events, then*

$$P(\cap_{n=1}^\infty A_n) = \prod_{n=1}^\infty P(A_n) := \lim_{N\to\infty} \prod_{n=1}^N P(A_n).$$

Proof. Since $\cap_{n=1}^N A_n \downarrow \cap_{n=1}^\infty A_n$, it follows that

$$P(\cap_{n=1}^\infty A_n) = \lim_{N\to\infty} P\left(\cap_{n=1}^N A_n\right) = \lim_{N\to\infty} \prod_{n=1}^N P(A_n),$$

where we have used the independence assumption for the last equality. ■

1.1 Borel Cantelli Lemmas

Definition 1.4. *Suppose that $\{A_n\}_{n=1}^\infty$ is a sequence of events. Let*

$$\{A_n \text{ i.o.}\} := \left\{\sum_{n=1}^\infty 1_{A_n} = \infty\right\}$$

denote the event where infinitely many of the events, A_n, occur. The abbreviation, "i.o." stands for infinitely often.

For example if X_n is H or T depending on weather a heads or tails is flipped at the n^{th} step, then $\{X_n = H \text{ i.o.}\}$ is the event where an infinite number of heads was flipped.

Lemma 1.5 (The First Borell – Cantelli Lemma). *If $\{A_n\}$ is a sequence of events such that $\sum_{n=0}^\infty P(A_n) < \infty$, then*

$$P(\{A_n \text{ i.o.}\}) = 0.$$

Proof. Since

$$\infty > \sum_{n=0}^\infty P(A_n) = \sum_{n=0}^\infty \mathbb{E} 1_{A_n} = \mathbb{E}\left[\sum_{n=0}^\infty 1_{A_n}\right]$$

it follows that $\sum_{n=0}^{\infty} 1_{A_n} < \infty$ almost surely (a.s.), i.e. with probability 1 only finitely many of the $\{A_n\}$ can occur. ∎

Under the additional assumption of independence we have the following strong converse of the first Borel-Cantelli Lemma.

Lemma 1.6 (Second Borel-Cantelli Lemma). *If $\{A_n\}_{n=1}^{\infty}$ are independent events, then*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \implies P(\{A_n \text{ i.o.}\}) = 1. \tag{1.1}$$

Proof. We are going to show $P(\{A_n \text{ i.o.}\}^c) = 0$. Since,

$$\{A_n \text{ i.o.}\}^c = \left\{\sum_{n=1}^{\infty} 1_{A_n} = \infty\right\}^c = \left\{\sum_{n=1}^{\infty} 1_{A_n} < \infty\right\}$$

we see that $\omega \in \{A_n \text{ i.o.}\}^c$ iff there exists $n \in \mathbb{N}$ such that $\omega \notin A_m$ for all $m \geq n$. Thus we have shown, if $\omega \in \{A_n \text{ i.o.}\}^c$ then $\omega \in B_n := \cap_{m\geq n} A_m^c$ for some n and therefore, $\{A_n \text{ i.o.}\}^c = \cup_{n=1}^{\infty} B_n$. As $B_n \uparrow \{A_n \text{ i.o.}\}^c$ we have

$$P(\{A_n \text{ i.o.}\}^c) = \lim_{n\to\infty} P(B_n).$$

But making use of the independence (see Lemmas 1.2 and 1.3) and the estimate, $1 - x \leq e^{-x}$, see Figure 1.1 below, we find

$$P(B_n) = P(\cap_{m\geq n} A_m^c) = \prod_{m\geq n} P(A_m^c) = \prod_{m\geq n} [1 - P(A_m)]$$
$$\leq \prod_{m\geq n} e^{-P(A_m)} = \exp\left(-\sum_{m\geq n} P(A_m)\right) = e^{-\infty} = 0.$$

∎

Combining the two Borel Cantelli Lemmas gives the following Zero-One Law.

Corollary 1.7 (Borel's Zero-One law). *If $\{A_n\}_{n=1}^{\infty}$ are independent events, then*

$$P(A_n \text{ i.o.}) = \begin{cases} 0 \text{ if } \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1 \text{ if } \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}.$$

Example 1.8. If $\{X_n\}_{n=1}^{\infty}$ denotes the outcomes of the toss of a coin such that $P(X_n = H) = p > 0$, then $P(X_n = H \text{ i.o.}) = 1$.

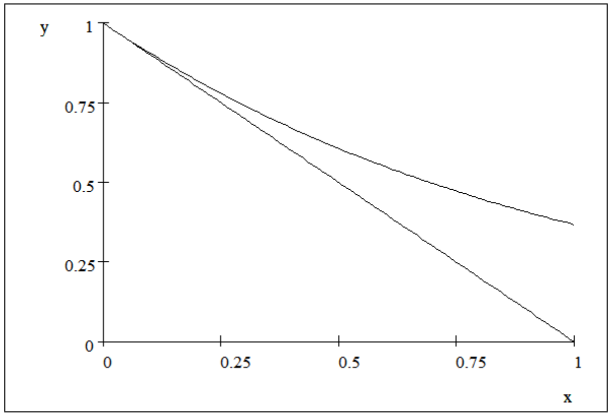


Fig. 1.1. Comparing e^{-x} and $1 - x$.

Example 1.9. If a monkey types on a keyboard with each stroke being independent and identically distributed with each key being hit with positive probability. Then eventually the monkey will type the text of the bible if she lives long enough. Indeed, let S be the set of possible key strokes and let $(s_1, \ldots, s_N)$ be the strokes necessary to type the bible. Further let $\{X_n\}_{n=1}^{\infty}$ be the strokes that the monkey types at time n. Then group the monkey's strokes as $Y_k := \left(X_{kN+1}, \ldots, X_{(k+1)N}\right)$. We then have

$$P(Y_k = (s_1, \ldots, s_N)) = \prod_{j=1}^{N} P(X_j = s_j) =: p > 0.$$

Therefore,

$$\sum_{k=1}^{\infty} P(Y_k = (s_1, \ldots, s_N)) = \infty$$

and so by the second Borel-Cantelli lemma,

$$P(\{Y_k = (s_1, \ldots, s_N)\} \text{ i.o. } k) = 1.$$

1.2 Independent Random Variables

Definition 1.10. *We say a collection of discrete random variables, $\{X_j\}_{j\in J}$, are **independent** if*

$$P(X_{j_1} = x_1, \ldots, X_{j_n} = x_n) = P(X_{j_1} = x_1) \cdots P(X_{j_n} = x_n) \tag{1.2}$$

for all possible choices of $\{j_1, \ldots, j_n\} \subset J$ and all possible values x_k of X_{j_k}.

Proposition 1.11. *A sequence of discrete random variables, $\{X_j\}_{j\in J}$, is independent iff*

$$\mathbb{E}\left[f_1\left(X_{j_1}\right)\dots f_n\left(X_{j_n}\right)\right]=\mathbb{E}\left[f_1\left(X_{j_1}\right)\right]\dots\mathbb{E}\left[f_n\left(X_{j_n}\right)\right] \tag{1.3}$$

for all choices of $\{j_1,\dots,j_n\}\subset J$ and all choice of bounded (or non-negative) functions, $f_1,\dots,f_n$. Here n is arbitrary.

Proof. ($\Longrightarrow$) If $\{X_j\}_{j\in J}$, are independent then

$$\begin{aligned}\mathbb{E}\left[f\left(X_{j_1},\dots,X_{j_n}\right)\right]&=\sum_{x_1,\dots,x_n}f\left(x_1,\dots,x_n\right)P\left(X_{j_1}=x_1,\dots,X_{j_n}=x_n\right)\\&=\sum_{x_1,\dots,x_n}f\left(x_1,\dots,x_n\right)P\left(X_{j_1}=x_1\right)\cdots P\left(X_{j_n}=x_n\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[f_1\left(X_{j_1}\right)\dots f_n\left(X_{j_n}\right)\right]&=\sum_{x_1,\dots,x_n}f_1\left(x_1\right)\dots f_n\left(x_n\right)P\left(X_{j_1}=x_1\right)\cdots P\left(X_{j_n}=x_n\right)\\&=\left(\sum_{x_1}f_1\left(x_1\right)P\left(X_{j_1}=x_1\right)\right)\cdots\left(\sum_{x_n}f\left(x_n\right)P\left(X_{j_n}=x_n\right)\right)\\&=\mathbb{E}\left[f_1\left(X_{j_1}\right)\right]\dots\mathbb{E}\left[f_n\left(X_{j_n}\right)\right].\end{aligned}$$

($\Longleftarrow$) Now suppose that Eq. (1.3) holds. If $f_j:=\delta_{x_j}$ for all j, then

$$\mathbb{E}\left[f_1\left(X_{j_1}\right)\dots f_n\left(X_{j_n}\right)\right]=\mathbb{E}\left[\delta_{x_1}\left(X_{j_1}\right)\dots\delta_{x_n}\left(X_{j_n}\right)\right]=P\left(X_{j_1}=x_1,\dots,X_{j_n}=x_n\right)$$

while

$$\mathbb{E}\left[f_k\left(X_{j_k}\right)\right]=\mathbb{E}\left[\delta_{x_k}\left(X_{j_k}\right)\right]=P\left(X_{j_k}=x_k\right).$$

Therefore it follows from Eq. (1.3) that Eq. (1.2) holds, i.e. $\{X_j\}_{j\in J}$ is an independent collection of random variables. ■

Using this as motivation we make the following definition.

Definition 1.12. *A collection of arbitrary random variables, $\{X_j\}_{j\in J}$, are* ***independent*** *iff*

$$\mathbb{E}\left[f_1\left(X_{j_1}\right)\dots f_n\left(X_{j_n}\right)\right]=\mathbb{E}\left[f_1\left(X_{j_1}\right)\right]\dots\mathbb{E}\left[f_n\left(X_{j_n}\right)\right]$$

for all choices of $\{j_1,\dots,j_n\}\subset J$ and all choice of bounded (or non-negative) functions, $f_1,\dots,f_n$.

Fact 1.13 *To check independence of a collection of real valued random variables, $\{X_j\}_{j\in J}$, it suffices to show*

$$P\left(X_{j_1}\le t_1,\dots,X_{j_n}\le t_n\right)=P\left(X_{j_1}\le t_1\right)\dots P\left(X_{j_n}\le t_n\right)$$

for all possible choices of $\{j_1,\dots,j_n\}\subset J$ and all possible $t_k\in\mathbb{R}$. Moreover, one can replace $\le$ by $<$ or reverse these inequalities in the the above expression.

One of the key theorems involving independent random variables is the strong law of large numbers. The other is the central limit theorem.

Theorem 1.14 (Kolmogorov's Strong Law of Large Numbers). *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables and let $S_n:=X_1+\dots+X_n$. Then there exists $\mu\in\mathbb{R}$ such that $\frac{1}{n}S_n\to\mu$ a.s. iff X_n is integrable and in which case $\mathbb{E}X_n=\mu$.*

Remark 1.15. If $\mathbb{E}\left|X_1\right|=\infty$ but $\mathbb{E}X_1^-<\infty$, then $\frac{1}{n}S_n\to\infty$ a.s. To prove this, for $M>0$ let

$$X_n^M:=\min\left(X_n,M\right)=\begin{cases}X_n \text{ if } X_n\le M\\ M \text{ if } X_n\ge M\end{cases}$$

and $S_n^M:=\sum_{i=1}^n X_i^M$. It follows from Theorem 1.14 that $\frac{1}{n}S_n^M\to\mu^M:=\mathbb{E}X_1^M$ a.s.. Since $S_n\ge S_n^M$, we may conclude that

$$\liminf_{n\to\infty}\frac{S_n}{n}\ge\liminf_{n\to\infty}\frac{1}{n}S_n^M=\mu^M \text{ a.s.}$$

Since $\mu^M\to\infty$ as $M\to\infty$, it follows that $\liminf_{n\to\infty}\frac{S_n}{n}=\infty$ a.s. and hence that $\lim_{n\to\infty}\frac{S_n}{n}=\infty$ a.s.

1.3 Conditioning

Suppose that X and Y are continuous random variables which have a joint density, $\rho_{(X,Y)}(x,y)$. Then by definition of $\rho_{(X,Y)}$, we have, for all bounded or non-negative, f, that

$$\mathbb{E}\left[f\left(X,Y\right)\right]=\int\int f\left(x,y\right)\rho_{(X,Y)}\left(x,y\right)dxdy. \tag{1.4}$$

The marginal density associated to Y is then given by

$$\rho_Y\left(y\right):=\int\rho_{(X,Y)}\left(x,y\right)dx. \tag{1.5}$$

Using this notation, we may rewrite Eq. (1.4) as:

$$\mathbb{E}\left[f\left(X,Y\right)\right]=\int\left[\int f\left(x,y\right)\frac{\rho_{(X,Y)}\left(x,y\right)}{\rho_Y\left(y\right)}dx\right]\rho_Y\left(y\right)dy. \tag{1.6}$$

The term in the bracket is formally the **conditional expectation of $f(X,Y)$ given $Y=y$**. (The technical difficulty here is the $P(Y=y)=0$ in this continuous setting. All of this can be made precise, but we will not do this here.) At any rate, we define,

$$\mathbb{E}\left[f\left(X,Y\right)|Y=y\right]=\mathbb{E}\left[f\left(X,y\right)|Y=y\right]:=\int f\left(x,y\right)\frac{\rho_{(X,Y)}\left(x,y\right)}{\rho_{Y}\left(y\right)}dx$$

in which case Eq. (1.6) may be written as

$$\mathbb{E}\left[f\left(X,Y\right)\right]=\int\mathbb{E}\left[f\left(X,Y\right)|Y=y\right]\rho_{Y}\left(y\right)dy. \tag{1.7}$$

This formula has obvious generalization to the case where X and Y are random vectors such that (X,Y) has a joint distribution, $\rho_{(X,Y)}$. For the purposes of Math 180C we need the following special case of Eq. (1.7).

Proposition 1.16. *Suppose that X and Y are independent random vectors with densities, $\rho_X(x)$ and $\rho_Y(y)$ respectively. Then*

$$\mathbb{E}\left[f\left(X,Y\right)\right]=\int\mathbb{E}\left[f\left(X,y\right)\right]\cdot\rho_{Y}\left(y\right)dy. \tag{1.8}$$

Proof. The independence assumption is equivalent of $\rho_{(X,Y)}(x,y)=\rho_X(x)\,\rho_Y(y)$. Therefore Eq. (1.4) becomes

$$\begin{aligned}\mathbb{E}\left[f\left(X,Y\right)\right]&=\int\int f\left(x,y\right)\rho_{X}\left(x\right)\rho_{Y}\left(y\right)dxdy\\&=\int\left[\int f\left(x,y\right)\rho_{X}\left(x\right)dx\right]\rho_{Y}\left(y\right)dy\\&=\int\mathbb{E}\left[f\left(X,y\right)\right]\cdot\rho_{Y}\left(y\right)dy.\end{aligned}$$

■

Remark 1.17. Proposition 1.16 should not be surprising based on our discussion leading up to Eq. (1.8). Indeed, because of the assumed independence of X and Y, we should have

$$\mathbb{E}\left[f\left(X,Y\right)|Y=y\right]=\mathbb{E}\left[f\left(X,y\right)|Y=y\right]=\mathbb{E}\left[f\left(X,y\right)\right].$$

Using this identity in Eq. (1.7) gives Eq. (1.8).

2

Some Distributions

2.1 Geometric Random Variables

Definition 2.1. *A integer valued random variable, N, is said to have a geometric distribution with parameter, $p \in (0,1)$ provided,*

$$P(N=k) = p(1-p)^{k-1} \text{ for } k \in \mathbb{N}.$$

If $|s| < \frac{1}{1-p}$, we find

$$\begin{aligned}\mathbb{E}\left[s^N\right] &= \sum_{k=1}^{\infty} p(1-p)^{k-1} s^k = ps \sum_{k=1}^{\infty} (1-p)^{k-1} s^{k-1} \\ &= \frac{ps}{1-s(1-p)}.\end{aligned}$$

Differentiating this equation in s implies,

$$\begin{aligned}\mathbb{E}\left[Ns^{N-1}\right] &= \frac{d}{ds}\frac{ps}{1-s(1-p)} \text{ and} \\ \mathbb{E}\left[N(N-1)s^{N-2}\right] &= \left(\frac{d}{ds}\right)^2 \frac{ps}{1-s(1-p)}.\end{aligned}$$

For $s = 1+\varepsilon$, we have

$$\begin{aligned}\frac{ps}{1-s(1-p)} &= \frac{p(1+\varepsilon)}{1-(1+\varepsilon)(1-p)} = \frac{p(1+\varepsilon)}{p(1+\varepsilon)-\varepsilon} = \frac{1}{1-\frac{\varepsilon}{p(1+\varepsilon)}} \\ &= \sum_{k=0}^{\infty} \frac{\varepsilon^k}{p^k(1+\varepsilon)^k} = 1 + \frac{\varepsilon}{p(1+\varepsilon)} + \frac{\varepsilon^2}{p^2(1+\varepsilon)^2} + O\left(\varepsilon^3\right) \\ &= 1 + \frac{\varepsilon(1-\varepsilon+\dots)}{p} + \frac{\varepsilon^2}{p^2} + O\left(\varepsilon^3\right) \\ &= 1 + \frac{\varepsilon}{p} + \varepsilon^2\left(\frac{1}{p^2} - \frac{1}{p}\right) + O\left(\varepsilon^3\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{d}{ds}|_{s=1}\frac{ps}{1-s(1-p)} &= \frac{1}{p} \text{ and} \\ \left(\frac{d}{ds}\right)^2 |_{s=1}\frac{ps}{1-s(1-p)} &= 2\left(\frac{1}{p^2} - \frac{1}{p}\right).\end{aligned}$$

Hence it follows that

$$\begin{aligned}\mathbb{E}N &= 1/p \text{ and} \\ \mathbb{E}N^2 - 1/p = \mathbb{E}\left[N(N-1)\right] &= 2\left(\frac{1}{p^2} - \frac{1}{p}\right)\end{aligned}$$

which shows,

$$\mathbb{E}N^2 = \frac{2}{p^2} - \frac{1}{p}$$

and therefore ,

$$\begin{aligned}\operatorname{Var}(N) = \mathbb{E}N^2 - (\mathbb{E}N)^2 &= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} \\ &= \frac{1-p}{p^2}.\end{aligned}$$

2.2 Exponential Times

Much of what follows is taken from [5].

Definition 2.2. *A random variable $T \geq 0$ has the* ***exponential distribution of parameter*** $\lambda \in [0,\infty)$ *provided, $P(T>t) = e^{-\lambda t}$ for all $t \geq 0$. We will write $T \sim E(\lambda)$ for short.*

If $\lambda > 0$, we have

$$P(T>t) = e^{-\lambda t} = \int_t^{\infty} \lambda e^{-\lambda \tau} d\tau$$

from which it follows that $P(T \in (t, t+dt)) = \lambda 1_{t \geq 0} e^{-\lambda t} dt$. Let us further observe that

$$\mathbb{E}T = \int_0^{\infty} t\lambda e^{-\lambda\tau} d\tau = \lambda\left(-\frac{d}{d\lambda}\right)\int_0^{\infty} e^{-\lambda\tau} d\tau = \lambda\left(-\frac{d}{d\lambda}\right)\lambda^{-1} = \lambda^{-1} \quad (2.1)$$

and similarly,

$$\mathbb{E}T^k = \int_0^\infty t^k \lambda e^{-\lambda\tau} d\tau = \lambda \left(-\frac{d}{d\lambda}\right)^k \int_0^\infty e^{-\lambda\tau} d\tau = \lambda \left(-\frac{d}{d\lambda}\right)^k \lambda^{-1} = k!\lambda^{-k}.$$

In particular we see that

$$\operatorname{Var}(T) = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \tag{2.2}$$

For later purposes, let us also compute,

$$\mathbb{E}\left[e^{-T}\right] = \int_0^\infty e^{-\tau} \lambda e^{-\lambda\tau} d\tau = \frac{\lambda}{1+\lambda} = \frac{1}{1+\lambda^{-1}}. \tag{2.3}$$

Theorem 2.3 (Memoryless property). *A random variable, $T \in (0,\infty]$ has an exponential distribution iff it satisfies the memoryless property:*

$$P(T > s+t | T > s) = P(T > t) \text{ for all } s,t \geq 0.$$

(Note that $T \sim E(0)$ means that $P(T > t) = e^{0t} = 1$ for all $t > 0$ and therefore that $T = \infty$ a.s.)

Proof. Suppose first that $T = E(\lambda)$ for some $\lambda > 0$. Then

$$P(T > s+t|T > s) = \frac{P(T > s+t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t).$$

For the converse, let $g(t) := P(T > t)$, then by assumption,

$$\frac{g(t+s)}{g(s)} = P(T > s+t|T > s) = P(T > t) = g(t)$$

whenever $g(s) \neq 0$ and $g(t)$ is a decreasing function. Therefore if $g(s) = 0$ for some $s > 0$ then $g(t) = 0$ for all $t > s$. Thus it follows that

$$g(t+s) = g(t)\, g(s) \text{ for all } s,t \geq 0.$$

Since $T > 0$, we know that $g(1/n) = P(T > 1/n) > 0$ for some n and therefore, $g(1) = g(1/n)^n > 0$ and we may write $g(1) = e^{-\lambda}$ for some $0 \leq \lambda < \infty$.

Observe for $p,q \in \mathbb{N}$, $g(p/q) = g(1/q)^p$ and taking $p = q$ then shows, $e^{-\lambda} = g(1) = g(1/q)^q$. Therefore, $g(p/q) = e^{-\lambda p/q}$ so that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{Q}_+$. Given $r,s \in \mathbb{Q}_+$ and $t \in \mathbb{R}$ such that $r \leq t \leq s$ we have since g is decreasing that

$$e^{-\lambda r} = g(r) \geq g(t) \geq g(s) = e^{-\lambda s}.$$

Hence letting $s \uparrow t$ and $r \downarrow t$ in the above equations shows that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{R}_+$ and therefore $T \sim E(\lambda)$. ■

Theorem 2.4. *Let I be a countable set and let $\{T_k\}_{k\in I}$ be independent random variables such that $T_k \sim E(q_k)$ with $q := \sum_{k\in I} q_k \in (0,\infty)$. Let $T := \inf_k T_k$ and let $K = k$ on the set where $T_j > T_k$ for all $j \neq k$. On the complement of all these sets, define $K = *$ where $*$ is some point not in I. Then $P(K = *) = 0$, K and T are independent, $T \sim E(q)$, and $P(K = k) = q_k/q$.*

Proof. Let $k \in I$ and $t \in \mathbb{R}_+$ and $\Lambda_n \subset_f I$ such that $\Lambda_n \uparrow I \setminus \{k\}$, then

$$\begin{aligned} P(K = k, T > t) &= P\left(\cap_{j\neq k} \{T_j > T_k\}, T_k > t\right) = \lim_{n\to\infty} P\left(\cap_{j\in\Lambda_n} \{T_j > T_k\}, T_k > t\right) \\ &= \lim_{n\to\infty} \int_{[0,\infty)^{\Lambda_n \cup \{k\}}} \prod_{j\in\Lambda_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n\left(\{t_j\}_{j\in\Lambda_n}\right) q_k e^{-q_k t_k} dt_k \end{aligned}$$

where μ_n is the joint distribution of $\{T_j\}_{j\in\Lambda_n}$. So by Fubini's theorem,

$$\begin{aligned} P(K = k, T > t) &= \lim_{n\to\infty} \int_t^\infty q_k e^{-q_k t_k} dt_k \int_{[0,\infty)^{\Lambda_n}} \prod_{j\in\Lambda_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n\left(\{t_j\}_{j\in\Lambda_n}\right) \\ &= \lim_{n\to\infty} \int_t^\infty P\left(\cap_{j\in\Lambda_n} \{T_j > t_k\}\right) q_k e^{-q_k t_k} dt_k \\ &= \int_t^\infty P\left(\cap_{j\neq k} \{T_j > \tau\}\right) q_k e^{-q_k \tau} d\tau \\ &= \int_t^\infty \prod_{j\neq k} e^{-q_j\tau} q_k e^{-q_k\tau} d\tau = \int_t^\infty \prod_{j\in I} e^{-q_j\tau} q_k d\tau \\ &= \int_t^\infty e^{-\sum_{j=1}^\infty q_j\tau} q_k d\tau = \int_t^\infty e^{-q\tau} q_k d\tau = \frac{q_k}{q} e^{-qt}. \end{aligned} \tag{2.4}$$

Taking $t = 0$ shows that $P(K = k) = \frac{q_k}{q}$ and summing this on k shows $P(K \in I) = 1$ so that $P(K = *) = 0$. Moreover summing Eq. (2.4) on k now shows that $P(T > t) = e^{-qt}$ so that T is exponential. Moreover we have shown that

$$P(K = k, T > t) = P(K = k)\, P(T > t)$$

proving the desired independence. ■

Theorem 2.5. *Suppose that $S \sim E(\lambda)$ and $R \sim E(\mu)$ are independent. Then for $t \geq 0$ we have*

$$\mu P(S \leq t < S + R) = \lambda P(R \leq t < R + S).$$

Proof. We have

$$\begin{aligned}\mu P(S \le t < S+R) &= \mu \int_0^t \lambda e^{-\lambda s} P(t < s+R)\, ds \\ &= \mu\lambda \int_0^t e^{-\lambda s} e^{-\mu(t-s)} ds \\ &= \mu\lambda e^{-\mu t} \int_0^t e^{-(\lambda-\mu)s} ds = \mu\lambda e^{-\mu t} \cdot \frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu} \\ &= \mu\lambda \cdot \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu}\end{aligned}$$

which is symmetric in the interchanged of μ and λ. ■

Example 2.6. Suppose T is a positive random variable such that $P(T \ge t+s | T \ge s) = P(T \ge t)$ for all $s, t \ge 0$, or equivalently

$$P(T \ge t+s) = P(T \ge t)\, P(T \ge s) \text{ for all } s,t \ge 0,$$

then $P(T \ge t) = e^{-at}$ for some $a > 0$. (Such exponential random variables are often used to model "waiting times.") The distribution function for T is $F_T(t) := P(T \le t) = 1 - e^{-a(t \vee 0)}$. Since $F_T(t)$ is piecewise differentiable, the law of T, $\mu := P \circ T^{-1}$, has a density,

$$d\mu(t) = F_T'(t)\, dt = ae^{-at} 1_{t \ge 0} dt.$$

Therefore,

$$\mathbb{E}\left[e^{iaT}\right] = \int_0^\infty ae^{-at} e^{i\lambda t} dt = \frac{a}{a - i\lambda} = \hat{\mu}(\lambda).$$

Since

$$\hat{\mu}'(\lambda) = i \frac{a}{(a-i\lambda)^2} \text{ and } \hat{\mu}''(\lambda) = -2 \frac{a}{(a-i\lambda)^3}$$

it follows that

$$\mathbb{E}T = \frac{\hat{\mu}'(0)}{i} = a^{-1} \text{ and } \mathbb{E}T^2 = \frac{\hat{\mu}''(0)}{i^2} = \frac{2}{a^2}$$

and hence $\mathrm{Var}(T) = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = a^{-2}$.

2.3 Gamma Distributions

Lemma 2.7. *Suppose that* $\{S_j\}_{j=1}^n$ *are independent exponential random variables with parameter,* θ*. and* $W_n = S_1 + \dots + S_n$*. Then*

$$P(W_n \le t) = 1 - e^{-\theta t} \left(\sum_{j=0}^{n-1} \frac{(\theta t)^j}{j!} \right) \tag{2.5}$$

$$= e^{-\theta t} \sum_{j=n}^{\infty} \frac{(\theta t)^j}{j!} \tag{2.6}$$

and the distribution function for W_n *is*

$$f_{W_n}(t) = \theta e^{-\theta t} \frac{(\theta t)^{n-1}}{(n-1)!}. \tag{2.7}$$

Proof. Let $W_k := S_1 + \dots + S_k$. We then have,

$$\begin{aligned}P(W_n \le t) &= P(W_{n-1} + S_n \le t) \\ &= \int_0^t P(W_{n-1} + S_n \le t | S_n = s)\, \theta e^{-\theta s} ds \\ &= \int_0^t P(W_{n-1} + s \le t)\, \theta e^{-\theta s} ds \\ &= \int_0^t P(W_{n-1} \le t - s)\, \theta e^{-\theta s} ds.\end{aligned}$$

We may now use this expression to compute $P(W_n \le t)$ inductively starting with

$$P(W_1 \le t) = P(S_1 \le t) = 1 - e^{-\theta t}.$$

For $n = 2$ we have,

$$\begin{aligned}P(W_2 \le t) &= \int_0^t \left(1 - e^{-\theta(t-s)}\right) \theta e^{-\theta s} ds = \theta \int_0^t \left(e^{-\theta s} - e^{-\theta t}\right) ds \\ &= 1 - e^{-\theta t} - \theta t e^{-\theta t} \\ &= 1 - e^{-\theta t}(1 + \theta t) \qquad (2.8) \\ &= e^{-\theta t}\left(e^{\theta t} - (1+\theta t)\right) = e^{-\theta t} \left(\frac{(\theta t)^2}{2!} + \frac{(\theta t)^3}{3!} + \dots \right) \\ &= e^{-\theta t} \sum_{n=2}^{\infty} \frac{(\theta t)^n}{n!}. \qquad (2.9)\end{aligned}$$

Differentiating Eq. (2.8) shows,

$$\begin{aligned}f_{W_2}(t) &= \frac{d}{dt} P(W_2 \le t) = \frac{d}{dt}\left[1 - e^{-\theta t}(1+\theta t)\right] \\ &= \theta e^{-\theta t}(1 + \theta t) - e^{-\theta t}\theta = \theta t e^{-\theta t}.\end{aligned}$$

For the general case, we find, assuming that Eq. (2.5) is correct,

$$\begin{aligned}
P\left(W_{n+1} \leq t\right) &= \theta \int_0^t \left[1 - e^{-\theta(t-s)} \left(\sum_{j=0}^{n-1} \frac{\left(\theta\left(t-s\right)\right)^j}{j!}\right)\right] e^{-\theta s} ds \\
&= \theta \int_0^t \left[e^{-\theta s} - e^{-\theta t} \left(\sum_{j=0}^{n-1} \frac{\left(\theta\left(t-s\right)\right)^j}{j!}\right)\right] ds \\
&= 1 - e^{-\theta t} - \theta e^{-\theta t} \sum_{j=0}^{n-1} \int_0^t \frac{\theta^j \left(t-s\right)^j}{j!} ds \\
&= 1 - e^{-\theta t} - \theta e^{-\theta t} \sum_{j=0}^{n-1} \frac{\theta^j t^{j+1}}{(j+1)!} \\
&= 1 - e^{-\theta t} - e^{-\theta t} \sum_{j=0}^{n-1} \frac{\theta^{j+1} t^{j+1}}{(j+1)!} = 1 - e^{-\theta t} \sum_{j=0}^{n} \frac{(\theta t)^j}{j!}
\end{aligned}$$

which completes the induction argument and proves Eq. (2.5). Since,

$$1 = e^{-\theta t} e^{\theta t} = e^{-\theta t} \sum_{j=0}^{\infty} \frac{(\theta t)^j}{j!}$$

we also have,

$$\begin{aligned}
P\left(W_n \leq t\right) &= e^{-\theta t} \sum_{j=0}^{\infty} \frac{(\theta t)^j}{j!} - e^{-\theta t} \left(\sum_{j=0}^{n-1} \frac{(\theta t)^j}{j!}\right) \\
&= e^{-\theta t} \sum_{j=n}^{\infty} \frac{(\theta t)^j}{j!}
\end{aligned}$$

which proves Eq. (2.6). The distribution function for W_n now be computed by,

$$\begin{aligned}
f_{W_n}(t) = \frac{d}{dt} P\left(W_n \leq t\right) &= \frac{d}{dt} \left[1 - e^{-\theta t} \left(\sum_{j=0}^{n-1} \frac{(\theta t)^j}{j!}\right)\right] \\
&= \theta e^{-\theta t} \left(\sum_{j=0}^{n-1} \frac{(\theta t)^j}{j!}\right) - e^{-\theta t} \sum_{j=1}^{n-1} \frac{\theta^j t^{j-1}}{(j-1)!} \\
&= \theta e^{-\theta t} \left[\sum_{j=0}^{n-1} \frac{(\theta t)^j}{j!} - \sum_{j=1}^{n-1} \frac{\theta^{j-1} t^{j-1}}{(j-1)!}\right] = \theta e^{-\theta t} \frac{(\theta t)^{n-1}}{(n-1)!}.
\end{aligned}$$

■

2.4 Beta Distribution

Lemma 2.8. *Let*

$$B(x, y) := \int_0^1 t^{x-1} (1-t)^{y-1} dt \text{ for } \operatorname{Re} x, \operatorname{Re} y > 0. \tag{2.10}$$

Then

$$B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}.$$

Proof. Let $u = \frac{t}{1-t}$ so that $t = u(1-t)$ or equivalently, $t = \frac{u}{1+u}$ and $1 - t = \frac{1}{1+u}$ and $dt = (1+u)^{-2} du$.

$$\begin{aligned}
B(x, y) &= \int_0^\infty \left(\frac{u}{1+u}\right)^{x-1} \left(\frac{1}{1+u}\right)^{y-1} \left(\frac{1}{1+u}\right)^2 du \\
&= \int_0^\infty u^{x-1} \left(\frac{1}{1+u}\right)^{x+y} du.
\end{aligned}$$

Recalling that

$$\Gamma(z) := \int_0^\infty e^{-t} t^z \frac{dt}{t}.$$

We find

$$\int_0^\infty e^{-\lambda t} t^z \frac{dt}{t} = \int_0^\infty e^{-t} \left(\frac{t}{\lambda}\right)^z \frac{dt}{t} = \frac{1}{\lambda^z} \Gamma(z),$$

i.e.

$$\frac{1}{\lambda^z} = \frac{1}{\Gamma(z)} \int_0^\infty e^{-\lambda t} t^z \frac{dt}{t}.$$

Taking $\lambda = (1+u)$ and $z = x + y$ shows

$$\begin{aligned}
B(x, y) &= \int_0^\infty u^{x-1} \frac{1}{\Gamma(x+y)} \int_0^\infty e^{-(1+u)t} t^{x+y} \frac{dt}{t} du \\
&= \frac{1}{\Gamma(x+y)} \int_0^\infty \frac{dt}{t}^x e^{-t} t^{x+y} \int_0^\infty \frac{du}{u} u^x e^{-ut} \\
&= \frac{1}{\Gamma(x+y)} \int_0^\infty \frac{dt}{t}^x e^{-t} t^{x+y} \frac{\Gamma(x)}{t^x} \\
&= \frac{\Gamma(x)}{\Gamma(x+y)} \int_0^\infty \frac{dt}{t}^x e^{-t} t^y = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}.
\end{aligned}$$

■

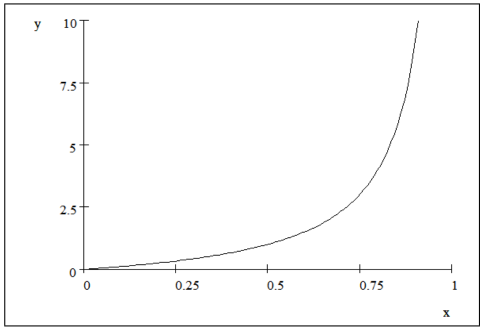


Fig. 2.1. Plot of $t/(1-t)$.

Definition 2.9. *The* β – *distribution is*

$$d\mu_{x,y}(t) = \frac{t^{x-1}(1-t)^{y-1}\,dt}{B(x,y)}.$$

Observe that

$$\int_0^1 t d\mu_{x,y}(t) = \frac{B(x+1,y)}{B(x,y)} = \frac{\frac{\Gamma(x+1)\Gamma(y)}{\Gamma(x+y+1)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{x}{x+y}$$

and

$$\int_0^1 t^2 d\mu_{x,y}(t) = \frac{B(x+2,y)}{B(x,y)} = \frac{\frac{\Gamma(x+2)\Gamma(y)}{\Gamma(x+y+2)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{(x+1)x}{(x+y+1)(x+y)}.$$

3

Markov Chains Basics

For this chapter, let S be a finite or at most countable **state space** and $p : S \times S \to [0,1]$ be a **Markov kernel,** i.e.

$$\sum_{y \in S} p(x,y) = 1 \text{ for all } i \in S. \tag{3.1}$$

A **probability** on S is a function, $\pi : S \to [0,1]$ such that $\sum_{x \in S} \pi(x) = 1$. Further, let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$,

$$\Omega := S^{\mathbb{N}_0} = \{\omega = (s_0, s_1, \dots) : s_j \in S\},$$

and for each $n \in \mathbb{N}_0$, let $X_n : \Omega \to S$ be given by

$$X_n(s_0, s_1, \dots) = s_n.$$

Definition 3.1. *A* ***Markov probability***[1], *P, on Ω with transition kernel, p, is probability on Ω such that*

$$\begin{aligned} &P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &\quad = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1}) \end{aligned} \tag{3.2}$$

where $\{x_j\}_{j=1}^{n+1}$ are allowed to range over S and n over $\mathbb{N}_0$. The identity in Eq. (3.2) is only to be checked on for those $x_j \in S$ such that $P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) > 0$.

If a Markov probability P is given we will often refer to $\{X_n\}_{n=0}^{\infty}$ as a Markov chain. The condition in Eq. (3.2) may also be written as,

$$\mathbb{E}[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] = \mathbb{E}[f(X_{n+1}) \mid X_n] = \sum_{y \in S} p(X_n, y) f(y) \tag{3.3}$$

for all $n \in \mathbb{N}_0$ and any bounded function, $f : S \to \mathbb{R}$.

Proposition 3.2. *If P is a Markov probability as in Definition 3.1 and $\pi(x) := P(X_0 = x)$, then for all $n \in \mathbb{N}_0$ and $\{x_j\} \subset S$,*

$$P(X_0 = x_0, \dots, X_n = x_n) = \pi(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n). \tag{3.4}$$

Conversely if $\pi : S \to [0,1]$ is a probability and $\{X_n\}_{n=0}^{\infty}$ is a sequence of random variables satisfying Eq. (3.4) for all n and $\{x_j\} \subset S$, then $(\{X_n\}, P, p)$ satisfies Definition 3.1.

Proof. ($\Longrightarrow$)We do the case $n = 2$ for simplicity. Here we have

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1, X_2 = x_2) &= P(X_2 = x_2 | X_0 = x_0, X_1 = x_1,) \cdot P(X_0 = x_0, X_1 = x \\ &= P(X_2 = x_2 | X_1 = x_1,) \cdot P(X_0 = x_0, X_1 = x_1) \\ &= p(x_1, x_2) \cdot P(X_1 = x_1 | X_0 = x_0) P(X_0 = x_0) \\ &= p(x_1, x_2) \cdot p(x_0, x_1) \pi(x_0). \end{aligned}$$

($\Longleftarrow$) By assumption we have

$$\begin{aligned} &P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &\quad = \frac{\pi(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) p(x_n, x_{n+1})}{\pi(x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n)} = p(x_n, x_{n+1}) \end{aligned}$$

provided the denominator is not zero. ■

Fact 3.3 *To each probability π on S there is a unique Markov probability, P_π, on Ω such that $P_\pi(X_0 = x) = \pi(x)$ for all $x \in X$. Moreover, P_π is uniquely determined by Eq. (3.4).*

Notation 3.4 *If*

$$\pi(y) = \delta_x(y) := \begin{cases} 1 \text{ if } x = y \\ 0 \text{ if } x \neq y \end{cases}, \tag{3.5}$$

we will write P_x for P_π. For a general probability, π, on S we have

$$P_\pi = \sum_{x \in S} \pi(x) P_x. \tag{3.6}$$

---

[1] The set Ω is sufficiently big that it is no longer so easy to give a rigorous definition of a probability on Ω. For the purposes of this class, a **probability on** Ω should be taken to mean an assignment, $P(A) \in [0,1]$ for all subsets, $A \subset \Omega$, such that $P(\emptyset) = 0$, $P(\Omega) = 1$, and

$$P(A) = \sum_{n=1}^{\infty} P(A_n)$$

whenever $A = \cup_{n=1}^{\infty} A_n$ with $A_n \cap A_m = \emptyset$ for all $m \neq n$. (There are technical problems with this definition which are addressed in a course on "measure theory." We may safely ignore these problems here.)

Notation 3.5 *Associated to a transition kernel, p, is a* ***jump graph (or jump diagram)*** *gotten by taking S as the set of vertices and then for $x, y \in S$, draw an arrow from x to y if $p(x,y) > 0$ and label this arrow by the value $p(x,y)$.*

Example 3.6. Suppose that $S = \{1,2,3\}$, then

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ P = & \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

has the jump graph given by 3.1.

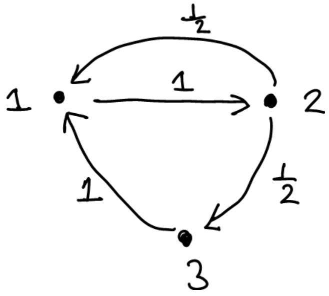


Fig. 3.1. A simple jump diagram.

Example 3.7. The transition matrix,

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ P = & \left[\begin{array}{ccc} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

is represented by the jump diagram in Figure 3.2.

If $q : S \times S \to [0,1]$ is another probability kernel we let $p \cdot q : S \times S \to [0,1]$ be defined by

$$(p \cdot q)(x,y) := \sum_{z \in S} p(x,z)\, q(z,y). \quad \text{(Matrix Multiplication!)} \tag{3.7}$$

We also let $p^n := \overbrace{p \cdot p \cdot \dots \cdot p}^{n \text{ - times}}$. If $\pi : S \to [0,1]$ is a probability we let $(\pi \cdot q) : S \to [0,1]$ be defined by

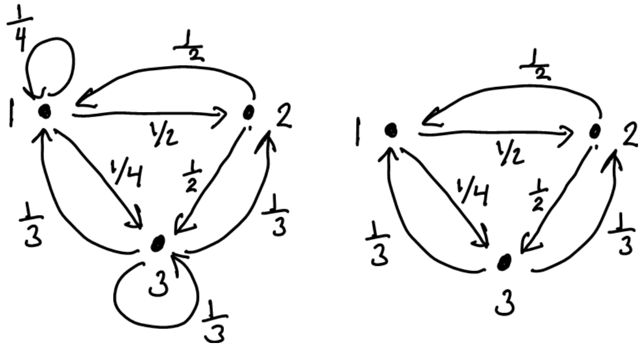


Fig. 3.2. The above diagrams contain the same information. In the one on the right we have dropped the jumps from a site back to itself since these can be deduced by conservation of probability.

$$(\pi \cdot q)(y) := \sum_{x \in S} \pi(x)\, q(x,y)$$

which again is matrix multiplication if we view π to be a row vector. It is easy to check that $\pi \cdot q$ is still a probability and $p \cdot q$ and p^n are Markov kernels.

A key point to keep in mind is that a Markov process is completely specified by its transition kernel, $p : S \times S \to [0,1]$. For example we have the following method for computing $P_x(X_n = y)$.

Lemma 3.8. *Keeping the above notation, $P_x(X_n = y) = p^n(x,y)$ and more generally,*

$$P_\pi(X_n = y) = \sum_{x \in S} \pi(x)\, p^n(x,y) = (\pi \cdot p^n)(y).$$

Proof. We have from Eq. (3.4) that

$$\begin{aligned} P_x(X_n = y) &= \sum_{x_0, \dots, x_{n-1} \in S} P_x(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = y) \\ &= \sum_{x_0, \dots, x_{n-1} \in S} \delta_x(x_0)\, p(x_0, x_1) \dots p(x_{n-2}, x_{n-1})\, p(x_{n-1}, y) \\ &= \sum_{x_1, \dots, x_{n-1} \in S} p(x, x_1) \dots p(x_{n-2}, x_{n-1})\, p(x_{n-1}, y) = p^n(x,y). \end{aligned}$$

The formula for $P_\pi(X_n = y)$ easily follows from this formula. ■

Definition 3.9. *We say that $\pi : S \to [0,1]$ is a* ***stationary*** *distribution for p, if*

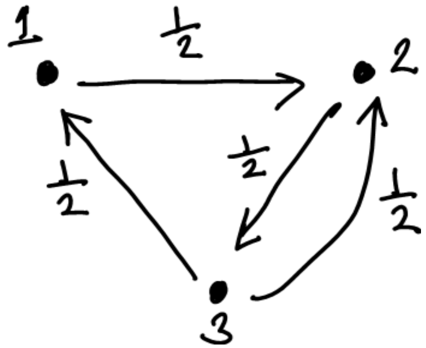
$$P_\pi(X_n = x) = \pi(x) \text{ for all } x \in S \text{ and } n \in \mathbb{N}.$$

Since $P_\pi(X_n = x) = (\pi \cdot p^n)(x)$, we see that π is a stationary distribution for p iff $\pi p^n = p$ for all $n \in \mathbb{N}$ iff $\pi p = p$ by induction.

Example 3.10. Consider the following example,

$$P = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left[\begin{array}{ccc} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

with jump diagram given in Figure 3.10. We have



$$P^2 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

and

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

To have a picture what is going on here, imaging that $\pi = (\pi_1, \pi_2, \pi_3)$ represents the amount of sand at the sites, 1, 2, and 3 respectively. During each time step we move the sand on the sites around according to the following rule. The sand at site j after one step is $\sum_i \pi_i p_{ij}$, namely site i contributes p_{ij} fraction its sand, π_i, to site j. Everyone does this to arrive at a new distribution. Hence π is an invariant distribution if each π_i remains unchanged, i.e. $\pi = \pi P$. (Keep in mind the sand is still moving around it is just that the size of the piles remains unchanged.)

As a specific example, suppose $\pi = (1, 0, 0)$ so that all of the sand starts at 1. After the first step, the pile at 1 is split into two and 1/2 is sent to 2 to get $\pi_1 = (1/2, 1/2, 0)$ which is the first row of P. At the next step the site 1 keeps 1/2 of its sand (= 1/4) and still receives nothing, while site 2 again receives the other 1/2 and keeps half of what it had (= 1/4 + 1/4) and site 3 then gets $(1/2 \cdot 1/2 = 1/4)$ so that $\pi_2 = \left[\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\right]$ which is the first row of P^2. It turns out in this case that this is the invariant distribution. Formally,

$$\left[\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\right] \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} = \left[\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\right].$$

In general we expect to reach the invariant distribution only in the limit as $n \to \infty$.

Notice that if π is any stationary distribution, then $\pi P^n = \pi$ for all n and in particular,

$$\pi = \pi P^2 = \left[\pi_1\ \pi_2\ \pi_3\right] \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \left[\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\right].$$

Hence $\left[\frac{1}{4}\ \frac{1}{2}\ \frac{1}{4}\right]$ is the unique stationary distribution for P in this case.

Example 3.11 (§3.2. p108 Ehrenfest Urn Model). Let a beaker filled with a particle fluid mixture be divided into two parts A and B by a semipermeable membrane. Let $X_n = (\#$ of particles in $A)$ which we assume evolves by choosing a particle at random from $A \cup B$ and then replacing this particle in the opposite bin from which it was found. Suppose there are N total number of particles in the flask, then the transition probabilities are given by,

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & \text{if } j \notin \{i-1, i+1\} \\ \frac{i}{N} & \text{if } \quad j = i - 1 \\ \frac{N-i}{N} & \text{if } \quad j = i + 1. \end{cases}$$

For example, if $N = 2$ we have

$$(p_{ij}) = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}$$

and if $N = 3$, then we have in matrix form,

$$(p_{ij}) = \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \left[\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}.$$

In the case $N = 2$,

$$\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

and when $N = 3$,

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{7}{9} & 0 & \frac{2}{9} \\ \frac{2}{9} & 0 & \frac{7}{9} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 0 & \frac{7}{9} & 0 & \frac{2}{9} \\ \frac{7}{27} & 0 & \frac{20}{27} & 0 \\ 0 & \frac{20}{27} & 0 & \frac{7}{27} \\ \frac{2}{9} & 0 & \frac{7}{9} & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{25} \cong \begin{bmatrix} 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{26} \cong \begin{bmatrix} 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \end{bmatrix}$$

:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{100} \cong \begin{bmatrix} 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \end{bmatrix}$$

We also have

$$(P - I)^{\mathrm{tr}} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ \frac{1}{3} & -1 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & -1 & \frac{1}{3} \\ 0 & 0 & 1 & -1 \end{bmatrix}^{\mathrm{tr}} = \begin{bmatrix} -1 & \frac{1}{3} & 0 & 0 \\ 1 & -1 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & -1 & 1 \\ 0 & 0 & \frac{1}{3} & -1 \end{bmatrix}$$

and

$$\mathrm{Nul}\left((P - I)^{\mathrm{tr}}\right) = \begin{bmatrix} 1 \\ 3 \\ 3 \\ 1 \end{bmatrix}.$$

Hence if we take, $\pi = \frac{1}{8}\begin{bmatrix} 1 & 3 & 3 & 1 \end{bmatrix}$ then

$$\pi P = \frac{1}{8}\begin{bmatrix} 1 & 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 1 & 3 & 3 & 1 \end{bmatrix} = \pi$$

is the stationary distribution. Notice that

$$\begin{aligned} \frac{1}{2}\left(P^{25} + P^{26}\right) &\cong \frac{1}{2}\begin{bmatrix} 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \\ 0.25 & 0.0 & 0.75 & 0.0 \\ 0.0 & 0.75 & 0.0 & 0.25 \end{bmatrix} \\ &= \begin{bmatrix} 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \pi \\ \pi \end{bmatrix}. \end{aligned}$$

3.1 First Step Analysis

We will need the following observation in the proof of Lemma 3.14 below. If T is a $\mathbb{N}_0 \cup \{\infty\}$ – valued random variable, then

$$\mathbb{E}_x T = \mathbb{E}_x \sum_{n=0}^{\infty} 1_{n<T} = \sum_{n=0}^{\infty} \mathbb{E}_x 1_{n<T} = \sum_{n=0}^{\infty} P_x\left(T > n\right). \tag{3.8}$$

Now suppose that S is a state space and assume that S is divided into two disjoint events, A and B. Let

$$T := \inf\{n \geq 0 : X_n \in B\}$$

be the **hitting time** of B. Let $Q := \left(p\left(x,y\right)\right)_{x,y \in A}$ and $R := \left(p\left(x,y\right)\right)_{x \in A,\ y \in B}$ so that the transition "matrix," $P = \left(p\left(x,y\right)\right)_{x,y \in S}$ may be written in the following block diagonal form;

$$P = \begin{bmatrix} Q & R \\ * & * \end{bmatrix} = \begin{array}{cc} \begin{array}{cc} A & B \end{array} & \\ \begin{bmatrix} Q & R \\ * & * \end{bmatrix} & \begin{array}{c} A \\ B \end{array} \end{array}.$$

Remark 3.12. To construct the matrix Q and R from P, let P' be P with the rows corresponding to B omitted. To form Q from P', remove the columns of P' corresponding to B and to form R from P', remove the columns of P' corresponding to A.

Example 3.13. Suppose that $S = \{1, 2, \ldots, 7\}$, $A = \{1, 2, 4, 5, 6\}$, $B = \{3, 7\}$, and

$$\begin{array}{rcl} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} & \\ P = & \left[\begin{array}{ccccccc} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right] & \begin{array}{l} 1 \\ 2 \\ 3 \\ 4. \\ 5 \\ 6 \\ 7 \end{array} \end{array}$$

Following the algorithm in Remark 3.12 leads to:

$$\begin{array}{rcl} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} & \\ P' = & \left[\begin{array}{ccccccc} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \end{array}\right] & \begin{array}{l} 1 \\ 2 \\ 4, \\ 5 \\ 6 \end{array} \end{array}$$

$$\begin{array}{rcl} & \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} & \\ Q = & \left[\begin{array}{ccccc} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{array}\right] & \begin{array}{l} 1 \\ 2 \\ 4, \\ 5 \\ 6 \end{array} \end{array} \text{ and } \begin{array}{rcl} & \begin{array}{cc} 3 & 7 \end{array} & \\ R = & \left[\begin{array}{cc} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{array}\right] & \begin{array}{l} 1 \\ 2 \\ 4. \\ 5 \\ 6 \end{array} \end{array}$$

Lemma 3.14. *Keeping the notation above we have*

$$\mathbb{E}_x T = \sum_{n=0}^{\infty} \sum_{y \in A} Q^n (x, y) \text{ for all } x \in A, \tag{3.9}$$

where $\mathbb{E}_x T = \infty$ *is possible.*

Proof. By definition of T we have for $x \in A$ and $n \in \mathbb{N}_0$ that,

$$\begin{aligned} P_x (T > n) &= P_x (X_1, \ldots, X_n \in A) \\ &= \sum_{x_1, \ldots, x_n \in A} p(x, x_1) p(x_1, x_2) \ldots p(x_{n-1}, x_n) \\ &= \sum_{y \in A} Q^n (x, y). \end{aligned} \tag{3.10}$$

Therefore Eq. (3.9) now follows from Eqs. (3.8) and (3.10). ■

Proposition 3.15. *Let us continue the notation above and let us further assume that A is a finite set and*

$$P_x (T < \infty) = P (X_n \in B \text{ for some } n) > 0 \ \forall \ x \in A. \tag{3.11}$$

Under these assumptions, $\mathbb{E}_x T < \infty$ for all $x \in A$ and in particular $P_x (T < \infty) = 1$ for all $x \in A$. In this case we may may write Eq. (3.9) as

$$(\mathbb{E}_x T)_{x \in A} = (I - Q)^{-1} \mathbf{1} \tag{3.12}$$

where $\mathbf{1}(x) = 1$ for all $x \in A$.

Proof. Since $\{T > n\} \downarrow \{T = \infty\}$ and $P_x (T = \infty) < 1$ for all $x \in A$ it follows that there exists an $m \in \mathbb{N}$ and $0 \le \alpha < 1$ such that $P_x (T > m) \le \alpha$ for all $x \in A$. Since $P_x (T > m) = \sum_{y \in A} Q^m (x, y)$ it follows that the row sums of Q^m are all less than $\alpha < 1$. Further observe that

$$\begin{aligned} \sum_{y \in A} Q^{2m} (x, y) &= \sum_{y, z \in A} Q^m (x, z) Q^m (z, y) = \sum_{z \in A} Q^m (x, z) \sum_{y \in A} Q^m (z, y) \\ &\le \sum_{z \in A} Q^m (x, z) \alpha \le \alpha^2. \end{aligned}$$

Similarly one may show that $\sum_{y \in A} Q^{km} (x, y) \le \alpha^k$ for all $k \in \mathbb{N}$. Therefore from Eq. (3.10) with m replaced by km, we learn that $P_x (T > km) \le \alpha^k$ for all $k \in \mathbb{N}$ which then implies that

$$\sum_{y \in A} Q^n (x, y) = P_x (T > n) \le \alpha^{\lfloor \frac{n}{k} \rfloor} \text{ for all } n \in \mathbb{N},$$

where $\lfloor t \rfloor = m \in \mathbb{N}_0$ if $m \le t < m + 1$, i.e. $\lfloor t \rfloor$ is the nearest integer to t which is smaller than t. Therefore, we have

$$\mathbb{E}_x T = \sum_{n=0}^{\infty} \sum_{y \in A} Q^n (x, y) \le \sum_{n=0}^{\infty} \alpha^{\lfloor \frac{n}{m} \rfloor} \le m \cdot \sum_{l=0}^{\infty} \alpha^l = m \frac{1}{1 - \alpha} < \infty.$$

So it only remains to prove Eq. (3.12). From the above computations we see that $\sum_{n=0}^{\infty} Q^n$ is convergent. Moreover,

$$(I - Q) \sum_{n=0}^{\infty} Q^n = \sum_{n=0}^{\infty} Q^n - \sum_{n=0}^{\infty} Q^{n+1} = I$$

and therefore $(I - Q)$ is invertible and $\sum_{n=0}^{\infty} Q^n = (I - Q)^{-1}$. Finally,

$$(I-Q)^{-1}\mathbf{1}=\sum_{n=0}^{\infty}Q^n\mathbf{1}=\left(\sum_{n=0}^{\infty}\sum_{y\in A}Q^n(x,y)\right)_{x\in A}=(\mathbb{E}_xT)_{x\in A}$$

as claimed. ■

Remark 3.16. Let $\{X_n\}_{n=0}^{\infty}$ denote the fair random walk on $\{0,1,2,\dots\}$ with 0 being an absorbing state. Using the first homework problems, see Remark 0.1, we learn that $\mathbb{E}_iT=\infty$ for all $i>0$. This shows that we can not in general drop the assumption that A ($A=\{1,2,\dots\}$ in this example) is a finite set the statement of Proposition 3.15.

For our next result we will make use of the following important version of the Markov property.

Theorem 3.17 (Markov Property II). *If $f(x_0,x_1,\dots)$ is a bounded random function of $\{x_n\}_{n=0}^{\infty}\subset S$ and $g(x_0,\dots,x_n)$ is a function on S^{n+1}, then*

$$\mathbb{E}_\pi[f(X_n,X_{n+1},\dots)g(X_0,\dots,X_n)]=\mathbb{E}_\pi[(\mathbb{E}_{X_n}[f(X_0,X_1,\dots)])g(X_0,\dots,X_n)] \tag{3.13}$$

$$\mathbb{E}_\pi[f(X_n,X_{n+1},\dots)|X_0=x_0,\dots,X_n=x_n]=\mathbb{E}_{x_n}f(X_0,X_1,\dots) \tag{3.14}$$

for all $x_0,\dots,x_n\in S$ such that $P_\pi(X_0=x_0,\dots,X_n=x_n)>0$. These results also hold when f and g are non-negative functions.

Proof. In proving this theorem, we will have to take for granted that it suffices to assume that f is a function of only finitely many $\{x_n\}$. In practice, any function, f, of the $\{x_n\}_{n=0}^{\infty}$ that we are going to deal with in this course may be written as a limit of functions depending on only finitely many of the $\{x_n\}$. With this as justification, we now suppose that f is a function of $(x_0,\dots,x_m)$ for some $m\in\mathbb{N}$. To simplify notation, let $F=f(X_0,X_1,\dots X_m)$, $\theta_nF:=f(X_n,X_{n+1},\dots X_{n+m})$, and $G=g(X_0,\dots,X_n)$.

We then have,

$$\begin{aligned}&\mathbb{E}_\pi[\theta_nF\cdot G]\\&\quad=\sum_{\{x_j\}_{j=0}^{m+n}\subset S}\pi(x_0)p(x_0,x_1)\dots p(x_{n+m-1},x_{m+n})f(x_n,x_{n+1},\dots x_{n+m})g(x_0,\dots,x_n)\end{aligned}$$

and

$$\begin{aligned}&\sum_{\{x_j\}_{j=n+1}^{m+n}\subset S}p(x_n,x_{n+1})\dots p(x_{n+m-1},x_{m+n})f(x_n,x_{n+1},\dots x_{n+m})g(x_0,\dots,x_n)\\&\qquad=g(x_0,\dots,x_n)\sum_{\{x_j\}_{j=n+1}^{m+n}\subset S}\left[\begin{array}{c}p(x_n,x_{n+1})\dots p(x_{n+m-1},x_{m+n})\cdot\\ \cdot f(x_n,x_{n+1},\dots x_{n+m})\end{array}\right]\\&\qquad=g(x_0,\dots,x_n)\mathbb{E}_{x_n}f(X_0,\dots,X_m)=g(x_0,\dots,x_n)\mathbb{E}_{x_n}F.\end{aligned}$$

Combining the last two equations implies,

$$\begin{aligned}&\mathbb{E}_\pi[\theta_nF\cdot G]\\&\quad=\sum_{\{x_j\}_{j=0}^{m}\subset S}\pi(x_0)p(x_0,x_1)\dots p(x_{n-1},x_n)g(x_0,\dots,x_n)\mathbb{E}_{x_n}F\\&\quad=\mathbb{E}_\pi[g(X_0,\dots,X_n)\cdot\mathbb{E}_{X_n}F]\end{aligned}$$

as was to be proved.

Taking $g(y_0,\dots,y_n)=\delta_{x_0,y_0}\dots\delta_{x_n,y_n}$ is Eq. (3.13) implies that

$$\begin{aligned}&\mathbb{E}_\pi[f(X_n,X_{n+1},\dots):X_0=x_0,\dots,X_n=x_n]\\&\quad=\mathbb{E}_{x_n}F\cdot P_\pi(X_0=x_0,\dots,X_n=x_n)\end{aligned}$$

which implies Eq. (3.14). The proofs of the remaining equivalence of the statements in the Theorem are left to the reader. ■

Here is a useful alternate statement of the Markov property. In words it states, if you know $X_n=x$ then the remainder of the chain $X_n,X_{n+1},X_{n+2},\dots$ forgets how it got to x and behave exactly like the original chain started at x.

Corollary 3.18. *Let $n\in\mathbb{N}_0$, $x\in S$ and π be any probability on S. Then relative to $P_\pi(\cdot|X_n=x)$, $\{X_{n+k}\}_{k\geq0}$ is independent of $\{X_0,\dots,X_n\}$ and $\{X_{n+k}\}_{k\geq0}$ has the same distribution as $\{X_k\}_{k=0}^{\infty}$ under P_x.*

Proof. According to Eq. (3.13),

$$\begin{aligned}&\mathbb{E}_\pi[g(X_0,\dots,X_n)f(X_n,X_{n+1},\dots):X_n=x]\\&\quad=\mathbb{E}_\pi[g(X_0,\dots,X_n)\delta_x(X_n)f(X_n,X_{n+1},\dots)]\\&\quad=\mathbb{E}_\pi[g(X_0,\dots,X_n)\delta_x(X_n)\mathbb{E}_{X_n}[f(X_0,X_1,\dots)]]\\&\quad=\mathbb{E}_\pi[g(X_0,\dots,X_n)\delta_x(X_n)\mathbb{E}_x[f(X_0,X_1,\dots)]]\\&\quad=\mathbb{E}_\pi[g(X_0,\dots,X_n):X_n=x]\mathbb{E}_x[f(X_0,X_1,\dots)].\end{aligned}$$

Dividing this equation by $P(X_n=x)$ shows,

$$\begin{aligned}&\mathbb{E}_\pi[g(X_0,\dots,X_n)f(X_n,X_{n+1},\dots)|X_n=x]\\&\quad=\mathbb{E}_\pi[g(X_0,\dots,X_n)|X_n=x]\mathbb{E}_x[f(X_0,X_1,\dots)].\end{aligned} \tag{3.15}$$

Taking $g=1$ in this equation then shows,

$$\mathbb{E}_\pi[f(X_n,X_{n+1},\dots)|X_n=x]=\mathbb{E}_x[f(X_0,X_1,\dots)]. \tag{3.16}$$

This shows that $\{X_{n+k}\}_{k\geq0}$ under $P_\pi(\cdot|X_n=x)$ has the same distribution as $\{X_k\}_{k=0}^{\infty}$ under P_x and, in combination, Eqs. (3.15) and (3.16) shows $\{X_{n+k}\}_{k\geq0}$ and $\{X_0,\dots,X_n\}$ are conditionally independent on $\{X_n=x\}$. ■

Theorem 3.19. *Let us continue the notation and assumption in Proposition 3.15 and further let* $g : A \to \mathbb{R}$ *and* $h : B \to \mathbb{R}$ *be two functions. Let* $\mathbf{g} := (g\left(x\right))_{x\in A}$ *and* $\mathbf{h} := \left(h\left(y\right)\right)_{y\in B}$ *to be thought of as column vectors. Then for all* $x \in A$,

$$\mathbb{E}_x\left[\sum_{n<T} g(X_n)\right] = x^{th} \text{ component of } (I-Q)^{-1}\mathbf{g} \tag{3.17}$$

and for all $x \in A$ *and* $y \in B$,

$$P_x\left(X_T = y\right) = \left[(I-Q)^{-1}R\right]_{x,y}. \tag{3.18}$$

Taking $g \equiv \mathbf{1}$ *(where* $\mathbf{1}\left(x\right) = 1$ *for all* $x \in A$*) in Eq. (3.17) shows that*

$$\mathbb{E}_x T = \text{ the } x^{th} \text{ component of } (I-Q)^{-1}\mathbf{1} \tag{3.19}$$

in agreement with Eq. (3.12). If we take $g\left(x'\right) = \delta_y\left(x'\right)$ *for some* $x \in A$*, then*

$$\mathbb{E}_x\left[\sum_{n<T} g(X_n)\right] = \mathbb{E}_x\left[\sum_{n<T} \delta_y(X_n)\right] = \mathbb{E}_x\left[\text{number of visits to } y \text{ before } T\right]$$

and by Eq. (3.17) it follows that

$$\mathbb{E}_x\left[\text{number of visits to } y \text{ before hitting } B\right] = (I-Q)^{-1}_{xy}. \tag{3.20}$$

Proof. Let

$$u\left(x\right) := \mathbb{E}_x\left[\sum_{0\le n<T} g(X_n)\right] = \mathbb{E}_x G$$

for $x \in A$ where $G := \sum_{0\le n<T} g(X_n)$. Then

$$u\left(x\right) = \mathbb{E}_x\left[\mathbb{E}_x\left[G|X_1\right]\right] = \sum_{y\in S} p\left(x,y\right)\mathbb{E}_x\left[G|X_1 = y\right].$$

For $y \in A$, by the Markov property[2] in Theorem 3.17 we have,

$$\begin{aligned}\mathbb{E}_x\left[G|X_1 = y\right] &= g\left(x\right) + \mathbb{E}_x\left[\sum_{1\le n<T} g(X_n)|X_1 = y\right] \\ &= g\left(x\right) + \mathbb{E}_y\left[\sum_{0\le n<T} g(X_n)\right] = g\left(x\right) + u\left(y\right)\end{aligned}$$

and for $y \in B$, $\mathbb{E}_x\left[G|X_1 = y\right] = g\left(x\right)$. Therefore

$$\begin{aligned}u\left(x\right) &= \sum_{y\in A} p\left(x,y\right)\left[g\left(x\right) + u\left(y\right)\right] + \sum_{y\in B} p\left(x,y\right)g\left(x\right) \\ &= g\left(x\right) + \sum_{y\in A} p\left(x,y\right)u\left(y\right).\end{aligned}$$

In matrix language this becomes, $\mathbf{u} = Q\mathbf{u}+\mathbf{g}$ and hence we have $\mathbf{u} = (I-Q)^{-1}\mathbf{g}$ which is precisely Eq. (3.17).

To prove Eq. (3.18), let $w\left(x\right) := \mathbb{E}_x\left[h\left(X_T\right)\right]$. Since X_T is the location of where $\left\{X_n\right\}_{n=0}^{\infty}$ first hits B if we are given $X_0 \in A$, then X_T is also the location where the sequence, $\left\{X_n\right\}_{n=1}^{\infty}$, first hits B and therefore $X_T \circ \theta_1 = X_T$ when $X_0 \in A$. Therefore, working as before and noting now that,

$$\begin{aligned}w\left(x\right) &= \sum_{y\in A} \mathbb{E}_x(h(X_T)|X_1 = y)p\left(x,y\right) + \sum_{y\in B} \mathbb{E}_x(h(X_T)|X_1 = y)p\left(x,y\right) \\ &= \sum_{y\in A} p\left(x,y\right)\mathbb{E}_x(h(X_T)\circ\theta_1|X_1 = y) + \sum_{y\in B} p\left(x,y\right)\mathbb{E}_x(h(X_T)|X_1 = y) \\ &= \sum_{y\in A} p\left(x,y\right)\mathbb{E}_y(h(X_T)) + \sum_{y\in B} p\left(x,y\right)h(y) \\ &= \sum_{y\in A} p\left(x,y\right)w\left(y\right) + \sum_{y\in B} p\left(x,y\right)h(y) = (Q\mathbf{w} + R\mathbf{h})_x.\end{aligned}$$

Writing this in matrix form gives, $\mathbf{w} =Q\mathbf{w} + Rh$ which we solve for $\mathbf{w}$ to find that $\mathbf{w} = (I-Q)^{-1}Rh$ and therefore,

$$\left(\mathbb{E}_x\left[h\left(X_T\right)\right]\right)_{x\in A} = x^{\text{th}} - \text{component of } (I-Q)^{-1}R\ \left(h\left(y\right)\right)_{y\in B}$$

Given $y_0 \in B$, the taking $h\left(y\right) = \delta_{y_0,y}$ in the above formula implies that

$$\begin{aligned}P_x\left(X_T = y_0\right) &= x^{\text{th}} - \text{component of } (I-Q)^{-1}R\ \left(\delta_{y_0,y}\right)_{y\in B} \\ &= \left[(I-Q)^{-1}R\right]_{x,y}.\end{aligned}$$

■

---

[2] In applying Theorem 3.17 we note that when $X_0 = x$, $T\left(X_0, X_1, \dots\right) \ge 1$, $T\left(X_1, X_2, \dots\right) = T\left(X_0, X_1, \dots\right) - 1$, and hence

$$\begin{aligned}&\theta_1\left(\sum_{0\le n<T(X_0,X_1,\dots)} g(X_n)\right) \\ &= \sum_{0\le n<T(X_1,X_2\dots)} g(X_{n+1}) = \sum_{0\le n<T(X_0,X_1,\dots)-1} g(X_{n+1}) \\ &= \sum_{1\le n+1<T(X_0,X_1,\dots)} g(X_{n+1}) = \sum_{1\le n<T(X_0,X_1,\dots)} g(X_n) = \sum_{1\le n<T} g(X_n).\end{aligned}$$

Remark 3.20. Here is a story to go along with the above scenario. Suppose that $g(x)$ is the toll you have to pay for visiting a site $x \in A$ while $h(y)$ is the amount of prize money you get when landing on a point in B. Then $\mathbb{E}_x\left[\sum_{0\le n<T} g(X_n)\right]$ is the expected toll you have to pay before your first exit from A while $\mathbb{E}_x[h(X_T)]$ is your expected winnings upon exiting B.

The next two results follow the development in Theorem 1.3.2 of Norris [5].

Theorem 3.21 (Hitting Probabilities). *Suppose that $A \subset S$ as above and now let $H := \inf\{n : X_n \in A\}$ be the first time that $\{X_n\}_{n=0}^\infty$ hits A with the convention that $H = \infty$ if X_n does not hit A. Let $h_i := P_i(H < \infty)$ be the* ***hitting*** *probability of A given $X_0 = i$, $v_i := \sum_{j\notin A} p(i,j)$ for all $i \notin A$, and $\{Q_{ij} := p(i,j)\}_{i,j\notin A}$. Then*

$$h_i = P_i(H<\infty) = 1_{i\in A} + 1_{i\notin A}\sum_{n=0}^\infty [Q^n v]_i \tag{3.21}$$

and h_i may also be characterized as the minimal non-negative solution to the following linear equations;

$$\begin{aligned} h_i &= 1 \text{ if } i \in A \text{ and} \\ h_i &= \sum_{j\in S} p(i,j)\, h_j = \sum_{j\in A^c} Q(i,j)\, h_j + v_i \text{ for all } i \in A^c. \end{aligned} \tag{3.22}$$

Proof. Let us first observe that $P_i(H=0) = P_i(X_0\in A) = 1_{i\in A}$. Also for any $n\in\mathbb{N}$

$$\{H=n\} = \{X_0\notin A,\dots,X_{n-1}\notin A, X_n\in A\}$$

and therefore,

$$\begin{aligned} P_i(H=n) &= 1_{i\notin A}\sum_{j_1,\dots,j_{n-1}\in A^c}\sum_{j_n\in A} p(i,j_1)\,p(j_1,j_2)\dots p(j_{n-2},j_{n-1})\,p(j_{n-1},j_n) \\ &= 1_{i\notin A}\left[Q^{n-1}v\right]_i. \end{aligned}$$

Since $\{H<\infty\} = \cup_{n=0}^\infty \{H=n\}$, it follows that

$$P_i(H<\infty) = 1_{i\in A} + \sum_{n=1}^\infty 1_{i\notin A}\left[Q^{n-1}v\right]_i$$

which is the same as Eq. (3.21). The remainder of the proof now follows from Lemma 3.22 below. Nevertheless, it is instructive to use the Markov property to show that Eq. (3.22) is valid. For this we have by the first step analysis; if $i\notin A$, then

$$\begin{aligned} h_i = P_i(H<\infty) &= \sum_{j\in S} p(i,j)\,P_i(H<\infty|X_1=j) \\ &= \sum_{j\in S} p(i,j)\,P_j(H<\infty) = \sum_{j\in S} p(i,j)\,h_j \end{aligned}$$

as claimed. ■

Lemma 3.22. *Suppose that Q_{ij} and v_i be as above. Then $h := \sum_{n=0}^\infty Q^n v$ is the unique non-negative minimal solution to the linear equations, $x = Qx + v$.*

Proof. Let us start with a heuristic proof that h satisfies, $h = Qh + v$. Formally we have $\sum_{n=0}^\infty Q^n = (1-Q)^{-1}$ so that $h = (1-Q)^{-1}v$ and therefore, $(1-Q)h = v$, i.e. $h = Qh + v$. The problem with this proof is that $(1-Q)$ may not be invertible.

Rigorous proof. We simply have

$$h - Qh = \sum_{n=0}^\infty Q^n v - \sum_{n=1}^\infty Q^n v = v.$$

Now suppose that $x = v + Qx$ with $x_i \ge 0$ for all i. Iterating this equation shows,

$$\begin{aligned} x &= v + Q(Qx+v) = v + Qv + Q^2x \\ x &= v + Qv + Q^2(Qx+v) = v + Qv + Q^2v + Q^3x \\ &\vdots \\ x &= \sum_{n=0}^N Q^n v + Q^{N+1}x \ge \sum_{n=0}^N Q^n v, \end{aligned}$$

where for the last inequality we have used $\left[Q^{N+1}x\right]_i \ge 0$ for all N and $i\in A^c$. Letting $N\to\infty$ in this last equation then shows that

$$x \ge \lim_{N\to\infty}\sum_{n=0}^N Q^n v = \sum_{n=0}^\infty Q^n v = h$$

so that $h_i \le x_i$ for all i. ■

3.2 First Step Analysis Examples

To simulate chains with at most 4 states, you might want to go to:

http://people.hofstra.edu/Stefan_Waner/markov/markov.html

Example 3.23. Consider the Markov chain determined by

$$P = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array}$$

$$P = \left[\begin{array}{cccc} 0 & 1/3 & 1/3 & 1/3 \\ 3/4 & 1/8 & 1/8 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}$$

Notice that 3 and 4 are absorbing states. Let $h_i = P_i\left(X_n \text{ hits } 3\right)$ for $i = 1, 2, 3, 4$. Clearly $h_3 = 1$ while $h_4 = 0$ and by the first step analysis we have

$$\begin{aligned} h_1 &= \frac{1}{3}h_2 + \frac{1}{3}h_3 + \frac{1}{3}h_4 = \frac{1}{3}h_2 + \frac{1}{3} \\ h_2 &= \frac{3}{4}h_1 + \frac{1}{8}h_2 + \frac{1}{8}h_3 = \frac{3}{4}h_1 + \frac{1}{8}h_2 + \frac{1}{8} \end{aligned}$$

i.e.

$$\begin{aligned} h_1 &= \frac{1}{3}h_2 + \frac{1}{3} \\ h_2 &= \frac{3}{4}h_1 + \frac{1}{8}h_2 + \frac{1}{8} \end{aligned}$$

which have solutions,

$$\begin{aligned} P_1\left(X_n \text{ hits } 3\right) &= h_1 = \frac{8}{15} \cong 0.533\,33 \\ P_2\left(X_n \text{ hits } 3\right) &= h_2 = \frac{3}{5}. \end{aligned}$$

Similarly if we let $h_i = P_i\left(X_n \text{ hits } 4\right)$ instead, from the above equations with $h_3 = 0$ and $h_4 = 1$, we find

$$\begin{aligned} h_1 &= \frac{1}{3}h_2 + \frac{1}{3} \\ h_2 &= \frac{3}{4}h_1 + \frac{1}{8}h_2 \end{aligned}$$

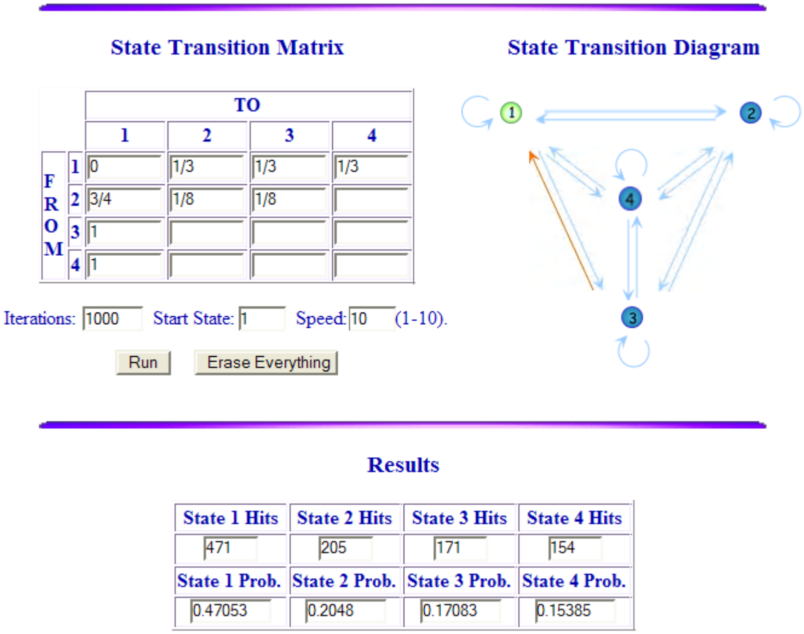
which has solutions,

$$\begin{aligned} P_1\left(X_n \text{ hits } 4\right) &= h_1 = \frac{7}{15} \text{ and} \\ P_2\left(X_n \text{ hits } 4\right) &= h_2 = \frac{2}{5}. \end{aligned}$$

Of course we did not really need to compute these, since

$$\begin{aligned} P_1\left(X_n \text{ hits } 3\right) + P_1\left(X_n \text{ hits } 4\right) &= 1 \text{ and} \\ P_2\left(X_n \text{ hits } 3\right) + P_2\left(X_n \text{ hits } 4\right) &= 1. \end{aligned}$$

The output of one simulation is in Figure 3.3 below.



State Transition Matrix

		TO			
		1	2	3	4
FROM	1	0	1/3	1/3	1/3
	2	3/4	1/8	1/8	
	3	1			
	4	1			

Iterations: 1000 Start State: 1 Speed: 10 (1-10).

State Transition Diagram

Results

State 1 Hits	State 2 Hits	State 3 Hits	State 4 Hits
471	205	171	154
State 1 Prob.	State 2 Prob.	State 3 Prob.	State 4 Prob.
0.47053	0.2048	0.17083	0.15385

Fig. 3.3. In this run, rather than making sites 3 and 4 absorbing, we have made them transition back to 1. I claim now to get an approximate value for $P_1\left(X_n \text{ hits } 3\right)$ we should compute: (State 3 Hits)/(State 3 Hits + State 4 Hits). In this example we will get $171/(171 + 154) = 0.526\,15$ which is a little lower than the predicted value of 0.533. You can try your own runs of this simulator.

3.2.1 A rat in a maze example Problem 5 on p.131.

Here is the maze

$$\begin{bmatrix} 1 & 2 & 3(\text{food}) \\ 4 & 5 & 6 \\ 7(\text{Shock}) & & \end{bmatrix}$$

in which the rat moves from nearest neighbor locations probability being $1/D$ where D is the number of doors in the room that the rat is currently in. The transition matrix is therefore,

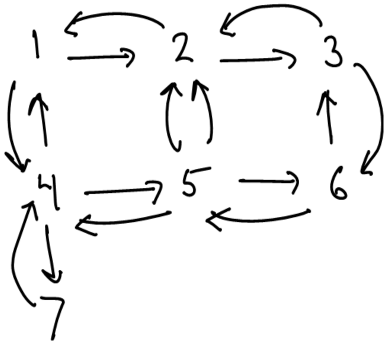


Fig. 3.4. Rat in a maze.

$$P = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}.$$

and the corresponding jump diagram is given in Figure 3.4

Given we want to stop when the rat is either shocked or gets the food, we first delete rows 3 and 7 from P and form Q and R from this matrix by taking columns $1, 2, 4, 5, 6$ and $3, 7$ respectively as in Remark 3.12. This gives,

$$Q = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{ccccc} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}$$

and

$$R = \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \left[\begin{array}{cc} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

Therefore,

$$I - Q = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{3} & 1 & 0 & -\frac{1}{3} & 0 \\ -\frac{1}{3} & 0 & 1 & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix},$$

$$(I - Q)^{-1} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} \\ \left[\begin{array}{ccccc} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array},$$

$$(I - Q)^{-1} \mathbf{1} = \begin{bmatrix} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{17}{3} \\ \frac{14}{3} \\ \frac{14}{3} \\ \frac{16}{3} \\ \frac{11}{3} \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array},$$

and

$$(I - Q)^{-1} R = \begin{bmatrix} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{bmatrix}$$

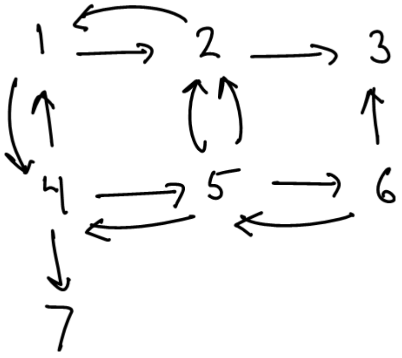
$$= \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \left[\begin{array}{cc} \frac{7}{12} & \frac{5}{12} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{5}{12} & \frac{7}{12} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{5}{6} & \frac{1}{6} \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}.$$

Hence we conclude, for example, that $\mathbb{E}_4 T = \frac{14}{3}$ and $P_4 \left(X_T = 3\right) = 5/12$ and the expected number of visits to site 5 starting at 4 is 1.

Let us now also work out the hitting probabilities,

$$h_i = P_i \left(X_n \text{ hits } 3 = \text{food before } 7 = \text{shock}\right),$$

in this example. To do this we make both 3 and 7 absorbing states so the jump diagram is in Figure 3.2.1. Therefore,



$$h_6 = \frac{1}{2}(1 + h_5)$$
$$h_5 = \frac{1}{3}(h_2 + h_4 + h_6)$$
$$h_4 = \frac{1}{2}h_1$$
$$h_2 = \frac{1}{3}(1 + h_1 + h_5)$$
$$h_1 = \frac{1}{2}(h_2 + h_4).$$

The solutions to these equations are,

$$h_1 = \frac{4}{9},\ h_2 = \frac{2}{3},\ h_4 = \frac{2}{9},\ h_5 = \frac{5}{9},\ h_6 = \frac{7}{9}. \tag{3.23}$$

Similarly if $h_i = P_i\,(X_n \text{ hits 7 before 3})$ we have $h_7 = 1$, $h_3 = 0$ and

$$h_6 = \frac{1}{2}h_5$$
$$h_5 = \frac{1}{3}(h_2 + h_4 + h_6)$$
$$h_4 = \frac{1}{2}(h_1 + 1)$$
$$h_2 = \frac{1}{3}(h_1 + h_5)$$
$$h_1 = \frac{1}{2}(h_2 + h_4)$$

whose solutions are

$$h_1 = \frac{5}{9},\ h_2 = \frac{1}{3},\ h_4 = \frac{7}{9},\ h_5 = \frac{4}{9},\ h_6 = \frac{2}{9}. \tag{3.24}$$

Notice that the sum of the hitting probabilities in Eqs. (3.23) and (3.24) add up to 1 as they should.

3.2.2 A modification of the previous maze

Here is the modified maze,

$$\begin{bmatrix} 1 & 2 & 3(\text{food}) \\ 4 & 5 & \\ 6(\text{Shock}) & & \end{bmatrix}.$$

The transition matrix with 3 and 6 made into absorbing states[3] is:

$$P = \begin{array}{cc} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \end{array},$$

$$Q = \begin{array}{cc} \begin{matrix} 1 & 2 & 4 & 5 \end{matrix} & \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \end{matrix} \end{array}, \quad R = \begin{array}{cc} \begin{matrix} 3 & 6 \end{matrix} & \\ \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \end{matrix} \end{array}$$

$$(I_4 - Q)^{-1} = \begin{array}{cc} \begin{matrix} 1 & 2 & 4 & 5 \end{matrix} & \\ \begin{bmatrix} 2 & \frac{3}{2} & \frac{3}{2} & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & \frac{3}{2} & \frac{3}{2} & 2 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \end{matrix} \end{array},$$

$$(I_4 - Q)^{-1} R = \begin{array}{cc} \begin{matrix} 3 & 6 \end{matrix} & \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 5 \end{matrix} \end{array},$$

---

[3] It is not necessary to make states 3 and 6 absorbing. In fact it does matter at all what the transition probabilites are for the chain for leaving either of the states 3 or 6 since we are going to stop when we hit these states. This is reflected in the fact that the first thing we will do in the first step analysis is to delete rows 3 and 6 from P. Making 3 and 6 absorbing simply saves a little ink.

$$(I_4 - Q)^{-1}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{array}{cc}\begin{bmatrix}6\\5\\5\\6\end{bmatrix} & \begin{matrix}1\\2\\4\\5\end{matrix}\end{array}.$$

So for example, $P_4(X_T = 3(\text{food})) = 1/3$, $E_4(\text{Number of visits to 1}) = 1$, $E_5(\text{Number of visits to 2}) = 3/2$ and $E_1T = E_5T = 6$ and $E_2T = E_4T = 5$.

4

Long Run Behavior of Discrete Markov Chains

For this chapter, X_n will be a Markov chain with a finite or countable state space, S. To each state $i \in S$, let

$$R_i := \min\{n \geq 1 : X_n = i\} \tag{4.1}$$

be the **first passage time of the chain to site** i, and

$$M_i := \sum_{n \geq 1} 1_{X_n = i} \tag{4.2}$$

be number of visits of $\{X_n\}_{n \geq 1}$ to site i.

Definition 4.1. *A state j is accessible from i (written $i \to j$) iff $P_i(R_j < \infty) > 0$ and $i \longleftrightarrow j$ (i communicates with j) iff $i \to j$ and $j \to i$. Notice that $i \to j$ iff there is a path, $i = x_0, x_1, \dots, x_n = j \in S$ such that $p(x_0, x_1)\, p(x_1, x_2) \dots p(x_{n-1}, x_n) > 0$.*

Definition 4.2. *For each $i \in S$, let $C_i := \{j \in S : i \longleftrightarrow j\}$ be the **communicating class of** i. The state space, S, is partitioned into a disjoint union of its communicating classes.*

Definition 4.3. *A communicating class $C \subset S$ is **closed** provided the probability that X_n leaves C given that it started in C is zero. In other words $P_{ij} = 0$ for all $i \in C$ and $j \notin C$. (Notice that if C is closed, then X_n restricted to C is a Markov chain.)*

Definition 4.4. *A state $i \in S$ is:*

1. ***transient** if $P_i(R_i < \infty) < 1$,*
2. ***recurrent** if $P_i(R_i < \infty) = 1$,*
    a) ***positive recurrent** if $1/(\mathbb{E}_i R_i) > 0$, i.e. $\mathbb{E}_i R_i < \infty$,*
    b) ***null recurrent** if it is recurrent ($P_i(R_i < \infty) = 1$) and $1/(\mathbb{E}_i R_i) = 0$, i.e. $\mathbb{E} R_i = \infty$.*

We let S_t, S_r, S_{pr}, and S_{nr} be the transient, recurrent, positive recurrent, and null recurrent states respectively.

The next two sections give the main results of this chapter along with some illustrative examples. The remaining sections are devoted to some of the more technical aspects of the proofs.

4.1 The Main Results

Proposition 4.5 (Class properties). *The notions of being recurrent, positive recurrent, null recurrent, or transient are all class properties. Namely if $C \subset S$ is a communicating class then either all $i \in C$ are recurrent, positive recurrent, null recurrent, or transient. Hence it makes sense to refer to C as being either recurrent, positive recurrent, null recurrent, or transient.*

Proof. See Proposition 4.13 for the assertion that being recurrent or transient is a class property. For the fact that positive and null recurrence is a class property, see Proposition 4.46 below. ■

Lemma 4.6. *Let $C \subset S$ be a communicating class. Then*

$$C \text{ not closed} \implies C \text{ is transient}$$

or equivalently put,

$$C \text{ is recurrent} \implies C \text{ is closed.}$$

Proof. If C is not closed and $i \in C$, there is a $j \notin C$ such that $i \to j$, i.e. there is a path $i = x_0, x_1, \dots, x_n = j$ with all of the $\{x_j\}_{j=0}^n$ being distinct such that

$$P_i(X_0 = i, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n = j) > 0.$$

Since $j \notin C$ we must have $j \nrightarrow C$ and therefore on the event,

$$A := \{X_0 = i, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n = j\},$$

$X_m \notin C$ for all $m \geq n$ and therefore $R_i = \infty$ on the event A which has positive probability. ■

Proposition 4.7. *Suppose that $C \subset S$ is a finite communicating class and $T = \inf\{n \geq 0 : X_n \notin C\}$ be the first exit time from C. If C is not closed, then not only is C transient but $\mathbb{E}_i T < \infty$ for all $i \in C$. We also have the equivalence of the following statements:*

1. *C is closed.*
2. *C is positive recurrent.*

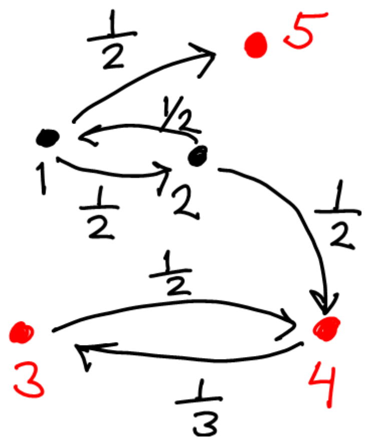
3. C is recurrent.

In particular if $\#(S) < \infty$, then the recurrent (= positively recurrent) states are precisely the union of the closed communication classes and the transient states are what is left over.

Proof. These results follow fairly easily from Proposition 3.15. Also see Corollary 4.20 for another proof. ■

Remark 4.8. Let $\{X_n\}_{n=0}^{\infty}$ denote the fair random walk on $\{0,1,2,\dots\}$ with 0 being an absorbing state. The communication classes are $\{0\}$ and $\{1,2,\dots\}$ with the latter class not being closed and hence transient. Using Remark 0.1, it follows that $\mathbb{E}_i T = \infty$ for all $i > 0$ which shows we can not drop the assumption that $\#(C) < \infty$ in the first statement in Proposition 4.7. Similarly, using the fair random walk example, we see that it is not possible to drop the condition that $\#(C) < \infty$ for the equivalence statements as well.

Example 4.9. Let P be the Markov matrix with jump diagram given in Figure 4.9. In this case the communication classes are $\{\{1,2\},\{3,4\},\{5\}\}$. The latter two are closed and hence positively recurrent while $\{1,2\}$ is transient.



Warning: if $C \subset S$ is closed and $\#(C) = \infty$, C could be recurrent or it could be transient. Transient in this case means the walk goes off to "infinity." The following proposition is a consequence of the strong Markov property in Corollary 4.42.

Proposition 4.10. *If $j \in S$, $k \in \mathbb{N}$, and $\nu : S \to [0,1]$ is any probability on S, then*

$$P_\nu(M_j \geq k) = P_\nu(R_j < \infty) \cdot P_j(R_j < \infty)^{k-1}. \tag{4.3}$$

Proof. Intuitively, $M_j \geq k$ happens iff the chain first visits j with probability $P_\nu(R_j < \infty)$ and then revisits j again $k-1$ times which the probability of each revisit being $P_j(R_j < \infty)$. Since Markov chains are forgetful, these probabilities are all independent and hence we arrive at Eq. (4.3). See Proposition 4.43 below for the formal proof based on the strong Markov property in Corollary 4.42. ■

Corollary 4.11. *If $j \in S$ and $\nu : S \to [0,1]$ is any probability on S, then*

$$P_\nu(M_j = \infty) = P_\nu(X_n = j \text{ i.o.}) = P_\nu(R_j < \infty)\, 1_{j \in S_r}, \tag{4.4}$$

$$P_j(M_j = \infty) = P_j(X_n = j \text{ i.o.}) = 1_{j \in S_r}, \tag{4.5}$$

$$\mathbb{E}_\nu M_j = \sum_{n=1}^{\infty} \sum_{i \in S} \nu(i)\, P_{ij}^n = \frac{P_\nu(R_j < \infty)}{1 - P_j(R_j < \infty)}, \tag{4.6}$$

and

$$\mathbb{E}_i M_j = \sum_{n=1}^{\infty} P_{ij}^n = \frac{P_i(R_j < \infty)}{1 - P_j(R_j < \infty)} \tag{4.7}$$

where the following conventions are used in interpreting the right hand side of Eqs. (4.6) and (4.7): $a/0 := \infty$ if $a > 0$ while $0/0 := 0$.

Proof. Since

$$\{M_j \geq k\} \downarrow \{M_j = \infty\} = \{X_n = j \text{ i.o. } n\} \text{ as } k \uparrow \infty,$$

it follows, using Eq. (4.3), that

$$P_\nu(X_n = j \text{ i.o. } n) = \lim_{k\to\infty} P_\nu(M_j \geq k) = P_\nu(R_j < \infty) \cdot \lim_{k\to\infty} P_j(R_j < \infty)^{k-1} \tag{4.8}$$

which gives Eq. (4.4). Equation (4.5) follows by taking $\nu = \delta_j$ in Eq. (4.4) and recalling that $j \in S_r$ iff $P_j(R_j < \infty) = 1$. Similarly Eq. (4.7) is a special case of Eq. (4.6) with $\nu = \delta_i$. We now prove Eq. (4.6).

Using the definition of M_j in Eq. (4.2),

$$\begin{aligned}\mathbb{E}_\nu M_j &= \mathbb{E}_\nu \sum_{n \geq 1} 1_{X_n = j} = \sum_{n \geq 1} \mathbb{E}_\nu 1_{X_n = j} \\ &= \sum_{n \geq 1} P_\nu(X_n = j) = \sum_{n=1}^{\infty} \sum_{j \in S} \nu(j)\, P_{jj}^n\end{aligned}$$

which is the first equality in Eq. (4.6). For the second, observe that

$$\sum_{k=1}^{\infty} P_\nu(M_j \geq k) = \sum_{k=1}^{\infty} \mathbb{E}_\nu 1_{M_j \geq k} = \mathbb{E}_\nu \sum_{k=1}^{\infty} 1_{k \leq M_j} = \mathbb{E}_\nu M_j.$$

On the other hand using Eq. (4.3) we have

$$\sum_{k=1}^{\infty} P_\nu(M_j \geq k) = \sum_{k=1}^{\infty} P_\nu(R_j < \infty) P_j(R_j < \infty)^{k-1} = \frac{P_\nu(R_j < \infty)}{1 - P_j(R_j < \infty)}$$

provided $a/0 := \infty$ if $a > 0$ while $0/0 := 0$. ■

It is worth remarking that if $j \in S_t$, then Eq. (4.6) asserts that

$$\mathbb{E}_\nu M_j = (\text{the expected number of visits to } j) < \infty$$

which then implies that M_j is a finite valued random variable almost surely. Hence, for almost all sample paths, X_n can visit j at most a finite number of times.

Theorem 4.12 (Recurrent States). *Let $j \in S$. Then the following are equivalent;*

1. *j is recurrent, i.e. $P_j(R_j < \infty) = 1$,*
2. *$P_j(X_n = j \text{ i.o. } n) = 1$,*
3. *$\mathbb{E}_j M_j = \sum_{n=1}^{\infty} P_{jj}^n = \infty$.*

Proof. The equivalence of the first two items follows directly from Eq. (4.5) and the equivalent of items 1. and 3. follows directly from Eq. (4.7) with $i = j$. ■

Proposition 4.13. *If $i \longleftrightarrow j$, then i is recurrent iff j is recurrent, i.e. the property of being recurrent or transient is a class property.*

Proof. Since i and j communicate, there exists α and β in $\mathbb{N}$ such that $P_{ij}^{\alpha} > 0$ and $P_{ji}^{\beta} > 0$. Therefore

$$\sum_{n \geq 1} P_{ii}^{n+\alpha+\beta} \geq \sum_{n \geq 1} P_{ij}^{\alpha} P_{jj}^{n} P_{ji}^{\beta}$$

which shows that $\sum_{n\geq 1} P_{jj}^n = \infty \Longrightarrow \sum_{n \geq 1} P_{ii}^n = \infty$. Similarly $\sum_{n \geq 1} P_{ii}^n = \infty \Longrightarrow \sum_{n\geq 1} P_{jj}^n = \infty$. Thus using item 3. of Theorem 4.12, it follows that i is recurrent iff j is recurrent. ■

Corollary 4.14. *If $C \subset S_r$ is a recurrent communication class, then*

$$P_i(R_j < \infty) = 1 \text{ for all } i, j \in C \tag{4.9}$$

and in fact

$$P_i(\cap_{j \in C}\{X_n = j \text{ i.o. } n\}) = 1 \text{ for all } i \in C. \tag{4.10}$$

More generally if $\nu : S \to [0,1]$ is a probability such that $\nu(i) = 0$ for $i \notin C$, then

$$P_\nu(\cap_{j \in C}\{X_n = j \text{ i.o. } n\}) = 1 \text{ for all } i \in C. \tag{4.11}$$

In words, if we start in C then every state in C is visited an infinite number of times. (Notice that $P_i(R_j < \infty) = P_i(\{X_n\}_{n \geq 1}$ hits j).)

Proof. Let $i, j \in C \subset S_r$ and choose $m \in \mathbb{N}$ such that $P_{ji}^m > 0$. Since $P_j(M_j = \infty) = 1$ and

$$\begin{aligned} &\{X_m = i \text{ and } X_n = j \text{ for some } n > m\} \\ &\quad = \sum_{n > m} \{X_m = i, X_{m+1} \neq j, \ldots, X_{n-1} \neq j, X_n = j\}, \end{aligned}$$

we have

$$\begin{aligned} P_{ji}^m &= P_j(X_m = i) = P_j(M_j = \infty, X_m = i) \\ &\leq P_j(X_m = i \text{ and } X_n = j \text{ for some } n > m) \\ &= \sum_{n > m} P_j(X_m = i, X_{m+1} \neq j, \ldots, X_{n-1} \neq j, X_n = j) \\ &= \sum_{n > m} P_{ji}^m P_i(X_1 \neq j, \ldots, X_{n-m-1} \neq j, X_{n-m} = j) \\ &= \sum_{n > m} P_{ji}^m P_i(R_j = n - m) = P_{ji}^m \sum_{k=1}^{\infty} P_i(R_j = k) \\ &= P_{ji}^m P_i(R_j < \infty). \end{aligned} \tag{4.12}$$

Because $P_{ji}^m > 0$, we may conclude from Eq. (4.12) that $1 \leq P_i(R_j < \infty)$, i.e. that $P_i(R_j < \infty) = 1$ and Eq. (4.9) is proved. Feeding this result back into Eq. (4.4) with $\nu = \delta_i$ shows $P_i(M_j = \infty) = 1$ for all $i, j \in C$ and therefore, $P_i(\cap_{j \in C}\{M_j = \infty\}) = 1$ for all $i \in C$ which is Eq. (4.10). Equation (4.11) follows by multiplying Eq. (4.10) by $\nu(i)$ and then summing on $i \in C$. ■

Theorem 4.15 (Transient States). *Let $j \in S$. Then the following are equivalent;*

1. *j is transient, i.e. $P_j(R_j < \infty) < 1$,*
2. *$P_j(X_n = j \text{ i.o. } n) = 0$, and*

3. $\mathbb{E}_j M_j = \sum_{n=1}^{\infty} P_{jj}^n < \infty$.

Moreover, if $i \in S$ *and* $j \in S_t$, *then*

$$\sum_{n=1}^{\infty} P_{ij}^n = \mathbb{E}_i M_j < \infty \implies \begin{cases} \lim_{n\to\infty} P_{ij}^n = 0 \\ P_i\left(X_n = j \text{ i.o. } n\right) = 0. \end{cases} \tag{4.13}$$

and more generally if $\nu : S \to [0,1]$ *is any probability, then*

$$\sum_{n=1}^{\infty} P_\nu\left(X_n = j\right) = \mathbb{E}_\nu M_j < \infty \implies \begin{cases} \lim_{n\to\infty} P_\nu\left(X_n = j\right) = 0 \\ P_\nu\left(X_n = j \text{ i.o. } n\right) = 0. \end{cases} \tag{4.14}$$

Proof. The equivalence of the first two items follows directly from Eq. (4.5) and the equivalent of items 1. and 3. follows directly from Eq. (4.7) with $i = j$. The fact that $\mathbb{E}_i M_j < \infty$ and $\mathbb{E}_\nu M_j < \infty$ for all $j \in S_t$ are consequences of Eqs. (4.7) and (4.6) respectively. The remaining implication in Eqs. (4.13) and (4.6) follow from the first Borel Cantelli Lemma 1.5 and the fact that n^{th} – term in a convergent series tends to zero as $n \to \infty$. ■

Corollary 4.16. *1) If the state space,* S, *is a finite set, then* $S_r \neq \emptyset$. *2) Any finite and closed communicating class* $C \subset S$ *is a recurrent.*

Proof. First suppose that $\#(S) < \infty$ and for the sake of contradiction, suppose $S_r = \emptyset$ or equivalently that $S = S_t$. Then by Theorem 4.15, $\lim_{n\to\infty} P_{ij}^n = 0$ for all $i, j \in S$. On the other hand, $\sum_{j\in S} P_{ij}^n = 1$ so that

$$1 = \lim_{n\to\infty} \sum_{j\in S} P_{ij}^n = \sum_{j\in S} \lim_{n\to\infty} P_{ij}^n = \sum_{j\in S} 0 = 0,$$

which is a contradiction. (Notice that if S were infinite, we could not interchange the limit and the above sum without some extra conditions.)

To prove the first statement, restrict X_n to C to get a Markov chain on a finite state space C. By what we have just proved, there is a recurrent state $i \in C$. Since recurrence is a class property, it follows that all states in C are recurrent. ■

Definition 4.17. *A function,* $\pi : S \to [0,1]$ *is a* ***sub-probability*** *if* $\sum_{j\in S} \pi(j) \leq 1$. *We call* $\sum_{j\in S} \pi(j)$ *the* ***mass*** *of* π. *So a probability is a sub-probability with mass one.*

Definition 4.18. *We say a sub-probability,* $\pi : S \to [0,1]$, *is* ***invariant*** *if* $\pi P = \pi$, *i.e.*

$$\sum_{i\in S} \pi(i) p_{ij} = \pi(j) \text{ for all } j \in S. \tag{4.15}$$

An invariant probability, $\pi : S \to [0,1]$, *is called an* ***invariant distribution.***

Theorem 4.19. *Suppose that* $P = (p_{ij})$ *is an irreducible Markov kernel and* $\pi_j := \frac{1}{\mathbb{E}_j R_j}$ *for all* $j \in S$. *Then:*

1. *For all* $i, j \in S$, *we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} 1_{X_n = j} = \pi_j \quad P_i - a.s. \tag{4.16}$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_i\left(X_n = j\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} P_{ij}^n = \pi_j. \tag{4.17}$$

2. *If* $\mu : S \to [0,1]$ *is an invariant sub-probability, then either* $\mu(i) > 0$ *for all* i *or* $\mu(i) = 0$ *for all* i.
3. P *has at most one invariant distribution.*
4. P *has a (necessarily unique) invariant distribution,* $\mu : S \to [0,1]$, *iff* P *is positive recurrent in which case* $\mu(i) = \pi(i) = \frac{1}{\mathbb{E}_i R_i} > 0$ *for all* $i \in S$.

(These results may of course be applied to the restriction of a general non-irreducible Markov chain to any one of its communication classes.)

Proof. These results are the contents of Theorem 4.45 and Propositions 4.46 and 4.47 below. ■

Using this result we can give another proof of Proposition 4.7.

Corollary 4.20. *If* C *is a closed finite communicating class then* C *is positive recurrent. (Recall that we already know that* C *is recurrent by Corollary 4.16.)*

Proof. For $i, j \in C$, let

$$\pi_j := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_i\left(X_n = j\right) = \frac{1}{\mathbb{E}_j R_j}$$

as in Theorem 4.21. Since C is closed,

$$\sum_{j\in C} P_i\left(X_n = j\right) = 1$$

1and therefore,

$$\sum_{j\in C} \pi_j = \lim_{N\to\infty} \frac{1}{N} \sum_{j\in C} \sum_{n=1}^{N} P_i\left(X_n = j\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{j\in C} P_i\left(X_n = j\right) = 1.$$

Therefore $\pi_j > 0$ for some $j \in C$ and hence all $j \in C$ by Theorem 4.19 with S replaced by C. Hence we have $\mathbb{E}_j R_j < \infty$, i.e. every $j \in C$ is a positive recurrent state. ■

Theorem 4.21 (General Convergence Theorem). *Let $\nu : S \to [0,1]$ be any probability, $i \in S$, C be the communicating class containing i,*

$$\{X_n \text{ hits } C\} := \{X_n \in C \text{ for some } n\},$$

and

$$\pi_i := \pi_i(\nu) = \frac{P_\nu(X_n \text{ hits } C)}{\mathbb{E}_i R_i}, \tag{4.18}$$

where $1/\infty := 0$. Then:

1. *P_ν – a.s.,*
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n = i} = \frac{1}{\mathbb{E}_i R_i} 1_{\{X_n \text{ hits } C\}}, \tag{4.19}$$
2. $$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{j\in S} \nu(j) P_{ji}^n = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\nu(X_n = i) = \pi_i, \tag{4.20}$$
3. *π is an invariant sub-probability for P, and*
4. *the mass of π is*
$$\sum_{i\in S} \pi_i = \sum_{C:\text{ pos. recurrent}} P_\nu(X_n \text{ hits } C) \le 1. \tag{4.21}$$

Proof. If $i \in S$ is a transient site, then according to Eq. (4.14), $P_\nu(M_i < \infty) = 1$ and therefore $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N 1_{X_n = i} = 0$ which agrees with Eq. (4.19) for $i \in S_t$.

So now suppose that $i \in S_r$ and let C be the communication class containing i and

$$T = \inf\{n \ge 0 : X_n \in C\}$$

be the first time when X_n enters C. It is clear that $\{R_i < \infty\} \subset \{T < \infty\}$. On the other hand, for any $j \in C$, it follows by the strong Markov property (Corollary 4.42) and Corollary 4.14 that, conditioned on $\{T < \infty, X_T = j\}$, $\{X_n\}$ hits i i.o. and hence $P(R_i < \infty | T < \infty, X_T = j) = 1$. Equivalently put,

$$P(R_i < \infty, T < \infty, X_T = j) = P(T < \infty, X_T = j) \text{ for all } j \in C.$$

Summing this last equation on $j \in C$ then shows

$$P(R_i < \infty) = P(R_i < \infty, T < \infty) = P(T < \infty)$$

and therefore $\{R_i < \infty\} = \{T < \infty\}$ modulo an event with P_ν – probability zero.

Another application of the strong Markov property (in Corollary 4.42), observing that $X_{R_i} = i$ on $\{R_i < \infty\}$, allows us to conclude that the $P_\nu(\cdot | R_i < \infty) = P_\nu(\cdot | T < \infty)$ – law of $(X_{R_i}, X_{R_i+1}, X_{R_i+2}, \dots)$ is the same as the P_i – law of $(X_0, X_1, X_2, \dots)$. Therefore, we may apply Theorem 4.19 to conclude that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n = i} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_{R_i+n} = i} = \frac{1}{\mathbb{E}_i R_i} \; P_\nu(\cdot | R_i < \infty) \text{ – a.s.}$$

On the other hand, on the event $\{R_i = \infty\}$ we have $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N 1_{X_n = i} = 0$. Thus we have shown P_ν – a.s. that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n = i} = \frac{1}{\mathbb{E}_i R_i} 1_{R_i < \infty} = \frac{1}{\mathbb{E}_i R_i} 1_{T < \infty} = \frac{1}{\mathbb{E}_i R_i} 1_{\{X_n \text{ hits } C\}}$$

which is Eq. (4.19). Taking expectations of this equation, using the dominated convergence theorem, gives Eq. (4.20).

Since $1/\mathbb{E}_i R_i = \infty$ unless i is a positive recurrent site, it follows that

$$\sum_{i\in S} \pi_i P_{ij} = \sum_{i\in S_{\text{pr}}} \pi_i P_{ij} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) \sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij}. \tag{4.22}$$

As each positive recurrent class, C, is closed; if $i \in C$ and $j \notin C$, then $P_{ij} = 0$. Therefore $\sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij}$ is zero unless $j \in C$. So if $j \notin S_{\text{pr}}$ we have $\sum_{i\in S} \pi_i P_{ij} = 0 = \pi_j$ and if $j \in S_{\text{pr}}$, then by Theorem 4.19,

$$\sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij} = 1_{j\in C} \cdot \frac{1}{\mathbb{E}_j R_j}.$$

Using this result in Eq. (4.22) shows that

$$\sum_{i\in S} \pi_i P_{ij} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) 1_{j\in C} \cdot \frac{1}{\mathbb{E}_j R_j} = \pi_j$$

so that π is an invariant distribution. Similarly, using Theorem 4.19 again,

$$\sum_{i\in S} \pi_i = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) \sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C).$$

■

Definition 4.22. *A state $i \in S$ is* ***aperiodic*** *if $P_{ii}^n > 0$ for all n sufficiently large.*

Lemma 4.23. *If $i \in S$ is aperiodic and $j \longleftrightarrow i$, then j is aperiodic. So being aperiodic is a class property.*

Proof. We have

$$P_{jj}^{n+m+k} = \sum_{w,z\in S} P_{j,w}^{n} P_{w,z}^{m} P_{z,j}^{k} \ge P_{j,i}^{n} P_{i,i}^{m} P_{i,j}^{k}.$$

Since $j \longleftrightarrow i$, there exists $n, k \in \mathbb{N}$ such that $P_{j,i}^{n} > 0$ and $P_{i,j}^{k} > 0$. Since $P_{i,i}^{m} > 0$ for all large m, it follows that $P_{jj}^{n+m+k} > 0$ for all large m and therefore, j is aperiodic as well. ∎

Lemma 4.24. *A state $i \in S$ is aperiodic iff 1 is the greatest common divisor of the set,*

$$\{n \in \mathbb{N} : P_i(X_n = i) = P_{ii}^{n} > 0\}.$$

Proof. Use the number theory Lemma 4.48 below. ∎

Theorem 4.25. *If P is an irreducible, aperiodic, and recurrent Markov chain, then*

$$\lim_{n\to\infty} P_{ij}^{n} = \pi_j = \frac{1}{\mathbb{E}_j(R_j)}. \tag{4.23}$$

More generally, if C is an aperiodic communication class, then

$$\lim_{n\to\infty} P_\nu(X_n = i) := \lim_{n\to\infty} \sum_{j\in S} \nu(j) P_{ji}^{n} = P_\nu(R_i < \infty) \frac{1}{\mathbb{E}_j(R_j)} \text{ for all } i \in C.$$

Proof. I will not prove this theorem here but refer the reader to Norris [5, Theorem 1.8.3] or Kallenberg [3, Chapter 8]. The proof given there is by a "coupling argument" is given. ∎

4.1.1 Finite State Space Remarks

For this subsection suppose that $S = \{1, 2, \dots, n\}$ and P_{ij} is a Markov matrix. Some of the previous results have fairly easy proofs in this setting.

Proposition 4.26. *The Markov matrix P has an invariant distribution.*

Proof. If $\mathbf{1} := \left[1\ 1\ \dots\ 1\right]^{\mathrm{tr}}$, then $P\mathbf{1} = \mathbf{1}$ from which it follows that

$$0 = \det(P - I) = \det\left(P^{\mathrm{tr}} - I\right).$$

Therefore there exists a non-zero row vector ν such that $P^{\mathrm{tr}}\nu^{\mathrm{tr}} = \nu^{\mathrm{tr}}$ or equivalently that $\nu P = \nu$. At this point we would be done if we knew that $\nu_i \ge 0$ for all i – but we don't. So let $\pi_i := |\nu_i|$ and observe that

$$\pi_i = |\nu_i| = \left|\sum_{k=1}^{n} \nu_k P_{ki}\right| \le \sum_{k=1}^{n} |\nu_k| P_{ki} \le \sum_{k=1}^{n} \pi_k P_{ki}.$$

We now claim that in fact $\pi = \pi P$. If this were not the case we would have $\pi_i < \sum_{k=1}^{n} \pi_k P_{ki}$ for some i and therefore

$$0 < \sum_{i=1}^{n} \pi_i < \sum_{i=1}^{n}\sum_{k=1}^{n} \pi_k P_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{n} \pi_k P_{ki} = \sum_{k=1}^{n} \pi_k$$

which is a contradiction. So all that is left to do is normalize π_i so $\sum_{i=1}^{n} \pi_i = 1$ and we are done. ∎

Proposition 4.27. *Suppose that P is irreducible. (In this case we may use Proposition 3.15 to show that $\mathbb{E}_i[R_j] < \infty$ for all i, j.) Then there is precisely one invariant distribution, π, which is given by $\pi_i = 1/\mathbb{E}_i R_i > 0$ for all $i \in S$.*

Proof. We begin by using the first step analysis to write equations for $\mathbb{E}_i[R_j]$ as follows:

$$\begin{aligned} \mathbb{E}_i[R_j] &= \sum_{k=1}^{n} \mathbb{E}_i[R_j | X_1 = k] P_{ik} = \sum_{k\ne j} \mathbb{E}_i[R_j | X_1 = k] P_{ik} + P_{ij} 1 \\ &= \sum_{k\ne j} (\mathbb{E}_k[R_j] + 1) P_{ik} + P_{ij} 1 = \sum_{k\ne j} \mathbb{E}_k[R_j] P_{ik} + 1. \end{aligned}$$

and therefore,

$$\mathbb{E}_i[R_j] = \sum_{k\ne j} P_{ik} \mathbb{E}_k[R_j] + 1. \tag{4.24}$$

Now suppose that π is any invariant distribution for P, then multiplying Eq. (4.24) by π_i and summing on i shows

$$\begin{aligned} \sum_{i=1}^{n} \pi_i \mathbb{E}_i[R_j] &= \sum_{i=1}^{n} \pi_i \sum_{k\ne j} P_{ik} \mathbb{E}_k[R_j] + \sum_{i=1}^{n} \pi_i 1 \\ &= \sum_{k\ne j} \pi_k \mathbb{E}_k[R_j] + 1 \end{aligned}$$

from which it follows that $\pi_j \mathbb{E}_j[R_j] = 1$. ∎

We may use Eq. (4.24) to compute $\mathbb{E}_i[R_j]$ in examples. To do this, fix j and set $v_i := \mathbb{E}_i R_j$. Then Eq. (4.24) states that $v = P^{(j)} v + \mathbf{1}$ where $P^{(j)}$ denotes P with the j^{th} – column replaced by all zeros. Thus we have

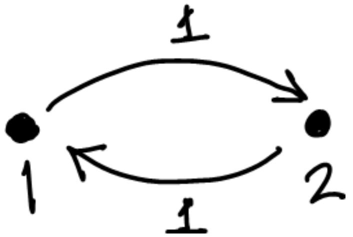
$$(\mathbb{E}_i R_j)_{i=1}^{n} = \left(I - P^{(j)}\right)^{-1} \mathbf{1}, \tag{4.25}$$

i.e.

$$\begin{bmatrix} \mathbb{E}_1 R_j \\ \vdots \\ \mathbb{E}_n R_j \end{bmatrix} = \left(I - P^{(j)}\right)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{4.26}$$

4.2 Examples

Example 4.28. Let $S = \{1, 2\}$ and $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with jump diagram in Figure 4.28. In this case $P^{2n} = I$ while $P^{2n+1} = P$ and therefore $\lim_{n\to\infty} P^n$ does not have a limit. On the other hand it is easy to see that the invariant distribution, π, for P is $\pi = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$. Moreover it is easy to see that



$$\frac{P + P^2 + \cdots + P^N}{N} \to \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \end{bmatrix}.$$

Let us compute

$$\begin{bmatrix} \mathbb{E}_1 R_1 \\ \mathbb{E}_2 R_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbb{E}_1 R_2 \\ \mathbb{E}_2 R_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

so that indeed, $\pi_1 = 1/\mathbb{E}_1 R_1$ and $\pi_2 = 1/\mathbb{E}_2 R_2$.

Example 4.29. Again let $S = \{1, 2\}$ and $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ with jump diagram in Figure 4.29. In this case the chain is not irreducible and every $\pi = [a\ b]$ with $a + b = 1$ and $a, b \geq 0$ is an invariant distribution.

Example 4.30. Suppose that $S = \{1, 2, 3\}$, and

$$P = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}$$

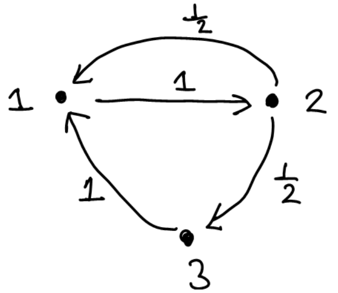


Fig. 4.1. A simple jump diagram.

has the jump graph given by 4.1. Notice that $P^2_{11} > 0$ and $P^3_{11} > 0$ that P is "aperiodic." We now find the invariant distribution,

$$\mathrm{Nul}\,(P - I)^{\mathrm{tr}} = \mathrm{Nul} \begin{bmatrix} -1 & \frac{1}{2} & 1 \\ 1 & -1 & 0 \\ 0 & \frac{1}{2} & -1 \end{bmatrix} = \mathbb{R} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

Therefore the invariant distribution is given by

$$\pi = \frac{1}{5}\begin{bmatrix} 2 & 2 & 1 \end{bmatrix}.$$

Let us now observe that

$$P^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$P^{20} = \begin{bmatrix} \frac{409}{1024} & \frac{205}{512} & \frac{205}{1024} \\ \frac{205}{512} & \frac{409}{1024} & \frac{205}{1024} \\ \frac{205}{512} & \frac{205}{512} & \frac{51}{256} \end{bmatrix} = \begin{bmatrix} 0.399\,41 & 0.400\,39 & 0.200\,20 \\ 0.400\,39 & 0.399\,41 & 0.200\,20 \\ 0.400\,39 & 0.400\,39 & 0.199\,22 \end{bmatrix}.$$

Let us also compute $\mathbb{E}_2 R_3$ via,

$$\begin{bmatrix}\mathbb{E}_1 R_3\\ \mathbb{E}_2 R_3\\ \mathbb{E}_3 R_3\end{bmatrix} = \left(\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} - \begin{bmatrix}0&1&0\\1/2&0&0\\1&0&0\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}4\\3\\5\end{bmatrix}$$

so that

$$\frac{1}{\mathbb{E}_3 R_3} = \frac{1}{5} = \pi_3.$$

Example 4.31. The transition matrix,

$$P = \begin{array}{c} \begin{array}{ccc}1 & 2 & 3\end{array} \\ \begin{bmatrix}1/4&1/2&1/4\\1/2&0&1/2\\1/3&1/3&1/3\end{bmatrix}\begin{array}{c}1\\2\\3\end{array}\end{array}$$

is represented by the jump diagram in Figure 4.2. This chain is aperiodic. We find the invariant distribution as,

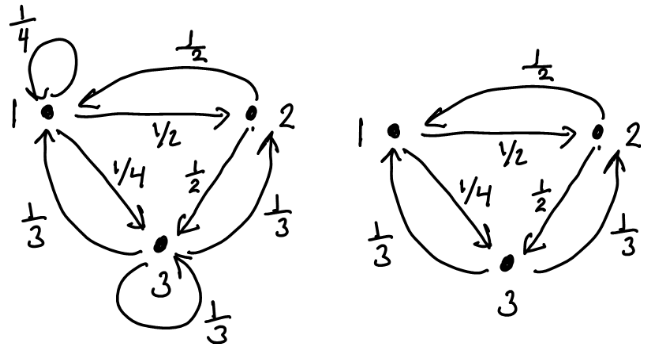


Fig. 4.2. The above diagrams contain the same information. In the one on the right we have dropped the jumps from a site back to itself since these can be deduced by conservation of probability.

$$\begin{aligned}\operatorname{Nul}(P-I)^{\mathrm{tr}} &= \operatorname{Nul}\left(\begin{bmatrix}1/4&1/2&1/4\\1/2&0&1/2\\1/3&1/3&1/3\end{bmatrix} - \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)^{\mathrm{tr}}\\ &= \operatorname{Nul}\left(\begin{bmatrix}-\frac{3}{4}&\frac{1}{2}&\frac{1}{3}\\ \frac{1}{2}&-1&\frac{1}{3}\\ \frac{1}{4}&\frac{1}{2}&-\frac{2}{3}\end{bmatrix}\right) = \mathbb{R}\begin{bmatrix}1\\ \frac{5}{6}\\ 1\end{bmatrix} = \mathbb{R}\begin{bmatrix}6\\5\\6\end{bmatrix}\end{aligned}$$

$$\pi = \frac{1}{17}\begin{bmatrix}6 & 5 & 6\end{bmatrix} = \begin{bmatrix}0.352\,94 & 0.294\,12 & 0.352\,94\end{bmatrix}.$$

In this case

$$P^{10} = \begin{bmatrix}1/4&1/2&1/4\\1/2&0&1/2\\1/3&1/3&1/3\end{bmatrix}^{10} = \begin{bmatrix}0.352\,98&0.294\,04&0.352\,98\\0.352\,89&0.294\,23&0.352\,89\\0.352\,95&0.294\,1&0.352\,95\end{bmatrix}.$$

Let us also compute

$$\begin{bmatrix}\mathbb{E}_1 R_2\\ \mathbb{E}_2 R_2\\ \mathbb{E}_3 R_2\end{bmatrix} = \left(\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} - \begin{bmatrix}1/4&0&1/4\\1/2&0&1/2\\1/3&0&1/3\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}\frac{11}{5}\\ \frac{17}{5}\\ \frac{13}{5}\end{bmatrix}$$

so that

$$1/\mathbb{E}_2 R_2 = 5/17 = \pi_2.$$

Example 4.32. Consider the following Markov matrix,

$$P = \begin{array}{c} \begin{array}{cccc}1 & 2 & 3 & 4\end{array} \\ \begin{bmatrix}1/4&1/4&1/4&1/4\\1/4&0&0&3/4\\1/2&1/2&0&0\\0&1/4&3/4&0\end{bmatrix}\begin{array}{c}1\\2\\3\\4\end{array}\end{array}$$

with jump diagram in Figure 4.3. Since this matrix is doubly stochastic, we know that $\pi = \frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}$. Let us compute $\mathbb{E}_3 R_3$ as follows

$$\begin{aligned}\begin{bmatrix}\mathbb{E}_1 R_3\\ \mathbb{E}_2 R_3\\ \mathbb{E}_3 R_3\\ \mathbb{E}_4 R_3\end{bmatrix} &= \left(\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} - \begin{bmatrix}1/4&1/4&0&1/4\\1/4&0&0&3/4\\1/2&1/2&0&0\\0&1/4&0&0\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\\ &= \begin{bmatrix}\frac{50}{17}\\ \frac{52}{17}\\ 4\\ \frac{30}{17}\end{bmatrix}\end{aligned}$$

so that $\mathbb{E}_3 R_3 = 4 = 1/\pi_4$ as it should. Similarly,

$$\begin{aligned}\begin{bmatrix}\mathbb{E}_1 R_2\\ \mathbb{E}_2 R_2\\ \mathbb{E}_3 R_2\\ \mathbb{E}_4 R_2\end{bmatrix} &= \left(\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} - \begin{bmatrix}1/4&0&1/4&1/4\\1/4&0&0&3/4\\1/2&0&0&0\\0&0&3/4&0\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\\ &= \begin{bmatrix}\frac{54}{17}\\ 4\\ \frac{44}{17}\\ \frac{50}{17}\end{bmatrix}\end{aligned}$$

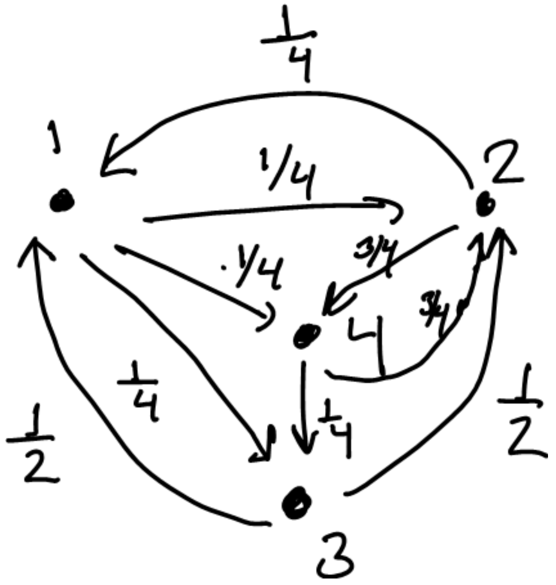


Fig. 4.3. The jump diagram for P.

and again $\mathbb{E}_2 R_2 = 4 = 1/\pi_2$.

Example 4.33 (Analyzing a non-irreducible Markov chain). In this example we are going to analyze the limiting behavior of the non-irreducible Markov chain determined by the Markov matrix,

$$P = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}.$$

Here are the steps to follow.

1. **Find the jump diagram for** P. In our case it is given in Figure 4.4.
2. **Identify the communication classes**. In our example they are $\{1,2\}$, $\{5\}$, and $\{3,4\}$. The first is not closed and hence transient while the second two are closed and finite sets and hence recurrent.
3. **Find the invariant distributions for the recurrent classes.** For $\{5\}$ it is simply $\pi'_{\{5\}} = [1]$ and for $\{3,4\}$ we must find the invariant distribution for the 2×2 Markov matrix,

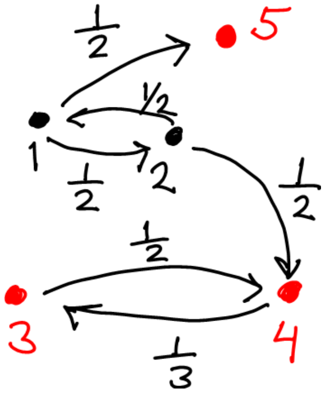


Fig. 4.4. The jump diagram for P above.

$$Q = \begin{array}{c} \begin{matrix} 3 & 4 \end{matrix} \\ \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix} \end{array} \begin{array}{c} \\ 3 \\ 4 \end{array}.$$

We do this in the usual way, namely

$$\operatorname{Nul}\left(I - Q^{\mathrm{tr}}\right) = \operatorname{Nul}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{2}{3} \end{bmatrix}\right) = \mathbb{R}\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

so that $\pi'_{\{3,4\}} = \frac{1}{5}\left[2\ 3\right]$.

4. We can turn $\pi'_{\{3,4\}}$ and $\pi'_{\{5\}}$ into invariant distributions for P by padding the row vectors with zeros to get

$$\pi_{\{3,4\}} = \left[0\ 0\ 2/5\ 3/5\ 0\right]$$
$$\pi_{\{5\}} = \left[0\ 0\ 0\ 0\ 1\right].$$

The general invariant distribution may then be written as;

$$\pi = \alpha\pi_{\{5\}} + \beta\pi_{\{3,4\}} \text{ with } \alpha, \beta \geq 0 \text{ and } \alpha + \beta = 1.$$

5. We can now work out the $\lim_{n\to\infty} P^n$. If we start at site i we are considering the i^{th} – row of $\lim_{n\to\infty} P^n$. If we start in the recurrent class $\{3,4\}$ we will simply get $\pi_{\{3,4\}}$ for these rows and we start in the recurrent class $\{5\}$ we will get $\pi_{\{5\}}$. However if start in the non-closed transient class, $\{1,2\}$ we have

$$\text{first row of } \lim_{n\to\infty} P^n = P_1\left(X_n \text{ hits } 5\right)\pi_{\{5\}} + P_1\left(X_n \text{ hits } \{3,4\}\right)\pi_{\{3,4\}} \tag{4.27}$$

and

$$\text{second row of } \lim_{n\to\infty} P^n = P_2\left(X_n \text{ hits } 5\right)\pi_{\{5\}} + P_2\left(X_n \text{ hits } \{3,4\}\right)\pi_{\{3,4\}}. \tag{4.28}$$

6. **Compute the required hitting probabilities.** Let us begin by computing the fraction of one pound of sand put at site 1 will end up at site 5, i.e. we want to find $h_1 := P_1\left(X_n \text{ hits } 5\right)$. To do this let $h_i = P_i\left(X_n \text{ hits } 5\right)$ for $i = 1, 2, \dots, 5$. It is clear that $h_5 = 1$, and $h_3 = h_4 = 0$. A first step analysis then shows

$$h_1 = \frac{1}{2} \cdot P_2\left(X_n \text{ hits } 5\right) + \frac{1}{2} P_5\left(X_n \text{ hits } 5\right)$$
$$h_2 = \frac{1}{2} \cdot P_1\left(X_n \text{ hits } 5\right) + \frac{1}{2} P_4\left(X_n \text{ hits } 5\right)$$

which leads to[1]

$$h_1 = \frac{1}{2}h_2 + \frac{1}{2}$$
$$h_2 = \frac{1}{2}h_1 + \frac{1}{2}0.$$

The solutions to these equations are

$$P_1\left(X_n \text{ hits } 5\right) = h_1 = \frac{2}{3} \text{ and } P_2\left(X_n \text{ hits } 5\right) = h_2 = \frac{1}{3}.$$

Since the process is either going to end up in $\{5\}$ or in $\{3,4\}$, we may also conclude that

$$P_1\left(X_n \text{ hits } \{3,4\}\right) = \frac{1}{3} \text{ and } P_2\left(X_n \text{ hits } \{3,4\}\right) = \frac{2}{3}.$$

7. Using these results in Eqs. (4.27) and (4.28) shows,

$$\begin{aligned}\text{first row of } \lim_{n\to\infty} P^n &= \frac{2}{3}\pi_{\{5\}} + \frac{1}{3}\pi_{\{3,4\}} \\ &= \left[0\ 0\ \tfrac{2}{15}\ \tfrac{1}{5}\ 2/3\right] \\ &= \left[0.0\ 0.0\ 0.133\,33\ 0.2\ 0.666\,67\right]\end{aligned}$$

and

$$\begin{aligned}\text{second row of } \lim_{n\to\infty} P^n &= \frac{1}{3}\pi_{\{5\}} + \frac{2}{3}\pi_{\{3,4\}} \\ &= \frac{1}{3}\left[0\ 0\ 0\ 0\ 1\right] + \frac{2}{3}\left[0\ 0\ 2/5\ 3/5\ 0\right] \\ &= \left[0\ 0\ \tfrac{4}{15}\ \tfrac{2}{5}\ \tfrac{1}{3}\right] \\ &= \left[0.0\ 0.0\ 0.266\,67\ 0.4\ 0.333\,33\right].\end{aligned}$$

These answers already compare well with

$$P^{10} = \begin{bmatrix} 9.7656\times10^{-4} & 0.0 & 0.132\,76 & 0.200\,24 & 0.666\,02 \\ 0.0 & 9.7656\times10^{-4} & 0.266\,26 & 0.399\,76 & 0.333\,01 \\ 0.0 & 0.0 & 0.4 & 0.600\,00 & 0.0 \\ 0.0 & 0.0 & 0.400\,00 & 0.6 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

4.3 The Strong Markov Property

In proving the results above, we are going to make essential use of a strong form of the Markov property which asserts that Theorem 3.17 continues to hold even when n is replaced by a random "stopping time."

Definition 4.35 (Stopping times). *Let τ be an $\mathbb{N}_0 \cup \{\infty\}$ - valued random variable which is a functional of a sequence of random variables, $\{X_n\}_{n=0}^{\infty}$ which we write by abuse of notation as, $\tau = \tau\left(X_0, X_1, \dots\right)$. We say that τ is a stopping time if for all $n \in \mathbb{N}_0$, the indicator random variable, $1_{\tau=n}$ is a functional of $\left(X_0, \dots, X_n\right)$. Thus for each $n \in \mathbb{N}_0$ there should exist a function, σ_n such that $1_{\tau=n} = \sigma_n\left(X_0, \dots, X_n\right)$. In other words, the event $\{\tau = n\}$ may be described using only $\left(X_0, \dots, X_n\right)$ for all $n \in \mathbb{N}$.*

Example 4.36. Here are some example of random times which are which are not stopping times. In these examples we will always use the convention that the minimum of the empty set is $+\infty$.

---

[1]

Example 4.34. **Note:** If we were to make use of Theorem 3.21 we would have not set $h_3 = h_4 = 0$ and we would have added the equations,

$$h_3 = \frac{1}{2}h_3 + \frac{1}{2}h_4$$
$$h_4 = \frac{1}{3}h_3 + \frac{2}{3}h_4,$$

to those above. The general solution to these equations is $c\,(1,1)$ for some $c \in \mathbb{R}$ and the non-negative minimal solution is the special case where $c = 0$, i.e. $h_3 = h_4 = 0$. The point is, since $\{3,4\}$ is a closed communication class there is no way to hit 5 starting in $\{3,4\}$ and therefore clearly $h_3 = h_4 = 0$.

1. The random time, $\tau = \min\{k : |X_k| \geq 5\}$ (the first time, k, such that $|X_k| \geq 5$) is a stopping time since
$$\{\tau = k\} = \{|X_1| < 5, \dots, |X_{k-1}| < 5, |X_k| \geq 5\}.$$
2. Let $W_k := X_1 + \dots + X_k$, then the random time,
$$\tau = \min\{k : W_k \geq \pi\}$$
is a stopping time since,
$$\{\tau = k\} = \left\{ \begin{array}{c} W_j = X_1 + \dots + X_j < \pi \text{ for } j = 1, 2, \dots, k-1, \\ \& \ X_1 + \dots + X_{k-1} + X_k \geq \pi \end{array} \right\}.$$
3. For $t \geq 0$, let $N(t) = \#\{k : W_k \leq t\}$. Then
$$\{N(t) = k\} = \{X_1 + \dots + X_k \leq t,\ X_1 + \dots + X_{k+1} > t\}$$
which shows that $N(t)$ is **not** a stopping time. On the other hand, since
$$\begin{aligned}\{N(t) + 1 = k\} &= \{N(t) = k-1\} \\ &= \{X_1 + \dots + X_{k-1} \leq t,\ X_1 + \dots + X_k > t\},\end{aligned}$$
we see that $N(t) + 1$ **is** a stopping time!
4. If τ is a stopping time then so is $\tau + 1$ because,
$$1_{\{\tau+1=k\}} = 1_{\{\tau=k-1\}} = \sigma_{k-1}(X_0, \dots, X_{k-1})$$
which is also a function of $(X_0, \dots, X_k)$ which happens not to depend on X_k.
5. On the other hand, if τ is a stopping time it is **not** necessarily true that $\tau - 1$ is still a stopping time.
6. One can also see that the last time, k, such that $|X_k| \geq \pi$ is typically **not** a stopping time. (Think about this.)

Remark 4.37. If τ is an $\{X_n\}_{n=0}^{\infty}$ - stopping time then
$$1_{\tau \geq n} = 1 - 1_{\tau < n} = 1 - \sum_{k<n} \sigma_k(X_0, \dots, X_k) =: u_n(X_0, \dots, X_{n-1}).$$
That is for a stopping time τ, $1_{\tau \geq n}$ is a function of $(X_0, \dots, X_{n-1})$ only for all $n \in \mathbb{N}_0$.

The following presentation of Wald's equation is taken from Ross [7, p. 59-60].

Theorem 4.38 (Wald's Equation). *Suppose that $\{X_n\}_{n=0}^{\infty}$ is a sequence of i.i.d. random variables, $f(x)$ is a non-negative function of $x \in \mathbb{R}$, and τ is a stopping time. Then*
$$\mathbb{E}\left[\sum_{n=0}^{\tau} f(X_n)\right] = \mathbb{E}f(X_0) \cdot \mathbb{E}\tau. \tag{4.29}$$
This identity also holds if $f(X_n)$ are real valued but integrable and τ is a stopping time such that $\mathbb{E}\tau < \infty$. (See Resnick for more identities along these lines.)

Proof. If $f(X_n) \geq 0$ for all n, then the the following computations need no justification,
$$\begin{aligned}\mathbb{E}\left[\sum_{n=0}^{\tau} f(X_n)\right] &= \mathbb{E}\left[\sum_{n=0}^{\infty} f(X_n) 1_{n \leq \tau}\right] = \sum_{n=0}^{\infty} \mathbb{E}[f(X_n) 1_{n \leq \tau}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[f(X_n) u_n(X_0, \dots, X_{n-1})] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[f(X_n)] \cdot \mathbb{E}[u_n(X_0, \dots, X_{n-1})] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[f(X_n)] \cdot \mathbb{E}[1_{n \leq \tau}] = \mathbb{E}f(X_0) \sum_{n=0}^{\infty} \mathbb{E}[1_{n \leq \tau}] \\ &= \mathbb{E}f(X_0) \cdot \mathbb{E}\left[\sum_{n=0}^{\infty} 1_{n \leq \tau}\right] = \mathbb{E}f(X_0) \cdot \mathbb{E}\tau.\end{aligned}$$

If $\mathbb{E}|f(X_n)| < \infty$ and $\mathbb{E}\tau < \infty$, the above computation with f replaced by $|f|$ shows all sums appearing above are equal $\mathbb{E}|f(X_0)| \cdot \mathbb{E}\tau < \infty$. Hence we may remove the absolute values to again arrive at Eq. (4.29). ■

Example 4.39. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. such that $P(X_n = 0) = P(X_n = 1) = 1/2$ and let
$$\tau := \min\{n : X_1 + \dots + X_n = 10\}.$$
For example τ is the first time we have flipped 10 heads of a fair coin. By Wald's equation (valid because $X_n \geq 0$ for all n) we find
$$10 = \mathbb{E}\left[\sum_{n=1}^{\tau} X_n\right] = \mathbb{E}X_1 \cdot \mathbb{E}\tau = \frac{1}{2}\mathbb{E}\tau$$
and therefore $\mathbb{E}\tau = 20 < \infty$.

Example 4.40 (Gambler's ruin). Let $\{X_n\}_{n=1}^\infty$ be i.i.d. such that $P(X_n = -1) = P(X_n = 1) = 1/2$ and let

$$\tau := \min\{n : X_1 + \cdots + X_n = 1\}.$$

So τ may represent the first time that a gambler is ahead by 1. Notice that $\mathbb{E}X_1 = 0$. If $\mathbb{E}\tau < \infty$, then we would have $\tau < \infty$ a.s. and by Wald's equation would give,

$$1 = \mathbb{E}\left[\sum_{n=1}^{\tau} X_n\right] = \mathbb{E}X_1 \cdot \mathbb{E}\tau = 0 \cdot \mathbb{E}\tau$$

which can not hold. Hence it must be that

$$\mathbb{E}\tau = \mathbb{E}[\text{first time that a gambler is ahead by 1}] = \infty.$$

Here is the analogue of

Theorem 4.41 (Strong Markov Property). *Let $\left(\{X_n\}_{n=0}^\infty, \{P_x\}_{x\in S}, p\right)$ be Markov chain as above and $\tau : \Omega \to [0,\infty]$ be a stopping time as in Definition 4.35. Then*

$$\begin{aligned}\mathbb{E}_\pi&\left[f(X_\tau, X_{\tau+1}, \dots) g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right] \\ &= \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right]. \end{aligned} \tag{4.30}$$

for all $f, g = \{g_n\} \geq 0$ or f and g bounded.

Proof. The proof of this deep result is now rather easy to reduce to Theorem 3.17. Indeed,

$$\begin{aligned}&\mathbb{E}_\pi\left[f(X_\tau, X_{\tau+1}, \dots) g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[f(X_n, X_{n+1}, \dots) g_n(X_0, \dots, X_n) 1_{\tau=n}\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[f(X_n, X_{n+1}, \dots) g_n(X_0, \dots, X_n) \sigma_n(X_0, \dots, X_n)\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_n} f(X_0, X_1, \dots)\right] g_n(X_0, \dots, X_n) \sigma_n(X_0, \dots, X_n)\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_n) 1_{\tau=n}\right] \\ &= \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_\tau) 1_{\tau<\infty}\right]\end{aligned}$$

wherein we have used Theorem 3.17 in the third equality. ■

The analogue of Corollary 3.18 in this more general setting states; conditioned on $\tau < \infty$ and $X_\tau = x$, $X_\tau, X_{\tau+1}, X_{\tau+2}, \dots$ is independent of $X_0, \dots, X_\tau$ and is distributed as $X_0, X_1, \dots$ under P_x.

Corollary 4.42. *Let τ be a stopping time, $x \in S$ and π be any probability on S. Then relative to $P_\pi(\cdot|\tau < \infty,\ X_\tau = x)$, $\{X_{\tau+k}\}_{k\geq 0}$ is independent of $\{X_0, \dots, X_\tau\}$ and $\{X_{\tau+k}\}_{k\geq 0}$ has the same distribution as $\{X_k\}_{k=0}^\infty$ under P_x.*

Proof. According to Eq. (4.30),

$$\begin{aligned}\mathbb{E}_\pi&\left[g(X_0, \dots, X_\tau) f(X_\tau, X_{\tau+1}, \dots) : \tau < \infty,\ X_\tau = x\right] \\ &= \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) 1_{\tau<\infty} \delta_x(X_\tau) f(X_\tau, X_{\tau+1}, \dots)\right] \\ &= \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) 1_{\tau<\infty} \delta_x(X_\tau) \mathbb{E}_{X_\tau}\left[f(X_0, X_1, \dots)\right]\right] \\ &= \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) 1_{\tau<\infty} \delta_x(X_\tau) \mathbb{E}_x\left[f(X_0, X_1, \dots)\right]\right] \\ &= \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) : \tau < \infty,\ X_\tau = x\right] \mathbb{E}_x\left[f(X_0, X_1, \dots)\right].\end{aligned}$$

Dividing this equation by $P(\tau < \infty,\ X_\tau = x)$ shows,

$$\begin{aligned}\mathbb{E}_\pi&\left[g(X_0, \dots, X_\tau) f(X_\tau, X_{\tau+1}, \dots) | \tau < \infty,\ X_\tau = x\right] \\ &= \mathbb{E}_\pi\left[g(X_0, \dots, X_\tau) | \tau < \infty,\ X_\tau = x\right] \mathbb{E}_x\left[f(X_0, X_1, \dots)\right].\end{aligned} \tag{4.31}$$

Taking $g = 1$ in this equation then shows,

$$\mathbb{E}_\pi\left[f(X_\tau, X_{\tau+1}, \dots) | \tau < \infty,\ X_\tau = x\right] = \mathbb{E}_x\left[f(X_0, X_1, \dots)\right]. \tag{4.32}$$

This shows that $\{X_{\tau+k}\}_{k\geq 0}$ under $P_\pi(\cdot|\tau < \infty,\ X_\tau = x)$ has the same distribution as $\{X_k\}_{k=0}^\infty$ under P_x and, in combination, Eqs. (4.31) and (4.32) shows $\{X_{\tau+k}\}_{k\geq 0}$ and $\{X_0, \dots, X_\tau\}$ are conditionally, on $\{\tau < \infty,\ X_\tau = x\}$, independent. ■

To match notation in the book, let

$$f_{ii}^{(n)} = P_i(R_i = n) = P_i(X_1 \neq i, \dots, X_{n-1} \neq i, X_n = i)$$

and $m_{ij} := \mathbb{E}_i(M_j)$ – the expected number of visits to j after $n = 0$.

Proposition 4.43. *Let $i \in S$ and $n \geq 1$. Then P_{ii}^n satisfies the "renewal equation,"*

$$P_{ii}^n = \sum_{k=1}^n P(R_i = k) P_{ii}^{n-k}. \tag{4.33}$$

Also if $j \in S$, $k \in \mathbb{N}$, and $\nu : S \to [0,1]$ is any probability on S, then Eq. (4.3) holds, i.e.

$$P_\nu(M_j \geq k) = P_\nu(R_j < \infty) \cdot P_j(R_j < \infty)^{k-1}. \tag{4.34}$$

Proof. To prove Eq. (4.33) we first observe for $n \geq 1$ that $\{X_n = i\}$ is the disjoint union of $\{X_n = i, R_i = k\}$ for $1 \leq k \leq n$ and therefore[2],

---

[2] Alternatively, we could use the Markov property to show,

$$\begin{aligned}P_{ii}^n &= P_i(X_n = i) = \sum_{k=1}^n P_i(R_i = k, X_n = i)\\ &= \sum_{k=1}^n P_i(X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i, X_n = i)\\ &= \sum_{k=1}^n P_i(X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i) P_{ii}^{n-k}\\ &= \sum_{k=1}^n P_{ii}^{n-k} P(R_i = k).\end{aligned}$$

For Eq. (4.34) we have $\{M_j \geq 1\} = \{R_j < \infty\}$ so that $P_i(M_j \geq 1) = P_i(R_j < \infty)$. For $k \geq 2$, since $R_j < \infty$ if $M_j \geq 1$, we have

$$P_i(M_j \geq k) = P_i(M_j \geq k | R_j < \infty) P_i(R_j < \infty).$$

Since, on $R_j < \infty$, $X_{R_j} = j$, it follows by the strong Markov property (Corollary 4.42) that;

$$\begin{aligned}P_i(M_j \geq k | R_j < \infty) &= P_i\left(M_j \geq k | R_j < \infty,\ X_{R_j} = j\right)\\ &= P_i\left(1 + \sum_{n \geq 1} 1_{X_{R_j + n} = j} \geq k | R_j < \infty,\ X_{R_j} = j\right)\\ &= P_j\left(1 + \sum_{n \geq 1} 1_{X_n = j} \geq k\right) = P_j(M_j \geq k - 1).\end{aligned}$$

By the last two displayed equations,

$$P_i(M_j \geq k) = P_j(M_j \geq k - 1) P_i(R_j < \infty) \tag{4.35}$$

Taking $i = j$ in this equation shows,

$$P_j(M_j \geq k) = P_j(M_j \geq k - 1) P_j(R_j < \infty)$$

and so by induction,

$$P_j(M_j \geq k) = P_j(R_j < \infty)^k. \tag{4.36}$$

Equation (4.34) now follows from Eqs. (4.35) and (4.36). ■

4.4 Irreducible Recurrent Chains

For this section we are going to assume that X_n is a irreducible recurrent Markov chain. Let us now fix a state, $j \in S$ and define,

$$\begin{aligned}\tau_1 &= R_j = \min\{n \geq 1 : X_n = j\},\\ \tau_2 &= \min\{n \geq 1 : X_{n+\tau_1} = j\},\\ &\vdots\\ \tau_n &= \min\{n \geq 1 : X_{n+\tau_{n-1}} = j\},\end{aligned}$$

so that τ_n is the time it takes for the chain to visit j after the $(n-1)$'st visit to j. By Corollary 4.14 we know that $P_i(\tau_n < \infty) = 1$ for all $i \in S$ and $n \in \mathbb{N}$. We will use strong Markov property to prove the following key lemma in our development.

Lemma 4.44. *We continue to use the notation above and in particular assume that X_n is an irreducible recurrent Markov chain. Then relative to any P_i with $i \in S$, $\{\tau_n\}_{n=1}^\infty$ is a sequence of independent random variables, $\{\tau_n\}_{n=2}^\infty$ are identically distributed, and $P_i(\tau_n = k) = P_j(\tau_1 = k)$ for all $k \in \mathbb{N}_0$ and $n \geq 2$.*

Proof. Let $T_0 = 0$ and then define T_k inductively by, $T_{k+1} = \inf\{n > T_k : X_n = j\}$ so that T_n is the time of the n'th visit of $\{X_n\}_{n=1}^\infty$ to site j. Observe that $T_1 = \tau_1$,

$$\tau_{n+1}(X_0, X_1, \dots) = \tau_1\left(X_{T_n}, X_{T_n+1}, X_{T_n+2}, \dots\right),$$

and $(\tau_1, \dots, \tau_n)$ is a function of $(X_0, \dots, X_{T_n})$. Since $P_i(T_n < \infty) = 1$ (Corollary 4.14) and $X_{T_n} = j$, we may apply the strong Markov property in the form of Corollary 4.42 to learn:

1. τ_{n+1} is independent of $(X_0, \dots, X_{T_n})$ and hence τ_{n+1} is independent of $(\tau_1, \dots, \tau_n)$, and
2. the distribution of τ_{n+1} under P_i is the same as the distribution of τ_1 under P_j.

The result now follows from these two observations and induction. ■

$$\begin{aligned}P_{ii}^n = P_i(X_n = i) &= \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k} \cdot 1_{X_n = i}) = \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k} \cdot \mathbb{E}_i 1_{X_{n-k} = i})\\ &= \sum_{k=1}^n \mathbb{E}_i(1_{R_i = k}) \mathbb{E}_i\left(1_{X_{n-k} = i}\right) = \sum_{k=1}^n P_i(R_i = k) P_i(X_{n-k} = i)\\ &= \sum_{k=1}^n P_{ii}^{n-k} P(R_i = k).\end{aligned}$$

Theorem 4.45. *Suppose that X_n is a irreducible recurrent Markov chain, and let $j \in S$ be a fixed state. Define*

$$\pi_j := \frac{1}{\mathbb{E}_j(R_j)}, \tag{4.37}$$

with the understanding that $\pi_j = 0$ if $\mathbb{E}_j(R_j) = \infty$. Then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} 1_{X_n=j} = \pi_j \quad P_i - a.s. \tag{4.38}$$

for all $i \in S$ and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} P_{ij}^n = \pi_j. \tag{4.39}$$

Proof. Let us first note that Eq. (4.39) follows by taking expectations of Eq. (4.38). So we must prove Eq. (4.38).

By Lemma 4.44, the sequence $\{\tau_n\}_{n\geq 2}$ is i.i.d. relative to P_i and $\mathbb{E}_i\tau_n = \mathbb{E}_j\tau_j = \mathbb{E}_jR_j$ for all $i \in S$. We may now use the strong law of large numbers (Theorem 1.14) to conclude that

$$\lim_{N\to\infty} \frac{\tau_1 + \tau_2 + \cdots + \tau_N}{N} = \mathbb{E}_i\tau_2 = \mathbb{E}_j\tau_1 = \mathbb{E}_jR_j \qquad (P_i\text{– a.s.}). \tag{4.40}$$

This may be expressed as follows, let $R_j^{(N)} = \tau_1 + \tau_2 + \cdots + \tau_N$, be the time when the chain first visits j for the N^{th} time, then

$$\lim_{N\to\infty} \frac{R_j^{(N)}}{N} = \mathbb{E}_jR_j \qquad (P_i\text{– a.s.}) \tag{4.41}$$

Let

$$\nu_N = \sum_{n=0}^{N} 1_{X_n} = j$$

be the number of time X_n visits j up to time N. Since j is visited infinitely often, $\nu_N \to \infty$ as $N \to \infty$ and therefore, $\lim_{N\to\infty} \frac{\nu_N+1}{\nu_N} = 1$. Since there were ν_N visits to j in the first N steps, the of the $\nu_N{}^{\text{th}}$ time j was hit is less than or equal to N, i.e. $R_j^{(\nu_N)} \leq N$. Similarly, the time, $R_j^{(\nu_N+1)}$, of the $(\nu_N+1)^{\text{st}}$ visit to j must be larger than N, so we have $R_j^{(\nu_N)} \leq N \leq R_j^{(\nu_N+1)}$. Putting these facts together along with Eq. (4.41) shows that

$$\begin{array}{cccccl} \frac{R_j^{(\nu_N)}}{\nu_N} & \leq & \frac{N}{\nu_N} & \leq & \frac{R_j^{(\nu_N+1)}}{\nu_N+1}\cdot\frac{\nu_N+1}{\nu_N} & \\ \downarrow & & \downarrow & & \downarrow & N\to\infty, \\ \mathbb{E}_jR_j & \leq & \lim_{N\to\infty}\frac{N}{\nu_N} & \leq & \mathbb{E}_jR_j\cdot 1 & \end{array}$$

i.e. $\lim_{N\to\infty}\frac{N}{\nu_N} = \mathbb{E}_jR_j$ for P_i – almost every sample path. Taking reciprocals of this last set of inequalities implies Eq. (4.38). ■

Proposition 4.46. *Suppose that X_n is a irreducible, recurrent Markov chain and let $\pi_j = \frac{1}{\mathbb{E}_j(R_j)}$ for all $j \in S$ as in Eq. (4.37). Then either $\pi_i = 0$ for all $i \in S$ (in which case X_n is null recurrent) or $\pi_i > 0$ for all $i \in S$ (in which case X_n is positive recurrent). Moreover if $\pi_i > 0$ then*

$$\sum_{i\in S}\pi_i = 1 \text{ and} \tag{4.42}$$

$$\sum_{i\in S}\pi_iP_{ij} = \pi_j \text{ for all } j \in S. \tag{4.43}$$

That is $\pi = (\pi_i)_{i\in S}$ is the unique stationary distribution for P.

Proof. Let us define

$$T_{ki}^n := \frac{1}{n}\sum_{l=1}^{n} P_{ki}^l \tag{4.44}$$

which, according to Theorem 4.45, satisfies,

$$\lim_{n\to\infty} T_{ki}^n = \pi_i \text{ for all } i,k \in S.$$

Observe that,

$$(T^nP)_{ki} = \frac{1}{n}\sum_{l=1}^{n} P_{ki}^{l+1} = \frac{1}{n}\sum_{l=1}^{n} P_{ki}^{l} + \frac{1}{n}\left[P_{ki}^{n+1} - P_{ki}\right] \to \pi_i \text{ as } n\to\infty.$$

Let $\alpha := \sum_{i\in S}\pi_i$. Since $\pi_i = \lim_{n\to\infty} T_{ki}^n$, Fatou's lemma implies for all $i,j \in S$ that

$$\alpha = \sum_{i\in S}\pi_i = \sum_{i\in S}\liminf_{n\to\infty} T_{ki}^n \leq \liminf_{n\to\infty}\sum_{i\in S} T_{ki}^n = 1$$

and

$$\sum_{i\in S}\pi_iP_{ij} = \sum_{i\in S}\lim_{n\to\infty} T_{li}^nP_{ij} \leq \liminf_{n\to\infty}\sum_{i\in S} T_{li}^nP_{ij} = \liminf_{n\to\infty} T_{lj}^{n+1} = \pi_j$$

where $l \in S$ is arbitrary. Thus

$$\sum_{i\in S}\pi_i =: \alpha \leq 1 \text{ and } \sum_{i\in S}\pi_iP_{ij} \leq \pi_j \text{ for all } j \in S. \tag{4.45}$$

By induction it also follows that

$$\sum_{i\in S}\pi_i P_{ij}^k \le \pi_j \text{ for all } j\in S. \tag{4.46}$$

So if $\pi_j = 0$ for some $j \in S$, then given any $i \in S$, there is a integer k such that $P_{ij}^k > 0$, and by Eq. (4.46) we learn that $\pi_i = 0$. This shows that either $\pi_i = 0$ for all $i \in S$ or $\pi_i > 0$ for all $i \in S$.

For the rest of the proof we assume that $\pi_i > 0$ for all $i \in S$. If there were some $j \in S$ such that $\sum_{i\in S}\pi_i P_{ij} < \pi_j$, we would have from Eq. (4.45) that

$$\alpha = \sum_{i\in S}\pi_i = \sum_{i\in S}\sum_{j\in S}\pi_i P_{ij} = \sum_{j\in S}\sum_{i\in S}\pi_i P_{ij} < \sum_{j\in S}\pi_j = \alpha,$$

which is a contradiction and Eq. (4.43) is proved.

From Eq. (4.43) and induction we also have

$$\sum_{i\in S}\pi_i P_{ij}^k = \pi_j \text{ for all } j\in S$$

for all $k \in \mathbb{N}$ and therefore,

$$\sum_{i\in S}\pi_i T_{ij}^k = \pi_j \text{ for all } j\in S. \tag{4.47}$$

Since $0 \le T_{ij} \le 1$ and $\sum_{i\in S}\pi_i = \alpha \le 1$, we may use the dominated convergence theorem to pass to the limit as $k \to \infty$ in Eq. (4.47) to find

$$\pi_j = \lim_{k\to\infty}\sum_{i\in S}\pi_i T_{ij}^k = \sum_{i\in S}\lim_{k\to\infty}\pi_i T_{ij}^k = \sum_{i\in S}\pi_i\pi_j = \alpha\pi_j.$$

Since $\pi_j > 0$, this implies that $\alpha = 1$ and hence Eq. (4.42) is now verified. ■

Proposition 4.47. *Suppose that P is an irreducible Markov kernel which admits a stationary distribution μ. Then P is positive recurrent and $\mu_j = \pi_j = \frac{1}{\mathbb{E}_j(R_j)}$ for all $j \in S$. In particular, an irreducible Markov kernel has at most one invariant distribution and it has exactly one iff P is positive recurrent.*

Proof. Suppose that $\mu = (\mu_i)$ is a stationary distribution for P, i.e. $\sum_{i\in S}\mu_i = 1$ and $\mu_j = \sum_{i\in S}\mu_i P_{ij}$ for all $j \in S$. Then we also have

$$\mu_j = \sum_{i\in S}\mu_i T_{ij}^k \text{ for all } k\in\mathbb{N} \tag{4.48}$$

where T_{ij}^k is defined above in Eq. (4.44). As in the proof of Proposition 4.46, we may use the dominated convergence theorem to find,

$$\mu_j = \lim_{k\to\infty}\sum_{i\in S}\mu_i T_{ij}^k = \sum_{i\in S}\lim_{k\to\infty}\mu_i T_{ij}^k = \sum_{i\in S}\mu_i\pi_j = \pi_j.$$

Alternative Proof. If P were not positive recurrent then P is either transient or null-recurrent in which case $\lim_{n\to\infty} T_{ij}^n = \frac{1}{\mathbb{E}_j(R_j)} = 0$ for all i, j. So letting $k \to \infty$, using the dominated convergence theorem, in Eq. (4.48) allows us to conclude that $\mu_j = 0$ for all j which contradicts the fact that μ was assumed to be a distribution. ■

Lemma 4.48 (A number theory lemma). *Suppose that 1 is the greatest common denominator of a set of positive integers, $\Gamma := \{n_1, \dots, n_k\}$. Then there exists $N \in \mathbb{N}$ such that the set,*

$$A = \{m_1 n_1 + \dots + m_k n_k : m_i \ge 0 \text{ for all } i\},$$

contains all $n \in \mathbb{N}$ with $n \ge N$.

Proof. (The following proof is from Durrett [2].) We first will show that A contains two consecutive positive integers, a and $a+1$. To prove this let,

$$k := \min\{|b-a| : a, b \in A \text{ with } a \ne b\}$$

and choose $a, b \in A$ with $b = a + k$. If $k > 1$, there exists $n \in \Gamma \subset A$ such that k does not divide n. Let us write $n = mk + r$ with $m \ge 0$ and $1 \le r < k$. It then follows that $(m+1)b$ and $(m+1)a + n$ are in A,

$$(m+1)b = (m+1)(a+k) > (m+1)a + mk + r = (m+1)a + n,$$

and

$$(m+1)b - (m+1)a + n = k - r < k.$$

This contradicts the definition of k and therefore, $k = 1$.

Let $N = a^2$. If $n \ge N$, then $n - a^2 = ma + r$ for some $m \ge 0$ and $0 \le r < a$. Therefore,

$$n = a^2 + ma + r = (a+m)a + r = (a+m-r)a + r(a+1) \in A.$$

■

Continuous Time Markov Chain Notions

In this chapter we are going to begin out study continuous time homogeneous Markov chains on discrete state spaces S. In more detail we will assume that $\{X_t\}_{t\geq 0}$ is a stochastic process whose sample paths are right continuous and have left hand limits, see Figures 5.1 and 5.2.

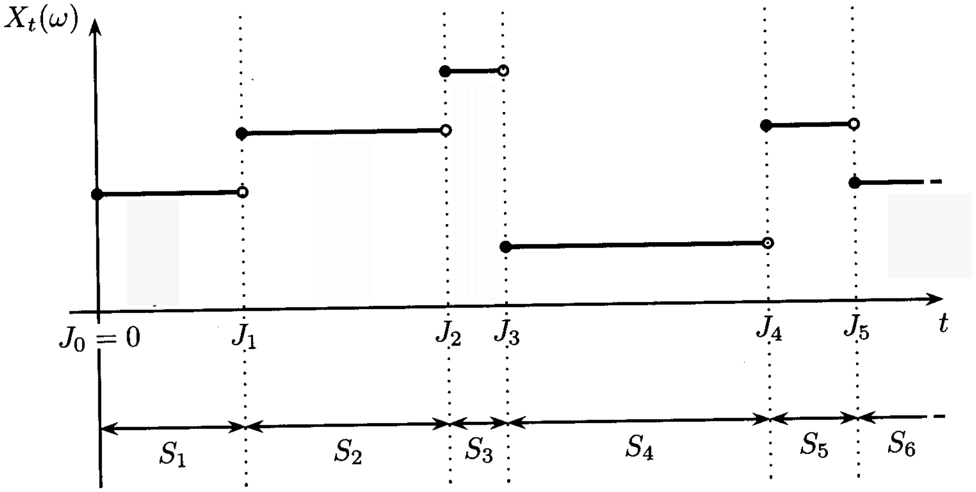


Fig. 5.1. Typical sample paths of a continuous time Markov chain in a discrete state space.

As in the discrete time Markov chain setting, to each $i \in S$, we will write $P_i(A) := P(A|X_0 = i)$. That is P_i is the probability associated to the scenario where the chain is forced to start at site i. We now define, for $i, j \in S$,

$$P_{ij}(t) := P_i(X(t) = j) \tag{5.1}$$

which is the probability of finding the chain at time t at site j given the chain starts at i.

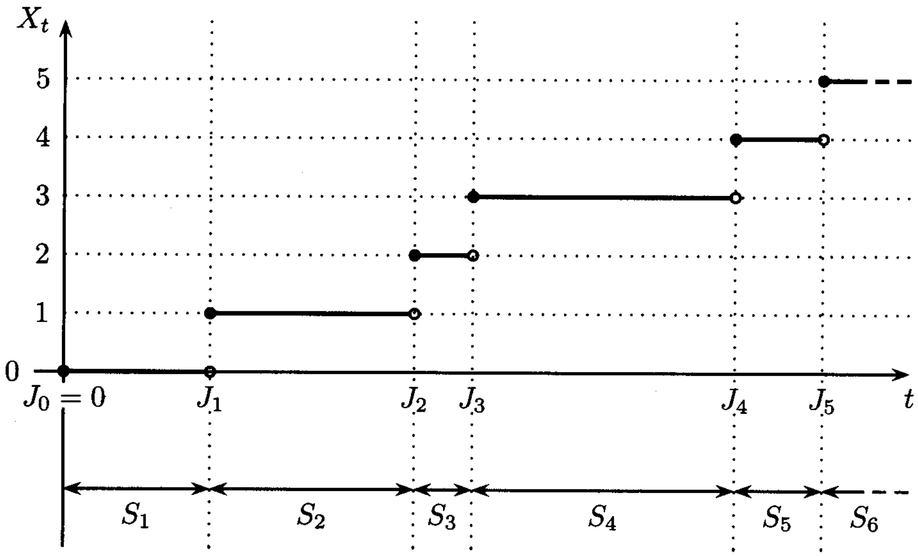


Fig. 5.2. A sample path of a birth process. Here the state space is $\{0, 1, 2, \dots\}$ to be thought of the possible population size.

Definition 5.1. *The* ***time homogeneous Markov property*** *states for every* $0 \leq s < t < \infty$ *and any choices of* $0 = t_0 < t_1 < \cdots < t_n = s < t$ *and* $i_1, \dots, i_n \in S$ *that*

$$P_i(X(t) = j|X(t_1) = i_1, \dots, X(t_n) = i_n) = P_{i_n,j}(t - s), \tag{5.2}$$

and consequently,

$$P_i(X(t) = j|X(s) = i_n) = P_{i_n,j}(t - s). \tag{5.3}$$

Roughly speaking the Markov property may be stated as follows; the probability that $X(t) = j$ given knowledge of the process up to time s is $P_{X(s),j}(t - s)$. In symbols we might express this last sentence as

$$P_i\left(X(t) = j | \{X(\tau)\}_{\tau \le s}\right) = P_i(X(t) = j | X(s)) = P_{X(s),j}(t-s).$$

So again a continuous time Markov process is forgetful in the sense what the chain does for $t \ge s$ depend only on where the chain is located, $X(s)$, at time s and not how it got there. See Fact 5.3 below for a more general statement of this property.

Definition 5.2 (Informal). *A stopping time, T, for $\{X(t)\}$, is a random variable with the property that the event $\{T \le t\}$ is determined from the knowledge of $\{X(s) : 0 \le s \le t\}$. Alternatively put, for each $t \ge 0$, there is a functional, f_t, such that*

$$1_{T \le t} = f_t(\{X(s) : 0 \le s \le t\}).$$

As in the discrete state space setting, the first time the chain hits some subset of states, $A \subset S$, is a typical example of a stopping time whereas the last time the chain hits a set $A \subset S$ is typically **not** a stopping time. Similar the discrete time setting, the Markov property leads to a strong form of forgetfulness of the chain. This property is again called the **strong Markov property** which we take for granted here.

Fact 5.3 (Strong Markov Property) *If $\{X(t)\}_{t \ge 0}$ is a Markov chain, T is a stopping time, and $j \in S$, then, conditioned on $\{T < \infty$ and $X_T = j\}$,*

$$\{X(s) : 0 \le s \le T\} \text{ and } \{X(t+T) : t \ge 0\}$$

are $\{X(t+T) : t \ge 0\}$ has the same distribution as $\{X(t)\}_{t \ge 0}$ under P_j.

We will use the above fact later in our discussions. For the moment, let us go back to more elementary considerations.

Theorem 5.4 (Finite dimensional distributions). *Let $0 < t_1 < t_2 < \dots < t_n$ and $i_0, i_1, i_2, \dots, i_n \in S$. Then*

$$\begin{aligned} P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2, \dots, X_{t_n} = i_n) \\ = P_{i_0,i_1}(t_1) P_{i_1,i_2}(t_2 - t_1) \dots P_{i_{n-1},i_n}(t_n - t_{n-1}). \end{aligned} \tag{5.4}$$

Proof. The proof is similar to that of Proposition 3.2. For notational simplicity let us suppose that $n = 3$. We then have

$$\begin{aligned} P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2, X_{t_3} = i_3) &= P_{i_0}(X_{t_3} = i_3 | X_{t_1} = i_1, X_{t_2} = i_2) P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2) \\ &= P_{i_2,i_3}(t_3 - t_2) P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2) \\ &= P_{i_2,i_3}(t_3 - t_2) P_{i_0}(X_{t_2} = i_2 | X_{t_1} = i_1) P_{i_0}(X_{t_1} = i_1) \\ &= P_{i_2,i_3}(t_3 - t_2) P_{i_1,i_2}(t_2 - t_1) P_{i_0,i_1}(t_1) \end{aligned}$$

wherein we have used the Markov property once in line 2 and twice in line 4. ■

Proposition 5.5 (Properties of P). *Let $P_{ij}(t) := P_i(X(t) = j)$ be as above. Then:*

1. *For each $t \ge 0$, $P(t)$ is a Markov matrix, i.e.*

$$\sum_{j \in S} P_{ij}(t) = 1 \text{ for all } i \in S \text{ and}$$

$$P_{ij}(t) \ge 0 \text{ for all } i, j \in S.$$

.

2. *$\lim_{t \downarrow 0} P_{ij}(t) = \delta_{ij}$ for all $i, j \in S$.*
3. *The **Chapman – Kolmogorov equation** holds:*

$$P(t+s) = P(t)P(s) \text{ for all } s, t \ge 0, \tag{5.5}$$

i.e.

$$P_{ij}(t+s) = \sum_{k \in S} P_{ik}(s) P_{kj}(t) \text{ for all } s, t \ge 0. \tag{5.6}$$

*We will call a matrix $\{P(t)\}_{t \ge 0}$ satisfying items 1. – 3. a **continuous time Markov semigroup.***

Proof. Most of the assertions follow from the basic properties of conditional probabilities. The assumed right continuity of X_t implies that $\lim_{t \downarrow 0} P(t) = P(0) = I$. From Equation (5.4) with $n = 2$ we learn that

$$\begin{aligned} P_{i_0,i_2}(t_2) &= \sum_{i_1 \in S} P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2) \\ &= \sum_{i_1 \in S} P_{i_0,i_1}(t_1) P_{i_1,i_2}(t_2 - t_1) \\ &= [P(t_1)P(t_2 - t_1)]_{i_0,i_2}. \end{aligned}$$

■

At this point it is not so clear how to find a non-trivial (i.e. $P(t) \ne I$ for all t) example of a continuous time Markov semi-group. It turns out the Poisson process provides such an example.

Example 5.6. In this example we will take $S = \{0, 1, 2, \dots\}$ and then define, for $\lambda > 0$,

$$P(t) = e^{-\lambda t} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & \dots \end{array} \\ \left[\begin{array}{ccccccc} 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \frac{(\lambda t)^4}{4!} & \frac{(\lambda t)^5}{5!} & \dots \\ 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \frac{(\lambda t)^4}{4!} & \dots \\ 0 & 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \dots \\ 0 & 0 & 0 & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{array} \right] \end{array} \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}.$$

In components this may be expressed as,

$$P_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} 1_{i \le j}$$

with the convention that 0! = 1. (See Exercise 0.12 of this weeks homework assignment to see where this example is coming from.)

If $i, j \in S$, then $P_{ik}(t) P_{kj}(s)$ will be zero unless $i \le k \le j$, therefore we have

$$\begin{aligned}\sum_{k \in S} P_{ik}(t) P_{kj}(s) &= 1_{i \le j} \sum_{i \le k \le j} P_{ik}(t) P_{kj}(s) \\ &= 1_{i \le j} e^{-\lambda(t+s)} \sum_{i \le k \le j} \frac{(\lambda t)^{k-i}}{(k-i)!} \frac{(\lambda s)^{j-k}}{(j-k)!}. \end{aligned} \tag{5.7}$$

Let $k = i + m$ with $0 \le m \le j - i$, then the above sum may be written as

$$\sum_{m=0}^{j-i} \frac{(\lambda t)^m}{m!} \frac{(\lambda s)^{j-i-m}}{(j-i-m)!} = \frac{1}{(j-i)!} \sum_{m=0}^{j-i} \binom{j-i}{m} (\lambda t)^m (\lambda s)^{j-i-m}$$

and hence by the Binomial formula we find,

$$\sum_{i \le k \le j} \frac{(\lambda t)^{k-i}}{(k-i)!} \frac{(\lambda s)^{j-k}}{(j-k)!} = \frac{1}{(j-i)!} (\lambda t + \lambda s)^{j-i}.$$

Combining this with Eq. (5.7) shows that

$$\sum_{k \in S} P_{ik}(t) P_{kj}(s) = P_{ij}(s+t).$$

Proposition 5.7. *Let $\{X_t\}_{t \ge 0}$ is the Markov chain determined by $P(t)$ of Example 5.6. Then relative to P_0, $\{X_t\}_{t \ge 0}$ is precisely the Poisson process on $[0, \infty)$ with intensity λ.*

Proof. Let $0 \le s < t$. Since $P_0(X_t = n | X_s = k) = P_{kn}(t-s) = 0$ if $n < k$, $\{X_t\}_{t \ge 0}$ is a non-decreasing integer value process. Suppose that $0 = s_0 < s_1 < s_2 < \cdots < s_n = s$ and $i_k \in S$ for $k = 0, 1, 2, \ldots, n$, then

$$\begin{aligned} &P_0\left(X_t - X_s = i_0 | X_{s_j} = i_j \text{ for } 1 \le j \le n\right) \\ &\quad = P_0\left(X_t = i_n + i_0 | X_{s_j} = i_j \text{ for } 1 \le j \le n\right) \\ &\quad = P_0\left(X_t = i_n + i_0 | X_{s_n} = i_n\right) \\ &\quad = e^{-\lambda(t-s)} \frac{(\lambda t)^{i_0}}{i_0!}. \end{aligned}$$

Since this answer is independent of $i_1, \ldots, i_n$ we also have

$$\begin{aligned} &P_0(X_t - X_s = i_0) \\ &\quad = \sum_{i_1, \ldots, i_n \in S} P_0\left(X_t - X_s = i_0 | X_{s_j} = i_j \text{ for } 1 \le j \le n\right) P_0\left(X_{s_j} = i_j \text{ for } 1 \le j \le n\right) \\ &\quad = \sum_{i_1, \ldots, i_n \in S} e^{-\lambda(t-s)} \frac{(\lambda t)^{i_0}}{i_0!} P_0\left(X_{s_j} = i_j \text{ for } 1 \le j \le n\right) = e^{-\lambda(t-s)} \frac{(\lambda t)^{i_0}}{i_0!}. \end{aligned}$$

Thus we may conclude that $X_t - X_s$ is Poisson random variable with intensity λ which is independent of $\{X_r\}_{r \le s}$. That is $\{X_t\}_{t \ge 0}$ is a Poisson process with rate λ. ■

The next example is generalization of the Poisson process example above. You will be asked to work this example out on a future homework set.

Example 5.8. In problems VI.6.P1 on p. 406, you will be asked to consider a discrete time Markov matrix, ρ_{ij}, on some discrete state space, S,with associate Markov chain $\{Y_n\}$. It is claimed in this problem that if $\{N(t)\}_{t \ge 0}$ is Poisson process which is independent of $\{Y_n\}$, then $X_t := Y_{N(t)}$ is a continuous time Markov chain. More precisely the claim is that Eq. (5.2) holds with

$$P(t) = e^{-t} \sum_{m=0}^{\infty} \frac{t^m}{m!} \rho^m =: e^{t(\rho - I)},$$

i.e.

$$P_{ij}(t) = e^{-t} \sum_{m=0}^{\infty} \frac{t^m}{m!} (\rho^m)_{ij}.$$

(We will see a little later, that this example can be used to construct all finite state continuous time Markov chains.)

Notice that in each of these examples, $P(t) = I + Qt + O(t^2)$ for some matrix Q. In the first example,

$$Q_{ij} = -\lambda \delta_{ij} + \lambda \delta_{i,i+1}$$

while in the second example, $Q = \rho - I$.

For a general Markov semigroup, $P(t)$, we are going to show (at least when $\#(S) < \infty$) that $P(t) = I + Qt + O(t^2)$ for some matrix Q which we call the **infinitesimal generator (or Markov generator)** of P. We will see that every infinitesimal generator must satisfy:

$$Q_{ij} \le 0 \text{ for all } i \ne j, \text{ and} \tag{5.8}$$

$$\sum_j Q_{ij} = 0, \text{ i.e. } -Q_{ii} = \sum_{j \ne i} Q_{ij} \text{ for all } i. \tag{5.9}$$

Moreover, to any such Q, the matrix

$$P(t) = e^{tQ} := \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n = I + tQ + \frac{t^2}{2!} Q^2 + \frac{t^3}{3!} Q^3 + \dots$$

will be a Markov semigroup.

One useful way to understand what is going on here is to choose an initial distribution, π on S and then define $\pi(t) := \pi P(t)$. We are going to interpret π_j as the amount of sand we have placed at each of the sites, $j \in S$. We are going to interpret $\pi_j(t)$ as the mass at site j at a later time t under the assumption that π satisfies, $\dot{\pi}(t) = \pi(t) Q$, i.e.

$$\dot{\pi}_j(t) = \sum_{i \neq j} \pi_i(t) Qij - q_j \pi_j(t), \tag{5.10}$$

where $q_j = -Q_{j\ j}$. (See Example 6.19 below.) Here is how to interpret each term in this equation:

$$\begin{aligned} \dot{\pi}_j(t) &= \text{ rate of change of the amount of sand at } j \text{ at time } t, \\ \pi_i(t) Qij &= \text{ rate at which sand is shoveled from site } i \text{ to } j, \\ q_j \pi_j(t) &= \text{ rate at which sand is shoveled out of site } i \text{ to all other sites.} \end{aligned}$$

With this interpretation Eq. 5.10 has the clear meaning: namely the rate of change of the mass of sand at j at time t should be equal to the rate at which sand is shoveled into site j form all other sites minus the rate at which sand is shoveled out of site i. With this interpretation, the condition,

$$-Q_{j\ j} := q_j = \sum_{k \neq j} Q_{j,k},$$

just states the total sand in the system should be conserved, i.e. this guarantees the rate of sand leaving j should equal the total rate of sand being sent to all of the other sites from j.

Warning: the book denotes Q by A but then denotes the entries of A by q_{ij}. I have just decided to write $A = Q$ and identify, Q_{ij} and q_{ij}. To avoid some technical details, in the next chapter we are mostly going to restrict ourselves to the case where $\#(S) < \infty$. Later we will consider examples in more detail where $\#(S) = \infty$.

6

Continuous Time M.C. Finite State Space Theory

For simplicity we will begin our study in the case where the state space is finite, say $S = \{1, 2, 3, \ldots, N\}$ for some $N < \infty$. It will be convenient to define,

$$\mathbf{1} := \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

be the column vector with all entries being 1.

Definition 6.1. *An $N \times N$ matrix function $P(t)$ for $t \geq 0$ is* ***Markov semi-group*** *if*

1. *$P(t)$ is Markov matrix for all $t \geq 0$, i.e. $P_{ij}(t) \geq 0$ for all i, j and*

$$\sum_{j \in S} P_{ij}(t) = 1 \text{ for all } i \in S. \tag{6.1}$$

The condition in Eq. (6.1) may be written in matrix notation as,

$$P(t)\mathbf{1} = \mathbf{1} \text{ for all } t \geq 0. \tag{6.2}$$

2. *$P(0) = I_{N \times N}$,*
3. *$P(t+s) = P(t)P(s)$ for all $s, t \geq 0$* ***(Chapman - Kolmogorov)****,*
4. *$\lim_{t \downarrow 0} P(t) = I$, i.e. P is continuous at $t = 0$.*

Definition 6.2. *An $N \times N$ matrix, Q, is an* ***infinitesimal generator*** *if $Q_{ij} \geq 0$ for all $i \neq j$ and*

$$\sum_{j \in S} Q_{ij} = 0 \text{ for all } i \in S. \tag{6.3}$$

The condition in Eq. (6.3) may be written in matrix notation as,

$$Q\mathbf{1} = 0. \tag{6.4}$$

6.1 Matrix Exponentials

In this section we are going to make use of the following facts from the theory of linear ordinary differential equations.

Theorem 6.3. *Let A and B be any $N \times N$ (real) matrices. Then there exists a unique $N \times N$ matrix function $P(t)$ solving the differential equation,*

$$\dot{P}(t) = AP(t) \text{ with } P(0) = B \tag{6.5}$$

which is in fact given by

$$P(t) = e^{tA}B \tag{6.6}$$

where

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!}A^n = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \ldots \tag{6.7}$$

The matrix function e^{tA} may be characterized as the unique solution Eq. (6.5) with $B = I$ and it is also the unique solution to

$$\dot{P}(t) = AP(t) \text{ with } P(0) = I.$$

Moreover, e^{tA} satisfies the semi-group property (Chapman Kolmogorov equation),

$$e^{(t+s)A} = e^{tA}e^{sA} \text{ for all } s, t \geq 0. \tag{6.8}$$

Proof. We will only prove Eq. (6.8) here assuming the first part of the theorem. Fix $s > 0$ and let $R(t) := e^{(t+s)A}$, then

$$\dot{R}(t) = Ae^{(t+s)A} = AR(t) \text{ with } R(0) = P(s).$$

Therefore by the first part of the theorem

$$e^{(t+s)A} = R(t) = e^{tA}R(0) = e^{tA}e^{sA}.$$

■

Example 6.4 (Thanks to Mike Gao!). If $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $A^n = 0$ for $n \geq 2$, so that

$$e^{tA} = I + tA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Similarly if $B = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$, then $B^n = 0$ for $n \geq 2$ and

$$e^{tB} = I + tB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -t & 1 \end{bmatrix}.$$

Now let $C = A + B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. In this case $C^2 = -I$, $C^3 = -C$, $C^4 = I$, $C^5 = C$ etc., so that

$$C^{2n} = (-1)^n I \text{ and } C^{2n+1} = (-1)^n C.$$

Therefore,

$$\begin{aligned} e^{tC} &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} C^{2n} + \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!} C^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} (-1)^n I + \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!} (-1)^n C \\ &= \cos(t) I + \sin(t) C = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \end{aligned}$$

which is the matrix representing rotation in the plan by t degrees.

Here is another way to compute e^{tC} in this example. Since $C^2 = -I$, we find

$$\frac{d^2}{dt^2} e^{tC} = C^2 e^{tC} = -e^{tC} \text{ with}$$
$$e^{0C} = I \text{ and } \frac{d}{dt} e^{tC}|_{t=0} = C.$$

It is now easy to verify the solution to this second order equation is given by,

$$e^{tC} = \cos t \cdot I + \sin t \cdot C$$

which agrees with our previous answer.

Remark 6.5. **Warning:** if A and B are two $N \times N$ matrices it is not generally true that

$$e^{(A+B)} = e^A e^B \tag{6.9}$$

as can be seen from Example 6.4.

However we have the following lemma.

Lemma 6.6. *If A and B commute, i.e. $AB = BA$, then Eq. (6.9) holds. In particular, taking $B = -A$, shows that $e^{-A} = \left[e^A\right]^{-1}$.*

Proof. First proof. Simply verify Eq. (6.9) using explicit manipulations with the infinite series expansion. The point is, because A and B compute, we may use the binomial formula to find;

$$(A+B)^n = \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k}.$$

(Notice that if A and B do not compute we will have

$$(A+B) = A^2 + AB + BA + B^2 \neq A^2 + 2AB + B^2.)$$

Therefore,

$$\begin{aligned} e^{(A+B)} &= \sum_{n=0}^{\infty} \frac{1}{n!} (A+B)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} \\ &= \sum_{0 \leq k \leq n < \infty} \frac{1}{k!} \frac{1}{(n-k)!} A^k B^{n-k} \qquad (\text{let } n-k=l) \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!} \frac{1}{l!} A^k B^l = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \cdot \sum_{l=0}^{\infty} \frac{1}{l!} B^l = e^A e^B. \end{aligned}$$

Second proof. Here is another proof which uses the ODE interpretation of e^{tA}. We will carry it out in a number of steps.

1. By Theorem 6.3 and the product rule
$$\frac{d}{dt} e^{-tA} B e^{tA} = e^{-tA} (-A) B e^{tA} + e^{-tA} B A e^{tA} = e^{-tA} (BA - AB) e^{tA} = 0$$
since A and B commute. This shows that $e^{-tA} B e^{tA} = B$ for all $t \in \mathbb{R}$.
2. Taking $B = I$ in 1. then shows $e^{-tA} e^{tA} = I$ for all t ,i.e. $e^{-tA} = \left[e^{tA}\right]^{-1}$. Hence we now conclude from Item 1. that $e^{-tA} B = B e^{-tA}$ for all t.
3. Using Theorem 6.3, Item 2., and the product rule implies
$$\begin{aligned} &\frac{d}{dt} \left[e^{-tB} e^{-tA} e^{t(A+B)}\right] \\ =& e^{-tB} (-B) e^{-tA} e^{t(A+B)} + e^{-tB} e^{-tA} (-A) e^{t(A+B)} \\ &+ e^{-tB} e^{-tA} (A+B) e^{t(A+B)} \\ =& e^{-tB} e^{-tA} (-B) e^{t(A+B)} + e^{-tB} e^{-tA} (-A) e^{t(A+B)} \\ &+ e^{-tB} e^{-tA} (A+B) e^{t(A+B)} = 0. \end{aligned}$$

Therefore,

$$e^{-tB}e^{-tA}e^{t(A+B)} = e^{-tB}e^{-tA}e^{t(A+B)}|_{t=0} = I \text{ for all } t,$$

and hence taking $t = 1$, shows

$$e^{-B}e^{-A}e^{(A+B)} = I. \tag{6.10}$$

Multiplying Eq. (6.10) on the left by $e^{A}e^{B}$ gives Eq. (6.9).

■

The next two results gives a practical method for computing e^{tQ} in many situations.

Proposition 6.7. *If Λ is a diagonal matrix,*

$$\Lambda := \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

then

$$e^{t\Lambda} = \begin{bmatrix} e^{t\lambda_1} & & & \\ & e^{t\lambda_2} & & \\ & & \ddots & \\ & & & e^{t\lambda_n} \end{bmatrix}.$$

Proof. One easily shows that

$$\Lambda^n := \begin{bmatrix} \lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_m^n \end{bmatrix}$$

for all n and therefore,

$$\begin{aligned} e^{t\Lambda} = \sum_{n=0}^{\infty} \frac{t^n}{n!}\Lambda^n &= \begin{bmatrix} \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_1^n & & & \\ & \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_2^n & & \\ & & \ddots & \\ & & & \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_m^n \end{bmatrix} \\ &= \begin{bmatrix} e^{t\lambda_1} & & & \\ & e^{t\lambda_2} & & \\ & & \ddots & \\ & & & e^{t\lambda_n} \end{bmatrix}. \end{aligned}$$

■

Theorem 6.8. *Suppose that Q is a diagonalizable matrix, i.e. there exists an invertible matrix, S, such that $S^{-1}QS = \Lambda$ with Λ being a diagonal matrix. In this case we have,*

$$e^{tQ} = Se^{t\Lambda}S^{-1} \tag{6.11}$$

Proof. We begin by observing that

$$\begin{aligned} \left(S^{-1}QS\right)^2 &= S^{-1}QSS^{-1}QS = S^{-1}Q^2S, \\ \left(S^{-1}QS\right)^3 &= S^{-1}Q^2SS^{-1}QS = S^{-1}Q^3S \\ &\vdots \\ \left(S^{-1}QS\right)^n &= S^{-1}Q^nS \quad \text{for all } n \geq 0. \end{aligned}$$

Therefore we find that

$$\begin{aligned} S^{-1}e^{tQ}S &= S^{-1}IS + \sum_{n=0}^{\infty} \frac{t^n}{n!}S^{-1}Q^nS \\ &= I + \sum_{n=0}^{\infty} \frac{t^n}{n!}\left(S^{-1}QS\right)^n \\ &= I + \sum_{n=0}^{\infty} \frac{t^n}{n!}\Lambda^n = e^{t\Lambda}. \end{aligned}$$

Solving this equation for e^{tQ} gives the desired result. ■

6.2 Characterizing Markov Semi-Groups

We now come to the main theorem of this chapter.

Theorem 6.9. *The collection of Markov semi-groups is in one to one correspondence with the collection of infinitesimal generators. More precisely we have;*

1. *$P(t) = e^{tQ}$ is Markov semi-group iff Q is an infinitesimal generator.*
2. *If $P(t)$ is a Markov semi-group, then $Q := \frac{d}{dt}|_{0+}P(t)$ exists, Q is an infinitesimal generator, and $P(t) = e^{tQ}$.*

Proof. The proof is completed by Propositions 6.10 – 6.13 below. (You might look at Example 6.4 to see what goes wrong if Q does not satisfy the properties of a Markov generator.) ■

We are now going to prove a number of results which in total will complete the proof of Theorem 6.9. The first result is technical and you may safely skip its proof.

Proposition 6.10 (Techinical proposition). *Every Markov semi-group, $\{P(t)\}_{t\geq 0}$ is continuously differentiable.*

Proof. First we want to show that $P(t)$ is continuous. For $t, h \geq 0$, we have

$$P(t+h) - P(t) = P(t)P(h) - P(t) = P(t)(P(h) - I) \to 0 \text{ as } h \downarrow 0.$$

Similarly if $t > 0$ and $0 \leq h < t$, we have

$$\begin{aligned} P(t) - P(t-h) &= P(t-h+h) - P(t-h) = P(t-h)P(h) - P(t-h) \\ &= P(t-h)[P(h) - I] \to 0 \text{ as } h \downarrow 0 \end{aligned}$$

where we use the fact that $P(t-h)$ has entries all bounded by 1 and therefore

$$\begin{aligned} \left|(P(t-h)[P(h)-I])_{ij}\right| &\leq \sum_k P_{ik}(t-h)\left|(P(h)-I)_{kj}\right| \\ &\leq \sum_k \left|(P(h)-I)_{kj}\right| \to 0 \text{ as } h \downarrow 0. \end{aligned}$$

Thus we have shown that $P(t)$ is continuous.

To prove the differentiability of $P(t)$ we use a trick due to Gärding. Choose $\varepsilon > 0$ such that

$$\Pi := \frac{1}{\varepsilon}\int_0^\varepsilon P(s)\,ds$$

is invertible. To see this is possible, observe that by the continuity of P, $\frac{1}{\varepsilon}\int_0^\varepsilon P(s)\,ds \to I$ as $\varepsilon \downarrow 0$. Therefore, by the continuity of the determinant function,

$$\det\left(\frac{1}{\varepsilon}\int_0^\varepsilon P(s)\,ds\right) \to \det(I) = 1 \text{ as } \varepsilon \downarrow 0.$$

With this definition of Π, we have

$$P(t)\Pi = \frac{1}{\varepsilon}\int_0^\varepsilon P(t)P(s)\,ds = \frac{1}{\varepsilon}\int_0^\varepsilon P(t+s)\,ds = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} P(s)\,ds.$$

So by the fundamental theorem of calculus, $P(t)\Pi$ is differentiable and

$$\frac{d}{dt}[P(t)\Pi] = \frac{1}{\varepsilon}(P(t+\varepsilon) - P(t)).$$

As Π is invertible, we may conclude that $P(t)$ is differentiable and that

$$\dot{P}(t) := \frac{1}{\varepsilon}(P(t+\varepsilon) - P(t))\Pi^{-1}.$$

Since the right hand side of this equation is continuous in t it follows that $\dot{P}(t)$ is continuous as well. ■

Proposition 6.11. *If $\{P(t)\}_{t\geq 0}$ is a Markov semi-group and $Q := \frac{d}{dt}|_{0+}P(t)$, then*

1. *$P(t)$ satisfies $P(0) = I$ and both,*
$$\dot{P}(t) = P(t)Q \qquad \textit{(Kolmogorov's forward Eq.)}$$
and
$$\dot{P}(t) = QP(t) \qquad \textit{(Kolmogorov's backwards Eq.)}$$
hold.
2. *$P(t) = e^{tQ}$.*
3. *Q is an infinitesimal generator.*

Proof. 1.-2. We may compute $\dot{P}(t)$ using

$$\dot{P}(t) = \frac{d}{ds}|_0 P(t+s).$$

We then may write $P(t+s)$ as $P(t)P(s)$ or as $P(s)P(t)$ and hence

$$\begin{aligned} \dot{P}(t) &= \frac{d}{ds}|_0[P(t)P(s)] = P(t)Q \text{ and} \\ \dot{P}(t) &= \frac{d}{ds}|_0[P(s)P(t)] = QP(t). \end{aligned}$$

This proves Item 1. and Item 2. now follows from Theorem 6.3.

3. Since $P(t)$ is continuously differentiable, $P(t) = I + tQ + O(t^2)$, and so for $i \neq j$,

$$0 \leq P_{ij}(t) = \delta_{ij} + tQ_{ij} + O(t^2) = tQ_{ij} + O(t^2).$$

Dividing this inequality by t and then letting $t \downarrow 0$ shows $Q_{ij} \geq 0$. Differentiating the Eq. (6.2), $P(t)\mathbf{1} = \mathbf{1}$, at $t = 0_+$ to show $Q\mathbf{1}$ =0. ■

Proposition 6.12. *Let Q be any matrix such that $Q_{ij} \geq 0$ for all $i \neq j$. Then $(e^{tQ})_{ij} \geq 0$ for all $t \geq 0$ and $i, j \in S$.*

Proof. Choose $\lambda \in \mathbb{R}$ such that $\lambda \geq -Q_{ii}$ for all $i \in S$. Then $\lambda I + Q$ has all non-negative entries and therefore $e^{t(\lambda I+Q)}$ has non-negative entries for all $t \geq 0$. (Think about the power series expansion for $e^{t(\lambda I+Q)}$.) By Lemma 6.6 we know that $e^{t(\lambda I+Q)} = e^{t\lambda I}e^{tQ}$ and since $e^{t\lambda I} = e^{t\lambda}I$ (you verify), we have[1]

$$e^{t(\lambda I+Q)} = e^{t\lambda}e^{tQ}.$$

Therefore, $e^{tQ} = e^{-t\lambda}e^{t(\lambda I+Q)}$ again has all non-negative entries and the proof is complete. ■

---

[1] Actually if you do not want to use Lemma 6.6, you may check that $e^{t(\lambda I+Q)} = e^{t\lambda}e^{tQ}$ by simply showing both sides of this equation satisfy the same ordinary differential equation.

Proposition 6.13. *Suppose that* Q *is any matrix such that* $\sum_{j\in S} Q_{ij} = 0$ *for all* $i \in S$, *i.e.* $Q\mathbf{1} = 0$. *Then* $e^{tQ}\mathbf{1} = \mathbf{1}$.

Proof. Since

$$\frac{d}{dt}e^{tQ}\mathbf{1} = e^{tQ}Q\mathbf{1} = 0,$$

it follows that $e^{tQ}\mathbf{1} = e^{tQ}\mathbf{1}|_{t=0} = \mathbf{1}$. ■

Lemma 6.14 (ODE Lemma). *If* $h(t)$ *is a given function and* $\lambda \in \mathbb{R}$, *then the solution to the differential equation,*

$$\dot{\pi}(t) = \lambda\pi(t) + h(t) \tag{6.12}$$

is

$$\pi(t) = e^{\lambda t}\left(\pi(0) + \int_0^t e^{-\lambda s}h(s)\,ds\right) \tag{6.13}$$

$$= e^{\lambda t}\pi(0) + \int_0^t e^{\lambda(t-s)}h(s)\,ds. \tag{6.14}$$

Proof. If $\pi(t)$ satisfies Eq. (6.12), then

$$\frac{d}{dt}\left(e^{-\lambda t}\pi(t)\right) = e^{-\lambda t}\left(-\lambda\pi(t) + \dot{\pi}(t)\right) = e^{-\lambda t}h(t).$$

Integrating this equation implies,

$$e^{-\lambda t}\pi(t) - \pi(0) = \int_0^t e^{-\lambda s}h(s)\,ds.$$

Solving this equation for $\pi(t)$ gives

$$\pi(t) = e^{\lambda t}\pi(0) + e^{\lambda t}\int_0^t e^{-\lambda s}h(s)\,ds \tag{6.15}$$

which is the same as Eq. (6.13). A direct check shows that $\pi(t)$ so defined solves Eq. (6.12). Indeed using Eq. (6.15) and the fundamental theorem of calculus shows,

$$\begin{aligned}\dot{\pi}(t) &= \lambda e^{\lambda t}\pi(0) + \lambda e^{\lambda t}\int_0^t e^{-\lambda s}h(s)\,ds + e^{\lambda t}e^{-\lambda t}h(t)\\ &= \lambda\pi(t) + h(t).\end{aligned}$$

■

Corollary 6.15. *Suppose* $\lambda \in \mathbb{R}$ *and* $\pi(t)$ *is a function which satisfies,* $\dot{\pi}(t) \geq \lambda\pi(t)$, *then*

$$\pi(t) \geq e^{\lambda t}\pi(0) \text{ for all } t \geq 0. \tag{6.16}$$

In particular if $\pi(0) \geq 0$ *then* $\pi(t) \geq 0$ *for all* t. *In particular if* Q *is a Markov generator and* $P(t) = e^{tQ}$, *then*

$$P_{ii}(t) \geq e^{-q_i t} \text{ for all } t > 0$$

where $q_i := -Q_{ii}$. *(If we put all of the sand at site* i *at time* 0, $e^{-q_i t}$ *represents the amount of sand at a later time* t *in the worst case scenario where no one else shovels sand back to site* i.*)*

Proof. Let $h(t) := \dot{\pi}(t) - \lambda\pi(t) \geq 0$ and then apply Lemma 6.14 to conclude that

$$\pi(t) = e^{\lambda t}\pi(0) + \int_0^t e^{\lambda(t-s)}h(s)\,ds. \tag{6.17}$$

Since $e^{\lambda(t-s)}h(s) \geq 0$, it follows that $\int_0^t e^{\lambda(t-s)}h(s)\,ds \geq 0$ and therefore if we ignore this term in Eq. (6.17) leads to the estimate in Eq. (6.16). ■

6.3 Examples

Example 6.16 (2×2 *case I).* The most general 2×2 rate matrix Q is of the form

$$Q = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left[\begin{array}{cc} -\alpha & \alpha \\ \beta & -\beta \end{array}\right] \end{array} \begin{array}{c} \\ 0 \\ 1 \end{array}$$

with rate diagram being given in Figure 6.1. We now find e^{tQ} using Theorem 6.8. To do this we start by observing that

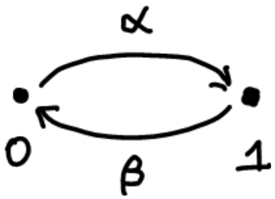


Fig. 6.1. Two state Markov chain rate diagram.

$$\begin{aligned}\det\left(Q-\lambda I\right) &= \det\left(\begin{bmatrix}-\alpha-\lambda & \alpha \\ \beta & -\beta-\lambda\end{bmatrix}\right) = \left(\alpha+\lambda\right)\left(\beta+\lambda\right)-\alpha\beta \\ &= \lambda^2+\tau\lambda = \lambda\left(\lambda+\tau\right).\end{aligned}$$

Thus the eigenvalues of Q are $\{0,-\tau\}$. The eigenvector for 0 is $\begin{bmatrix}1 & 1\end{bmatrix}^{\mathrm{tr}}$. Moreover,

$$Q-\left(-\tau\right)I = \begin{bmatrix}\beta & \alpha \\ \beta & \alpha\end{bmatrix}$$

which has $\begin{bmatrix}\alpha & -\beta\end{bmatrix}^{\mathrm{tr}}$ and therefore we let

$$S = \begin{bmatrix}1 & \alpha \\ 1 & -\beta\end{bmatrix} \text{ and } S^{-1} = \frac{1}{\tau}\begin{bmatrix}\beta & \alpha \\ 1 & -1\end{bmatrix}.$$

We then have

$$S^{-1}QS = \begin{bmatrix}0 & 0 \\ 0 & -\tau\end{bmatrix} =: \Lambda.$$

So in our case

$$S^{-1}e^{tQ}S = e^{t\Lambda} = \begin{bmatrix}e^{0t} & 0 \\ 0 & e^{-\tau t}\end{bmatrix} = \begin{bmatrix}1 & 0 \\ 0 & e^{-\tau t}\end{bmatrix}.$$

Hence we must have,

$$\begin{aligned}e^{tQ} &= S\begin{bmatrix}1 & 0 \\ 0 & e^{-\tau t}\end{bmatrix}S^{-1} \\ &= \frac{1}{\tau}\begin{bmatrix}1 & \alpha \\ 1 & -\beta\end{bmatrix}\begin{bmatrix}1 & 0 \\ 0 & e^{-\tau t}\end{bmatrix}\begin{bmatrix}\beta & \alpha \\ 1 & -1\end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix}\beta+\alpha e^{-\tau t} & \alpha-\alpha e^{-\tau t} \\ \beta-\beta e^{-\tau t} & \alpha+\beta e^{-\tau t}\end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix}\beta+\alpha e^{-\tau t} & \alpha\left(1-e^{-\tau t}\right) \\ \beta\left(1-e^{-\tau t}\right) & \alpha+\beta e^{-\tau t}\end{bmatrix}.\end{aligned}$$

Example 6.17 (2 × 2 case II). If $P\left(t\right)=e^{tQ}$ and $\pi\left(t\right)=\pi\left(0\right)P\left(t\right)$, then

$$\begin{aligned}\dot{\pi}\left(t\right) = \pi\left(t\right)Q &= \left[\pi_0\left(t\right),\pi_1\left(t\right)\right]\begin{bmatrix}-\alpha & \alpha \\ \beta & -\beta\end{bmatrix} \\ &\left[-\alpha\pi_0\left(t\right)+\beta\pi_1\left(t\right) \;\; \alpha\pi_0\left(t\right)-\beta\pi_1\left(t\right)\right],\end{aligned}$$

i.e

$$\dot{\pi}_0\left(t\right) = -\alpha\pi_0\left(t\right)+\beta\pi_1\left(t\right) \tag{6.18}$$

$$\dot{\pi}_1\left(t\right) = \alpha\pi_0\left(t\right)-\beta\pi_1\left(t\right). \tag{6.19}$$

The latter pair of equations is easy to write down using the jump diagram and the movement of sand interpretation. If we assume that $\pi_0\left(0\right)+\pi_1\left(0\right)=1$ then we know $\pi_0\left(t\right)+\pi_1\left(t\right)=1$ for all later times and therefore we may rewrite Eq. (6.18) as

$$\begin{aligned}\dot{\pi}_0\left(t\right) &= -\alpha\pi_0\left(t\right)+\beta\left(1-\pi_0\left(t\right)\right) \\ &= -\tau\pi_0\left(t\right)+\beta\end{aligned}$$

where $\tau:=\alpha+\beta$. We may use Lemma 6.14 below to find

$$\begin{aligned}\pi_0\left(t\right) &= e^{-\tau t}\pi_0\left(0\right)+\int_0^t e^{-\tau\left(t-s\right)}\beta ds \\ &= e^{-\tau t}\pi_0\left(0\right)+\frac{\beta}{\tau}\left(1-e^{-\tau t}\right).\end{aligned}$$

We may also conclude that

$$\begin{aligned}\pi_1\left(t\right) = 1-\pi_0\left(t\right) &= 1-e^{-\tau t}\pi_0\left(0\right)-\frac{\beta}{\tau}\left(1-e^{-\tau t}\right) \\ &= 1-e^{-\tau t}\left(1-\pi_1\left(0\right)\right)-\frac{\beta}{\tau}\left(1-e^{-\tau t}\right) \\ &= e^{-\tau t}\pi_1\left(0\right)+\left(1-e^{-\tau t}\right)-\frac{\beta}{\tau}\left(1-e^{-\tau t}\right) \\ &= e^{-\tau t}\pi_1\left(0\right)+\frac{\alpha}{\tau}\left(1-e^{-\tau t}\right).\end{aligned}$$

By taking $\pi_0\left(0\right)=1$ and $\pi_1\left(0\right)=0$ we get the first row of $P\left(t\right)$ is equal to

$$\left[e^{-\tau t}1+\tfrac{\beta}{\tau}\left(1-e^{-\tau t}\right) \;\; \tfrac{\alpha}{\tau}\left(1-e^{-\tau t}\right)\right] = \frac{1}{\tau}\left[e^{-\tau t}\alpha+\beta \;\; \alpha\left(1-e^{-\tau t}\right)\right]$$

and similarly the second row of $P\left(t\right)$ is found by taking $\pi_0\left(0\right)=0$ and $\pi_1\left(0\right)=1$ to find

$$\left[\tfrac{\beta}{\tau}\left(1-e^{-\tau t}\right) \;\; e^{-\tau t}+\tfrac{\alpha}{\tau}\left(1-e^{-\tau t}\right)\right] = \frac{1}{\tau}\left[\beta\left(1-e^{-\tau t}\right) \;\; \beta e^{-\tau t}+\alpha\right].$$

Hence we have found

$$\begin{aligned}P\left(t\right) &= \frac{1}{\tau}\begin{bmatrix}e^{-\tau t}\alpha+\beta & \alpha\left(1-e^{-\tau t}\right) \\ \beta\left(1-e^{-\tau t}\right) & \beta e^{-\tau t}+\alpha\end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix}\left(e^{-\tau t}-1\right)\alpha+\beta+\alpha & \alpha\left(1-e^{-\tau t}\right) \\ \beta\left(1-e^{-\tau t}\right) & \beta\left(e^{-\tau t}-1\right)+\alpha+\beta\end{bmatrix} \\ &= I+\frac{1}{\tau}\left(1-e^{-\tau t}\right)\begin{bmatrix}-\alpha & \alpha \\ \beta & -\beta\end{bmatrix} \\ &= I+\frac{1}{\tau}\left(1-e^{-\tau t}\right)Q.\end{aligned}$$

Let us verify that this is indeed the correct solution. It is clear that $P(0) = I$,

$$\dot{P}(t) = e^{-\tau t} \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

while on the other hand,

$$Q^2 = \begin{bmatrix} \alpha\beta + \alpha^2 & -\alpha\beta - \alpha^2 \\ -\alpha\beta - \beta^2 & \alpha\beta + \beta^2 \end{bmatrix} = \tau \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} = -\tau Q$$

and therefore,

$$P(t) Q = Q - \left(1 - e^{-\tau t}\right) Q = e^{-\tau t} Q$$

as desired.

We also have

$$\begin{aligned} P(s) P(t) &= \left(I + \frac{1}{\tau}\left(1 - e^{-\tau s}\right) Q\right)\left(I + \frac{1}{\tau}\left(1 - e^{-\tau t}\right) Q\right) \\ &= I + \frac{1}{\tau}\left(2 - e^{-\tau s} - e^{-\tau t}\right) Q + \frac{1}{\tau}\left(1 - e^{-\tau s}\right)\frac{1}{\tau}\left(1 - e^{-\tau t}\right)(-\tau) Q \\ &= I + \frac{1}{\tau}\left[\left(2 - e^{-\tau s} - e^{-\tau t}\right) - \left(1 - e^{-\tau s}\right)\left(1 - e^{-\tau t}\right)\right] Q \\ &= I + \frac{1}{\tau}\left[1 - e^{-\tau(s+t)}\right] Q = P(s+t) \end{aligned}$$

as it should be. Lastly let us observe that

$$\begin{aligned} \lim_{t\to\infty} P(t) &= I + \frac{1}{\tau}\lim_{t\to\infty}\left(1 - e^{-\tau t}\right)\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \\ &= I - \frac{1}{\tau}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}. \end{aligned}$$

Moreover we have

$$\lim_{t\to\infty} \dot{P}(t) = \lim_{t\to\infty} e^{-\tau t}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} = 0.$$

Suppose that π is any distribution, then

$$\lim_{t\to\infty} \pi P(t) = \frac{1}{\tau}\left[\pi_0 \ \pi_1\right]\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} = \frac{1}{\tau}\left[\beta \ \alpha\right]$$

independent of π. Moreover, since

$$\begin{aligned} \frac{1}{\tau}\left[\beta \ \alpha\right] P(s) &= \lim_{t\to\infty} \pi P(t) P(s) = \lim_{t\to\infty} \pi P(t+s) \\ &= \lim_{t\to\infty} \pi P(t) = \frac{1}{\tau}\left[\beta \ \alpha\right] \end{aligned}$$

which shows that the limiting distribution is also an invariant distribution. If π is any invariant distribution for P, we must have

$$\pi = \lim_{t\to\infty} \pi P(t) = \frac{1}{\tau}\left[\beta \ \alpha\right] = \left[\frac{\beta}{\alpha+\beta} \ \frac{\alpha}{\alpha+\beta}\right] \tag{6.20}$$

and moreover,

$$0 = \frac{d}{dt}|_0 \pi = \frac{d}{dt}|_0 \pi P(t) = \pi Q.$$

The solutions of $\pi Q = 0$ correspond to the null space of Q^{tr} which implies

$$\mathrm{Nul}\, Q^{\mathrm{tr}} = \mathrm{Nul}\begin{bmatrix} -\alpha & \beta \\ \alpha & -\beta \end{bmatrix} = \mathbb{R}\cdot\begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

and hence we have again recovered $\pi = \frac{1}{\tau}\left[\beta \ \alpha\right]$.

Example 6.18 (2×2 case III). We now compute e^{tQ} by the power series method as follows. A simple computation shows that

$$Q^2 = \begin{bmatrix} \alpha\beta + \alpha^2 & -\alpha\beta - \alpha^2 \\ -\alpha\beta - \beta^2 & \alpha\beta + \beta^2 \end{bmatrix} = \tau \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} = -\tau Q.$$

Hence it follows by induction that $Q^n = (-\tau)^{n-1} Q$ and therefore,

$$\begin{aligned} P(t) = e^{tQ} &= I + \sum_{n=1}^{\infty}\frac{t^n}{n!}(-\tau)^{n-1} Q \\ &= I - \frac{1}{\tau}\sum_{n=1}^{\infty}\frac{t^n}{n!}(-\tau)^n Q = I - \frac{1}{\tau}\left(e^{-\tau t} - 1\right) Q \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{\tau}\left(e^{-\tau t} - 1\right)\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \\ &= \begin{bmatrix} \frac{\alpha}{\tau}\left(e^{-t\tau} - 1\right) + 1 & -\frac{\alpha}{\tau}\left(e^{-t\tau} - 1\right) \\ -\frac{\beta}{\tau}\left(e^{-t\tau} - 1\right) & \frac{\beta}{\tau}\left(e^{-t\tau} - 1\right) + 1 \end{bmatrix} \\ &= \frac{1}{\tau}\begin{bmatrix} \alpha e^{-t\tau} + \beta & \alpha\left(1 - e^{-t\tau}\right) \\ \beta\left(1 - e^{-t\tau}\right) & \beta e^{-t\tau} + \alpha \end{bmatrix} \end{aligned}$$

: Let us again verify that this answer is correct;

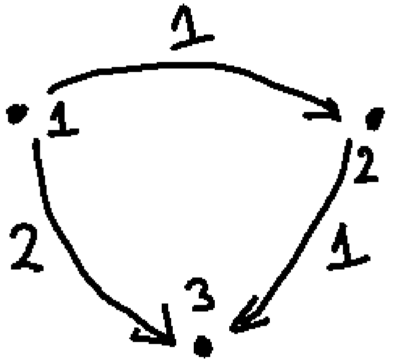
$$\dot{P}(t) = e^{-\tau t} Q \text{ while}$$

$$P(t) Q = Q - \frac{1}{\tau}\left(e^{-\tau t} - 1\right)(-\tau) Q = Q + \left(e^{-\tau t} - 1\right) Q = \dot{P}(t).$$

Example 6.19. Let $S = \{1, 2, 3\}$ and

$$Q=\begin{array}{cc} \begin{matrix}1 & 2 & 3\end{matrix} & \\ \begin{bmatrix}-3&1&2\\0&-1&1\\0&0&0\end{bmatrix} & \begin{matrix}1\\2\\3\end{matrix}\end{array}$$

which we represent by Figure 6.19. Let $\pi=(\pi_1,\pi_2,\pi_3)$ be a given initial (at $t=0$) distribution (of sand say) on S and let $\pi(t):=\pi e^{tQ}$ be the distribution at time t. Then



$$\dot{\pi}(t)=\pi e^{tQ}Q=\pi(t)\,Q.$$

In this particular example this gives,

$$\begin{aligned}\left[\dot{\pi}_1\ \dot{\pi}_2\ \dot{\pi}_3\right]&=\left[\pi_1\ \pi_2\ \pi_3\right]\begin{bmatrix}-3&1&2\\0&-1&1\\0&0&0\end{bmatrix}\\&=\left[-3\pi_1\ \pi_1-\pi_2\ 2\pi_1+\pi_2\right],\end{aligned}$$

or equivalently,

$$\dot{\pi}_1=-3\pi_1 \tag{6.21}$$
$$\dot{\pi}_2=\pi_1-\pi_2 \tag{6.22}$$
$$\dot{\pi}_3=2\pi_1+\pi_2. \tag{6.23}$$

Notice that these equations are easy to read off from Figure 6.19. For example, the second equation represents the fact that rate of change of sand at site 2 is equal to the rate which sand is entering site 2 (in this case from 1 with rate $1\pi_1$) minus the rate at which sand is leaving site 2 (in this case $1\pi_2$ is the rate that sand is being transported to 3). Similarly, site 3 is greedy and never gives up any of its sand while happily receiving sand from site 1 at rate $2\pi_1$ and from site 2 are rate $1\pi_2$. Solving Eq. (6.21) gives,

$$\pi_1(t)=e^{-3t}\pi_1(0)$$

and therefore Eq. (6.22) becomes

$$\dot{\pi}_2=e^{-3t}\pi_1(0)-\pi_2$$

which, by Lemma 6.14 below, has solution,

$$\begin{aligned}\pi_2(t)&=e^{-t}\pi_2(0)+e^{-t}\int_0^t e^{\tau}e^{-3\tau}\pi_1(0)\,d\tau\\&=\frac{1}{2}\left(e^{-t}-e^{-3t}\right)\pi_1(0)+e^{-t}\pi_2(0).\end{aligned}$$

Using this back in Eq. (6.23) then shows

$$\begin{aligned}\dot{\pi}_3&=2e^{-3t}\pi_1(0)+\frac{1}{2}\left(e^{-t}-e^{-3t}\right)\pi_1(0)+e^{-t}\pi_2(0)\\&=\left(\frac{1}{2}e^{-t}+\frac{3}{2}e^{-3t}\right)\pi_1(0)+e^{-t}\pi_2(0)\end{aligned}$$

which integrates to

$$\begin{aligned}\pi_3(t)&=\left(\frac{1}{2}\left[1-e^{-t}\right]+\frac{1}{2}\left(1-e^{-3t}\right)\right)\pi_1(0)+\left(1-e^{-t}\right)\pi_2(0)+\pi_3(0)\\&=\left(1-\frac{1}{2}\left[e^{-t}+e^{-3t}\right]\right)\pi_1(0)+\left(1-e^{-t}\right)\pi_2(0)+\pi_3(0).\end{aligned}$$

Thus we have

$$\begin{bmatrix}\pi_1(t)\\\pi_2(t)\\\pi_3(t)\end{bmatrix}=\begin{bmatrix}e^{-3t}\pi_1(0)\\\frac{1}{2}\left(e^{-t}-e^{-3t}\right)\pi_1(0)+e^{-t}\pi_2(0)\\\left(1-\frac{1}{2}\left[e^{-t}+e^{-3t}\right]\right)\pi_1(0)+\left(1-e^{-t}\right)\pi_2(0)+\pi_3(0)\end{bmatrix}$$

$$=\begin{bmatrix}e^{-3t}&0&0\\\frac{1}{2}\left(e^{-t}-e^{-3t}\right)&e^{-t}&0\\1-\frac{1}{2}\left[e^{-t}+e^{-3t}\right]&1-e^{-t}&1\end{bmatrix}\begin{bmatrix}\pi_1(0)\\\pi_2(0)\\\pi_3(0)\end{bmatrix}.$$

From this we may conclude that

$$P(t) = e^{tQ} = \begin{bmatrix} e^{-3t} & 0 & 0 \\ \frac{1}{2}\left(e^{-t} - e^{-3t}\right) & e^{-t} & 0 \\ 1 - \frac{1}{2}\left[e^{-t} + e^{-3t}\right] & 1 - e^{-t} & 1 \end{bmatrix}^{\mathrm{tr}}$$

$$= \begin{bmatrix} e^{-3t} & \left(\frac{1}{2}e^{-t} - \frac{1}{2}e^{-3t}\right) & \left(1 - \frac{1}{2}e^{-t} - \frac{1}{2}e^{-3t}\right) \\ 0 & e^{-t} & -e^{-t} + 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

6.4 Construction of continuous time Markov processes

Theorem 6.20. *Let $\{\rho_{ij}\}_{i,j\in S}$ be a discrete time Markov matrix over a discrete state space, S and $\{Y_n\}_{n=0}^{\infty}$ be the corresponding Markov chain. Also let $\{N_t\}_{t\geq 0}$ be a Poisson process with rate $\lambda > 0$ which is independent of $\{Y_n\}$. Then $X_t := Y_{N_t}$ is a continuous time Markov chain with transition semi-group given by,*

$$P(t) = e^{t\lambda(\rho - I)} = e^{-\lambda t} e^{t\lambda\rho}.$$

Proof. (To be supplied later.) STOP ∎

6.5 Jump and Hold Description

We would now like to make a direct connection between Q and the Markov process X_t. To this end, let τ denote the first time the process makes a jump between two states. In this section we are going to write x and y for typical element in the state space, S.

Theorem 6.21. *Let $Q_x := -Q_{x,x} \geq 0$. Then $P_x(S > t) = e^{-Q_x t}$, which shows that relative P_x, S is exponentially distributed with parameter Q_x. Moreover, X_S is independent of S and*

$$P_x(X_S = y) = Q_{x,y}/Q_x.$$

Proof. For the first assertion we let

$$A_n := \left\{X\left(\frac{i}{2^n}t\right) = x \text{ for } i = 1, 2, \dots, 2^n - 1, 2^n\right\}.$$

Then

$$A_n \downarrow \{X(s) = x \text{ for } s \leq t\} = \{S > t\}$$

and therefore, $P_x(A_n) \downarrow P_x(S > t)$. Since,

$$P(A_n) = \left[P_{x,x}(t/2^n)\right]^{2^n} = \left[1 - \frac{tQ_x}{2^n} + O\left(\left(1/2^n\right)^2\right)\right]^{2^n} \\ \to e^{-tQ_x} \text{ as } n \to \infty,$$

we have shown $P_x(S > t) = e^{-tQ_x}$.

First proof of the second assertion. Let T be the time between the second and first jump of the process. Then by the strong Markov property, for any $t \geq 0$ and $\Delta > 0$ small, we have,

$$\begin{aligned} P_x(t < S \leq t + \Delta,\ T \leq \Delta) &= \sum_{y\in S} P_x(t < S \leq t + \Delta,\ T \leq \Delta,\ X_S = y) \\ &= \sum_{y\in S} P_x(t < S \leq t + \Delta,\ X_S = y) \cdot P_y(T \leq \Delta) \\ &= \sum_{y\in S} P_x(t < S \leq t + \Delta,\ X_S = y) \cdot \left(1 - e^{-Q_y\Delta}\right) \\ &\leq \min_{y\in S}\left(1 - e^{-Q_y\Delta}\right) \sum_{y\in S} P_x(t < S \leq t + \Delta,\ X_S = y) \\ &= \min_{y\in S}\left(1 - e^{-Q_y\Delta}\right) P_x(t < S \leq t + \Delta) \\ &= \min_{y\in S}\left(1 - e^{-Q_y\Delta}\right) \int_t^{t+\Delta} Q_x e^{-Q_x\tau} d\tau = O\left(\Delta^2\right). \end{aligned}$$

(Here we have used that the rates, $\{Q_y\}_{y\in S}$ are bounded which is certainly the case when $\#(S) < \infty$.) Therefore the probability of two jumps occurring in the time interval, $[t, t + \Delta]$, may be ignored and we have,

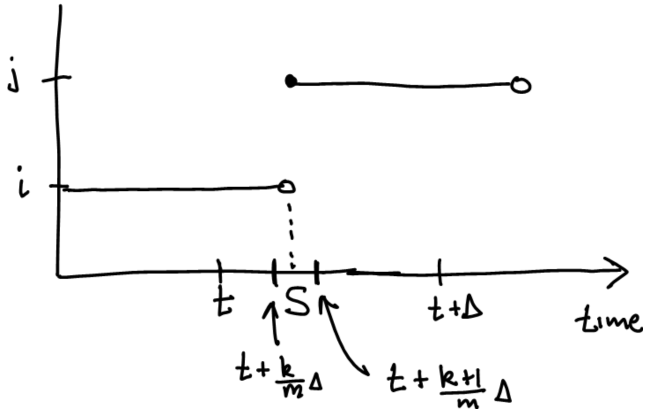
$$\begin{aligned} P_x(X_S = y, t < S \leq t + \Delta) &= P_x(X_{t+\Delta} = y, S > t) + o(\Delta) \\ &= P_x(X_{t+\Delta} = y, X_t = x, S > t) + o(\Delta) \\ &= \lim_{n\to\infty}\left[1 - \frac{tQ_x}{n} + O(n^{-2})\right]^n P_{x,y}(\Delta) + o(\Delta) \\ &= e^{-tQ_x} P_{x,y}(\Delta) + o(\Delta). \end{aligned}$$

Also

$$P_x(t < S \leq t + \Delta) = \int_t^{t+\Delta} Q_x e^{-Q_x s} ds = e^{-Q_x t} - e^{-Q_x(t+\Delta)} = Q_x e^{-Q_x t}\Delta + o(\Delta).$$

Therefore,

$$\begin{aligned} P_x(X_S = y | S = t) &= \lim_{\Delta\downarrow 0} \frac{P_x(X_S = y, t < S \leq t + \Delta)}{P_x(t < S \leq t + \Delta)} \\ &= \lim_{\Delta\downarrow 0} \frac{e^{-tQ_x} P_{x,y}(\Delta) + o(\Delta)}{Q_x e^{-Q_x t}\Delta + o(\Delta)} = \frac{1}{Q_x} \lim_{\Delta\downarrow 0} \frac{P_{x,y}(\Delta)}{\Delta} = Q_{x,y}/Q_x. \end{aligned}$$



This shows that S and X_S are independent and that $P_x(X_S = y) = Q_{x,y}/Q_x$.

Second Proof. For $t > 0$ and $\delta > 0$, we have that

$$\begin{aligned}P_x(S > t, X_{t+\delta} = y) &= \lim_{n\to\infty} P_x(X_{t+\delta} = y \text{ and } X\left(\frac{i}{2^n}t\right) = x \text{ for } i = 1, 2, \dots, 2^n)\\ &= \lim_{n\to\infty} \left[P_{x,x}(t/2^n)\right]^{2^n} P_{xy}(\delta)\\ &= P_{xy}(\delta) \lim_{n\to\infty}\left[1 - \frac{tQ_x}{2^n} + O\left(2^{-2n}\right)\right]^{2^n} = P_{xy}(\delta)e^{-tQ_x}.\end{aligned}$$

With this computation in hand, we may now compute $P_x(X_S = y, t < S \le t + \Delta)$ using the Figure 6.5 as our guide

So according Figure 6.5, we must have $X_S = y \;\&\; t < S \le t + \Delta$ iff for all large n there exists $0 \le k < n$ such that $S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y$ and therefore

$$\begin{aligned}&P_x(X_S = y \;\&\; t < S \le t + \Delta)\\ &\quad= \lim_{n\to\infty} P_x\begin{pmatrix} S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y\\ \text{for some } 0 \le k < n\end{pmatrix}\\ &\quad= \lim_{n\to\infty}\sum_{k=0}^{n-1} P_x(S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y)\\ &\quad= \lim_{n\to\infty}\sum_{k=0}^{n-1} P_{xy}(\Delta/n)e^{-(t+k\Delta/n)Q_x}\\ &\quad= \lim_{n\to\infty}\sum_{k=0}^{n-1} e^{-(t+k\Delta/n)Q_x}(\frac{\Delta}{n}Q_{xy} + o(n^{-1}))\\ &\quad= Q_{xy}\int_t^{t+\Delta} e^{-Q_x s}ds = \frac{Q_{x,y}}{Q_x}\int_t^{t+\Delta} Q_x e^{-Q_x s}ds\\ &\quad= \frac{Q_{x,y}}{Q_x}P_x(t < S \le t + \Delta).\end{aligned}$$

Letting $t \downarrow 0$ and $\Delta \uparrow \infty$ in this equation we learn that

$$P_x(X_S = y) = \frac{Q_{x,y}}{Q_x}$$

and hence

$$P_x(X_S = y, t < S \le t + \Delta) = P_x(X_S = y) \cdot P_x(t < S \le t + \Delta).$$

This proves also that X_S and S are independent random variables. ■

Remark 6.22. Technically in the proof above, we have used the identity,

$$\begin{aligned}&\{X_S = y \;\&\; t < S \le t + \Delta\}\\ &\quad= \cup_{N=1}^{\infty} \cap_{n\ge N} \cup_{0\le m<n} \left\{S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y\right\}.\end{aligned}$$

Using Theorem 6.21 along with Fact 5.3 leads to the following description of the Markov process associated to Q. Define a Markov matrix, $\tilde{P}$, by

$$\tilde{P}_{xy} := \begin{cases} \frac{Q_{x,y}}{-Q_{x,x}} & \text{if } x \ne y\\ 0 & \text{if } x = y\end{cases} \text{ for all } x, y \in S. \tag{6.24}$$

The process X starting at x may be described as follows: 1) stay at x for an $\exp(Q_x)$ – amount of time, S_1, then jump to x_1 with probability $\tilde{P}_{x,x_1}$. Stay at x_1 for an $\exp(Q_{x_1})$ amount of time, S_2, independent of S_1 and then jump to x_2 with probability $\tilde{P}_{x_1,x_2}$. Stay at x_2 for an $\exp(Q_{x_2})$ amount of time, S_3, independent of S_1 and S_2 and then jump to x_3 with probability $\tilde{P}_{x_2,x_3}$, etc. etc. etc. etc. The next corollary formalizes these rules.

Corollary 6.23. *Let Q be the infinitesimal generator of a Markov semigroup $P(t)$. Then the Markov chain, $\{X_t\}$, associated to $P(t)$ may be described as follows. Let $\{Y_k\}_{k=0}^{\infty}$ denote the discrete time Markov chain with Markov matrix $\tilde{P}$ as in Eq. (6.24). Let $\{S_j\}_{j=1}^{\infty}$ be random times such that given $\{Y_j = x_j : j \leq n\}$, $S_j \overset{d}{=} \exp(Q_{x_{j-1}})$ and the $\{S_j\}_{j=1}^{n}$ are independent for $1 \leq j \leq n$.*[2] *Now let $N_t = \max\{j : S_1 + \cdots + S_j \leq t\}$ (see Figure 6.2) and $X_t := Y_{N t}$. Then $\{X_t\}_t \geq 0$ is the Markov process starting at x with Markov semi-group, $P(t) = e^t Q$.*

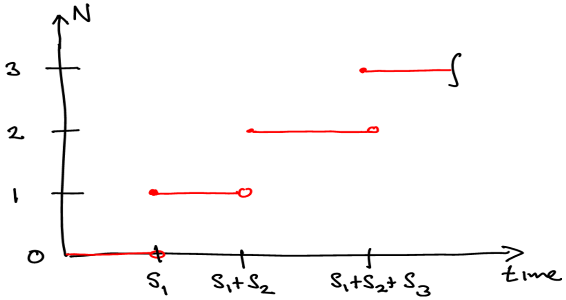


Fig. 6.2. Defining N_t.

In a manner somewhat similar to the proof of Example 5.8 one shows the description in Corollary 6.23 is defines a Markov process with the correct semi-group, $P(t)$. For the much more on the details the reader is referred to Norris [5, See Theorems 2.8.2 and 2.8.4].

---

[2] A concrete way to chose the $\{S_j\}_{j=1}^{\infty}$ is as follows. Given a sequence, $\{T_j\}_{j=1}^{\infty}$, of i.i.d. $\exp(1)$ – random variables which are independent of $\{Y\}$, define $S_j := T_j / Q_{Y_{j-1}}$.

7

Continuous Time M.C. Examples

7.1 Birth and Death Process basics

A **birth and death process** is a continuous time Markov chain with state space being $S = \{0, 1, 2, \dots\}$ and transitions rates of the form;

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3 \dots \underset{\mu_{n-1}}{\overset{\lambda_{n-2}}{\rightleftarrows}} (n-1) \underset{\mu_n}{\overset{\lambda_{n-1}}{\rightleftarrows}} n \underset{\mu_{n+1}}{\overset{\lambda_n}{\rightleftarrows}} (n+1) \dots$$

The associated Q matrix for this chain is given by

$$Q = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \dots \\ \hline 0 & -\lambda_0 & \lambda_0 & & & \\ 1 & \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & \\ 2 & & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \\ 3 & & & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 \\ 4 & & & & \ddots & \ddots \ \ddots \\ \vdots & & & & & \end{array}.$$

If $\pi_n(t) = P(X(t) = n)$, then $\pi(t) = (\pi_n(t))_{n \geq 0}$ satisfies, $\dot{\pi}(t) = \pi(t) Q$ which written out in components is the system of differential equations;

$$\begin{aligned} \dot{\pi}_0(t) &= -\lambda_0 \pi_0(t) + \mu_1 \pi_1(t) \\ \dot{\pi}_1(t) &= \lambda_0 \pi_0(t) - (\lambda_1 + \mu_1) \pi_1(t) + \mu_2 \pi_2(t) \\ &\vdots \\ \dot{\pi}_n(t) &= \lambda_{n-1} \pi_{n-1}(t) - (\lambda_n + \mu_n) \pi_n(t) + \mu_{n+1} \pi_{n+1}(t). \\ &\vdots \end{aligned}$$

The associated discrete time chain is described by the jump diagram,

$$0 \underset{\frac{\mu_1}{\lambda_1+\mu_1}}{\overset{1}{\rightleftarrows}} 1 \underset{\frac{\mu_2}{\lambda_2+\mu_2}}{\overset{\frac{\lambda_1}{\lambda_1+\mu_1}}{\rightleftarrows}} 2 \underset{\frac{\mu_3}{\lambda_3+\mu_3}}{\overset{\frac{\lambda_2}{\lambda_2+\mu_2}}{\rightleftarrows}} 3 \cdots \rightleftarrows (n-1) \underset{\frac{\mu_n}{\lambda_n+\mu_n}}{\overset{\frac{\lambda_{n-1}}{\lambda_{n-1}+\mu_{n-1}}}{\rightleftarrows}} n \underset{\frac{\mu_{n+1}}{\lambda_{n+1}+\mu_{n+1}}}{\overset{\frac{\lambda_n}{\lambda_n+\mu_n}}{\rightleftarrows}} \dots$$

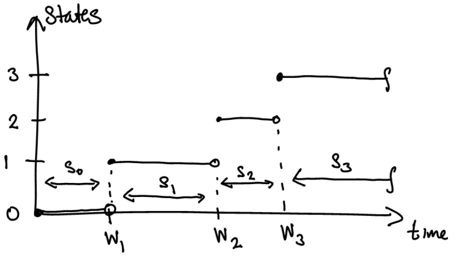
In the jump hold description, a particle follows this discrete time chain. When it arrives at a site, say n, it stays there for an $\exp(\lambda_n + \mu_\nu)$ – time and then jumps to either $n-1$ or n with probability $\frac{\mu_n}{\lambda_n+\mu_n}$ or $\frac{\lambda_n}{\lambda_n+\mu_n}$ respectively. Given your homework problem we may also describe these transitions by assuming at each site we have a death clock $D_n = \exp(\mu_n)$ and a Birth clock $B_n = \exp(\lambda_n)$ with B_n and D_n being independent. We then stay at site n until either B_n or D_n rings, i.e. for $\min(B_n, D_n) = \exp(\lambda_n + \mu_n)$ – amount of time. If B_n rings first we go to $n+1$ while if D_n rings first we go to $n-1$. When we are at 0 we go to 1 after waiting $\exp(\lambda_0)$ – amount of time.

7.2 Pure Birth Process:

The infinitesimal generator for a pure Birth process is described by the following rate diagram

$$0 \xrightarrow{\lambda_0} 1 \xrightarrow{\lambda_1} 2 \xrightarrow{\lambda_2} \dots \xrightarrow{\lambda_{n-1}} (n-1) \xrightarrow{\lambda_{n-1}} n \xrightarrow{\lambda_n} \dots.$$

For simplicity we are going to assume that we start at state 0. We will examine this model is both the sojourn description and the infinitesimal description. The typical sample path is shown in Figure 7.2.



7.2.1 Infinitesimal description

The matrix Q is this case is given by

$$Q_{i,i+1} = \lambda_i \text{ and } Q_{ii} = -\lambda_i \text{ for all } i = 0, 1, 2, \dots$$

with all other entries being zero. Thus we have

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \dots & \end{array} \\ \left[\begin{array}{cccccc} -\lambda_0 & \lambda_0 & & & & \\ & -\lambda_1 & \lambda_1 & & & \\ & & -\lambda_2 & \lambda_2 & & \\ & & & -\lambda_3 & \lambda_3 & \\ & & & & \ddots & \ddots \\ & & & & & \end{array}\right. \end{array}.$$

If we now let

$$\pi_j(t) = P_0(X(t) = j) = \left[\pi(0)\, e^{tQ}\right]_j$$

then $\pi_j(t)$ satisfies the system of differential equations;

$$\begin{aligned} \dot{\pi}_0(t) &= -\lambda_0 \pi_0(t) \\ \dot{\pi}_1(t) &= \lambda_0 \pi_0(t) - \lambda_1 \pi_1(t) \\ &\vdots \\ \dot{\pi}_n(t) &= \lambda_{n-1} \pi_{n-1}(t) - \lambda_n \pi_n(t) \\ &\vdots \end{aligned}$$

The solution to the first equation is given by

$$\pi_0(t) = e^{-\lambda_0 t} \pi(0) = e^{-\lambda_0 t}$$

and the remaining may now be obtained inductively, see the ODE Lemma 6.14, using

$$\pi_n(t) = \lambda_{n-1} e^{-\lambda_n t} \int_0^t e^{\lambda_n \tau} \pi_{n-1}(\tau)\, d\tau. \tag{7.1}$$

So for example

$$\begin{aligned} \pi_1(t) &= \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 \tau} \pi_0(\tau)\, d\tau = \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 \tau} e^{-\lambda_0 \tau} d\tau \\ &= \frac{\lambda_0}{\lambda_1 - \lambda_0} e^{-\lambda_1 t} e^{(\lambda_1 - \lambda_0)\tau} |_{\tau=0}^{\tau=t} = \frac{\lambda_0}{\lambda_1 - \lambda_0} \left[e^{-\lambda_1 t} e^{(\lambda_1 - \lambda_0)t} - e^{-\lambda_1 t} \right] \\ &= \frac{\lambda_0}{\lambda_1 - \lambda_0} \left[e^{-\lambda_0 t} - e^{-\lambda_1 t} \right]. \end{aligned}$$

If $\lambda_1 = \lambda_0$, this becomes, $\pi_1(t) = (\lambda_0 t)\, e^{-\lambda_0 t}$ instead. In principle one can compute all of these integrals (you have already done the case where $\lambda_j = \lambda$ for all j) to find all of the $\pi_n(t)$. The formula for the solution is given as

$$\pi_n(t) = P(X(t) = n | X(0) = 0) = \lambda_0 \dots \lambda_{n-1} \left[\sum_{k=0}^{n} B_{k,n} e^{-\lambda_k t} \right]$$

where the $B_{k,n}$ are given on p. 338 of the book.

To see that this form of the answer is reasonable, if we look at the equations for $n = 0, 1, 2, 3$, we have

$$\begin{aligned} \dot{\pi}_0(t) &= -\lambda_0 \pi_0(t) \\ \dot{\pi}_1(t) &= \lambda_0 \pi_0(t) - \lambda_1 \pi_1(t) \\ \dot{\pi}_2(t) &= \lambda_1 \pi_1(t) - \lambda_2 \pi_2(t) \\ \dot{\pi}_3(t) &= \lambda_2 \pi_2(t) - \lambda_3 \pi_3(t) \end{aligned}$$

and the matrix associated to this system is

$$Q' = \begin{bmatrix} -\lambda_0 & \lambda_0 & & \\ & -\lambda_1 & \lambda_1 & \\ & & -\lambda_2 & \lambda_2 \\ & & & -\lambda_3 \end{bmatrix}$$

so that $(\pi_0(t), \dots, \pi_3(t)) = (1, 0, 0, 0)\, e^{tQ'}$. If all of the λ_j are distinct, then Q' has $\{\lambda_j\}_{j=0}^3$ as its distinct eigenvalues and hence is diagonalizable. Therefore we will have

$$(\pi_0(t), \dots, \pi_3(t)) = (1, 0, 0, 0)\, S \begin{bmatrix} e^{-t\lambda_0} & & & \\ & e^{-t\lambda_1} & & \\ & & e^{-t\lambda_2} & \\ & & & e^{-t\lambda_3} \end{bmatrix} S^{-1}$$

for some invertible matrix S. In particular it follows that $\pi_3(t)$ must be a linear combination of $\left\{e^{-t\lambda_j}\right\}_{j=0}^3$. Generalizing this argument shows that there must be constants, $\{C_{k,n}\}_{k=0}^n$ such that

$$\pi_n(t) = \sum_{k=0}^{n} C_{kn} e^{-t\lambda_k}.$$

We may now plug these expressions into the differential equations,

$$\dot{\pi}_n(t) = \lambda_{n-1} \pi_{n-1}(t) - \lambda_n \pi_n(t),$$

to learn

$$-\sum_{k=0}^{n} \lambda_k C_{kn} e^{-t\lambda_k} = \lambda_{n-1} \sum_{k=0}^{n-1} C_{k,n-1} e^{-t\lambda_k} - \lambda_n \sum_{k=0}^{n} C_{kn} e^{-t\lambda_k}.$$

Since one may show $\left\{e^{-t\lambda_k}\right\}_{k=0}^{n}$ are linearly independent, we conclude that

$$-\lambda_k C_{kn} = \lambda_{n-1} C_{k,n-1} \cdot 1_{k\le n-1} - \lambda_n C_{kn} \text{ for } k = 0, 1, 2, \dots, n.$$

This equation gives no information for $k = n$, but for $k < n$ it implies,

$$C_{k,n} = \frac{\lambda_{n-1}}{\lambda_n - \lambda_k} C_{k,n-1} \text{ for } k \le n - 1.$$

To discover the value of $C_{n,n}$ we use the fact that $\sum_{k=0}^{n} C_{kn} = \pi_n(0) = 0$ for $n \ge 1$ to learn,

$$C_{n,n} = -\sum_{k=0}^{n-1} C_{k,n} = -\sum_{k=0}^{n-1} \frac{\lambda_{n-1}}{\lambda_n - \lambda_k} C_{k,n-1}.$$

One may determine all of the coefficients from these equations. For example, we know that $C_{00} = 1$ and therefore,

$$C_{0,1} = \frac{\lambda_0}{\lambda_1 - \lambda_0} \text{ and } C_{1,1} = -C_{0,1} = -\frac{\lambda_0}{\lambda_1 - \lambda_0}.$$

Thus we learn that

$$\pi_1(t) = \frac{\lambda_0}{\lambda_1 - \lambda_0}\left(e^{-\lambda_0 t} - e^{-\lambda_1 t}\right)$$

as we have seen from above.

Remark 7.1. It is interesting to observe that

$$\begin{aligned}
\frac{d}{dt}\left(\pi_0(t), \dots, \pi_3(t)\right)\begin{bmatrix}1\\1\\1\\1\end{bmatrix} &= \frac{d}{dt}(1,0,0,0)\, e^{tQ'}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \\
&= (1,0,0,0)\, e^{tQ'} Q' \begin{bmatrix}1\\1\\1\\1\end{bmatrix}
\end{aligned}$$

where

$$Q'\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}-\lambda_0 & \lambda_0 & & \\ & -\lambda_1 & \lambda_1 & \\ & & -\lambda_2 & \lambda_2 \\ & & & -\lambda_3\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\\-\lambda_3\end{bmatrix}$$

and therefore,

$$\frac{d}{dt}\left(\pi_0(t), \dots, \pi_3(t)\right)\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \le 0.$$

This shows that $\sum_{j=0}^{3} \pi_j(t) \le \sum_{j=0}^{3} \pi_j(0) = 1$. Similarly one shows that

$$\sum_{j=0}^{n} \pi_j(t) \le 1 \text{ for all } t \ge 0 \text{ and } n.$$

Letting $n \to \infty$ in this estimate then implies

$$\sum_{j=0}^{\infty} \pi_j(t) \le 1.$$

It is possible that we have a strict inequality here! We will discuss this below.

Remark 7.2. We may iterate Eq. (7.1) to find,

$$\begin{aligned}
\pi_1(t) &= \lambda_0 e^{-\lambda_1 t}\int_0^t e^{\lambda_1 \tau}\pi_0(\tau)\, d\tau = \lambda_0 e^{-\lambda_1 t}\int_0^t e^{\lambda_1\tau} e^{-\lambda_0\tau} d\tau \\
\pi_2(t) &= \lambda_1 e^{-\lambda_2 t}\int_0^t e^{\lambda_2\sigma}\pi_1(\sigma)\, d\sigma \\
&= \lambda_1 e^{-\lambda_2 t}\int_0^t e^{\lambda_2\sigma}\left[\lambda_0 e^{-\lambda_1\sigma}\int_0^\sigma e^{\lambda_1\tau} e^{-\lambda_0\tau} d\tau\right] d\sigma \\
&= \lambda_0\lambda_1 e^{-\lambda_2 t}\int_0^t d\sigma e^{(\lambda_2-\lambda_1)\sigma}\int_0^\sigma e^{(\lambda_1-\lambda_0)\tau} d\tau \\
&= \lambda_0\lambda_1 e^{-\lambda_2 t}\int_{0\le\tau\le\sigma\le t} e^{(\lambda_2-\lambda_1)\sigma + (\lambda_1-\lambda_0)\tau} d\sigma d\tau
\end{aligned}$$

and continuing on this way we find,

$$\pi_n(t) = \lambda_0\lambda_1\dots\lambda_{n-1} e^{-\lambda_n t}\int_{0\le s_1\le s_2\le\dots\le s_n\le t} e^{\sum_{j=1}^{n}(\lambda_j-\lambda_{j-1})s_j} ds_1\dots ds_n. \tag{7.2}$$

In the special case where $\lambda_j = \lambda$ for all j, this gives, by Lemma 7.3 below with $f(s) = 1$,

$$\pi_n(t) = \lambda^n e^{-\lambda t}\int_{0\le s_1\le s_2\le\dots\le s_n\le t} ds_1\dots ds_n = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \tag{7.3}$$

Another special case of interest is when $\lambda_j = \beta(j+1)$ for all $j \ge 0$. This will be the Yule process discussed below. In this case,

$$\begin{aligned}\pi_n(t) &= n!\beta^n e^{-(n+1)\beta t}\int_{0\le s_1\le s_2\le\dots\le s_n\le t} e^{\beta\sum_{j=1}^n s_j} ds_1\dots ds_n \\ &= n!\beta^n e^{-(n+1)\beta t}\frac{1}{n!}\left(\int_0^t e^{\beta s}ds\right)^n = \beta^n e^{-(n+1)\beta t}\left(\frac{e^{\beta t}-1}{\beta}\right)^n \\ &= e^{-\beta t}\left(1-e^{-\beta t}\right)^n, \qquad (7.4)\end{aligned}$$

wherein we have used Lemma 7.3 below for the the second equality.

Lemma 7.3. *Let $f(t)$ be a continuous function, then for all $n\in\mathbb{N}$ we have*

$$\int_{0\le s_1\le s_2\le\dots\le s_n\le t} f(s_1)\dots f(s_n)\,ds_1\dots ds_n = \frac{1}{n!}\left(\int_0^t f(s)\,ds\right)^n.$$

Proof. Let $F(t) := \int_0^t f(s)\,ds$. The proof goes by induction on n. The statement is clearly true when $n=1$ and if it holds at level n, then

$$\begin{aligned}&\int_{0\le s_1\le s_2\le\dots\le s_n\le s_{n+1}\le t} f(s_1)\dots f(s_n)f(s_{n+1})\,ds_1\dots ds_n ds_{n+1} \\ &= \int_0^t\left(\int_{0\le s_1\le s_2\le\dots\le s_n\le s_{n+1}} f(s_1)\dots f(s_n)\,ds_1\dots ds_n\right)f(s_{n+1})\,ds_{n+1} \\ &= \int_0^t\left(\frac{1}{n!}\left(F(s_{n+1})\right)^n\right)F'(s_{n+1})\,ds_{n+1} = \int_0^{F(t)}\left(\frac{1}{n!}u^n\right)du \\ &= \frac{F(t)^{n+1}}{(n+1)!}\end{aligned}$$

as required. ∎

7.2.2 Yule Process

Suppose that each member of a population gives birth independently to one offspring at an exponential time with rate β. If there are k members of the population with birth times, $T_1,\dots,T_k$, then the time of the birth for this population is $\min(T_1,\dots,T_k) = S_k$ where S_k is now an exponential random variable with parameter, βk. This description gives rise to a pure Birth process with parameters $\lambda_k = \beta k$. In this case we start with initial distribution, $\pi_j(0) = \delta_{j,1}$. We have already solved for $\pi_k(t)$ in this case. Indeed from from Eq. (7.4) after a shift of the index by 1, we find,

$$\pi_n(t) = e^{-\beta t}\left(1-e^{-\beta t}\right)^{n-1} \text{ for } n\ge 1.$$

7.2.3 Sojourn description

Let $\{S_n\}_{n=0}^\infty$ be independent exponential random variables with $P(S_n>t) = e^{-\lambda_n t}$ for all n and let

$$W_k := S_0+\dots+S_{k-1}$$

be the time of the k^{th} – birth, see Figure 7.2 where the graph of $X(t)$ is shown as determined by the sequence $\{S_n\}_{n=0}^\infty$. With this notation we have

$$\begin{aligned}P(X(t)=0) &= P(S_0>t) = e^{-\lambda_0 t} \\ P(X(t)=1) &= P(S_0\le t<S_0+S_1) = P(W_1\le t<W_2) \\ P(X(t)=2) &= P(W_2\le t<W_3) \\ &\vdots \\ P(X(t)=j) &= P(W_j\le t<W_{j+1})\end{aligned}$$

where $\{W_j\le t<W_{j+1}\}$ represents the event where the j^{th} – birth has occurred by time t but the j^{th} – birth as not. Consider,

$$P(W_1\le t<W_2) = \lambda_0\lambda_1\int_{0\le x_0\le t<x_0+x_1} e^{-\lambda_0 x_0}e^{-\lambda_1 x_1}dx_0dx_1.$$

Doing the x_1 -integral first gives,

$$\begin{aligned}P(X(t)=1) &= P(W_1\le t<W_2) \\ &= \lambda_0\int_{0\le x_0\le t<x_0+x_1} e^{-\lambda_0 x_0}\left[-e^{-\lambda_1 x_1}\right]_{x_1=t-x_0}^\infty dx_0 \\ &= \lambda_0\int_{0\le x_0\le t} e^{-\lambda_0 x_0}e^{-\lambda_1(t-x_0)}dx_0 \\ &= \lambda_0 e^{-\lambda_1 t}\int_{0\le x_0\le t} e^{(\lambda_1-\lambda_0)x_0}dx_0 \\ &= \frac{\lambda_0}{\lambda_1-\lambda_0}e^{-\lambda_1 t}\left[e^{(\lambda_1-\lambda_0)t}-1\right] \\ &= \frac{\lambda_0}{\lambda_1-\lambda_0}\left[e^{-\lambda_0 t}-e^{-\lambda_1 t}\right].\end{aligned}$$

There is one point which we have not yet addressed in this model, namely does it make sense without further information. In terms of the Sojourn description this comes down to the issue as to whether $P\left(\sum_{j=1}^\infty S_j = \infty\right) = 1$. Indeed, if this is not the case, we will only have $X(t)$ defined for $t<\sum_{j=1}^\infty S_j$ which may be less than infinity. The next theorem tells us precisely when this phenomenon can happen.

Theorem 7.4. *Let* $\{S_j\}_{j=1}^{\infty}$ *be independent random variables such that* $S_j \overset{d}{=} \exp(\lambda_j)$ *with* $0 < \lambda_j < \infty$ *for all* j. *Then:*

1. *If* $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$ *then* $P\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$.
2. *If* $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ *then* $P\left(\sum_{n=1}^{\infty} S_n = \infty\right) = 1$.

Proof. 1. Since

$$\mathbb{E}\left[\sum_{n=1}^{\infty} S_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[S_n] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^{\infty} S_n < \infty$ a.s.

2. By the DCT, independence, and Eq. (2.3),

$$\begin{aligned}\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} S_n}\right] &= \lim_{N\to\infty} \mathbb{E}\left[e^{-\sum_{n=1}^{N} S_n}\right] = \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{E}\left[e^{-S_n}\right] \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} \left(\frac{1}{1+\lambda_n^{-1}}\right) = \lim_{N\to\infty} \exp\left(-\sum_{n=1}^{N} \ln\left(1+\lambda_n^{-1}\right)\right) \\ &= \exp\left(-\sum_{n=1}^{\infty} \ln\left(1+\lambda_n^{-1}\right)\right).\end{aligned}$$

If λ_n does not go to infinity, then the latter sum is infinite and $\lambda_n \to \infty$ and $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ then $\sum_{n=1}^{\infty} \ln\left(1+\lambda_n^{-1}\right) = \infty$ as $\ln\left(1+\lambda_n^{-1}\right) \cong \lambda_n^{-1}$ for large n. In any case we have shown that $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} S_n}\right] = 0$ which can happen iff $e^{-\sum_{n=1}^{\infty} S_n} = 0$ a.s. or equivalently $\sum_{n=1}^{\infty} S_n = \infty$ a.s. ■

Remark 7.5. If $\sum_{k=1}^{\infty} 1/\lambda_k < \infty$ so that $P\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$, one may define $X(t) = \infty$ on $\left\{t \geq \sum_{n=1}^{\infty} S_n\right\}$. With this definition, $\{X(t)\}_{t\geq 0}$ is again a Markov process. However, most of the examples we study will satisfy $\sum_{k=1}^{\infty} 1/\lambda_k = \infty$.

7.3 Pure Death Process

A pure death process is described by the following rate diagram,

$$0 \underset{\mu_1}{\longleftarrow} 1 \underset{\mu_2}{\longleftarrow} 2 \underset{\mu_3}{\longleftarrow} 3 \dots \underset{\mu_{N-1}}{\longleftarrow} (N-1) \underset{\mu_N}{\longleftarrow} N.$$

If $\pi_j(t) = P(X(t) = j | X(0) = \pi_j(0))$, we have that

$$\begin{aligned}\dot{\pi}_N(t) &= -\mu_N \pi_N(t) \\ \dot{\pi}_{N-1}(t) &= \mu_N \pi_N(t) - \mu_{N-1}\pi_{N-1}(t) \\ &\vdots \\ \dot{\pi}_n(t) &= \mu_{n+1}\pi_{n+1}(t) - \mu_n \pi_n(t) \\ &\vdots \\ \dot{\pi}_1(t) &= \mu_2 \pi_2(t) - \mu_1 \pi_1(t) \\ \dot{\pi}_0(t) &= -\mu_1 \pi_1(t).\end{aligned}$$

Let us now suppose that $\pi_j(t) = P(X(t) = j | X(0) = N)$. A little thought shows that we may find $\pi_j(t)$ for $j = 1, 2, \dots, N$ by using the solutions for the pure Birth process with $0 \to N$, $1 \to (N-1), 2 \to (N-2), \dots,$ and $(N-1) \to 1$. We may then compute

$$\pi_0(t) := 1 - \sum_{j=1}^{N} \pi_j(t).$$

The explicit formula for these solutions may be found in the book on p. 346 in the special case where all of the death parameters are distinct.

7.3.1 Cable Failure Model

Suppose that a cable is made up of N individual strands with the life time of each strand being a $\exp(K(l))$ – random variable where $K(l) > 0$ is some function of the load, l, on the strand. We suppose that the cable starts with N – fibers and is put under a total load of NL that L is the load applied per fiber when all N fibers are unbroken. If there are k – fibers in tact, the load per fiber is NL/k and the exponential life time of each fiber is now $K(NL/k)$. Thus when k - fibers are in tact the time to the next fiber breaking is $\exp(kK(NL/k))$. So if $\{S_j\}_{j=N}^{1}$ are the Sojourn times at state j, the time to failure of the cable is $T = \sum_{j=1}^{N} S_j$ and the expected time to failure is

$$\mathbb{E}T = \sum_{j=1}^{N} \mathbb{E}S_j = \sum_{j=1}^{N} \frac{1}{kK(NL/k)} = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{\frac{k}{N}K\left(\frac{N}{k}L\right)} \cong \int_0^1 \frac{1}{xK(L/x)}dx$$

if K is a nice enough function and N is large. For example, if $K(l) = l^{\beta}/A$ for some $\beta > 0$ and $A > 0$, we find

$$\mathbb{E}T = \int_0^1 \frac{A}{x(L/x)^{\beta}}dx = \frac{A}{L^{\beta}}\int_0^1 x^{\beta-1}dx = \frac{A}{L^{\beta}\beta}.$$

Where as the expected life, at the start, of any one strand is $1/K(L) = A/L^{\beta}$. Thus the cable last only $\frac{1}{\beta}$ times the average strand life. It is actually better to let L_0 be the total load applied so that $L = L_0/N$, then the above formula becomes,

$$\mathbb{E}T = \frac{A}{L_0^{\beta}} \frac{N^{\beta}}{\beta}.$$

7.3.2 Linear Death Process basics

Similar to the Yule process, suppose that each individual in a population has a life expectancy, $T \overset{d}{=} \exp(\alpha)$. Thus if there are k members in the population at time t, using the memoryless property of the exponential distribution, we the time of the next death is has distribution, $\exp(k\alpha)$. Thus the $\mu_k = \alpha k$ in this case. Using the formula in the book on p. 346, we then learn that if we start with an population of size N, then

$$\begin{aligned}\pi_n(t) &= P(X(t) = n | X(0) = N) \\ &= \binom{N}{n} e^{-n\alpha t}\left(1 - e^{-\alpha t}\right)^{N-n} \text{ for } n = 0, 1, 2, \dots, N. \qquad (7.5)\end{aligned}$$

So $\{\pi_n(t)\}_{n=0}^{N}$ is the binomial distribution with parameter $e^{-\alpha t}$. This may be understood as follows. We have $\{X(t) = n\}$ iff there are exactly n members out of the original N still alive. Let ξ_j be the life time of the j^{th} member of the population, so that $\{\xi_j\}_{j=1}^{N}$ are i.i.d. $\exp(\mu)$ – distributed random variables. We then have the probability that a particular choice, $A \subset \{1, 2, \dots, N\}$ of n - members are alive with the others being dead is given by

$$P\left(\left(\cap_{j\in A}\{\xi_j > t\}\right) \cap \left(\cap_{j\notin A}\{\xi_j \le t\}\right)\right) = \left(e^{-\alpha t}\right)^{n}\left(1 - e^{-\alpha t}\right)^{N-n}.$$

As there are $\binom{N}{n}$ – ways to choose such subsets, $A \subset \{1, 2, \dots, N\}$, with n – members, we arrive at Eq. (7.5).

7.3.3 Linear death process in more detail

(You may safely skip this subsection.) In this subsection, we suppose that we start with a population of size N with ξ_j being the life time of the j^{th} member of the population. We assume that $\{\xi_j\}_{j=1}^{N}$ are i.i.d. $\exp(\mu)$ – distributed random variables and let $X(t)$ denote the number of people alive at time t, i.e.

$$X(t) = \#\{j : \xi_j > t\}.$$

Theorem 7.6. *The process, $\{X(t)\}_{t\ge 0}$ is the linear death Markov process with parameter, α.*

We will begin with the following lemma.

Lemma 7.7. *Suppose that B and $\{A_j\}_{j=1}^{n}$ are events such that: 1) $\{A_j\}_{j=1}^{n}$ are pairwise disjoint, 2) $P(A_j) = P(A_1)$ for all j, and 3) $P(B \cap A_j) = P(B \cap A_1)$ for all j. Then*

$$P\left(B | \cup_{j=1}^{N} A_j\right) = P(B|A_1). \qquad (7.6)$$

We also use the identity, that

$$P(B|A \cap C) = P(B|A) \qquad (7.7)$$

whenever C is independent of $\{A, B\}$.

Proof. The proof is the following simple computation,

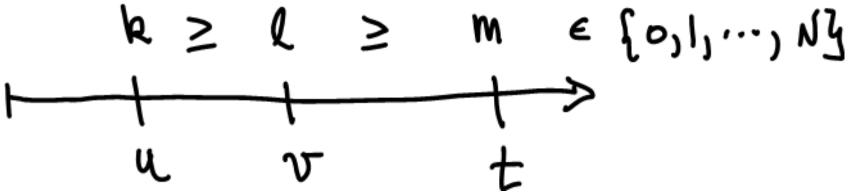
$$\begin{aligned}P\left(B | \cup_{j=1}^{n} A_j\right) &= \frac{P\left(B \cap \left(\cup_{j=1}^{n} A_j\right)\right)}{P\left(\cup_{j=1}^{n} A_j\right)} = \frac{P\left(\cup_{j=1}^{n} B \cap A_j\right)}{P\left(\cup_{j=1}^{n} A_j\right)} \\ &= \frac{\sum_{j=1}^{n} P(B \cap A_j)}{\sum_{j=1}^{n} P(A_j)} = \frac{nP(B \cap A_1)}{nP(A_1)} = P(B|A_1).\end{aligned}$$

For the second assertion, we have

$$\begin{aligned}P(B|A \cap C) &= \frac{P(B \cap A \cap C)}{P(A \cap C)} = \frac{P(B \cap A) \cdot P(C)}{P(A) \cdot P(C)} \\ &= \frac{P(B \cap A)}{P(A)} = P(B|A).\end{aligned}$$

■

Proof. Sketch of the proof of Theorem 7.6. Let $0 < u < v < t$ and $k \ge l \ge m$ as in Figure 7.3.3. Given $V \subset U \subset \{1, 2, \dots, N\}$ with $\#V = l$ and $\#U = k$, let



$$A_{U,V} = \cap_{j\in U}\{\xi_j > u\} \cap \cap_{j\notin U}\{\xi_j \le u\} \cap \cap_{j\in V}\{\xi_j > v\} \cap \cap_{j\notin V}\{\xi_j \le v\}$$

so that $\{X_u = k, X_v = l\}$ is the disjoint union of $\{A_{U,V}\}$ over all such choices of $V \subset U$ as above. Notice that $P(A_{U,V})$ is independent of how $U \subset V$ as is $P(\{X_t = m\} \cap A_{U,V})$. Therefore by Lemma 7.7, we have, with $V = \{1, 2, \dots, l\} \subset U = \{1, 2, \dots, k\}$, that

$$\begin{aligned}
&P(X_t = m | X_u = k, X_v = l) = P(X_t = m | A_{U,V}) \\
&\quad = P(\text{Exactly } m \text{ of } \xi_1, \dots, \xi_l \ > t | \xi_1 > v, \dots, \xi_l > v, v \geq \xi_{l+1} > u, \dots, v \geq \xi_k > u) \\
&\quad = P(\text{Exactly } m \text{ of } \xi_1, \dots, \xi_l \ > t | \xi_1 > v, \dots, \xi_l > v) \\
&\quad = \binom{l}{m} P(\xi_1 > t, \dots, \xi_m > t, \xi_{m+1} \leq t, \dots \xi_l \leq t | \xi_1 > v, \dots, \xi_l > v) \\
&\quad = \binom{l}{m} \frac{P(\xi_1 > t)^m \cdot P(v < \xi_1 \leq t)^{l-m}}{P(v < \xi_1)^l} \\
&\quad = \binom{l}{m} \frac{\left(e^{-\alpha t}\right)^m \cdot \left(e^{-vt} - e^{-\alpha t}\right)^{l-m}}{e^{-\alpha v l}} = \binom{l}{m} e^{-\alpha m (t-v)} \left(1 - e^{-\alpha(t-v)}\right)^{l-m}.
\end{aligned}$$

Similar considerations show that X_t has the Markov property and we have just found the transition matrix for this process to be,

$$P(X_t = m | X_v = l) = 1_{l \geq m} \binom{l}{m} e^{-\alpha m (t-v)} \left(1 - e^{-\alpha(t-v)}\right)^{l-m}.$$

So

$$P_{lm}(t) := P(X_t = m | X_0 = l) = 1_{\geq m} \binom{l}{m} e^{-\alpha m t} \left(1 - e^{-\alpha t}\right)^{l-m}.$$

Differentiating this equation at $t = 0$ implies $\frac{d}{dt}|_{0+} P_{lm}(t) = 0$ unless $m = l$ or $m = l - 1$ and

$$\begin{aligned}
\frac{d}{dt}|_{0+} P_{l\,l}(t) &= -\alpha l \text{ and} \\
\frac{d}{dt}|_{0+} P_{l\,,l-1}(t) &= \binom{l}{l-1} \alpha = \alpha l.
\end{aligned}$$

These are precisely the transition rate of the linear death process with parameter α. ■

Let us now also work out the Sojourn description in this model.

Theorem 7.8. *Suppose that* $\{\xi_j\}_{j=1}^N$ *are independent exponential random variables with parameter,* α *as in the above model for the life times of a population. Let* $W_1 < W_2 < \cdots < W_N$ *be the order statistics of* $\{\xi_j\}_{j=1}^N$, *i.e.* $\{W_1 < W_2 < \cdots < W_N\} = \{\xi_j\}_{j=1}^N$. *Hence* W_j *is the time of the* j^{th} *– death. Further let* $S_1 = W_1$, $S_2 = W_2 - W_1, \dots, S_N = W_N - W_{N-1}$ *are times between successive deaths. Then* $\{S_j\}_{j=1}^N$ *are exponential random variables with* $S_j \overset{d}{=} \exp((N - j)\alpha)$.

Proof. Since $W_1 = S_1 = \min(\xi_1, \dots, \xi_N)$, by a homework problem, $S_1 \overset{d}{=} \exp(N\alpha)$. Let

$$A_j := \left\{\xi_j < \min(\xi_k)_{k \neq j}\right\} \cap \{\xi_j = t\}.$$

We then have

$$\{W_1 = t\} = \cup_{j=1}^N A_j$$

and

$$A_j \cap \{W_2 > s + t\} = \left\{s + t < \min(\xi_k)_{k \neq j}\right\} \cap \{\xi_j = t\}.$$

By symmetry we have (this is the informal part)

$$\begin{aligned}
P(A_j) &= P(A_1) \text{ and} \\
P(A_j \cap \{W_2 > s + t\}) &= P(A_1 \cap \{W_2 > s + t\}),
\end{aligned}$$

and hence by Lemma 7.7,

$$P(W_2 > s + t | W_1 = t) = P$$

Now consider

$$\begin{aligned}
W_2 &= P(A_1 \cap \{W_2 > s + t\} | A_1) \\
&= P\left(\{\xi_1 = t\} \cap \left\{\min(\xi_k)_{k \neq 1} > s + t\right\} | \min(\xi_k)_{k \neq 1} > \xi_1 = t\right) \\
&= \frac{P\left(\min(\xi_k)_{k \neq 1} > s + t, \ \xi_1 = t\right)}{P\left(\min(\xi_k)_{k \neq 1} > t, \ \xi_1 = t\right)} \\
&= \frac{P\left(\min(\xi_k)_{k \neq 1} > s + t\right)}{P\left(\min(\xi_k)_{k \neq 1} > t\right)} = e^{-(N-1)\alpha s}
\end{aligned}$$

since $\min(\xi_k)_{k \neq 1} \overset{d}{=} \exp((N-1)\alpha)$ and the memoryless property of exponential random variables. This shows that $S_2 := W_2 - W_1 \overset{d}{=} \exp((N-1)\alpha)$.

Let us consider the next case, namely $P(W_3 - W_2 > t | W_1 = a, W_2 = a + b)$. In this case we argue as above that

$$\begin{aligned}
&P(W_3 - W_2 > t | W_1 = a, W_2 = a + b) \\
&= P(\min(\xi_3, \dots, \xi_N) - \xi_2 > t | \xi_1 = a, \ \xi_2 = a + b, \ \min(\xi_3, \dots, \xi_N) > \xi_2) \\
&= \frac{P(\min(\xi_3, \dots, \xi_N) > t + a + b, \ \xi_1 = a, \ \xi_2 = a + b, \ \min(\xi_3, \dots, \xi_N) > \xi_2)}{P(\xi_1 = a, \ \xi_2 = a + b, \ \min(\xi_3, \dots, \xi_N) > a + b)} \\
&= \frac{P(\min(\xi_3, \dots, \xi_N) > t + a + b)}{P(\min(\xi_3, \dots, \xi_N) > a + b)} = e^{-(N-2)\alpha t}.
\end{aligned}$$

We continue on this way to get the result. This proof is not rigorous, since $P(\xi_j = t) = 0$ but the spirit is correct.

Rigorous Proof. (Probably should be skipped.) In this proof, let g be a bounded function and $T_k := \min(\xi_l : l \neq k)$. We then have that T_k and ξ_k are independent, $T_k \overset{d}{=} \exp((N-1)\alpha)$, and hence

$$\begin{aligned}
\mathbb{E}\left[1_{W_2 - W_1 > t}\; g(W_1)\right] &= \sum_k \mathbb{E}\left[1_{W_2 - W_1 > t}\; g(W_1) : \xi_k < T_k\right] \\
&= \sum_k \mathbb{E}\left[1_{T_k - \xi_k > t}\; g(\xi_k) : \xi_k < T_k\right] \\
&= \sum_k \mathbb{E}\left[1_{T_k - \xi_k > t}\; g(\xi_k)\right] \\
&= \sum_k \mathbb{E}\left[\exp\left(-(N-1)\alpha(t + \xi_k)\right) g(\xi_k)\right] \\
&= \exp(-(N-1)\alpha t) \sum_k \mathbb{E}\left[\exp(-(N-1)\alpha\xi_k)\, g(\xi_k)\right] \\
&= \exp(-(N-1)\alpha t) \sum_k \mathbb{E}\left[1_{T_k - \xi_k > 0}\; g(\xi_k)\right] \\
&= \exp(-(N-1)\alpha t) \sum_k \mathbb{E}\left[1_{T_k - \xi_k > 0}\; g(W_1)\right] \\
&= \exp(-(N-1)\alpha t) \cdot \mathbb{E}\left[g(W_1)\right].
\end{aligned}$$

It follows from this calculation that $W_2 - W_1$ and W_1 are independent, $W_2 - W_1 = \exp(\alpha(N-1))$.

The general case may be done similarly. To see how this goes, let us show that $W_3 - W_2 \overset{d}{=} \exp((N-2)\alpha)$ and is independent of W_1 and W_2. To this end, let $T_{jk} := \min\{\xi_l : l \neq j \text{ or } k\}$ for $j \neq k$ in which case $T_{jk} \overset{d}{=} \exp((N-2)\alpha)$ and is independent of $\{\xi_j, \xi_k\}$. We then have

$$\begin{aligned}
\mathbb{E}\left[1_{W_3 - W_2 > t}\; g(W_1, W_2)\right] &= \sum_{j \neq k} \mathbb{E}\left[1_{W_3 - W_2 > t}\; g(W_1, W_2) : \xi_j < \xi_k < T_{jk}\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > t}\; g(\xi_j, \xi_k) : \xi_j < \xi_k < T_{jk}\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > t}\; g(\xi_j, \xi_k) : \xi_j < \xi_k\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[\exp\left(-(N-2)\alpha(t + \xi_k)\right) g(\xi_j, \xi_k) : \xi_j < \xi_k\right] \\
&= \exp(-(N-2)\alpha t) \sum_{j \neq k} \mathbb{E}\left[\exp(-(N-2)\alpha\xi_k)\, g(\xi_j, \xi_k) : \xi_j <\right. \\
&= \exp(-(N-2)\alpha t) \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > 0}\; g(\xi_j, \xi_k) : \xi_j < \xi_k\right] \\
&= \exp(-(N-2)\alpha t) \sum_{j \neq k} \mathbb{E}\left[g(W_1, W_2) : \xi_j < \xi_k < T_{jk}\right] \\
&= \exp(-(N-2)\alpha t) \cdot \mathbb{E}\left[g(W_1, W_2)\right].
\end{aligned}$$

This again shows that $W_3 - W_2$ is independent of $\{W_1, W_2\}$ and $W_3 - W_2 \overset{d}{=} \exp((N-2)\alpha)$. We leave the general argument to the reader. ■

8

Long time behavior

In this section, suppose that $\{X(t)\}_{t\geq 0}$ is a continuous time Markov chain with infinitesimal generator, Q, so that

$$P(X(t+h)=j|X(t)=i)=\delta_{ij}+Q_{ij}h+o(h).$$

We further assume that Q completely determines the chain.

Definition 8.1. $\{X(t)\}$ *is irreducible iff the underlying discrete time jump chain,* $\{Y_n\}$, *determined by the Markov matrix,* $\tilde{P}_{ij}:=\frac{Q_{ij}}{q_i}1_{i\neq j}$, *is irreducible, where*

$$q_i:=-Q_{ii}=\sum_{j\neq i}Q_{ij}.$$

Remark 8.2. Using the Sojourn time description of $X(t)$ it is easy to see that $P_{ij}(t)=\left(e^{tQ}\right)_{ij}>0$ for all $t>0$ and $i,j\in S$ if $X(t)$ is irreducible. Moreover, if for all $i,j\in S$, $P_{ij}(t)>0$ for some $t>0$ then, for the chain $\{Y_n\}$, $i\to j$ and hence $X(t)$ is irreducible. In short the following are equivalent:

1. $\{X(t)\}$ is irreducible,
2. or all $i,j\in S$, $P_{ij}(t)>0$ for some $t>0$, and
3. $P_{ij}(t)>0$ for all $t>0$ and $i,j\in S$.

In particular, in continuous time all chains are "aperiodic."

The next theorem gives the basic limiting behavior of irreducible Markov chains. Before stating the theorem we need to introduce a little more notation.

Notation 8.3 *Let* S_1 *be the time of the first jump of* $X(t)$, *and*

$$R_i:=\min\{t\geq S_1:X(t)=i\},$$

is the first time hitting the site i *after the first jump, and set*

$$\pi_i=\frac{1}{q_i\cdot\mathbb{E}_iR_i}\text{ where }q_i:=-Q_{ii}.$$

Theorem 8.4 (Limiting behavior). *Let* $\{X(t)\}$ *be an irreducible Markov chain. Then*

1. *for all initial staring distributions,* $\nu(j):=P(X(0)=j)$ *for all* $j\in S$, *and all* $j\in S$,
$$P_\nu\left(\lim_{T\to\infty}\frac{1}{T}\int_0^T 1_{X(t)=j}dt=\pi_j\right)=1. \tag{8.1}$$
2. $\lim_{t\to\infty}P_{ij}(t)=\pi_j$ *independent of* i.
3. $\pi=(\pi_j)_{j\in S}$ *is stationary, i.e.* $0=\pi Q$, *i.e.*
$$\sum_{i\in S}\pi_iQ_{ij}=0\text{ for all }j\in S,$$
which is equivalent to $\pi P(t)=\pi$ *for all* t *and to* $P_\pi(X(t)=j)=\pi(j)$ *for all* $t>0$ *and* $j\in S$.
4. *If* $\pi_i>0$ *for some* $i\in S$, *then* $\pi_i>0$ *for all* $i\in S$ *and* $\sum_{i\in S}\pi_i=1$.
5. *The* π_i *are all positive iff there exists a solution,* $\nu_i\geq 0$ *to*
$$\sum_{i\in S}\nu_iQ_{ij}=0\text{ for all }j\in S\text{ with }\sum_{i\in S}\nu_i=1.$$
If such a solution exists it is unique and $\nu=\pi$.

Proof. We refer the reader to [5, Theorems 3.8.1.] for the full proof. Let us make a few comments on the proof taking for granted that $\lim_{t\to\infty}P_{ij}(t)=:\pi_j$ exists.

1. Suppose we assume that and that ν is a stationary distribution, i.e. $\nu P(t)=\nu$, then (by dominated convergence theorem),

$$\nu_j=\lim_{t\to\infty}\sum_i\nu_iP_{ij}(t)=\sum_i\lim_{t\to\infty}\nu_iP_{ij}(t)=\left(\sum_i\nu_i\right)\pi_j=\pi_j.$$

Thus $\nu_j=\pi_j$. If $\pi_j=0$ for all j we must conclude there is not stationary distribution.

2. If we are in the finite state setting, the following computation is justified:

$$\begin{aligned}\sum_{j\in S}\pi_jP_{jk}(s)&=\sum_{j\in S}\lim_{t\to\infty}P_{ij}(t)P_{jk}(s)=\lim_{t\to\infty}\sum_{j\in S}P_{ij}(t)P_{jk}(s)\\&=\lim_{t\to\infty}P_{ik}(t+s)=\pi_k.\end{aligned}$$

This show that $\pi P(s) = \pi$ for all s and differentiating this equation at $s = 0$ then shows, $\pi Q = 0$.

3. Let us now explain why

$$\frac{1}{T}\int_0^T 1_{X(t)=j}dt \to \frac{1}{q_j \cdot \mathbb{E}_j R_j}.$$

The idea is that, because the chain is irreducible, no matter how we start the chain we will eventually hit the site j. Once we hit j, the (strong) Markov property implies the chain forgets how it got there and behaves as if it started at j. Since what happens for the initial time interval of hitting j in computing the average time spent at j, namely $\lim_{T\to\infty}\frac{1}{T}\int_0^T 1_{X(t)=j}dt$, we may as well have started our chain at j in the first place.

Now consider one typical cycle in the chain staring at j jumping away at time S_1 and then returning to j at time R_j. The average first jump time is $\mathbb{E}S_1 = 1/q_j$ while the average length of such as cycle is $\mathbb{E}R_j$. As the chain repeats this procedure over and over again with the same statistics, we expect (by a law of large numbers) that the average time spent at site j is given by

$$\frac{\mathbb{E}S_1}{\mathbb{E}R_j} = \frac{1/q_j}{\mathbb{E}_j R_j} = \frac{1}{q_j \cdot \mathbb{E}_j R_j}.$$

■

8.1 Birth and Death Processes

We have already discussed the basics of the Birth and death processes. To have the existence of the process requires some restrictions on the Birth and Death parameters which are discussed on p. 359 of the book. In general, we are not able to find solve for the transition semi-group, e^{tQ}, in this case. We will therefore have to ask more limited questions about more limited models. This is what we will consider in the rest of this section. We will also consider some interesting situations which one might model by a Birth and Death process.

Recall that the functions, $\pi_j(t) = P(X(t) = j)$, satisfy the differential equations

$$\begin{aligned}
\dot\pi_0(t) &= -\lambda_0\pi_0(t) + \mu_1\pi_1(t)\\
\dot\pi_1(t) &= \lambda_0\pi_0(t) - (\lambda_1+\mu_1)\pi_1(t) + \mu_2\pi_2(t)\\
\dot\pi_2(t) &= \lambda_1\pi_1(t) - (\lambda_2+\mu_2)\pi_2(t) + \mu_3\pi_3(t)\\
&\vdots\\
\dot\pi_n(t) &= \lambda_{n-1}\pi_{n-1}(t) - (\lambda_n+\mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t).\\
&\vdots
\end{aligned}$$

Hence if are going to look for a stationary distribution, we must set $\dot\pi_j(t) = 0$ for all t and solve the system of algebraic equations:

$$\begin{aligned}
0 &= -\lambda_0\pi_0 + \mu_1\pi_1\\
0 &= \lambda_0\pi_0 - (\lambda_1+\mu_1)\pi_1 + \mu_2\pi_2\\
0 &= \lambda_1\pi_1 - (\lambda_2+\mu_2)\pi_2 + \mu_3\pi_3\\
&\vdots\\
0 &= \lambda_{n-1}\pi_{n-1} - (\lambda_n+\mu_n)\pi_n + \mu_{n+1}\pi_{n+1}.\\
&\vdots
\end{aligned}$$

We solve these equations in order to find,

$$\begin{aligned}
\pi_1 &= \frac{\lambda_0}{\mu_1}\pi_0,\\
\pi_2 &= \frac{\lambda_1+\mu_1}{\mu_2}\pi_1 - \frac{\lambda_0}{\mu_2}\pi_0 = \frac{\lambda_1+\mu_1}{\mu_2}\frac{\lambda_0}{\mu_1}\pi_0 - \frac{\lambda_0}{\mu_2}\pi_0\\
&= \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0\\
\pi_3 &= \frac{\lambda_2+\mu_2}{\mu_3}\pi_2 - \frac{\lambda_1}{\mu_3}\pi_1 = \frac{\lambda_2+\mu_2}{\mu_3}\frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0 - \frac{\lambda_1}{\mu_3}\frac{\lambda_0}{\mu_1}\pi_0\\
&= \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\pi_0\\
&\vdots\\
\pi_n &= \frac{\lambda_0\lambda_1\lambda_2\ldots\lambda_{n-1}}{\mu_1\mu_2\mu_3\ldots\mu_n}\pi_0.
\end{aligned}$$

This leads to the following proposition.

Proposition 8.5. *Let* $\theta_n := \frac{\lambda_0\lambda_1\lambda_2\ldots\lambda_{n-1}}{\mu_1\mu_2\mu_3\ldots\mu_n}$ *for* $n = 1, 2, \ldots$ *and* $\theta_0 := 1$. *Then the birth and death process,* $\{X(t)\}$ *with birth rates* $\{\lambda_j\}_{j=0}^\infty$ *and death rates* $\{\mu_j\}_{j=1}^\infty$ *has a stationary distribution,* π, *iff* $\Theta := \sum_{n=0}^\infty \theta_n < \infty$ *in which case,*

$$\pi_n = \frac{\theta_n}{\Theta} \text{ for all } n.$$

Lemma 8.6 (Detail balance). *In general, if we can find a distribution,* π, *satisfying the* ***detail balance equation,***

$$\pi_i Q_{ij} = \pi_j Q_{ji} \text{ for all } i \neq j, \tag{8.2}$$

then π *is a stationary distribution, i.e.* $\pi Q = 0$.

Proof. First proof. Intuitively, Eq. (8.2) states that sites i and j are always exchanging sand back and forth at equal rates. Hence if all sites are doing this the size of the piles of sand at each site must remain unchanged.

Second Proof. Summing Eq. (8.2) on i making use of the fact that $\sum_i Q_{ji} = 0$ for all j implies, $\sum_i \pi_i Q_{ij} = 0$. ■

We could have used this result on our birth death processes to find the stationary distribution as well. Indeed, looking at the rate diagram,

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3 \dots \underset{\mu_{n-1}}{\overset{\lambda_{n-2}}{\rightleftarrows}} (n-1) \underset{\mu_n}{\overset{\lambda_{n-1}}{\rightleftarrows}} n \underset{\mu_{n+1}}{\overset{\lambda_n}{\rightleftarrows}} (n+1),$$

we see the conditions for detail balance between n and $n=1$ are,

$$\pi_n \lambda_n = \pi_{n+1} \mu_{n+1}$$

which implies $\frac{\pi_{n+1}}{\pi_n} = \frac{\lambda_n}{\mu_{n+1}}$. Therefore it follows that,

$$\begin{aligned}
\frac{\pi_1}{\pi_0} &= \frac{\lambda_0}{\mu_1}, \\
\frac{\pi_2}{\pi_0} &= \frac{\pi_2}{\pi_1}\frac{\pi_1}{\pi_0} = \frac{\lambda_1}{\mu_2}\frac{\lambda_0}{\mu_1}, \\
&\vdots \\
\frac{\pi_n}{\pi_0} &= \frac{\pi_n}{\pi_{n-1}}\frac{\pi_{n-1}}{\pi_{n-2}} \dots \frac{\pi_1}{\pi_0} = \frac{\lambda_{n-1}}{\mu_n} \dots \frac{\lambda_1}{\mu_2}\frac{\lambda_0}{\mu_1} \\
&= \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} = \theta_n
\end{aligned}$$

as before.

Lemma 8.7. *For* $|x| < 1$ *and* $\alpha \in \mathbb{R}$ *we have,*

$$(1-x)^{-\alpha} = \sum_{k=0}^{\infty} \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{k!} x^k, \tag{8.3}$$

where $\frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{k!} := 1$ *when* $k = 0$.

Proof. This is a consequence of Taylor's theorem with integral remainder. The main point is to observe that

$$\begin{aligned}
\frac{d}{dx}(1-x)^{-\alpha} &= \alpha(1-x)^{-(\alpha+1)} \\
\left(\frac{d}{dx}\right)^2 (1-x)^{-\alpha} &= \alpha(\alpha+1)(1-x)^{-(\alpha+2)} \\
&\vdots \\
\left(\frac{d}{dx}\right)^k (1-x)^{-\alpha} &= \alpha(\alpha+1)\dots(\alpha+k-1)(1-x)^{-(\alpha+k)} \\
&\vdots
\end{aligned}$$

and hence,

$$\left(\frac{d}{dx}\right)^k (1-x)^{-\alpha}|_{x=0} = \alpha(\alpha+1)\dots(\alpha+k-1). \tag{8.4}$$

Therefore by Taylor's theorem,

$$(1-x)^{-\alpha} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{d}{dx}\right)^k (1-x)^{-\alpha}|_{x=0} \cdot x^k$$

which combined with Eq. (8.4) gives Eq. (8.3). ■

Example 8.8 (Exercise 4.5 on p. 377). Suppose that $\lambda_n = \theta < 1$ and $\mu_n = \frac{n}{n+1}$. In this case,

$$\theta_n = \frac{\theta^n}{\frac{1}{2}\frac{2}{3}\dots\frac{n}{n+1}} = (n+1)\theta^n$$

and we must have,

$$\pi_n = \frac{(n+1)\theta^n}{\sum_{n=0}^{\infty}(n+1)\theta^n}.$$

We can simplify this answer a bit by noticing that

$$\sum_{n=0}^{\infty}(n+1)\theta^n = \frac{d}{d\theta}\sum_{n=0}^{\infty}\theta^{n+1} = \frac{d}{d\theta}\frac{\theta}{1-\theta} = \frac{(1-\theta)+\theta}{(1-\theta)^2} = \frac{1}{(1-\theta)^2}.$$

(Alternatively, apply Lemma 8.7 with $\alpha = 2$ and $x = \theta$.)Thus we have,

$$\pi_n = (1-\theta)^2 (n+1)\theta^n.$$

Example 8.9 (Exercise 4.4 on p. 377). Two machines operate with failure rate μ and there is a repair facility which can repair one machine at a time with rate λ. Let $X(t)$ be the number of operational machines at time t. The state space is thus, $\{0, 1, 2\}$ with the transition diagram,

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2$$

where $\lambda_0 = \lambda$, $\lambda_1 = \lambda$, $\mu_2 = 2\mu$ and $\mu_1 = \mu$. Thus we find,

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0 = \frac{\lambda}{\mu}\pi_0$$
$$\pi_2 = \frac{\lambda^2}{2\mu^2}\pi_0 = \frac{1}{2}\frac{\lambda^2}{\mu^2}\pi_0.$$

so that

$$1 = \pi_0 + \pi_1 + \pi_2 = \left(1 + \frac{\lambda}{\mu} + \frac{1}{2}\frac{\lambda^2}{\mu^2}\right)\pi_0.$$

So the long run probability that all machines are broken is given by

$$\pi_0 = \left(1 + \frac{\lambda}{\mu} + \frac{1}{2}\frac{\lambda^2}{\mu^2}\right)^{-1}.$$

If we now suppose that only one machine can be in operation at a time (perhaps there is only one plug), the new rates become, $\lambda_0 = \lambda$, $\lambda_1 = \lambda$, $\mu_2 = \mu$ and $\mu_1 = \mu$ and working as above we have:

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0 = \frac{\lambda}{\mu}\pi_0$$
$$\pi_2 = \frac{\lambda^2}{\mu^2}\pi_0 = \frac{\lambda^2}{\mu^2}\pi_0.$$

so that

$$1 = \pi_0 + \pi_1 + \pi_2 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2}\right)\pi_0.$$

So the long run probability that all machines are broken is given by

$$\pi_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2}\right)^{-1}.$$

Example 8.10 (Problem VI.4.7, p. 379). A system consists of 3 machines and 2 repairmen. At most 2 machines can operate at any time. The amount of time that an operating machine works before breaking down is exponentially distributed with mean 5 hours. The amount of time that it takes a single repairman to fix a machine is exponentially distributed with mean 4 hours. Only one repairman can work on a failed machine at any given time. Let $X(t)$ be the number of machines in operating condition at time t.

a) Calculate the long run probability distribution of $X(t)$.

b) If an operating machine produces 100 units of output per hour, what is the long run output per hour from the factory.

Solution to Exercise (Problem VI.4.7, p. 379). The state space of operating machines is $S = \{0, 1, 2, 3\}$ and the system is modeled by a birth death process with rate diagram,

$$0 \underset{1/5}{\overset{2/4}{\rightleftarrows}} 1 \underset{2/5}{\overset{2/4}{\rightleftarrows}} 2 \underset{2/5}{\overset{1/4}{\rightleftarrows}} 3.$$

a) We then have $\theta_0 = 1$,

$$\theta_1 = \frac{1/2}{1/5} = \frac{5}{2}$$
$$\theta_2 = \frac{1/2}{1/5}\frac{1/2}{2/5} = \frac{5^2}{2^3}$$
$$\theta_3 = \frac{1/2}{1/5}\frac{1/2}{2/5}\frac{1/4}{2/5} = \frac{5^3}{2^4}\frac{1}{4} = \frac{5^3}{2^6}$$

and

$$\Theta = \sum_{j=0}^{3}\theta_i = 1 + \frac{5}{2} + \frac{5^2}{2^3} + \frac{5^3}{2^6} = \frac{549}{64}.$$

: $\frac{549}{64}$Therefore $\pi_i = \theta_i/\Theta$ gives,

$$\begin{aligned}(\pi_0, \pi_1, \pi_2, \pi_3) &= \frac{64}{549}\left(1, \frac{5}{2}, \frac{5^2}{2^3}, \frac{5^3}{2^6}\right)\\ &= \left(\tfrac{64}{549}\ \tfrac{160}{549}\ \tfrac{200}{549}\ \tfrac{125}{549}\right)\\ &= \left(0.116\,58\ 0.291\,44\ 0.364\,30\ 0.227\,69\right).\end{aligned}$$

b) If the operating machines can produce 100 units per hour, the long run output per hour is,

$$100{\cdot}\pi_1 + 200(\pi_2 + \pi_3) = 100{\cdot}0.291\,44 + 200(0.364\,30 + 0.227\,69) \cong 147.54\ /\text{hour}.$$

Solution to Exercise (Problem VI.4.7, p. 379 but only one repair person.). Here is the same problems with only one repair person. The state space of operating machines is $S = \{0, 1, 2, 3\}$ and the system is modeled by a birth death process with rate diagram,

$$0 \underset{\mu_1}{\overset{\lambda_0}{\rightleftarrows}} 1 \underset{\mu_2}{\overset{\lambda_1}{\rightleftarrows}} 2 \underset{\mu_3}{\overset{\lambda_2}{\rightleftarrows}} 3$$

where, $\lambda_0 = \lambda_1 = \lambda_2 = 1/4$ and $\mu_1 = 1/5$, $\mu_2 = \mu_3 = 2/5$, so the rate diagram is,

$$0 \underset{1/5}{\overset{1/4}{\rightleftarrows}} 1 \underset{2/5}{\overset{1/4}{\rightleftarrows}} 2 \underset{2/5}{\overset{1/4}{\rightleftarrows}} 3.$$

a) We then have $\theta_0 = 1$,

$$\begin{aligned} \theta_1 &= \frac{1/4}{1/5} = \frac{5}{4} \\ \theta_2 &= \frac{1/4}{1/5}\frac{1/4}{2/5} = \frac{5^2}{2}\frac{1}{4^2} \\ \theta_3 &= \frac{1/4}{1/5}\frac{1/4}{2/5}\frac{1/4}{2/5} = \frac{5^3}{2^2}\frac{1}{4^3} \end{aligned}$$

and

$$\Theta = \sum_{j=0}^{3} \theta_i = 1 + \frac{5}{4} + \frac{5^2}{2}\frac{1}{4^2} + \frac{5^3}{2^2}\frac{1}{4^3} = \frac{901}{256}.$$

Therefore $\pi_i = \theta_i / \Theta$ gives,

$$\begin{aligned} (\pi_0, \pi_1, \pi_2, \pi_3) &= \frac{256}{901}\left(1, \frac{5}{4}, \frac{5^2}{2}\frac{1}{4^2}, \frac{5^3}{2^2}\frac{1}{4^3}\right) \\ &= \left(\tfrac{256}{901}\ \tfrac{320}{901}\ \tfrac{200}{901}\ \tfrac{125}{901}\right) \\ &= \left(0.284\ 0.355\ 0.222\ 0.139\right). \end{aligned}$$

b) If the operating machines can produce 100 units per hour, the long run output per hour is,

$$100 \cdot \pi_1 + 200\left(\pi_2 + \pi_3\right) = 100 \cdot 0.355 + 200\left(0.222 + 0.139\right) \cong 108 \text{ /hour.}$$

Example 8.11 (Telephone Exchange). Consider as telephone exchange consisting of K out going lines. The mean call time is $1/\mu$ and new call requests arrive at the exchange at rate λ. If all lines are occupied, the call is lost. Let $X(t)$ be the number of outgoing lines which are in service at time t – see Figure 8.1. We model this as a birth death process with state space, $\{0, 1, 2, \ldots, K\}$ and birth parameters, $\lambda_k = \lambda$ for $k = 0, 1, 2, \ldots, K-1$ and death rates, $\mu_k = k\mu$ for $k = 1, 2, \ldots, K$, see Figure 8.2. In this case,

$$\theta = 1,\ \theta_1 = \frac{\lambda}{\mu},\ \theta_2 = \frac{\lambda^2}{2\mu^2},\ \theta_3 = \frac{\lambda^3}{3! \cdot \mu^3}, \ldots, \theta_K = \frac{\lambda^K}{K!\mu^K}$$

so that

$$\Theta := \sum_{k=0}^{K} \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \cong e^{\lambda/\mu} \text{ for large } K.$$

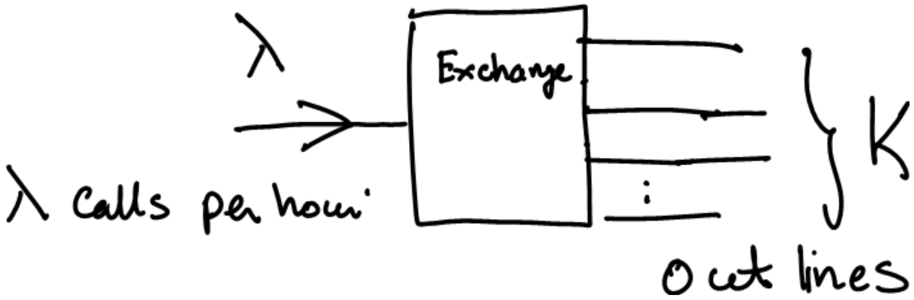


Fig. 8.1. Schematic of a telephone exchange.

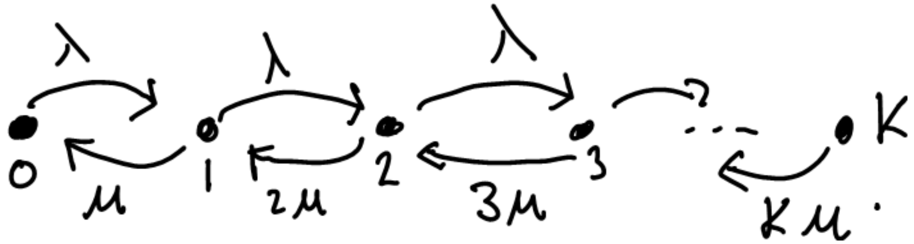


Fig. 8.2. Rate diagram for the telephone exchange.

and hence

$$\pi_k = \Theta^{-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \cong \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k e^{-\lambda/\mu}.$$

For example, suppose $\lambda = 100$ calls / hour and average duration of a connected call is 1/4 of an hour, i.e. $\mu = 4$. Then we have

$$\pi_{25} = \frac{\frac{1}{25!}\left(25\right)^{25}}{\sum_{k=0}^{25}\frac{1}{k!}\left(25\right)^k} \cong 0.144.$$

so the exchange is busy 14.4% of the time. On the other hand if there are 30 or even 35 lines, then we have,

$$\pi_{30} = \frac{\frac{1}{30!}\left(25\right)^{30}}{\sum_{k=0}^{30}\frac{1}{k!}\left(25\right)^k} \cong 0.053$$

and

$$\pi_{35} = \frac{\frac{1}{35!}\left(25\right)^{35}}{\sum_{k=0}^{35}\frac{1}{k!}\left(25\right)^k} \cong .012$$

and hence the exchange is busy 5.3% and 1.2% respectively.

8.1.1 Linear birth and death process with immigration

Suppose now that $\lambda_n = n\lambda + a$ and $\mu_n = n\mu$ for some $\lambda, \mu > 0$ where λ and μ represent the birth rates and deaths of each individual in the population and a represents the rate of migration into the population. In this case,

$$\begin{aligned}\theta_n &= \frac{a\left(a+\lambda\right)\left(a+2\lambda\right)\dots\left(a+\left(n-1\right)\lambda\right)}{n!\mu^n} \\ &= \left(\frac{\lambda}{\mu}\right)^n \frac{\frac{a}{\lambda}\left(\frac{a}{\lambda}+1\right)\left(\frac{a}{\lambda}+2\right)\dots\left(\frac{a}{\lambda}+\left(n-1\right)\right)}{n!}.\end{aligned}$$

Using Lemma 8.7 with $\alpha = a/\lambda$ and $x = \lambda/\mu$ which we need to assume is less than 1, we find

$$\varTheta := \sum_{n=0}^{\infty} \theta_n = \left(1 - \frac{\lambda}{\mu}\right)^{-a/\lambda}$$

and therefore,

$$\pi_n = \left(1 - \frac{\lambda}{\mu}\right)^{a/\lambda} \frac{\frac{a}{\lambda}\left(\frac{a}{\lambda}+1\right)\left(\frac{a}{\lambda}+2\right)\dots\left(\frac{a}{\lambda}+\left(n-1\right)\right)}{n!}\left(\frac{\lambda}{\mu}\right)^n$$

In this case there is an invariant distribution iff $\lambda < \mu$ and $a > 0$. Notice that if $a = 0$, then 0 is an absorbing state so when $\lambda < \mu$, the process actually dies out.

Now that we have found the stationary distribution in this case, let us try to compute the expected population of this model at time t.

Theorem 8.12. *If*

$$M\left(t\right) := \mathbb{E}\left[X\left(t\right)\right] = \sum_{n=1}^{\infty} nP\left(X\left(t\right) = n\right) = \sum_{n=1}^{\infty} n\pi_n\left(t\right)$$

be the expected population size for our linear birth and death process with immigration, then

$$M\left(t\right) = \frac{a}{\lambda - \mu}\left(e^{t(\lambda-\mu)} - 1\right) + M\left(0\right)e^{t(\lambda-\mu)}$$

which when $\lambda = \mu$ should be interpreted as

$$M\left(t\right) = at + M\left(0\right).$$

Proof. In this proof we take for granted the fact that it is permissible to interchange the time derivative with the infinite sum. Assuming this fact we find,

$$\begin{aligned}\dot{M}\left(t\right) &= \sum_{n=1}^{\infty} n\dot{\pi}_n\left(t\right) \\ &= \sum_{n=1}^{\infty} n\begin{pmatrix} \left(a+\lambda\left(n-1\right)\right)\pi_{n-1}\left(t\right) \\ -\left(a+\lambda n+\mu n\right)\pi_n\left(t\right)+\mu\left(n+1\right)\pi_{n+1}\left(t\right)\end{pmatrix} \\ &= \sum_{n=1}^{\infty} n\left(a+\lambda\left(n-1\right)\right)\pi_{n-1}\left(t\right) \\ &\quad - \sum_{n=1}^{\infty} n\left(a+\lambda n+\mu n\right)\pi_n\left(t\right) + \sum_{n=1}^{\infty}\mu n\left(n+1\right)\pi_{n+1}\left(t\right) \\ &= \sum_{n=0}^{\infty}\left(n+1\right)\left(a+\lambda n\right)\pi_n\left(t\right) \\ &\quad - \sum_{n=1}^{\infty} n\left(a+\lambda n+\mu n\right)\pi_n\left(t\right) + \sum_{n=2}^{\infty}\mu\left(n-1\right)n\pi_n\left(t\right) \\ &= a\pi_0\left(t\right) + \left[2\left(a+\lambda\right)-\left(a+\lambda+\mu\right)\right]\pi_1\left(t\right) \\ &\quad + \sum_{n=2}^{\infty}\left[\left(n+1\right)\left(a+\lambda n\right)+\mu\left(n-1\right)n - n\left(a+\lambda n+\mu n\right)\right]\pi_n\left(t\right) \\ &= a\pi_0\left(t\right) + \left[a+\lambda-\mu\right]\pi_1\left(t\right) + \sum_{n=2}^{\infty}\left[\left(a+\lambda n\right)-\mu n\right]\pi_n\left(t\right) \\ &= a\pi_0\left(t\right) + \sum_{n=1}^{\infty}\left[a+\lambda n-\mu n\right]\pi_n\left(t\right) \\ &= \sum_{n=0}^{\infty}\left[a+\lambda n-\mu n\right]\pi_n\left(t\right) = a + \left(\lambda-\mu\right)M\left(t\right).\end{aligned}$$

Thus we have shown that

$$\dot{M}\left(t\right) = a + \left(\lambda - \mu\right)M\left(t\right) \text{ with } M\left(0\right) = \sum_{n=1}^{\infty} n\pi_n\left(0\right),$$

where $M\left(0\right)$ is the mean size of the initial population. Solving this simple differential equation gives the results. ■

8.2 What you should know for the first midterm

1. Basics of discrete time Markov chain theory.
   a) You should be able to compute $P(X_0 = x_0, \ldots, X_n = x_n)$ given the transition matrix, P, and the initial distribution as in Proposition 3.2.
   b) You should be able to go back and forth between P and its jump diagram.
   c) Use the jump diagram to find all of the communication classes of the chain.
   d) Know how to compute hitting probabilities and expected hitting times using the first step analysis.
   e) Know how to find the invariant distributions of the chain.
   f) Understand how to use hitting probabilities and the invariant distributions of the recurrent classes in order to compute the long time behavior of the chain.
   g) **Mainly** study the examples in Section 3.2 and the related homework problems. Especially see Example 4.33 and Exercises 0.6 – 0.9.
2. Basics of continuous time Markov chain theory:
   a) You should be able to compute $P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2, \ldots, X_{t_n} = i_n)$ given the Markov semi-group, $P(t)$, as in Theorem 5.4.
   b) You should understand the relationship of $P(t)$ to its infinitesimal generator, Q. Namely $P(t) = e^{tQ}$ and

   $$\frac{d}{dt}|_0 P(t) = Q.$$

   For example, if

   $$P(t) = \begin{array}{cccc} 1 & 2 & 3 & \\ \left[\begin{array}{c} \frac{1}{3}e^{-3t} + \frac{2}{3} \\ -e^{-2t} + \frac{1}{3}e^{-3t} + \frac{2}{3} \\ -\frac{2}{3}e^{-3t} + \frac{2}{3} \end{array}\right. & \begin{array}{c} 0 \\ e^{-2t} \\ 0 \end{array} & \left.\begin{array}{c} -\frac{1}{3}e^{-3t} + \frac{1}{3} \\ -\frac{1}{3}e^{-3t} + \frac{1}{3} \\ \frac{2}{3}e^{-3t} + \frac{1}{3} \end{array}\right] & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

   then

   $$Q = \dot{P}(0) = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \\ 2 & 0 & -2 \end{bmatrix}.$$

   Note: you will not be asked to compute $P(t)$ from Q but you **should** be able to find Q from $P(t)$ as in the above example.
   c) You should know how to go between the generator Q and its rate diagram.
   d) You should understand the jump hold description of a continuous time Markov chain explained in Section 6.5. In particular in the example above, if $S_1 = \inf\{t > 0 : X(t) \neq X(0)\}$ is the first jump time of the chain, you should know that, if the chain starts at sites $1, 2$, or 3, then S_1 is exponentially distributed with parameter $q_1 = 1$, $q_2 = 2$, $q_3 = 2$ respectively, i.e.

   $$P(S_1 > t | X(0) = i) = e^{-q_i t},$$

   where $q_i = -Q_{ii}$.
   e) You should also understand that

   $$P(X_{S_1} = j | X(0) = i) = \frac{Q_{ij}}{q_i}$$

   so that in this example, $P(X_{S_1} = 3 | X(0) = 2) = 1/2$ in the above example.
   f) You should understand how to associate a rate diagram to Q, see the example section 6.3.
   g) You should be familiar with the basics of birth and death processes.
      i. Know how to compute the invariant distribution, Proposition 8.5.
      ii. Know the relationship of the invariant distribution to the long time behavior of the chain, Theorem 8.4.
      iii. Understand the basics of the repairman models. In particular see Example 8.9 and homework Problem VI.4 (p. 377 –) P4.1.

Let us look more carefully at Q above and its rate diagram:

$$Q = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \\ 2 & 0 & -2 \end{bmatrix} \Longrightarrow \begin{array}{ccc} 1 & \overset{1}{\underset{2}{\rightleftarrows}} & 3 \\ {}_1\uparrow & \underset{1}{\nearrow} & \\ 2 & & \end{array}.$$

The associated embedded Markov chain jump matrix and its rate diagram is given by

$$\tilde{P} := \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} \Longrightarrow \begin{array}{ccc} 1 & \overset{1}{\underset{1}{\rightleftarrows}} & 3 \\ {}_{1/2}\uparrow & \underset{1/2}{\nearrow} & \\ 2 & & \end{array}.$$

The communication classes are $\{1, 3\}$ and $\{2\}$ with $\{2\}$ being transient and $\{1, 3\}$ being closed and hence recurrent. The equations for the invariant distributions of the Q and $\tilde{P}$ restricted to $\{1, 3\}$ are;

$$\pi_1 = 2\pi_3 \text{ or } (\pi_1, \pi_3) = \frac{1}{3}(2, 1) \text{ and}$$
$$\pi_1 = \pi_3 \text{ or } (\pi_1, \pi_3) = \frac{1}{2}(1, 1)$$

respectively. These are different – can you explain why?

9

Hitting and Expected Return times and Probabilities

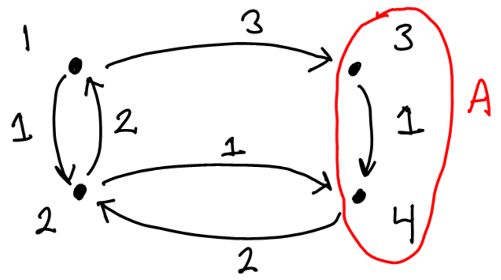


Fig. 9.1. A rate diagram for a four state Markov chain.

Let $\{X(t)\}_{t\ge 0}$ be a continuous time Markov chain described by its infinitesimal generator, $Q = (Q_{ij})_{i,j\in S}$ where S is the state space. Further let

$$S_1 = \inf\{t > 0 : X(t) \neq X(0)\}$$

be the first jump time of the chain and $q_j := -Q_{j\,j}$ for all $j \in S$. Recall $P(S_1 > t|X(0) = j) = e^{-q_j t}$ for all $t > 0$ and $\mathbb{E}[S_1|X(0) = j] = 1/q_j$. Given a subset, A, of the state space, S, let

$$T_A := \inf\{t \ge 0 : X(t) \in A\}$$

be the first time the process, $X(t)$, hits A. By convention, $T_A = \infty$ if $X(t) \notin A$ for all t, i.e. if $X(t)$ does not hit A.

Example 9.1. Let $S = \{1, 2, 3, 4\}$ and $X(t)$ be the continuous time Markov chain determined by the rate diagram, Further let $A = \{3, 4\}$. We would like to compute, $h_i = P_i(X(t) \text{ hits } A)$ for $i = 1, 2$. If $\{Y_n\}_{n=0}^{\infty}$ is the embedded discrete time chain, this is the same as computing, $h_i = P_i(Y_n \text{ hits } A)$ which we know how to do. We now carry out the details. First off the infinitesimal generator, Q, is given by

$$Q = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & -4 & 1 & 3 & 0 \\ 2 & 2 & -3 & 0 & 1 \\ 3 & 1 & 0 & -1 & 0 \\ 4 & 0 & 2 & 0 & -2 \end{array}$$

and hence the Markov matrix for $\{Y_n\}$ is given by,

$$\tilde{P} := \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 1/4 & 3/4 & 0 \\ 2 & 2/3 & 0 & 0 & 1/3 \\ 3 & 1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 & 0 \end{array}.$$

The first step analysis for the hitting probabilities then implies,

$$\begin{aligned} h_1 =& P_1(X(t) \text{ hits } A|X_{S_1} = 3)\,P_1(X_{S_1} = 3) \\ &+ P_1(X(t) \text{ hits } A|X_{S_1} = 2)\,P_1(X_{S_1} = 2) \\ =& \frac{3}{4} + h_2\frac{1}{4} \end{aligned}$$

and

$$\begin{aligned} h_2 =& P_2(X(t) \text{ hits } A|X_{S_1} = 1)\,P_2(X_{S_1} = 1) \\ &+ P_2(X(t) \text{ hits } A|X_{S_1} = 4)\,P_2(X_{S_1} = 4) \\ =& \frac{2}{3}h_1 + \frac{1}{3} \end{aligned}$$

which have solutions, $h_1 = h_2 = 1$ as we know should be the case since this is an irreducible Markov chain.

Example 9.2. Continuing the set up in Example 9.1, we are going to compute $w_i = \mathbb{E}_i T_A$ for $i = 1, 2$. Again by a first step analysis we have,

$$\begin{aligned} w_1 &= \mathbb{E}_1(T_A|X_{S_1} = 3)\,P_1(X_{S_1} = 3) + \mathbb{E}_1(T_A|X_{S_1} = 2)\,P_1(X_{S_1} = 2) \\ &= \frac{1}{4}\frac{3}{4} + \left(\frac{1}{4} + w_2\right)\frac{1}{4} = \frac{1}{4} + \frac{1}{4}w_2 \end{aligned}$$

and

$$\begin{aligned} w_2 &= \mathbb{E}_2(T_A|X_{S_1} = 1)\,P_2(X_{S_1} = 1) + \mathbb{E}_2(T_A|X_{S_1} = 4)\,P_2(X_{S_1} = 4) \\ &= \left(\frac{1}{3} + w_1\right)\frac{2}{3} + \frac{1}{3}\frac{1}{3} = \frac{1}{3} + \frac{2}{3}w_1, \end{aligned}$$

where $\frac{1}{4} = \mathbb{E}_1(S_1)$ and $\frac{1}{3} = \mathbb{E}_2(S_1)$. The solutions to these equations are:

$$\mathbb{E}_1(T_A) = w_1 = \frac{2}{5} \text{ and } \mathbb{E}_2(T_A) = w_2 = \frac{3}{5}.$$

With this example as background, let us now work out the general formula for these hitting times.

Proposition 9.3. *Let Q be the infinitesimal generator of a continuous time Markov chain, $\{X(t)\}_{t\geq 0}$, with state space, S. Suppose that $A \subset S$ and $T_A := \inf\{t \geq 0 : X(t) \in A\}$. If we let $w_i := \mathbb{E}_i T_A$ for all $i \notin A$, then $\{w_i\}_{i\in A^c}$ satisfy the system of linear equations,*

$$w_i = \frac{1}{q_i} + \sum_{j\notin A} \tilde{P}_{ij} w_j = \frac{1}{q_i} + \sum_{j\notin A} \frac{Q_{ij}}{q_i} w_j$$

where as usual, $q_i = -Q_{ii} = \sum_{j\neq i} Q_{ij}$.

Proof. By the first step analysis we have, for $i \notin A$,

$$\begin{aligned} w_i &= \sum_{j\neq i} \mathbb{E}_i\left[T_A | X_{S_1} = j\right] P_i\left(X_{S_1} = j\right) \\ &= \sum_{j\neq i} \tilde{P}_{ij}\ \mathbb{E}_i\left[T_A | X_{S_1} = j\right]. \end{aligned}$$

By the strong Markov property,

$$\mathbb{E}_i\left[T_A | X_{S_1} = j\right] = \mathbb{E}_i S_1 + \mathbb{E}_j T_A = \frac{1}{q_i} + w_j$$

where $w_j := \mathbb{E}_j T_A = 0$ if $j \in A$. Therefore we have,

$$\begin{aligned} w_i &= \sum_{j\neq i} \tilde{P}_{ij}\ \left(\frac{1}{q_i} + w_j\right) = \frac{1}{q_i} + \sum_{j\neq i} \tilde{P}_{ij}\ w_j \\ &= \frac{1}{q_i} + \sum_{j\notin A} \tilde{P}_{ij}\ w_j \end{aligned}$$

as claimed. ■

Notation 9.4 *Now let*

$$R_j := \inf\{t > S_1 : X_t = j\}$$

be the first return time to j.

Our next goal is to find a formula for $\mathbb{E}_i R_j$ for all $i, j \in S$. Before going to the general case, let us work out an example.

Example 9.5. Let us do an example of a two state Markov chain. Say

$$0 \xrightarrow{\alpha} 1 \xrightarrow{\beta} 0.$$

Let $m_0 = \mathbb{E}_0 R_0$ and $m_1 = \mathbb{E}_1 R_0$, then

$$\begin{aligned} m_0 &= \mathbb{E}_0\left[R_0 | X_{S_1} = 1\right] P\left(X_{S_1} = 1\right) = \frac{1}{\alpha} + m_1 \\ m_1 &= \mathbb{E}_0\left[R_0 | X_{S_1} = 0\right] P\left(X_{S_1} = 0\right) = \frac{1}{\beta} \end{aligned}$$

and therefore, $m_0 = \frac{1}{\alpha} + \frac{1}{\beta}$ which is clearly the correct answer in this case. The long run fraction of the time we are in state 0 is therefore

$$\frac{1/\alpha}{m_0} = \frac{\beta}{\alpha + \beta}.$$

This is the same as computing $\lim_{t\to\infty} P(X(t) = 0) = \pi_0$. Indeed for this case,

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

has invariant distribution, $\pi = (\beta, \alpha)/(\alpha + \beta)$. Therefore,

$$\pi_0 = \frac{\beta}{\alpha + \beta} \text{ and } \pi_1 = \frac{\alpha}{\alpha + \beta}. \tag{9.1}$$

as argued above.

Proposition 9.6 (Expected return times). *If $m_{ij} := \mathbb{E}_i R_j$ for all $j \in S$, then*

$$m_{ij} = \frac{1}{q_i} + \sum_{k\neq i \text{ or } j} \frac{Q_{ik}}{q_i} m_{kj}. \tag{9.2}$$

Proof. By a first step analysis we have,

$$\begin{aligned} m_{ij} = \mathbb{E}_i R_j &= \sum_{k\neq i} \mathbb{E}_i\left[R_j | X_{S_1} = k\right] P\left(X_{S_1} = k\right) \\ &= \sum_{k\neq i} \mathbb{E}_i\left[R_j | X_{S_1} = k\right] \frac{Q_{ik}}{q_i}. \end{aligned}$$

Since

$$\begin{aligned}\mathbb{E}_i\left[R_j|X_{S_1}=k\right] &= \begin{cases}\mathbb{E}_iS_1+\mathbb{E}_kR_j & \text{if } k\neq j\\ \mathbb{E}_iS_1 & \text{if } k=j\end{cases}\\ &= \begin{cases}\frac{1}{q_i}+m_{kj} & \text{if } k\neq j\\ \frac{1}{q_i} & \text{if } k=j\end{cases}.\end{aligned}$$

we arrive at the

$$\begin{aligned}m_{ij} &= \sum_{k\neq i}\mathbb{E}_i\left[R_j|X_{S_1}=k\right]\frac{Q_{ik}}{q_i}\\ &= \frac{1}{q_i^2}Q_{ij}+\sum_{k\neq i \text{ and } k\neq j}\left(\frac{1}{q_i}+m_{kj}\right)\frac{Q_{ik}}{q_i}\\ &= \sum_{k\neq i}\frac{Q_{ik}}{q_i^2}+\sum_{k\neq i \text{ and } k\neq j}m_{kj}\frac{Q_{ik}}{q_i}\\ &= \frac{1}{q_i}+\sum_k 1_{k\neq i,k\neq j}m_{kj}\frac{Q_{ik}}{q_i}.\end{aligned}$$

■

Corollary 9.7. *Let* $\{X(t)\}_{t\geq 0}$ *be a finite state irreducible Markov chain with generator,* $Q=(Q_{ij})_{i,j\in S}$. *If* $\pi=(\pi_i)$ *is an invariant distribution, then*

$$\pi_i=\frac{1}{q_im_{ii}}=\frac{1}{q_i\mathbb{E}_i\left[R_i\right]}. \tag{9.3}$$

Proof. Suppose that π_j is an invariant distribution for the chain, so that $\sum_i \pi_iQ_{ik}=0$ or equivalently,

$$\sum_{i\neq k}\pi_iQ_{ik}=-\pi_kQ_{kk}=\pi_kq_k.$$

It follows from Eq. (9.2) that

$$\begin{aligned}\sum_i\pi_iq_im_{ij} &= \sum_i\pi_iq_i\frac{1}{q_i}+\sum_i\pi_iq_i\sum_k 1_{k\neq i,k\neq j}\frac{Q_{ik}}{q_i}m_{kj}\\ &= 1+\sum_{i,k}\pi_i1_{i\neq k,k\neq j}Q_{ik}m_{kj}\\ &= 1+\sum_k 1_{k\neq j}\pi_kq_km_{kj}\\ &= 1+\sum_i 1_{i\neq j}\pi_iq_im_{ij}.\end{aligned}$$

Hence it follows that $\pi_iq_im_{ii}=1$ which proves Eq. (9.3). ■

10

Renewal Processes

Renewal process. Suppose we have a box of identical components, each numbered by $1, 2, 3, \ldots$. Let X_i denote the lifetime of the i^{th} component and assume that $\{X_i\}_{i=1}^{\infty}$ are i.i.d. non-negative random variables with distribution function, F. We assume at very least that $F(0) < 1$, i.e. there is a positive probability that each component is in fact good. At time zero we put the first component into service, when it fails we **immediately** replace it by the second, when the second fails we **immediately** replace it by the third, and so on. Based on this scenario we make the following definition in which the reader should refer to Figure 10.1.

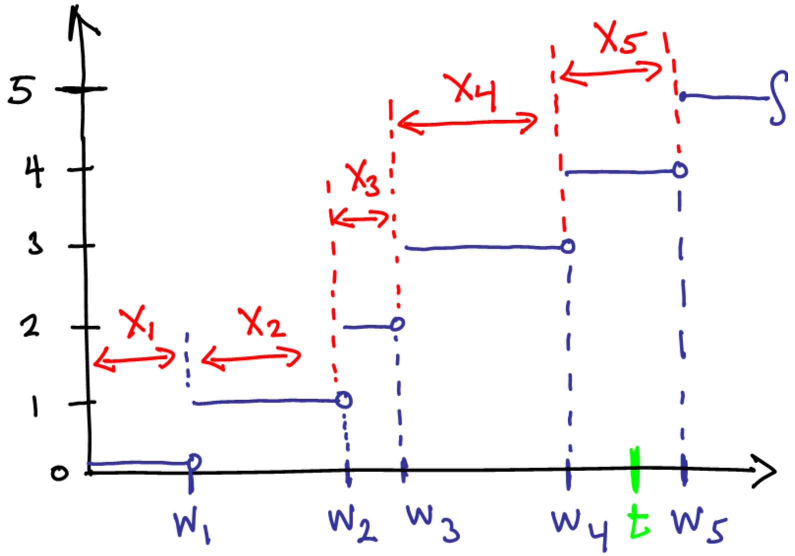


Fig. 10.1. The graph of $N(t)$ given a realization of the $\{X_i\}_{i=1}^{\infty}$. From this picture it should be clear that $\{N(t) \geq k\} = \{W_k \leq t\}$.

10.1 Basic Definitions and Properties

Definition 10.1 (Renewal Process). *Let $\{X_k\}_{k=1}^{\infty}$ be i.i.d. random variables, assume $X_k > 0$ a.s. and $\mu := \mathbb{E}X_1 > 0$ with $\mu = \infty$ being an allowed value. Further let*

$$W_n := X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

be the time of the n^{th} "renewal." The ***renewal process*** *is the counting process defined by*

$$N(t) = \#\{n : W_n \leq t\} = \max\{n : W_n \leq t\}.$$

More generally, for $0 \leq a < b < \infty$, let

$$N((a, b]) = N(b) - N(a) = \#\{n : a < W_n \leq b\}.$$

So $N(t)$ counts the number of **renewals** which have occurred at time t or less and $N((a, b])$ counts the number of renewals in $(a, b]$. The random variable, W_n, is the time of the n^{th} renewal whereas X_n is the time between the $(n-1)^{\text{th}}$ and the n^{th} renewals. Since the **inter-renewal times**, $\{X_n\}_{n=1}^{\infty}$, are i.i.d., the process probabilistically restarts at each renewal.

Example 10.2 (Poisson Process). If X_k is exponential with parameter λ then we know that $N(\cdot)$ is the Poisson Process with parameter λ. In particular we know that $N(t)$ is a Poisson random variable so that

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

and N has independent increments.

Example 10.3 (Markov Chain). Suppose that $\{Y_n\}_{n=0}^{\infty}$ is a recurrent Markov chain on some state space S. Suppose the chain starts at some site, $x \in S$, and let $\{X_k\}_{k=1}^{\infty}$ be the subsequent return times for the chain to x. It follows by the strong Markov property, that $\{X_k\}_{k=1}^{\infty}$ are i.i.d. random variables. In this case $N(t)$ counts the number of returns to x before or equal to time t. This example has a analogue for continuous time Markov chains as well.

Referring to Figure 10.1 we see that the following important relationship holds:

$$\{N(t) \geq k\} = \{W_k \leq t\}. \tag{10.1}$$

Moreover if t is as in Figure 10.1 we see that $N(t) = 4$, $N(t) + 1 = 5$ and $W_{N(t)} = W_4 \leq t < W_5 = W_{N(t)+1}$. In general we always have,

$$W_{N(t)} \leq t < W_{N(t)+1}. \tag{10.2}$$

Notation 10.4 *Referring to Figure 10.2 we introduce the following terminology:*

1. $\gamma_t = W_{N(t)+1} - t =$ *(**residual life** of the part in service at time t).*
2. $\delta_t = t - W_{N(t)} =$ *(**age (or current life)** of the part in service).*
3. $\beta_t = \delta_t + \gamma_t = W_{N(t)+1} - W_{N(t)} =$ ***total life time** of the part in service at time t.*

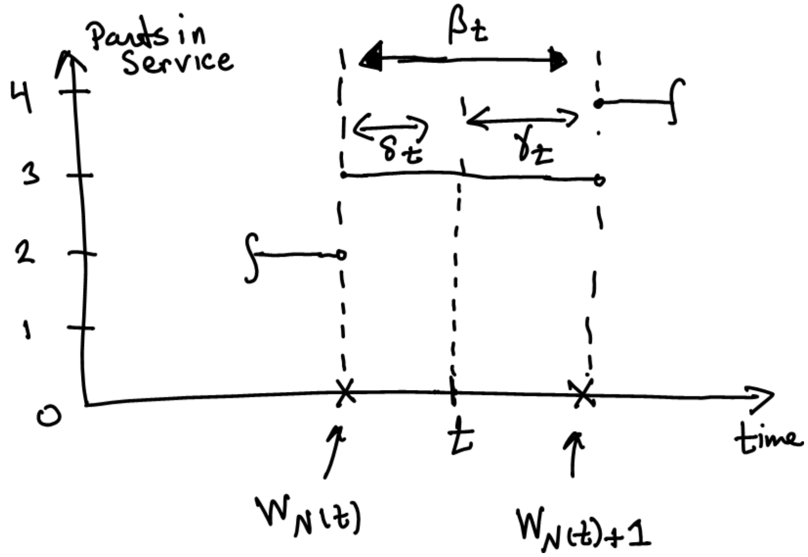


Fig. 10.2. The geometry of a renewal process.

For future reference let us note that:

1. For $y \geq 0$, $\gamma_t > y$ iff there are no renewals in $(t, t+y]$ and hence

$$\{\gamma_t > y\} = \{N((t, t+y]) = 0\} = \{N(t+y) = N(t)\}.$$

2. For $t \geq x$, $\delta_t \geq x$ iff there were no renewals in $(t-x, t]$, i.e.

$$\{\delta_t \geq x\} = \{N((t-x, t]) = 0\} = \{N(t) = N(t-x)\}.$$

Example 10.5 (Poisson Process). Suppose that X_k is exponential with parameter λ so that $N(t)$ is the Poisson Process with parameter λ. Then for $x, y \geq 0$ we have,

$$\begin{aligned} P(\delta_t \geq x, \gamma_t > y) &= P(N((t-x, t]) = 0, N((t, t+y]) = 0) \\ &= P(N((t-x, t]) = 0) \cdot P(N((t, t+y]) = 0) \\ &= 1_{x \leq t} e^{-\lambda x} \cdot e^{-\lambda y}. \end{aligned}$$

This shows γ_t and δ_t are independent, $\gamma_t \overset{d}{=} \exp(\lambda)$ and δ_t is a truncated exponential,

$$P(\delta_t < x) = 1 - 1_{x \leq t} e^{-\lambda x} = \begin{cases} 1 & \text{if } x > t \\ 1 - e^{-\lambda x} & \text{if } x \leq t \end{cases}$$

which is equivalent to

$$P(\delta_t \leq x) = \begin{cases} 1 & \text{if } x \geq t \\ 1 - e^{-\lambda x} & \text{if } x < t \end{cases},$$

see Figure 10.3.

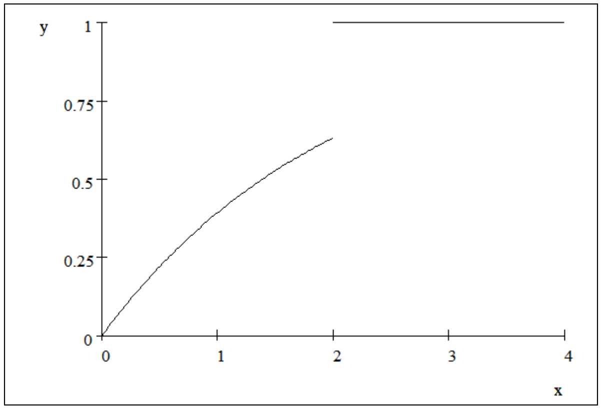


Fig. 10.3. The plot of the distribution function of a truncated exponential with $t = 2$.

Let us observe that $\mathbb{E}\gamma_t = \frac{1}{\lambda}$ while

$$\mathbb{E}\delta_t = \int_0^\infty P(\delta_t \ge x)\,dx = \int_0^\infty 1_{x\le t} e^{-\lambda x} dx$$
$$= \int_0^t e^{-\lambda x} dx = \frac{1}{\lambda}\left(1 - e^{-\lambda t}\right).$$

Therefore,

$$\mathbb{E}\beta_t = \mathbb{E}\gamma_t + \mathbb{E}\delta_t = \frac{1}{\lambda} + \frac{1}{\lambda}\left(1 - e^{-\lambda t}\right).$$

Compare this with $\mathbb{E}X_j = 1/\lambda$. Notice that the lifetime of the part at service at time t is in the limit as $t \to \infty$, twice as long as the mean lifetime of any individual part. Why is this. The point is that t is fixed and hence the we are more likely to see life intervals of a part which encompass t rather than the shorter ones. This also represents the memoryless property of the exponential.

Definition 10.6 (Renewal function). *We call, $M(t) := \mathbb{E}N(t)$, the* ***renewal function.***

Example 10.7. If $X_j \overset{d}{=} \exp(\lambda)$, i.e. $P(X_j > t) = e^{-\lambda t}$ for all $t \ge 0$, then $M(t) = \mathbb{E}N(t) = \lambda t$.

For later purposes it is useful to observe that $M(t)$ may be computed in terms of the $\{W_n\}$ by,

$$M(t) = \mathbb{E}N(t) = \mathbb{E}\left(\sum_{n=1}^\infty 1_{n\le N(t)}\right) = \mathbb{E}\left(\sum_{n=1}^\infty 1_{W_n\le t}\right) = \sum_{n=1}^\infty P(W_n \le t). \tag{10.3}$$

Remark 10.8 (Convolution). Recall that if X and Y are two independent random variables, $F_X(t) = P(X \le t)$, $F_Y(t) = P(Y \le t)$ and $F_{X+Y}(t) = P(X + Y \le t)$, then

$$F_{X+Y}(t) = \int_{-\infty}^\infty P(X + Y \le t | X = x)\,dF_X(x)$$
$$= \int_{-\infty}^\infty P(x + Y \le t | X = x)\,dF_X(x)$$
$$= \int_{-\infty}^\infty P(Y \le t - x)\,dF_X(x)$$
$$= \int_{-\infty}^\infty F_Y(t - x)\,dF_X(x).$$

When X and Y are non-negative, then $F_X(t) = F_Y(t) = F_{X+Y}(t) = 0$ if $t \le 0$ and for $t \ge 0$ we have,

$$F_{X+Y}(t) = \int_0^t F_Y(t - x)\,dF_X(x). \tag{10.4}$$

Notation 10.9 (Convolution Notation) *If F is a (generalized) distribution function and g is a function, let*

$$g * F(t) := \int_0^t g(t - x)\,dF(x)$$

and if f is a density, let

$$g * f(t) := \int_0^t g(t - x) f(x)\,dx.$$

With this notation we may write Eq. (10.4) more succinctly as $F_{X+Y} = F_X * F_Y$.

Definition 10.10. *Let $F_n(t) := F_{W_n}(t) = P(W_n \le t)$ with $F(t) := F_1(t) = P(X_1 \le t)$. We may also write this as*

$$F_n = \overbrace{F * F * \dots * F}^{n\ -\ times}. \tag{10.5}$$

According to Eq. (10.4), we have,

$$F_{n+1}(t) = P(W_n + X_{n+1} \le t)$$
$$= \int_0^t F_n(t - x)\,dF(x) = \int_0^t F(t - x)\,dF_n(x). \tag{10.6}$$

With this notation, it follows from Eq. (10.1) that

$$P(N(t) = k) = P(N(t) \ge k) - P(N(t) \ge k + 1)$$
$$= P(W_k \le t) - P(W_{k+1} \le t) = F_k(t) - F_{k+1}(t).$$

It now follows from Eq. (10.3) that

$$M(t) = \sum_{n=1}^\infty P(W_n \le t) = \sum_{n=1}^\infty F_n(t) = \sum_{n=1}^\infty F^{*n}. \tag{10.7}$$

Example 10.11. When $N(t)$ is a Poisson as in Example 10.7, we know that $M(t) = \lambda t$. As a check, let us compute the right side of Eq. (10.7) and verify that it gives λt in this case. From Lemma 2.7, we know that

$$F_n(t) = P(W_n \le t) = e^{-\lambda t} \sum_{j=n}^\infty \frac{(\lambda t)^j}{j!}.$$

Therefore,

$$\begin{aligned} M(t) &= e^{-\lambda t}\sum_{n=1}^{\infty}\sum_{j=n}^{\infty}\frac{(\lambda t)^j}{j!} = e^{-\lambda t}\sum_{1\le n\le j<\infty}\frac{(\lambda t)^j}{j!} \\ &= e^{-\lambda t}\sum_{1\le j<\infty}\frac{(\lambda t)^j}{j!}j = \lambda t\cdot e^{-\lambda t}\sum_{1\le j<\infty}\frac{(\lambda t)^{j-1}}{(j-1)!} = \lambda t. \end{aligned}$$

Proposition 10.12 (Renewal Equation). *The renewal function satisfies the **renewal equation**,*

$$M(t) = F(t) + \int_0^t M(t-x)\,dF(x). \tag{10.8}$$

or written more succinctly,

$$M = F + M * F. \tag{10.9}$$

Proof. First proof. Recall that

$$M = \sum_{n=1}^{\infty} F_n = \sum_{n=1}^{\infty} F^{*n}$$

and therefore,

$$M * F = \sum_{n=1}^{\infty} F^{*n} * F = \sum_{n=1}^{\infty} F^{*(n+1)} = M - F,$$

which is Eq. (10.9). Written out in more detail, using the definition of M and F_n, we have

$$\int_0^t M(t-x)\,dF(x) = \sum_{n=1}^{\infty}\int_0^t F_n(t-x)\,dF(x) = \sum_{n=1}^{\infty} F_{n+1}(t) = M(t) - F(t),$$

which is Eq. (10.8).

Second proof based on conditioning on X_1. We start with a **informal proof** using,

$$M(t) = \int_0^{\infty}\mathbb{E}[N(t) : X_1 = x]\,dF(x)$$

and

$$\mathbb{E}[N(t)|X_1 = x] = \begin{cases} 0 & \text{if } t < x \\ \mathbb{E}[1 + N(t-x)] & \text{if } x \le t \end{cases}. \tag{10.10}$$

Therefore,

$$M(t) = \int_0^t \mathbb{E}[N(t) : X_1 = x]\,dF(x) = \int_0^t (1 + M(t-x))\,dF(x)$$

which is Eq. (10.8). The point in Eq. (10.10) is that if $t < X_1$ then $N(t) = 0$ while if $t \ge X_1$, then

$$N(t) = 1 + N(t - X_1 : X_2, X_3, \dots)$$

where we write

$$N(t : y_1, y_2, \dots) := \#\{n : y_1 + \dots + y_n \le t\},$$

see Figure 10.4.

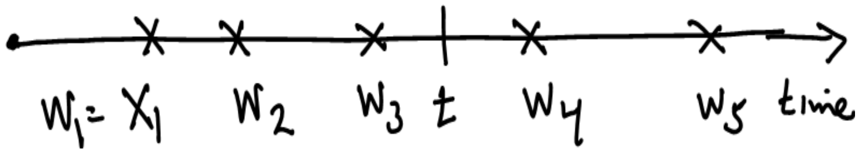


Fig. 10.4. In this example, $N(t) = 3$ while $N(t - X_1 : X_2, X_3, \dots) = 2$.

(Rigorous version of second proof.) We have to compute $\mathbb{E}[N(t)|X_1 = x]$ with more care. To this end, suppose that $g(x)$ is a given bounded function. Then

$$\begin{aligned} \mathbb{E}[N(t)g(X_1)] &= \mathbb{E}[N(t)g(X_1)1_{t<X_1}] + \mathbb{E}[N(t)g(X_1)1_{X_1\le t}] \\ &= \mathbb{E}[(1 + N(t - X_1 : X_2, X_3, \dots))g(X_1)1_{X_1\le t}] \\ &= \mathbb{E}[(1 + \mathbb{E}N(t-x)|_{x=X_1})1_{X_1\le t}g(X_1)] \end{aligned} \tag{10.11}$$

where in; 1) the first equality we have used $N(t) = 0$ if $t < X_1$ and

$$N(t : X_1, X_2, \dots) = 1 + N(t - X_1 : X_2, X_3, \dots) \text{ for } X_1 \le t,$$

and 2) the second equality we used the independence X_1 from $(X_2, X_3, \dots)$. It follows from Eq. (10.11) that

$$\mathbb{E}[N(t)|X_1] = (1 + M(t - X_1))1_{X_1\le t}$$

and therefore,

$$\begin{aligned} \mathbb{E}[N(t)] &= \mathbb{E}(\mathbb{E}[N(t)|X_1]) \\ &= \mathbb{E}[(1 + M(t - X_1))1_{X_1\le t}] \\ &= \int_{[0,t]}[1 + M(t-x)]\,dF(x) \\ &= F(t) + \int_{[0,t]} M(t-x)\,dF(x). \end{aligned}$$

■

Remark 10.13. We will often assume that the distribution of X_j is continuous so that $\dot{F}(t) = f(t)$ exists and is a probability density so that $F(t) = \int_0^t f(x)\,dx$. In this case for each n, the distribution, $F_n(t) = P(W_n \le t)$ also has a density. For example,

$$\begin{aligned}\frac{d}{dt}F_2(t) &= \frac{d}{dt}\int_0^t F(t-x)\,f(x)\,dx \\ &= F(t-t)\,f(t) + \int_0^t F'(t-x)\,f(x)\,dx = \int_0^t f(t-x)\,f(x)\,dx \\ &= f * f(t).\end{aligned}$$

In general we have $\dot{F}_n(t) = \dot{F}_{n-1} * f$ from which it follows that

$$\dot{F}_n = \overbrace{f * f * \cdots * f}^{n\text{ - times}}.$$

We now define

$$m(t) := \sum_{n=1}^{\infty} f^{*n}$$

so that $M(t) = \int_0^t m(x)\,dx$. By the same reasoning as in the first proof of Proposition 10.12 we find that m satisfies the renewal equation,

$$m = f + m * f. \tag{10.12}$$

This may also be seen by differentiating the identity,

$$M(t) = F(t) + \int_0^t M(t-x)\,f(x)\,dx, \tag{10.13}$$

to find

$$\begin{aligned}m(t) &= f(t) + \int_0^t m(t-x)\,f(x)\,dx + M(0)\,f(t) \\ &= f(t) + \int_0^t m(x)\,f(t-x)\,dx. \qquad (10.14)\end{aligned}$$

wherein we have used $P(X_1 = 0) = 0$ and therefore, $N(0) = 0$ and hence $M(0) = \mathbb{E}N(0) = 0$.

Example 10.14 (Posisson process again). If $X_j \overset{d}{=} \exp(\lambda)$, then $f(t) = \lambda e^{-\lambda t}$, and therefore

$$\begin{aligned}m(t) &= \lambda e^{-\lambda t} + \int_0^t m(x)\,\lambda e^{-\lambda(t-x)}dx \\ &= \lambda e^{-\lambda t}\left[1 + \int_0^t m(x)\,e^{\lambda x}dx\right].\end{aligned}$$

From this equation it follows that $m(0) = \lambda$ and

$$\frac{d}{dt}\left(e^{\lambda t}m(t)\right) = \lambda\left(e^{\lambda t}m(t)\right)$$

which has solution, $e^{\lambda t}m(t) = \lambda e^{\lambda t}$, i.e. $m(t) = \lambda$ and therefore,

$$M(t) = \int_0^t m(\tau)\,d\tau = \lambda t$$

which is consistent with the results in Example 10.11.

The next theorem gives an indication as to why the renewal function $M(t)$ is an important quantity.

Theorem 10.15. *The joint distribution of γ_t and δ_t is determined by, for $t \ge x \ge 0$ and $y > 0$,*

$$P(\gamma_t > y, \delta_t \ge x) = 1 - F(y+t) + \int_0^{t-x}(1 - F(y+t-z))\,dM(z). \tag{10.15}$$

In particular, taking $x = 0$ implies,

$$P(\gamma_t > y) = 1 - F(y+t) + \int_0^t (1 - F(y+t-z))\,dM(z)$$

and taking $y = 0$,

$$P(\delta_t \ge x) = 1 - F(t) + \int_0^{t-x}(1 - F(t-z))\,dM(z). \tag{10.16}$$

Proof. You will show in Problem VII.P1.1 that

$$P(\gamma_t > y, \delta_t \ge x) = 1 - F(y+t) + \sum_{k=1}^{\infty}\int_0^{t-x}(1 - F(y+t-z))\,dF_k(z).$$

Recalling the $M(t) := \sum_{k=1}^{\infty} F_k(t)$, we may write this last equation as in Eq. (10.16). ■

Notice that when $x = y = 0$, we should have

$$1 = 1 - F(t) + \int_0^t (1 - F(t-z))\,dM(z)$$

which is the case since,

$$\int_0^t (1 - F(t - z))\, dM(z) = M(t) - F * M(t) = F(t).$$

Goals: Find the limiting behavior of $N(t)$, $M(t)$, γ_t, and δ_t and apply these results to reliability models.

The first step in this program is to show the renewal function, $M(t) = \mathbb{E}N(t)$, is finite.

Lemma 10.16. *Suppose that* $X_1 \geq 0$ *with* $X_1 \neq 0$ *a.s. (We do not have to assume that* $P(X = 0) = 0$ *here only that* $P(X > 0) > 0$*.) Then* $M(t) < \infty$ *for all* t *and consequently* $N(t) < \infty$ *a.s. for all* $t \geq 0$.

Proof. Choose $\alpha > 0$ such that $p := P(X_1 \geq \alpha) > 0$ and hence

$$q := 1 - p = P(X_1 < \alpha) < 1.$$

Then $X_k \geq \alpha 1_{X_k \geq \alpha}$ and hence $\sum_{k=1}^n \alpha 1_{X_k \geq \alpha} \leq W_n$ from which it follows that

$$\{W_n \leq t\} \subset \left\{\sum_{k=1}^n \alpha 1_{X_k \geq \alpha} \leq t\right\} = \left\{\sum_{k=1}^n 1_{X_k \geq \alpha} \leq \frac{t}{\alpha}\right\}.$$

For large n, the latter event happens iff no more than $m := [t/\alpha]$ of the X_k are greater than or equal to α. The probability of this event is

$$\begin{aligned}\sum_{k=0}^m \binom{n}{k} p^k q^{n-k} &= \sum_{k=0}^m \frac{n \cdot (n-1) \dots (n-k+1)}{k!} p^k q^{n-k} \\ &\leq m \cdot n^m q^{n-m} = \frac{m}{q^m} \cdot n^m q^n.\end{aligned}$$

So we have shown for large n, that

$$P(W_n \leq t) \leq \frac{m}{q^m} \cdot n^m q^n$$

from which it follows (by the ratio test) that $M(t) = \sum_{n=1}^\infty P(W_n \leq t) < \infty$. ■

10.2 The Elementary Renewal Theorem

The following proposition is a consequence of Wald's formula. We will also give a proof based on renewal equation theory. For this second proof we will need the following proposition.

Proposition 10.17. *Let* $\mu := \mathbb{E}X_1 = \mathbb{E}X_k$, *then*

$$\mathbb{E}\left[W_{(N(t)+1)}\right] = \mathbb{E}X_1 \cdot \mathbb{E}[N(t) + 1] = \mu(1 + M(t)). \tag{10.17}$$

Proof. This is a direct consequence of Wald's formula in Theorem 4.38, which is applicable because $N(t) + 1$ is a stopping time as we saw in item 3. of Example 4.36. Let us recall the main points here. Recall that

$$\{N(t) = k\} = \{W_k \leq t,\ W_{k+1} > t\}$$

which shows $N(t)$ is **not** a stopping time. However,

$$\{N(t) + 1 = k\} = \{N(t) = k - 1\} = \{W_{k-1} \leq t,\ W_k > t\},$$

from which it follows that $N(t)+1$ **is** a stopping time and hence Wald's formula applies. ■

Lemma 10.18. $\lim_{t\to\infty} N(t) = \infty$ *a.s., i.e.* $P\left(\lim_{t\to\infty} N(t) = \infty\right) = 1$.

Proof. Since $N(t)$ is increasing in t, $N(\infty) := \lim_{t\to\infty} N(t)$ exists and represents the total number of events occurring at any time $t \geq 0$. Therefore $N(\infty) < \infty$ iff $W_n = \infty$ for some n which happens iff $X_n = \infty$ for some n. Thus we conclude that

$$\begin{aligned}P(N(\infty) < \infty) &= P(X_n = \infty \text{ for some } n) \\ &= P(\cup_n \{X_n = \infty\}) \leq \sum_n P(X_n = \infty) = \sum_n 0 = 0.\end{aligned}$$

■

Theorem 10.19 (Pointwise renewal theorem). *Let* $\mu := \mathbb{E}X_1$ *and* $N(t)$ *be as above, then*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu} \quad a.s.$$

Proof. By the strong law of large numbers

$$\frac{W_n}{n} = \frac{X_1 + \dots + X_n}{n} \to \mu \quad a.s.$$

Case 1: If $\mu = \infty$, then for any $0 < M < \infty$, we will have $W_n \geq Mn$ for large n (depending on ω). Therefore, for large t,

$$\begin{aligned}N(t) &= \max\{n : W_n \leq t\} \\ &\leq \max\{n : Mn \leq t\} \cong t/M\end{aligned}$$

and hence we may conclude that

$$\frac{N(t)}{t} \leq \frac{1}{M} \text{ for large } t.$$

Since $M < \infty$ was arbitrary it follows that $\lim_{t\to\infty} \frac{N(t)}{t} = 0 = 1/\mu$.

Case 2: If $\mu = \mathbb{E}X_1 < \infty$, then $\frac{W_n}{n} \cong \mu$ for large n, i.e. $W_n \cong \mu n$. Therefore, for large t,

$$N(t) = \max\{n : W_n \leq t\} \cong \max\{\mu n \leq t\} \cong t/\mu.$$

Dividing this equation by t and then letting $t \uparrow \infty$ shows,

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$$

■

From Theorem 10.19, we expect,

$$\lim_{t\to\infty} \frac{\mathbb{E}N(t)}{t} = \lim_{t\to\infty} \mathbb{E}\left(\frac{N(t)}{t}\right) = \mathbb{E}\left(\lim_{t\to\infty} \frac{N(t)}{t}\right) = \mathbb{E}\left(\frac{1}{\mu}\right) = \frac{1}{\mu}$$

provided it is permissible to interchange the limit and expectation operation in this case. Assuming this is OK, we are lead to the elementary renewal Theorem 10.21 below. Before stating this theorem let us consider the following example which shows that interchanging limits and expectations is not always permissible.

Example 10.20. Suppose that $Y_n(x) = n \cdot 1_{x \leq 1/n}$ for $0 \leq x \leq 1$ as in Figure 10.5. If $\mathbb{E}f := \int_0^1 f(x)\,dx$, then we have $\mathbb{E}Y_n = 1$ for all n while $\lim_{n\to\infty} Y_n = 0$

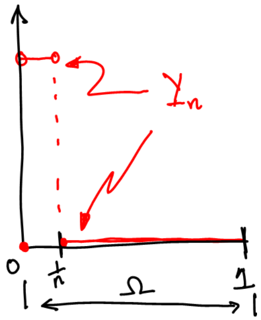


Fig. 10.5. A sequence of approximate δ – functions.

and therefore,

$$\lim_{n\to\infty} \mathbb{E}Y_n = 1 \neq 0 = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right].$$

Theorem 10.21 (Elementary renewal theorem). *If $M(t) := \mathbb{E}N(t)$ is the renewal function, then*

$$\lim_{t\to\infty} \frac{M(t)}{t} = \frac{1}{\mu}. \tag{10.18}$$

Proof. First proof for X_j bounded. In this proof we will assume that the X_j are bounded, i.e. there exists some $T < \infty$ such that $P(X_1 \geq T) = 0$. (This is a rather reasonable assumption for a man-made object.) The key point is to observe that $W_{N(t)+1} = t + \gamma_t$. Taking expectations of this identity and using Proposition 10.17 implies

$$\mu(M(t) + 1) = t + \mathbb{E}\gamma_t. \tag{10.19}$$

After a little algebra this gives,

$$\frac{M(t)}{t} = \frac{1}{\mu} + \frac{1}{t}\left(\frac{\mathbb{E}\gamma_t}{\mu} - 1\right). \tag{10.20}$$

Since X_1 is bounded we must have $\gamma_t \leq \beta_t \leq T$ for all t and therefore, $\mathbb{E}\gamma_t \leq T$. This then implies,

$$\left|\frac{1}{t}\left(\frac{\mathbb{E}\gamma_t}{\mu} - 1\right)\right| \leq \frac{1}{t}\left(\frac{T}{\mu} + 1\right) \to 0 \text{ as } t \to \infty,$$

which combined with Eq. (10.20) gives Eq. (10.18).

Second Proof. Recall that $N(t) = k$ iff $W_k \leq t < W_{k+1}$ and hence we have

$$W_{N(t)} \leq t < W_{N(t)+1}. \tag{10.21}$$

Taking expectations of this equation and then making use of Proposition 10.17 implies,

$$t \leq \mathbb{E}\left[W_{N(t)+1}\right] = \mathbb{E}X_1 \cdot \mathbb{E}(N(t) + 1) = \mu \cdot (M(t) + 1).$$

Dividing this equation by t and then letting $t \to \infty$ implies,

$$\liminf_{t\to\infty} \frac{M(t)}{t} \geq \frac{1}{\mu}. \tag{10.22}$$

To prove the opposite inequality, $c > 0$, $X_i^c = X_i \wedge c := \min(X_i, c)$, and $W_k^c := \sum_{i=1}^k X_i^c$. If $N^c(t) = k$ then

$$W_k^c \leq t < W_{k+1}^c \leq W_k^c + c \leq t + c$$

from which it follows that

$$W_{N^c(t)+1}^c \leq t + c.$$

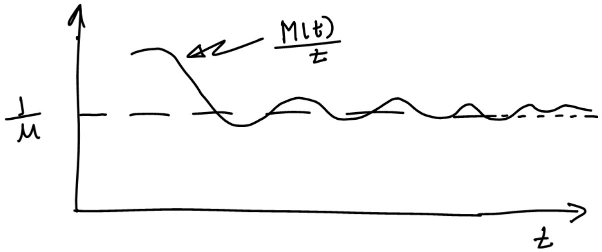


Fig. 10.6. A possible graph of $M(t)/t$.

Taking expectation of this equation again making use of Proposition 10.17 implies,

$$\mathbb{E}X_1^c \cdot (\mathbb{E}N^c(t) + 1) \le t + c.$$

Dividing this inequality by t and then letting $t \uparrow \infty$ implies,

$$\limsup_{t\to\infty} \frac{\mathbb{E}N^c(t)}{t} \le \frac{1}{\mathbb{E}X_1^c}. \tag{10.23}$$

Since $X_i^c \le X_i$, $W_k^c \le W_k$ and hence $N^c(t) \ge N(t)$ which implies $\mathbb{E}N(t) \le \mathbb{E}N^c(t)$. Therefore it follows form Eq. (10.23) that

$$\limsup_{t\to\infty} \frac{\mathbb{E}N(t)}{t} \le \limsup_{t\to\infty} \frac{\mathbb{E}N^c(t)}{t} \le \frac{1}{\mathbb{E}X_1^c}.$$

We may now let $c \uparrow \infty$ in which case $\mathbb{E}X_1^c \uparrow \mu$ and hence we may conclude that

$$\limsup_{t\to\infty} \frac{\mathbb{E}N(t)}{t} \le \frac{1}{\mu}. \tag{10.24}$$

Combining Eqs. (10.22) and (10.24) completes the proof of the theorem. ■

Besides these theorems we also have a central limit theorem for the renewal process, $N(t)$.

Theorem 10.22 (Renewal central limit theorem). *Let* $\mu := \mathbb{E}X_1 < \infty$ *and* $\sigma^2 := \operatorname{Var}(X_1)$, *then*

$$N(t) \overset{d}{\approx} \frac{t}{\mu} + \frac{\sigma}{\mu^{3/2}}\sqrt{t}Z \text{ as } t \to \infty. \tag{10.25}$$

To be more precise,

$$\frac{N(t) - \frac{t}{\mu}}{\frac{\sigma}{\mu^{3/2}}\sqrt{t}} \implies Z, \tag{10.26}$$

where Z *is a standard normal random variable.*

Proof. I will not attempt to give a rigorous proof here but let us try to understand the formula in Eq. (10.25). In doing so well will follow [4, page 110]. (For another "proof," see subsection 10.5.1 below.) By the usual central limit theorem,

$$W_n = \mu n + \sqrt{n}\sigma Z_n$$

where Z_n is approaching in distribution a standard normal random variable, Z. be a standard normal random variable, then by the standard central limit theorem,from which it follows that

$$N\left(\mu n + \sqrt{n}\sigma Z_n\right) = N(W_n) = n.$$

If we write t for μn, then

$$n = N\left(\mu n + \sqrt{n}\sigma Z_n\right) = \begin{cases} N(t) + N((t, t + \sqrt{n}\sigma Z_n]) \text{ if } Z_n > 0 \\ N(t) - N((t + \sqrt{n}\sigma Z_n, t]) \text{ if } Z_n \le 0 \end{cases}. \tag{10.27}$$

However, the number of renewals in an interval of size $\sqrt{n}\sigma|Z_n|$ near large times t, should be approximately, $\sqrt{n}\sigma|Z_n|/\mu$ and therefore we may write Eq. (10.27) as

$$n \cong N(t) + \sqrt{n}\sigma Z_n/\mu$$

which gives,

$$\begin{aligned} N(t) &\cong n - \sqrt{n}\sigma Z_n/\mu = \frac{t}{\mu} - \sqrt{\frac{t}{\mu}}\sigma Z_n/\mu \\ &\overset{d}{\approx} \frac{t}{\mu} - \frac{\sigma}{\mu^{3/2}}\sqrt{t}Z \overset{d}{\approx} \frac{t}{\mu} + \frac{\sigma}{\mu^{3/2}}\sqrt{t}Z \end{aligned}$$

where for the last approximation we have used $Z \overset{d}{=} -Z$. ■

10.3 Applications of the elementary renewal theorem

10.3.1 Age Replacement Policies

Let $X_1, X_2, X_3, \ldots$ be the i.i.d. life times of some device and $\mu = \mathbb{E}X_k$ be their common means. The device is to be replaced upon failure or at some time $T < \infty$ whichever comes first. With this replacement policy, the new effective lifetime of the j^{th} device is X_j^T where $X_j^T = T \wedge X_j = \min(T, X_j)$. Observe that

$$\begin{aligned} \mathbb{E}X_1^T &= \int_0^\infty P(X_1^T > x)dx = \int_0^T P(X_1^T > x)dx \\ &= \int_0^T P(X_1 > x)dx = \int_0^T (1 - F(x))dx =: \mu_T < \mu. \end{aligned} \tag{10.28}$$

Given the above replacement policy, let Y_1 denote the time of the first replacement of a part which occurred before time T, i.e. the first time that a part is replaced because it has failed. We begin by working out the distribution of Y_1. To understand Y_1 better we see by its definition, that

$$\begin{aligned}
&Y_1 = X_1 \text{ if } X_1 \leq T,\\
&Y_1 = T + X_2 \text{ if } X_1 > T \text{ but } X_2 \leq T,\\
&Y_1 = 2T + X_3 \text{ if } X_1 > T,\ X_2 > T, \text{ and } X_3 \leq T,\\
&\quad\vdots\\
&Y_1 = nT + X_{n+1} \text{ if } X_1 > T,\ X_2 > T, \dots,\ X_n > T \text{ and } X_{n+1} \leq T.
\end{aligned}$$

Thus we see that we may write $Y_1 = NT + Z$, where N is the $\mathbb{N}_0$ − valued random variable Z takes values in $[0, T]$ such that $N = 0$ and $Z = X_1$ if $X_1 \leq T$ and

$$N = n \text{ and } Z = X_{n+1} \text{ on } \{X_1 > T, \dots, X_n > T \text{ and } X_{n+1} \leq T\}.$$

Hence it follows that

$$\begin{aligned}
\{N \geq n\} &= \{X_1 > T,\ X_2 > T, \dots,\ X_n > T\} \text{ and}\\
\{N = n\} &= \{X_1 > T,\ X_2 > T, \dots,\ X_n > T, X_{n+1} \leq T\}.
\end{aligned}$$

Therefore for all $n \in \mathbb{N}_0$,

$$\begin{aligned}
P(N \geq n) &= P(X_1 > T)^n = (1 - F(T))^n \text{ and} && (10.29)\\
P(N = n) &= (1 - F(T))^n F(T). && (10.30)
\end{aligned}$$

Moreover, for $0 \leq z \leq T$ we have

$$\begin{aligned}
P(Z \leq z, N = n) &= P(X_{n+1} \leq z, X_1 > T, \dots, X_n > T, X_{n+1} \leq T)\\
&= P(X_1 > T, \dots, X_n > T, X_{n+1} \leq z)\\
&= P(X_{n+1} \leq z) P(X_1 > T)^n = F(z) P(X_1 > T)^n\\
&= \frac{F(z)}{F(T)} P(X_1 > T)^n F(T) = \frac{F(z)}{F(T)} P(N = n). && (10.31)
\end{aligned}$$

Summing this equation on n, shows,

$$P(Z \leq z) = \frac{F(z)}{F(T)} \text{ for } 0 \leq z \leq T. \quad (10.32)$$

Using this information it follows that

$$\begin{aligned}
\mathbb{E}N &= \mathbb{E}\sum_{n=1}^{\infty} 1_{n \leq N} = \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} P(X_1 > T)^n\\
&= \frac{P(X_1 > T)}{1 - P(X_1 > T)} = \frac{1 - F(T)}{F(T)}. && (10.33)
\end{aligned}$$

Similarly,

$$\mathbb{E}Z = \mathbb{E}\int_0^T 1_{z<Z} dz = \int_0^T P(Z > z)\, dz = \int_0^T \left[1 - \frac{F(z)}{F(T)}\right] dz \quad (10.34)$$

and since $Y_1 = NT + Z$ we find,

$$\begin{aligned}
\mathbb{E}Y_1 &= T \cdot \mathbb{E}N + \mathbb{E}Z\\
&= \frac{1 - F(T)}{F(T)} \cdot T + \int_0^T \left[1 - \frac{F(z)}{F(T)}\right] dz\\
&= \int_0^T \left[\frac{1 - F(T)}{F(T)} + \frac{F(T) - F(z)}{F(T)}\right] dz\\
&= \frac{1}{F(T)} \int_0^T (1 - F(z))\, dz = \frac{\mu_T}{F(T)} = \frac{\mathbb{E}[X_1 \wedge T]}{P(X_1 \leq T)},
\end{aligned}$$

where μ_T was the mean used life of the part under the replacement scheme in Eq. (10.28).

Summary: If

$$\mu_T := \int_0^T (1 - F(x)) dx, \quad (10.35)$$

then

$$\mathbb{E}[X_1 \wedge T] = \mu_T \text{ and } \mathbb{E}Y_1 = \frac{\mu_T}{F(T)}. \quad (10.36)$$

We now let $\{Y_i\}_{i=1}^{\infty}$ be the times between actual successive failures of a part using the replacement scheme above.

Proposition 10.23. *The sequence of random variables,* $\{Y_i\}_{i=1}^{\infty}$ *are i.i.d.*

Proof. We are going to show more, namely if we write $Y_i = N_i T + Z_i$ then the sequence of random variables, $\{(N_i, Z_i)\}_{i=1}^{\infty}$ are i.i.d. I will actually only work out the joint distribution of (N_1, Z_1) and (N_2, Z_2) here. Let $n_1, n_2 \in \mathbb{N}_0$ and $z_1, z_2 \in [0, T]$, then

$$\begin{aligned}
&\{N_1 = n_1, Z_1 \leq z_1, N_2 = n_2, Z_2 \leq z_2\}\\
&= \{X_1 > T, \dots, X_{n_1} > T, X_{n_1+1} \leq z_1, X_{n_1+2} > T, \dots, X_{n_1+n_2+2} > T, X_{n_1+n_2+3} \leq z_2\}
\end{aligned}$$

and therefore,

$$\begin{aligned}
&P(N_1 = n_1, Z_1 \leq z_1, N_2 = n_2, Z_2 \leq z_2)\\
&= P(X_1 > T, \dots, X_{n_1} > T, X_{n_1+1} \leq z_1)\\
&\quad \cdot P(X_{n_1+2} > T, \dots, X_{n_1+n_2+2} > T, X_{n_1+n_2+3} \leq z_2)\\
&= P(N_1 = n_1, Z_1 \leq z_1) \cdot P(X_1 > T, \dots, X_{n_2} > T, X_{n_2+1} \leq z_2)\\
&= P(N_1 = n_1, Z_1 \leq z_1) \cdot P(N_2 = n_2, Z_2 \leq z_2).
\end{aligned}$$

This shows that (N_1, Z_1) and (N_2, Z_2) are independent and have the same distribution. The general case follows similarly. ■

The elementary renewal theorem then implies that the failure rate for this replacement scheme is

$$\lim_{t\to\infty}\frac{M_{eff}(t)}{t}=\frac{F(T)}{\mu_T}=\frac{F(T)}{\int_0^T(1-F(z))dz}=\frac{P(X_1\le T)}{\mathbb{E}[X_1\wedge T]}.$$

The long time replacement rate for replacing on failure is given by

$$\lim_{t\to\infty}\frac{M(t)}{t}=\frac{1}{\mathbb{E}X_1}=\frac{1}{\int_0^\infty(1-F(z))dz}$$

Example 10.24. Suppose that X_i has the uniform on $[0,1]$ and $0<T<1$. Then $F(x)=P(X_i\le x)=x\wedge 1$. Thus effective long run failure rate is given by

$$\frac{F(T)}{\mu_T}=\frac{F(T)}{\int_0^T(1-F(x))\,dx}=\frac{T}{\int_0^T(1-x)dx}$$
$$=\frac{T}{T-\frac{T^2}{2}}=\frac{1}{1-\frac{T}{2}}=\frac{2}{2-T},$$

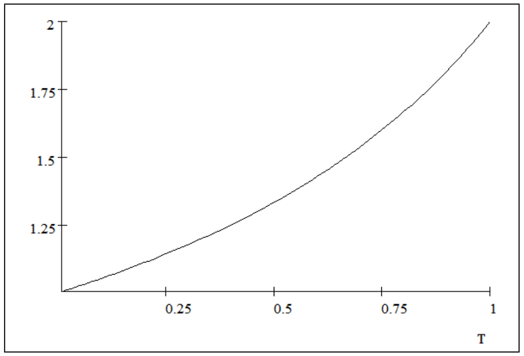
that is :

$$\lim_{t\to\infty}\frac{M_{eff}(t)}{t}=\frac{F(T)}{\mu_T}=\frac{2}{2-T},\tag{10.37}$$

while

$$\lim_{t\to\infty}\frac{M(t)}{t}=\frac{1}{\mu}=\frac{1}{\frac{1}{2}}=2.\tag{10.38}$$

(Observe that letting $T\uparrow 1$ in Eq. (10.37) gives back Eq. (10.38).) The original failure rate was 2 while the effective failure rate as a function of T is $\frac{2}{2-T}$ which is plotted in Figure 10.24 below.



The effective failure rate as function of the forced replacement time, T.

By making T small we can reduce the long run failure rate to close to 1. Keep in mind that we are making replacements on average at rate,

$$\frac{1}{\mu_T}=\frac{1}{\mathbb{E}[X_1\wedge T]}=\frac{1}{\int_0^T xdx+T(1-T)}=\frac{1}{T^2/2+T(1-T)}$$
$$=\frac{1}{T-T^2/2}$$

which is graphed in Figure 10.7. For example if we take $T=1/2$, we get a failure rate of 4/3 and a replacement rate of

$$\frac{1}{1/2-1/8}=\frac{8}{3}\cong 2.67$$

versus the original failure rate of 2 and replacement rate of 2.

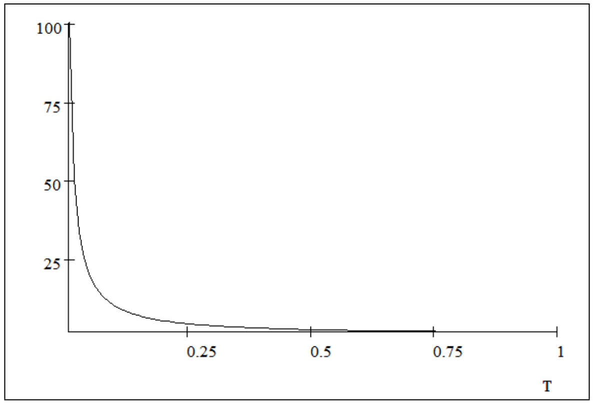


Fig. 10.7. Replacement rate as a function of T.

Example 10.25. Let us continue the above analysis by associating different cost to replacement and to failures. We are now going to assume the cost to replace a part (failed or not) is K dollars and each failed part incurs and additional cost of c dollars. The total cost up to time t to the factory is then

$$V(t):=K\cdot N\left(t:X_1^T,X_2^T,\dots\right)+c\cdot N(t:Y_1,Y_2,\dots)$$

and so the expected cost is

$$\mathbb{E}V(t)=K\cdot\mathbb{E}N\left(t:X_1^T,X_2^T,\dots\right)+c\cdot\mathbb{E}N(t:Y_1,Y_2,\dots)$$

and the limiting cost per unit time becomes, by two applications of the elementary renewal theorem,

$$\begin{aligned}\lim_{t\to\infty}\frac{\mathbb{E}V(t)}{t} &= K\cdot\lim_{t\to\infty}\mathbb{E}N\left(t:X_1^T,X_2^T,\dots\right)+c\cdot\lim_{t\to\infty}\mathbb{E}N\left(t:Y_1,Y_2,\dots\right)\\ &=\frac{K}{\mathbb{E}X_1^T}+\frac{c}{\mathbb{E}Y_1}.\end{aligned}$$

Recalling form Eqs. (10.35) and (10.36) that

$$\mathbb{E}\left[X_1^T\right]=\mu_T \text{ and } \mathbb{E}Y_1=\frac{\mu_T}{F(T)},\text{ where}$$
$$\mu_T:=\int_0^T(1-F(x))dx,$$

we have,

$$\begin{aligned}C(T):=\lim_{t\to\infty}\frac{\mathbb{E}V(t)}{t}&=\frac{K}{\mu_T}+\frac{c}{\mu_T/F(T)}=\frac{K+cF(T)}{\mu_T}\\ &=\frac{K+cF(T)}{\int_0^T(1-F(t))\,dt}.\end{aligned}$$

which represents the long time cost per unit time of running the factory with this replacement strategy.

Goal: given K, c, and F, we would like to choose T so as to minimize $C(T)$.

Let us now be more specific. work this out in an example or two.

Example 10.26. Suppose that X_n have the uniform distribution on $[0,1]$, i.e. $F(x)=1\wedge x=\min(1,x)$. In this case we should keep $T<1$ and then we have,

$$C(T)=\frac{K+cT}{\int_0^T(1-x)dx}=\frac{K+cT}{T-\frac{T^2}{2}}.$$

We now the first derivative test to try to find the best choice for T so as to minimize the cost function, $C(T)$;

$$\begin{aligned}0\overset{\text{set}}{=}C'(T)&\propto c\left(T-\frac{T^2}{2}\right)-(1-T)(K+cT)\\ &=\frac{c}{2}T^2+KT-K.\end{aligned}$$

The quadratic formula then gives,

$$T=\frac{-K\pm\sqrt{K^2+4K\frac{c}{2}}}{c}=\frac{-K\pm K\sqrt{1+2c/k}}{c}.$$

Only the plus sign gives a positive root and hence the optimal time, T^*, is given by

$$T^*=\frac{K}{c}\left(\sqrt{1+\frac{2c}{k}}-1\right).$$

For example if $K=1$ **and** $c=4$ we find,

$$C(T)=\frac{1+4T}{T(1-T/2)}$$

$$T^*=\frac{1}{2}\text{ and }C(T^*)=8.$$

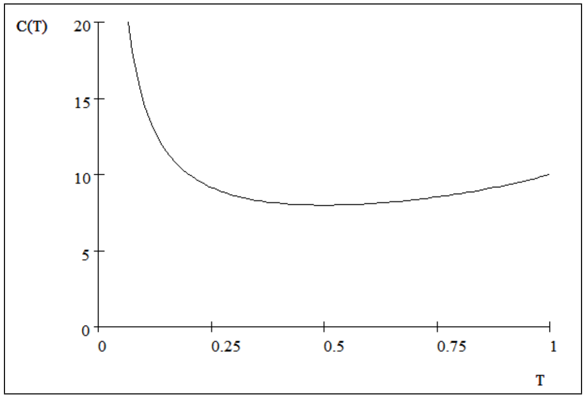


Fig. 10.8. A plot of $C(T)$ for $K=1$ and $c=4$.

Example 10.27. Let us work out the above scenario under the assumption that $X_j\overset{d}{=}\exp(\lambda)$, so that $F(t)=1-e^{-\lambda t}$. In this case

$$\mu_T=\int_0^T e^{-\lambda t}dt=\frac{1}{\lambda}\left(1-e^{-\lambda T}\right)=\frac{1}{\lambda}F(T).$$

Therefore $\mathbb{E}X_1^T=\mu_T=\frac{1}{\lambda}F(T)<\frac{1}{\lambda}=\mu$ while $\mathbb{E}Y_1=\mu_T/F(T)=\frac{1}{\lambda}$. So in this case the actual failure rate is the same no matter what forced replacement time, T, we use. Because of this, the best replacement strategy is to take $T=\infty$ as can also be seen by looking at the cost function,

$$C(T)=K\frac{1}{\mathbb{E}X_1}+c\frac{1}{\mathbb{E}Y_1}=\frac{\lambda K}{F(T)}+\lambda c\downarrow\lambda(K+c)\text{ as }t\uparrow\infty.$$

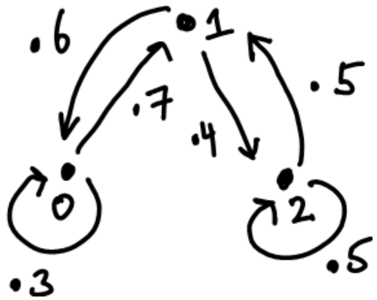
To see this is reasonable, one might think the parts are failing because of some rare catastrophic event, like a satellite being hit by a piece of space debris. In this case the satellite is not wearing out, so if it is not hit at some time T it is as in good shape as it was the day it was put into service. Therefore there is no good reason to replace it early.

10.3.2 Comments on Problem VII.4.5

In this problem one is dealing with the Markov chain determined by the Markov matrix

$$P = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cc} \begin{array}{ccc} 0 & 1 & 2 \end{array} & \\ \begin{bmatrix} .3 & .7 & 0 \\ .6 & 0 & .4 \\ 0 & .5 & .5 \end{bmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \end{array}$$

with jump diagram



Suppose that

$$(X_0, X_1, \dots) = (1, 0, 0, 0, 1, 2, 2, 2, 2, 1, \dots)$$

is a sample path of the system. The three consecutive zeros is said to be a Sojourn at 0 with duration time, $S_0 = 3$ and the four consecutive twos is said to be a Sojourn at 2 with duration time $S_2 = 4$. In this problem a renewal cycle consists of the times between visits to 1. So in this case the first inter-renewal time, X_1, is $4 = 1 + S_0$ while the second inter-renewal time is $X_2 = 5 = 1 + S_2$. The mean inter-renewal time is thus,

$$\mu = \mathbb{E}_1 [X_1 | \text{first visit } 0] P_1 (\text{first visit } 0) + \mathbb{E}_1 [X_1 | \text{first visit } 2] P_1 (\text{first visit } 2)$$

where

$$\mathbb{E}_1 [X_1 | \text{first visit } 0] = \mathbb{E}_0 S_0 + 1 \text{ and}$$
$$\mathbb{E}_1 [X_1 | \text{first visit } 2] = \mathbb{E}_2 S_2 + 1.$$

In order to check your final answer, let us compute the invariant distribution for this chain:

$$\begin{aligned} \operatorname{Nul} (P - I)^{\mathrm{tr}} &= \operatorname{Nul} \left(\begin{bmatrix} .3 & .7 & 0 \\ .6 & 0 & .4 \\ 0 & .5 & .5 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)^{\mathrm{tr}} \\ &= \operatorname{Nul} \left(\begin{bmatrix} -0.7 & 0.6 & 0 \\ 0.7 & -1 & 0.5 \\ 0 & 0.4 & -0.5 \end{bmatrix} \right) = \begin{bmatrix} 0.556\,22 \\ 0.648\,93 \\ 0.519\,14 \end{bmatrix}, \end{aligned}$$

hence the invariant distribution is given by

$$\begin{aligned} \pi &= \frac{1}{0.556\,22 + 0.648\,93 + 0.519\,14} \left[0.556\,22 \; 0.648\,93 \; 0.519\,14 \right] \\ &= \left[0.322\,58 \; 0.376\,35 \; 0.301\,07 \right]. \end{aligned}$$

10.4 The Key Renewal Theorem

Let F be the distribution function of the interarrival random variables, $\{X_j\}$ ($f = \dot{F}$ if F can be described by a density) and $h : \mathbb{R}_+ \to \mathbb{R}$ be a given a function. We are interested in solving the **renewal equation**

$$g = h + g * F \tag{10.39}$$

for the unknown function g.

Example 10.28. We have see in Proposition 10.12 that $g = M$ satisfies the renewal Eq. (10.39) with $h = F$. We also have noted in Eq. (10.12) that if $f = \dot{F}$ exists then $m = \dot{M}$ exists and $g = m$ satisfies the renewal equation with $h = f$.

Example 10.29. Let $g(t) := \mathbb{E}\left[W_{N(t)+1}\right]$, then

$$g(t) = \int_0^\infty \mathbb{E}\left[W_{(N(t)+1)} | X_1 = x\right] dF(x). \tag{10.40}$$

If $t < X_1$, then $N(t) = 0$ and $N(t) + 1 = 1$ so that $W_{N(t)+1} = X_1$, while if $t \geq X_1$, we have and $W_{N(t)+1}(X_1, X_2, X_3, \dots) = X_1 + W_{N(t-X_1)+1}(X_2, X_3, \dots)$ and therefore,

$$\mathbb{E}\left[W_{(N(t)+1)} | X_1 = x\right] = \begin{cases} x & \text{if } t < x \\ x + g(t - x) & \text{if } t \geq x \end{cases}.$$

Using this in Eq. (10.40) shows,

$$\begin{aligned} g(t) &= \int_0^\infty x dF(x) + \int_0^t g(t-x)\, dF(x) = \mu + \int_0^t g(t-x)\, dF(x) \\ &= \mu + g * F(t), \end{aligned}$$

i.e. g satisfies the renewal Eq. (10.39) with $h(t) = \mu$.

Theorem 10.30. *Suppose that $h : \mathbb{R}_+ \to \mathbb{R}$ is function which is bounded on bounded intervals, then among the functions, g, which are bounded on bounded intervals, there is a unique solution to Eq. (10.39). Moreover this solution is given*[1] *by,*

$$g(t) = h(t) + h * M(t) = h(t) + \int_0^t h(t-x)\, dM(x). \tag{10.41}$$

Proof. Uniqueness. Suppose that g_1 and g_2 are two such solutions to Eq. (10.39). Then there difference, $k := g_2 - g_1$ solves, $k = k * F$. Iterating this equation then shows,

$$\begin{aligned} |k(t)| = |k * F_n(t)| &= \left| \int_0^t k(t-x)\, dF_n(x) \right| \\ &\le 2K_t \int_0^t dF_n(x) = 2K_t P(W_n \le t). \end{aligned}$$

where K_t is a bounded on $g_1(s)$ and $g_2(s)$ for $s \le t$. Since $\frac{1}{n} W_n \to \mu > 0$ as $n \to \infty$, i.e. $W_n \cong n\mu$ for n large, it follows that $P(W_n \le t) \to 0$ as $n \to \infty$. This shows that $k(t) = 0$ and hence $g_1(t) = g_2(t)$. Thus the solution is unique.

Existence. Notice that

$$|h * M|(t) = \left| \int_0^t h(t-x)\, dM(x) \right| \le \int_0^t |h(t-x)|\, dM(x) \le K_t M(t)$$

where $K_t = \sup_{s \le t} |h(s)| < \infty$. Therefore $g := h + h * M$ is bounded on bounded intervals. Moreover,

$$\begin{aligned} h + g * F &= h + (h + h * M) * F \\ &= h + h * F + (h * M) * F \\ &= h + h * F + h * (M * F). \end{aligned}$$

Recalling, see Eq. (10.12) that M satisfies the renewal equation, $M * F = M - F$, it follows that

$$h + g * F = h + h * F + h * (M - F) = h + h * M = g$$

as desired. ■

Example 10.31. In this example we give a second proof of the identity in Eq. (10.17) in Proposition 10.17, namely that

$$\mathbb{E}\left[W_{(N(t)+1)}\right] = \mu(1 + M(t)).$$

By Example 10.29, $g(t) := \mathbb{E}\left[W_{(N(t)+1)}\right]$, satisfies the renewal equation, $g = \mu + g * F$. The solution to this equation, by Proposition 10.47, is $g(t) = \mu + \mu * M(t)$. This complete the proof since,

$$\mu * M(t) = \int_0^t \mu dM(x) = \mu M(t) - M(0) = \mu M(t).$$

To make use of these renewal equations and solutions, we need the following stronger version of the elementary renewal theorem.

Theorem 10.32 (Blackwell's renewal theorem). *Suppose $\{X_i\}_{i=1}^\infty$ are i.i.d. random times which have continuous distributions. Then (a bit informally),*

$$m(t) = \dot{M}(t) \to \frac{1}{\mu} \text{ as } t \to \infty.$$

Proof. Heuristic proof. The elementary renewal theorem states, $M(t) \cong \frac{1}{\mu} t$ for t large. Hence if we differentiate this relation, we suspect that $m(t) = \dot{M}(t) \cong \frac{1}{\mu}$ for t large.

For a proof and the correct statement of this theorem along with its generalizations to "non-lattice" random variables, the reader is referred to Durrett [2, Theorem 4.3 on p. 206] and the references therein. ■

---

[1] If we let $C_F g := g * F$, then the renewal equation may be stated as,

$$(I - C_F) g = h.$$

The formal solution to this linear equation is therefore,

$$g = (I - C_F)^{-1} h.$$

Motivated by geometric series ideas, we should expect,

$$(I - C_F)^{-1} = \sum_{n=0}^\infty C_F^n = I + \sum_{n=1}^\infty C_F^n$$

and therefore that

$$g = h + \sum_{n=1}^\infty C_F^n h = h + \sum_{n=1}^\infty h * F_n = h + h * M.$$

This motivates the formula for g in Eq. (10.41).

Theorem 10.33 (Key Renewal Theorem). *Suppose that h is a function which is bounded on bounded intervals, $\lim_{t\to\infty} h(t) = 0$, and $\int_0^\infty |h(t)|\,dt < \infty$. Let us further suppose that $\dot{F} = f$ exists. If $g = h + h * M$ is the solution to the renewal Eq. (10.39), then*

$$\lim_{t\to\infty} g(t) = \frac{1}{\mu}\int_0^\infty h(t)\,dt. \tag{10.42}$$

Proof. By Theorem 10.30, we know that $g(t) = h(t) + h * M(t)$ and therefore,

$$\lim_{t\to\infty} g(t) = \lim_{t\to\infty} h * M(t) = \lim_{t\to\infty}\int_0^t h(x)\,m(t-x)\,dx.$$

By Blackwell's renewal Theorem 10.32, $m(x) \to \frac{1}{\mu}$ as $x \to \infty$, it follows that

$$\lim_{t\to\infty}\int_0^{t/2} h(x)\,m(t-x)\,dx = \frac{1}{\mu}\int_0^\infty h(x)\,dx.$$

On the other hand (assuming that m is bounded by some $K < \infty$), then

$$\left|\int_{t/2}^t h(x)\,m(t-x)\,dx\right| \le K\int_{t/2}^t |h(x)|\,dx \le K\int_{t/2}^\infty |h(x)|\,dx \to \infty \text{ as } t \to \infty.$$

We may relax the assumption on the boundedness of K as follows. For large t we know that $m(t) \to \frac{1}{\mu}$ and hence there exists a $c < \infty$ such that $m(t) \le K$ for $t \ge c$. Therefore,

$$\begin{aligned}\left|\int_{t/2}^t h(x)\,m(t-x)\,dx\right| &\le \int_{t/2}^t |h(x)|\,m(t-x)\,dx \\ &= \int_{t/2}^{t-c} |h(x)|\,m(t-x)\,dx + \int_{t-c}^t |h(x)|\,m(t-x)\,dx.\end{aligned}$$

The first integral goes to zero by the previous argument. For the latter integral we have,

$$\begin{aligned}\int_{t-c}^t |h(x)|\,m(t-x)\,dx &\le \sup_{x\ge t-c}|h(x)|\cdot\int_{t-c}^t m(t-x)\,dx \\ &= \sup_{x\ge t-c}|h(x)|\cdot\int_0^c m(u)\,du \\ &= M(c)\cdot\sup_{x\ge t-c}|h(x)| \to 0 \text{ as } t\to\infty.\end{aligned}$$

■

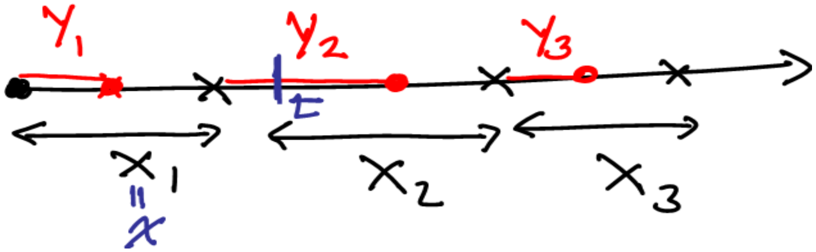


Fig. 10.9. The picture you should have in mind for a two part renewal process. For this sample path and choice of t, t is in the Y phase of the renewal process.

10.5 Examples using the key renewal theorem

Let us now work out a number of examples of this theory. We will always assume that $\dot{F}(t) = f(t)$ exists so that $m(t) = \dot{M}(t)$ exists.

Proposition 10.34 (Two component renewals). *Suppose Y_i represents a portion of the duration of X_i. To be precise, we assume that $\{(Y_i, X_i)\}_{i=1}^\infty$ are i.i.d., $0 \le Y_i \le X_i$, and X_i is a continuous random variable. Then*

$$\lim_{t\to\infty} P(\textit{in the } Y \textit{ phase at time } t) = \frac{\mathbb{E}Y_1}{\mathbb{E}X_1}, \tag{10.43}$$

where $P(\text{in the } Y \text{ phase at time } t)$ represents the probability that t falls in a Y – portion at time t, see Figure 10.9. (The result in Eq. (10.43) is a intuitively reasonable.)

Proof. If we let $A_t := \{\text{in the } Y \text{ phase at time } t\}$ and $g(t) := P(A_t)$, then as usual,

$$g(t) = \int_0^\infty P[A_t|X_1 = x]\,dF(x).$$

If $x > t$, then

$$P[A_t|X_1 = x] = P(Y_1 > t|X_1 = x)$$

while if $x < t$ (see Figure 10.9), then

$$P[A_t|X_1 = x] = P(A_{t-x}) = g(t-x).$$

Therefore,

$$g(t) = \int_t^\infty P(Y_1 > t|X_1 = x)\,dF(x) + \int_0^t g(t-x)\,dF(x).$$

i.e. $g = h + g * F$ where

$$h(t) := \int_t^\infty P(Y_1 > t | X_1 = x)\, dF(x). \tag{10.44}$$

To evaluate h more explicitly, observe that $P(Y_t > t|X_1 = x) = 0$ if $x \le t$, therefore we may write Eq. (10.44) as

$$h(t) := \int_0^\infty P(Y_1 > t|X_1 = x)\, dF(x) = P(Y_1 > t).$$

An application of the key renewal Theorem 10.33 then gives,

$$\lim_{t\to\infty} P(A_t) = \frac{1}{\mu}\int_0^\infty P(Y_1 > t)\, dt = \frac{\mathbb{E}Y_1}{\mathbb{E}X_1}.$$

■

Example 10.35 (Peter Principle, see page 450-451 of Karlen and Taylor.). A person is selected at random from an infinite population containing a fraction p of competent people and $1-p$ of incompetent people. If the person selected is competent she/he remains in the job for a random time, T_c, before being promoted. While if the person is incompetent, he or she remains for a random time, T_i, and then retires. Once the job is vacated, another person is selected at random and the process repeats (i.e. renews). Let $\mu := \mathbb{E}T_c$ and $\nu = \mathbb{E}T_i$.

Question. In the long run what fraction, f, of the time is the job held by an incompetent person.

Answer. A renewal interval is a random time,

$$X = \begin{cases} T_c & \text{if a competent person is chosen} \\ T_i & \text{if an incompetent person is chosen} \end{cases}.$$

If Y represents the incompetent phase of the renewal interval, then

$$Y = \begin{cases} 0 & \text{if a competent person is chosen} \\ T_i & \text{if an incompetent person is chosen} \end{cases}.$$

We then have

$$f = \frac{\mathbb{E}Y}{\mathbb{E}X} = \frac{(1-p)\,\mathbb{E}T_i}{p\mathbb{E}T_c + (1-p)\,\mathbb{E}T_i} = \frac{(1-p)\,\nu}{p\mu + (1-p)\,\nu},$$

wherein we have used,

$$\mathbb{E}X = \mathbb{E}[X|\text{competent}]\, P(\text{competent}) + \mathbb{E}[X|\text{incompetent}]\, P(\text{incompetent})$$

and

$$\mathbb{E}Y = \mathbb{E}[Y|\text{competent}]\, P(\text{competent}) + \mathbb{E}[Y|\text{incompetent}]\, P(\text{incompetent}).$$

As a specific example, suppose that $p = 1/2$, $\nu = 10$ and $\mu = 1$. Then

$$f = \frac{\frac{1}{2}10}{\frac{1}{2} + \frac{1}{2}10} = \frac{10}{11} \cong .91.$$

Theorem 10.36 (Distribution of $(\gamma_\infty, \delta_\infty)$.). *Suppose that F is the cumulative distribution function for X_j. Then for all $x, y \ge 0$,*

$$\lim_{t\to\infty} P(\gamma_t > y, \delta_t \ge x) = \frac{1}{\mu}\int_{x+y}^\infty (1 - F(w))\, dw. \tag{10.45}$$

In particular, if we let $(\gamma_\infty, \delta_\infty)$ be random variables with

$$\begin{aligned} P(\gamma_\infty > y, \delta_\infty \ge x) &:= \lim_{t\to\infty} P(\gamma_t > y, \delta_t \ge x) \\ &= \frac{1}{\mu}\int_{x+y}^\infty (1 - F(w))\, dw, \end{aligned}$$

then γ_∞ and δ_∞ have distributions with densities given by $\frac{1}{\mu}(1 - F(t))$ for $0 \le t < \infty$.

Proof. First Proof. Recall from your homework (see Theorem 10.15) that

$$P(\gamma_t > y, \delta_t \ge x) = 1 - F(y+t) + \int_0^{t-x} (1 - F(y+t-z))\, m(z)\, dz.$$

Making the change of variables, $w = y + t - z$ in the above integral shows,

$$P(\gamma_t > y, \delta_t \ge x) = 1 - F(y+t) + \int_{x+y}^{y+t} (1 - F(w))\, m(y+t-w)\, dw.$$

Writing $m(z) = \frac{1}{\mu} + \varepsilon(z)$ where $\varepsilon(z) \to 0$ as $z \to \infty$ we learn that

$$\begin{aligned} \lim_{t\to\infty} P(\gamma_t > y, \delta_t \ge x) &= \lim_{t\to\infty}\int_{x+y}^{y+t} (1 - F(w))\, m(y+t-w)\, dw \\ &= \frac{1}{\mu}\int_{x+y}^\infty (1 - F(w))\, dw \\ &\quad + \lim_{t\to\infty}\int_{x+y}^{y+t} (1 - F(w))\, \varepsilon(y+t-w)\, dw \\ &= \frac{1}{\mu}\int_{x+y}^\infty (1 - F(w))\, dw. \end{aligned}$$

Second Proof. We start by considering $g(t) := P(\gamma_t > y)$ for some fixed $y \ge 0$. Then by conditioning on $X_1 = x$ we find,

$$g(t) = \int_0^\infty P(\gamma_t > y|X_1 = x)\, f(x)\, dx$$

where

$$P(\gamma_t > y | X_1 = x) = \begin{cases} 1_{t+y<x} & \text{if } t < x \\ P(\gamma_{t-x} > y) = g(t-x) & \text{if } t \geq x \end{cases}.$$

Therefore,

$$g(t) = \int_t^\infty 1_{t+y<x} f(x)\,dx + \int_0^t g(t-x) f(x)\,dx$$
$$= 1 - F(y+t) + \int_0^t g(t-x) f(x)\,dx,$$

which shows $g(t)$ satisfies the renewal equation with $h(t) = 1 - F(y+t)$. Therefore by the key renewal Theorem 10.33,

$$\lim_{t\to\infty} P(\gamma_t > y) = \frac{1}{\mu}\int_0^\infty (1 - F(y+t))\,dt$$
$$= \frac{1}{\mu}\int_y^\infty (1 - F(t))\,dt.$$

Notice that if $y = 0$, then

$$\int_0^\infty (1 - F(t))\,dt = \int_0^\infty P(X_1 > t)\,dt = \mathbb{E}X_1 = \mu,$$

so that

$$G(y) := 1 - \frac{1}{\mu}\int_y^\infty (1 - F(t))\,dt$$

To finish the second proof, we observe that

$$\{\gamma_t > y, \delta_t \geq x\} = \{\gamma_{t-x} > x + y\}.$$

Therefore we may conclude that

$$\lim_{t\to\infty} P(\gamma_t > y, \delta_t \geq x) = \lim_{t\to\infty} P(\gamma_{t-x} > x + y) = \frac{1}{\mu}\int_{x+y}^\infty (1 - F(t))\,dt.$$

■

Example 10.37. If F is the exponential distribution with parameter, $\lambda = 1/\mu$ so that $1 - F(t) = e^{-\lambda t}$, then

$$\lim_{t\to\infty} P(\gamma_t > y, \delta_t \geq x) = \frac{1}{\mu}\int_{x+y}^\infty e^{-\lambda t} dt = \frac{1}{\mu\lambda} e^{-\mu(x+y)} = e^{-\mu(x+y)}$$

which is a result we know to be true even without taking the limit as $t \to \infty$ as we saw in Example 10.5.

Example 10.38 (Earthquakes in California). The inter-earthquake time distribution in California is $U(0,1)$ years. What is the long run probability that an earthquake will hit California within 6 months? What is the long run probability that it has been at most 6 months since an earthquake last hit California?

Solution: Since,

$$\lim_{t\to\infty} P(\gamma_t \leq 0.5) = \lim_{t\to\infty} P(\delta_t \leq 0.5) = 1 - \lim_{t\to\infty} P(\gamma_t > 0.5)$$

the answer to both questions is:

$$1 - \lim_{t\to\infty} P(\gamma_t > 0.5) = 1 - \frac{1}{1/2}\int_{.5}^\infty (1 - F(t))\,dt$$
$$= 1 - \frac{1}{1/2}\int_{.5}^1 (1 - t)\,dt = 0.75.$$

In general,

$$P(\gamma_\infty > T) = 2\int_T^1 (1 - t)\,dt = -(1 - t)^2 |_T^1$$
$$= (1 - T)^2 \text{ for } 0 \leq T \leq 1.$$

Proposition 10.39. *Let $o(1)$ denote a function of t which tends to zero as $t \to \infty$, then*

$$\mathbb{E}\gamma_t = \frac{1}{2\mu}\left(\sigma^2 + \mu^2\right) + o(1), \tag{10.46}$$

$$\mathbb{E}\delta_t = \frac{1}{2\mu}\left(\sigma^2 + \mu^2\right) + o(1) \text{ and} \tag{10.47}$$

$$\mathbb{E}\beta_t = \frac{1}{\mu}\left(\sigma^2 + \mu^2\right) + o(1) > \mu. \tag{10.48}$$

Proof. By Theorem 10.36, we know that

$$\lim_{t\to\infty} \mathbb{E}[f(\gamma_t, \delta_t)] = \mathbb{E}[f(\gamma_\infty, \delta_\infty)]$$

where

$$P(\gamma_\infty > y, \delta_\infty \geq x) := \frac{1}{\mu}\int_{x+y}^\infty (1 - F(w))\,dw.$$

Notice that

$$P(\gamma_\infty > y) = P(\delta_\infty > y) = \frac{1}{\mu}\int_y^\infty (1 - F(t))\,dt \tag{10.49}$$

so that γ_∞ and δ_∞ have the same distribution[2]. Therefore we have,

---

[2] In general γ_∞ and δ_∞ are independent iff X_j is exponentially distributed.

$$\begin{aligned}\mathbb{E}\delta_\infty &= \mathbb{E}\gamma_\infty = \int_0^\infty dy \frac{1}{\mu}\int_y^\infty \left(1-F\left(t\right)\right) dt \\ &= \frac{1}{\mu}\int\int 1_{0\le y\le t<\infty}\left(1-F\left(t\right)\right) dy\ dt \\ &= \frac{1}{\mu}\int 1_{0\le t<\infty} t\left(1-F\left(t\right)\right) dt = \frac{1}{\mu}\int_0^\infty tP\left(X_1>t\right) dt \\ &= \frac{1}{\mu}\mathbb{E}\int_0^\infty t1_{X_1>t}dt = \frac{1}{\mu}\mathbb{E}\left[\frac{1}{2}X_1^2\right] = \frac{1}{2\mu}\left(\sigma^2+\mu^2\right),\end{aligned}$$

from which Eqs. (10.46) and (10.47) follows. Equation (10.48) is now a simple consequence of Eqs. (10.46) and (10.47) and the fact that $\beta_t = \delta_t + \gamma_t$. ■

Example 10.40 (Earthquakes in California continued). Let us continue the notation in Example 10.38. We now want to compute the long run expected time to the next earth quake, i.e.

$$\lim_{t\to\infty}\mathbb{E}\gamma_t = \frac{1}{2\mu}\left(\sigma^2+\mu^2\right).$$

For the uniform distribution on $(0,1)$, $\mu = \frac{1}{2}$, and

$$\sigma^2 = \int_0^1 x^2 dx - \mu^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Therefore,

$$\lim_{t\to\infty}\mathbb{E}\gamma_t = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}.$$

The long run expected time between earthquakes is $\lim_{t\to\infty}\mathbb{E}\beta_t = 2/3 > 1/2 = \mathbb{E}X_1$.

Using Theorem 10.36 we can give the following improvement on the elementary renewal theorem.

Proposition 10.41. *Suppose that $f(t) = \dot{F}(t)$ exists (i.e. X_n, are continuous random variables) and suppose that $\mu = \mathbb{E}X_1$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then*

$$M\left(t\right) = \frac{t}{\mu} + \frac{\sigma^2-\mu^2}{2\mu^2} + o\left(1\right) \tag{10.50}$$

where $o(1)$ denotes a function of t which tends to zero as $t\to\infty$. This shows that we may recover μ and σ from the large t behavior of $M(t)$.

Proof. From Eq. (10.19) and Eq. (10.46) we have,

$$\left(M\left(t\right)+1\right)\mu = t + \mathbb{E}\gamma_t = t + \frac{1}{2\mu}\left(\sigma^2+\mu^2\right) + o\left(1\right).$$

Solving this identity for $M(t)$ gives Eq. (10.50) upon observing;

$$\frac{1}{2\mu^2}\left(\sigma^2+\mu^2\right) - 1 = \frac{1}{2\mu^2}\left(\sigma^2+\mu^2\right) - \frac{\mu^2}{\mu^2} = \frac{\sigma^2-\mu^2}{2\mu^2}.$$

■

Proposition 10.42. *Suppose that $f = \dot{F}(t)$ exists, $\mu = \mathbb{E}X_1$ and $\sigma^2 = \mathrm{Var}(X_1) = \mathbb{E}X_1^2 - \mu^2$, then*

$$\mathrm{Var}\left(N\left(t\right)\right) = \frac{\sigma^2}{\mu^3}t + o\left(t\right), \tag{10.51}$$

where $o(t)$ represents a function of t such that $\lim_{t\to\infty}\frac{o(t)}{t} = 0$.

Proof. Let $g(t) = \mathbb{E}\left[N^2(t)\right]$, then by the usual conditioning arguments,

$$\begin{aligned}g\left(t\right) &= \int_0^\infty \mathbb{E}\left[N^2\left(t\right)|X_1=x\right]dF\left(x\right) = \int_0^t \mathbb{E}\left[N^2\left(t\right)|X_1=x\right]dF\left(x\right) \\ &= \int_0^t \mathbb{E}\left[\left(N\left(t-x\right)+1\right)^2\right]dF\left(x\right) = \int_0^t \left(g\left(t-x\right)+2M\left(t-x\right)+1\right)dF\left(x\right) \\ &= \left(g*F+2M*F+F\right)\left(t\right).\end{aligned}$$

Since $M = F + M*F$, it follows that

$$g = g*F + 2\left(M-F\right) + F = g*F + 2M - F$$

and hence g satisfies the renewal equation with $h = 2M - F$. The solution to this equation is

$$\begin{aligned}g &= h + h*M = 2M - F + \left(2M-F\right)*M \\ &= 2M - F + 2M*M - F*M \\ &= 2M - F + 2M*M - \left(M-F\right) = M + 2M*M.\end{aligned}$$

Let us now consider,

$$\frac{1}{t}M*M\left(t\right) = \frac{1}{t}\int_0^t M\left(x\right)m\left(t-x\right)dx.$$

For t large, the contributions from the integral near $x=0$ is not relevant and so we may replace $M(x)$ by

$$M\left(x\right) = \frac{x}{\mu} + c + o\left(1\right)$$

where

$$c = \frac{1}{2\mu^2}\left(\sigma^2 - \mu^2\right). \tag{10.52}$$

Thus we have,

$$\begin{aligned}
\frac{1}{t} M * M(t) &= \frac{1}{t}\int_0^t \left(\frac{x}{\mu} + c + o(1)\right) m(t-x)\,dx \\
&= \frac{1}{t}\int_0^t \left(\frac{x}{\mu} + c + o(1)\right) m(t-x)\,dx \\
&= [c + o(1)]\frac{M(t)}{t} + \frac{1}{\mu t}\int_0^t x m(t-x)\,dx \\
&= [c + o(1)]\frac{M(t)}{t} + \frac{1}{\mu t}\int_0^t (t-x)\,dM(x) \\
&= [c + o(1)]\frac{M(t)}{t} + \frac{1}{\mu t}\int_0^t M(x)\,dx.
\end{aligned}$$

Again in the last integral, we need not worry about the contribution of the integral near zero because of the $1/t$ factor, and therefore,

$$\begin{aligned}
\frac{1}{t} M * M(t) &= [c + o(1)]\frac{M(t)}{t} + \frac{1}{\mu t}\int_0^t \left(\frac{x}{\mu} + c + o(1)\right) dx \\
&= \frac{c}{\mu} + \frac{1}{\mu t}\left(\frac{t^2}{2\mu} + ct + o(t)\right) = 2\frac{c}{\mu} + \frac{1}{2\mu^2}t + o(1).
\end{aligned}$$

Putting this all together shows,

$$\frac{g(t)}{t} = \frac{M(t)}{t} + \frac{2}{t}M * M(t) = \frac{1}{\mu} + 4\frac{c}{\mu} + \frac{1}{\mu^2}t + o(1).$$

Let us also notice that

$$\frac{1}{t}M^2(t) = \frac{1}{t}\left(\frac{1}{\mu}t + c + o(1)\right)^2 = \frac{1}{\mu^2}t + 2\frac{c}{\mu} + o(1)$$

From the previous two equations along with Eq. (10.52) for c, we find,

$$\begin{aligned}
\frac{\operatorname{Var}(N(t))}{t} &= \frac{g(t)}{t} - \frac{1}{t}M^2(t) \\
&= \frac{1}{\mu} + 2\frac{c}{\mu} + o(t) = \frac{1}{\mu} + 2\frac{c}{\mu} + o(1) \\
&= \frac{1}{\mu}\left[1 + \frac{1}{\mu^2}\left(\sigma^2 - \mu^2\right)\right] + o(1) \\
&= \frac{\sigma^2}{\mu^3} + o(1)
\end{aligned}$$

which is equivalent to Eq. (10.51). ■

10.5.1 Second Proof of Theorem 10.22

Using this result we can give another proof or the renewal central limit Theorem 10.22, namely that

$$N(t) \overset{d}{\approx} \frac{t}{\mu} + \frac{\sigma}{\mu^{3/2}}\sqrt{t}Z,$$

where Z is a standard normal random variable. To do this let $\sigma^2(t) := \operatorname{Var}(N(t)) \cong \frac{\sigma}{\mu^{3/2}}t$, and then start with the basic relationship, $N(t) \le k$ iff $W_k \le t$. This then implies,

$$\begin{aligned}
\frac{N(t) - M(t)}{\sigma(t)} \ge x \iff N(t) \ge M(t) + x\frac{\sigma}{\mu^{3/2}}\sqrt{t} =: k(t) \\
\iff W_{k(t)} \le t \iff \frac{W_k(t) - \mu k(t)}{\sigma\sqrt{k(t)}} \le \frac{t - \mu k(t)}{\sigma\sqrt{k(t)}}.
\end{aligned}$$

Notice that

$$k(t) \cong \frac{1}{\mu}t + c + x\frac{\sigma}{\mu^{3/2}}\sqrt{t} + o\left(\sqrt{t}\right) \to \infty \text{ as } t \to \infty, \tag{10.53}$$

therefore by the usual central limit theorem, $\frac{W_k(t) - \mu k(t)}{\sigma\sqrt{k(t)}}$ is close to a standard normal random variable, Z. Therefore we have

$$P\left(\frac{N(t) - M(t)}{\sigma(t)} \ge x\right) \cong P\left(Z \le \frac{t - \mu k(t)}{\sigma\sqrt{k(t)}}\right).$$

From Eq. (10.53),

$$\frac{t - \mu k(t)}{\sigma\sqrt{k(t)}} \cong \frac{-x\frac{\sigma}{\mu^{1/2}}\sqrt{t}}{\sigma\sqrt{\frac{1}{\mu}t}} = -x$$

and therefore,

$$P\left(\frac{N(t) - M(t)}{\frac{\sigma}{\mu^{3/2}}\sqrt{t}} \ge x\right) \cong P(Z \le -x) = P(Z > x).$$

wherein we have used $Z \overset{d}{=} -Z$ in the last equality. This shows that

$$\frac{N(t) - M(t)}{\frac{\sigma}{\mu^{3/2}}\sqrt{t}} \implies Z \text{ as } t \to \infty.$$

10.6 Renewal Theory Extras

You should **ignore** this section.

Example 10.43 (Another proof of Proposition 10.41). We begin by observing that

$$\begin{aligned}\left(t^k * F\right)(t) &= \int_0^t (t-x)^k \, dF(x) \\ &= (t-x)^k F(x) |_{x=0}^{x=t} + \int_0^t k (t-x)^{k-1} F(x) \, dx \\ &= \int_0^t k (t-x)^{k-1} F(x) \, dx.\end{aligned}$$

Taking $k = 0$ and $k = 1$, we find,

$$(1 * F)(t) = \int_0^t 1 dF(x) = F(t)$$

and

$$\begin{aligned}(t * F)(t) &= \int_0^t (t-x) \, dF(x) = \int_0^t F(x) \, dx = \int_0^t P(X_1 \le x) \, dx \\ &= \mathbb{E} \int 1_{X_1 \le x \le t} dx = \mathbb{E}\left[(t - X_1)_+\right].\end{aligned}$$

Hence if we let

$$g(t) = M(t) - \frac{t}{\mu} + 1,$$

then

$$\begin{aligned}g * F - g &= M * F - \frac{1}{\mu} t * F + 1 * F - \left(M - \frac{t}{\mu} + 1\right) \\ &= M - F - \frac{1}{\mu} t * F + F - \left(M - \frac{t}{\mu} + 1\right) \\ &= -\frac{1}{\mu} t * F + \frac{t}{\mu} - 1.\end{aligned}$$

Now,

$$t - t * F = t - \int_0^t (t-x) \, dF(x) = t - \int_0^t F(x) \, dx = \int_0^t (1 - F(x)) \, dx$$

and hence

$$\begin{aligned}\frac{1}{\mu} \int_0^t (1 - F(x)) \, dx - 1 &= \frac{1}{\mu} \int_0^t (1 - F(x)) \, dx - \frac{1}{\mu} \int_0^\infty (1 - F(x)) \, dx \\ &= -\frac{1}{\mu} \int_t^\infty (1 - F(x)) \, dx.\end{aligned}$$

Thus we have shown that

$$g * F - g = -\frac{1}{\mu} \int_t^\infty (1 - F(x)) \, dx$$

i.e.,

$$g = g * F + \frac{1}{\mu} \int_t^\infty (1 - F(x)) \, dx.$$

where,

$$\int_t^\infty P(X_1 > x) \, dx = \mathbb{E} \int 1_{X_1 > x \ge t} dx = \mathbb{E}(X_1 - t)_+ \to 0 \text{ as } t \to \infty.$$

Therefore by the key renewal theorem,

$$\begin{aligned}\lim_{t \to \infty} g(t) &= \frac{1}{\mu} \frac{1}{\mu} \int_0^\infty dt \int_t^\infty (1 - F(x)) \, dx = \frac{1}{\mu^2} \int \int_{0 \le t \le x < \infty} (1 - F(x)) \, dx dt \\ &= \frac{1}{\mu^2} \int_0^\infty x (1 - F(x)) \, dx = \frac{1}{\mu^2} \int_0^\infty x P(X_1 > x) \, dx \\ &= \frac{1}{\mu^2} \int_0^\infty x \mathbb{E} 1_{X_1 > x} dx = \frac{1}{2\mu^2} \mathbb{E} X_1^2 = \frac{1}{2\mu^2} \left(\sigma^2 + \mu^2\right).\end{aligned}$$

Thus we have shown

$$\lim_{t \to \infty} \left(M(t) - \frac{t}{\mu} + 1\right) = \frac{1}{2\mu^2} \left(\sigma^2 + \mu^2\right)$$

i.e.

$$\begin{aligned}\lim_{t \to \infty} \left(M(t) - \frac{t}{\mu}\right) &= \frac{1}{2\mu^2} \left(\sigma^2 + \mu^2\right) - 1 = \frac{1}{2\mu^2} \left(\sigma^2 + \mu^2\right) - \frac{\mu^2}{\mu^2} \\ &= \frac{1}{2\mu^2} \left(\sigma^2 - \mu^2\right).\end{aligned}$$

which is to say

$$M(t) = \frac{t}{\mu} + \frac{1}{2\mu^2} \left(\sigma^2 - \mu^2\right) + o(1).$$

10.6.1 Laplace transform considerations

The Laplace (and Fourier) transform is often a useful tool in renewal theory. This subsection introduces the Laplace transform in this setting.

Notation 10.44 (Laplace Transform) *If F is a (generalized) distribution function we define the Laplace transform of F (for all $\lambda > 0$ sufficiently large) by*

$$\tilde{F}(\lambda) := \int_0^\infty e^{-\lambda x} dF(x).$$

If f is a density function, we define the Laplace transform of f (for all $\lambda > 0$ sufficiently large) by

$$\tilde{f}(\lambda) := \int_0^\infty e^{-\lambda x} f(x)\, dx.$$

Fact 10.45 *Under fairly general conditions, if $\tilde{F}(\lambda) = 0$ for all large λ then $F = 0$ and $\tilde{f}(\lambda) = 0$ for all large λ then $f = 0$.*

Theorem 10.46 (Laplace Transform). *If h is a function admitting a Laplace transform and F is a (generalized) distribution and f is a density, then*

$$(h * F)^{\sim} = \tilde{h} \cdot \tilde{F} \text{ and } (h * f)^{\sim} = \tilde{h} \cdot \tilde{f}.$$

That is the Laplace transform takes convolution to multiplication (a much simpler operation).

Proof. Let us prove the first equation as the second follows by taking $F := \int_0^x f(y)\, dy$. By the definitions we have,

$$\begin{aligned}
(h * F)^{\sim}(\lambda) &= \int_0^\infty (h * F)(x)\, e^{-\lambda x} dx = \int_0^\infty \int_0^x h(x-y)\, dF(y)\, e^{-\lambda x} dx \\
&= \int\int 1_{0 \le y \le x < \infty} h(x-y)\, dF(y)\, e^{-\lambda x} dx \\
&= \int\int 1_{0 \le y \le x < \infty} h(x-y)\, e^{-\lambda x} dx dF(y) \\
&= \int\int 1_{0 \le y \le x < \infty} h(x)\, e^{-\lambda(x+y)} dx dF(y) \\
&= \int_0^\infty \left(\int_0^\infty h(x)\, e^{-\lambda x} dx\right) e^{-\lambda y} dF(y) \\
&= \int_0^\infty \tilde{h}(\lambda)\, e^{-\lambda y} dF(y) = \tilde{h}(\lambda)\, \tilde{F}(\lambda).
\end{aligned}$$

■

Proposition 10.47 (Solving Renewal Equations). *Suppose that g satisfies,*

$$g(t) = h(t) + \int_0^t g(t-x)\, dF(x), \text{ i.e. } g = h + g * F. \tag{10.54}$$

Then under "reasonable" growth restrictions on g, the unique solution to this equation is given by,

$$g(t) = h(t) + \int_0^t h(t-x)\, dM(x), \text{ i.e. } g = h + h * M. \tag{10.55}$$

Proof. We do not give the full proof here, just enough to understand where the solution is coming from. To simplify notation, let $g * F(t) := \int_0^t g(t-x)\, dF(x)$ so that Eq. (10.54) becomes,

$$g = h + g * F \tag{10.56}$$

and $F_n = \overbrace{F * F * \cdots * F}^{n \text{ times}}$. Feeding Eq. (10.56) back into itself implies,

$$g = h + (h + g * F) * F = h + h * F + g * F_2 \tag{10.57}$$

and then feeding Eq. (10.56) back into Eq. (10.57) implies,

$$\begin{aligned}
g &= h + h * F + (h + g * F) * F_2 \\
&= h + h * F + h * F_2 + g * F_3.
\end{aligned}$$

Continuing on this way shows,

$$g = h + \sum_{k=1}^{n-1} h * F_k + g * F_n. \tag{10.58}$$

The remainder term, $g * F_n$ may be written as,

$$g * F_n(t) = \int_0^t F_n(t-x)\, dg(x) = \int_0^t P(W_n \le t - x)\, dg(x)$$

where by the strong law of large numbers, $\frac{W_n}{n} \to \mu$ an $n \to \infty$, so that

$$P\left(\frac{W_n}{n} \le \frac{t-x}{n}\right) \to P(\mu \le 0) = 0.$$

On these grounds we might expect $g * F_n \to 0$ and hence may expect $g * F_n \to 0$. Thus letting $n \to \infty$ in Eq. (10.58) gives g is given by

$$g = h + \sum_{k=1}^{\infty} h * F_k = h + h * \sum_{k=1}^{\infty} F_k = h + h * M.$$

Conversely if we define g by this equation, we have

$$\begin{aligned} g * F &= h * F + h * M * F = h * F + h * (M - F) \\ &= h * M = g - h, \end{aligned}$$

so that g solves the desired renewal equation.

Another way to understand the uniqueness assertion is by making use of the Laplace transform. Taking the Laplace transform of Eq. (10.54) shows,

$$\tilde{g}(\lambda) = \tilde{h}(\lambda) + \tilde{g}(\lambda) \tilde{F}(\lambda)$$

and therefore,

$$\tilde{g}(\lambda) = \frac{\tilde{h}(\lambda)}{1 - \tilde{F}(\lambda)}$$

while taking the Laplace transform of Eq. (10.55) implies,

$$\tilde{g}(\lambda) = \tilde{h}(\lambda) + \tilde{h}(\lambda) \tilde{M}(\lambda)$$

but

$$\tilde{M}(\lambda) = \sum_{n=1}^{\infty} \tilde{F}_n(\lambda) = \sum_{n=1}^{\infty} \left[\tilde{F}(\lambda)\right]^n = \frac{\tilde{F}(\lambda)}{1 - \tilde{F}(\lambda)}$$

and therefore,

$$\tilde{g}(\lambda) = \tilde{h}(\lambda) + \tilde{h}(\lambda) \frac{\tilde{F}(\lambda)}{1 - \tilde{F}(\lambda)} = \tilde{h}(\lambda) \frac{1}{1 - \tilde{F}(\lambda)}.$$

Since both formulas give the same Laplace transform for g they must define the same function g by Fact 10.45. ■

11

What you need to know for the Final

11.1 Continuous Time Markov Chain Review

See the part of Section 8.2 pertaining to continuous time Markov chains. Besides what is there you should also know how to compute hitting probabilities and expected hitting times using first step analysis, see Examples 9.1 and 9.2 and Proposition 9.3 for the general theory. You should also be familiar with long time limiting behavior of continuos time Markov chains in Theorem 8.4.

11.2 Formula for $\mathbb{E}X^p$

It is worth remembering that if $X \geq 0$ is a random variable, then

$$\mathbb{E}X = \int_0^\infty P(X > x)\,dx = \int_0^\infty (1 - F(x))\,dx. \tag{11.1}$$

More generally, if $1 \leq p < \infty$,

$$\begin{aligned}\mathbb{E}X^p &= \mathbb{E}\int_0^X px^{p-1}dx = \mathbb{E}\int_0^\infty 1_{x<X}px^{p-1}dx \\ &= \int_0^\infty \mathbb{E}1_{x<X} \cdot px^{p-1}dx \\ &= p\int_0^\infty P(X > x)\,x^{p-1}dx. \end{aligned} \tag{11.2}$$

Taking $p = 1$ gives Eq. (11.1) and taking $p = 2$ gives,

$$\mathbb{E}X^2 = 2\int_0^\infty P(X > x)\,xdx = 2\int_0^\infty (1 - F(x))\,xdx. \tag{11.3}$$

11.3 Renewal Theory Review

11.3.1 Renewal Theory Setup

Recall the setup: $\{X_i\}_{i=1}^\infty$ i.i.d. sequence of random times, $W_0 = 0$,

$$\begin{aligned} W_n &= X_1 + \cdots + X_n, \\ N(t) &= \#\{n : W_n \leq t\} \\ M(t) &:= \mathbb{E}N(t) \ \text{– the renewal function} \\ \gamma_t &:= W_{N(t)+1} - t \ = \ \text{excess life process} \\ \delta_t &:= t - W_{N(t)} = \ \text{are or current life process} \\ \beta_t &:= \gamma_t + \delta_t \ = \ \text{total lifetime process.} \end{aligned}$$

Let $F(t) := P(X_j \leq t)$ be the cumulative distribution function for the inter-renewal times, $\{X_i\}$.

11.3.2 Renewal Theorems

We now suppose that $F(t) = \int_0^t f(x)\,dx$, i.e. the distribution of X_j is described by a probability density, f. Also let

$$\mu = \mathbb{E}X_1 \text{ and } \sigma^2 = \mathrm{Var}(X_1).$$

Here is a listing of a number of the key renewal results:

1. $M(t) = \frac{t}{\mu} + \frac{\sigma^2 - \mu^2}{2\mu^2} + o(1)$
2. $\lim_{t\to\infty} P(\gamma_t > y, \delta_t \geq x) = \frac{1}{\mu}\int_{x+y}^\infty (1 - F(w))\,dw$. In particular,
$$\lim_{t\to\infty} P(\gamma_t > x) = \lim_{t\to\infty} P(\delta_t \geq x) = \frac{1}{\mu}\int_x^\infty (1 - F(w))\,dw.$$
3. From the previous item we derived,
$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}\gamma_t &= \lim_{t\to\infty}\mathbb{E}\delta_t = \frac{1}{2\mu}\left(\sigma^2 + \mu^2\right), \text{ and} \\ \lim_{t\to\infty}\mathbb{E}\beta_t &= \frac{1}{\mu}\left(\sigma^2 + \mu^2\right).\end{aligned}$$
4. You should also be familiar with the alternating renewal theorem (see Proposition 10.34 and Example 10.35) which states
$$\lim_{t\to\infty} P(\text{in the } Y \text{ phase at time } t) = \frac{\mathbb{E}Y_1}{\mathbb{E}X_1}.$$

12

Brownian Motion

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of independent Bernoulli random variables with $P(X_j = \pm 1) = \frac{1}{2}$ and let $W_0 = 0$, $W_n = X_1 + \cdots + X_n$ be the random walk on $\mathbb{Z}$. For each $\varepsilon > 0$, we would like to consider W_n at $n = t/\varepsilon$. We can not expect $W_{t/\varepsilon}$ to have a limit as $\varepsilon \to 0$ without further scaling. To see what scaling is needed, recall that

$$\operatorname{Var}(X_1) = \mathbb{E}X_1^2 = \frac{1}{2}1^2 + \frac{1}{2}(-1)^2 = 1$$

and therefore, $\operatorname{Var}(W_n) = n$. Thus we have

$$\operatorname{Var}\left(W_{t/\varepsilon}\right) = t/\varepsilon$$

and hence to get a limit we should scale $W_{t/\varepsilon}$ by $\sqrt{\varepsilon}$. These considerations motivate the following theorem.

Theorem 12.1. *For all $\varepsilon > 0$, let $\{B_\varepsilon(t)\}_{t\geq 0}$ be the continuous time process, defined as follows:*

1. *If $t = n\varepsilon$ for some $n \in \mathbb{N}_0$, let $B_\varepsilon(n\varepsilon) := \sqrt{\varepsilon}W_n$ and*
2. *if $n\varepsilon < t < (n+1)\varepsilon$, let $B_\varepsilon(t)$ be given by*

$$\begin{aligned} B_\varepsilon(t) &= B_\varepsilon(n\varepsilon) + \frac{t-n\varepsilon}{\varepsilon}\left(B_\varepsilon((n+1)\varepsilon) - B_\varepsilon(n\varepsilon)\right) \\ &= \sqrt{\varepsilon}W_n + \frac{t-n\varepsilon}{\varepsilon}\left(\sqrt{\varepsilon}W_{n+1} - \sqrt{\varepsilon}W_n\right) \\ &= \sqrt{\varepsilon}W_n + \frac{t-n\varepsilon}{\varepsilon}\sqrt{\varepsilon}X_{n+1}, \end{aligned}$$

i.e. $B_\varepsilon(t)$ is the linear interpolation between $(n\varepsilon, \sqrt{\varepsilon}W_n)$ and $((n+1)\varepsilon, \sqrt{\varepsilon}W_{n+1})$, see Figure 12.1. Then $B_\varepsilon \implies B$ ("weak convergence") as $\varepsilon \downarrow 0$, where B is a continuous random process.

The next proposition gives some of the basic facts about Brownian motion.

Proposition 12.2. *The law of the process, B, is uniquely determined by the following properties:*

1. *$B(0) = 0$.*

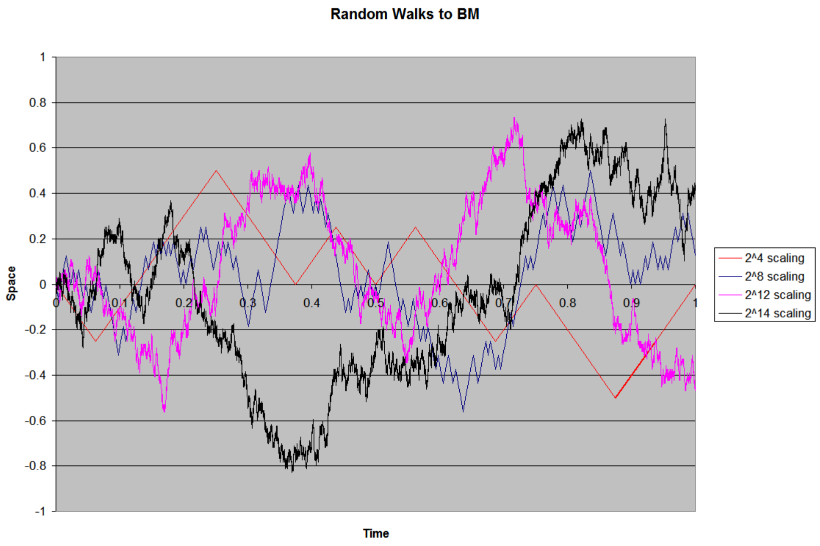


Fig. 12.1. The four graphs are constructed (in Excel) from a single realization of a random walk. Each graph corresponds to a different scaling parameter, namely, $\varepsilon \in \left\{2^{-4}, 2^{-8}, 2^{-12}, 2^{-14}\right\}$. It is clear from these pictures that $B_\varepsilon(t)$ is not converging to $B(t)$ for each realization. The convergence is only in law.

2. *For all $0 \leq s < t < \infty$, $B(t) - B(s)$ is a Gaussian random variable with variance $t - s$.*
3. *The increments of B are independent. To be more specific, if $0 = t_0 < t_1 < \cdots < t_n < \infty$, then $\{B(t_i) - B(t_{i-1})\}_{i=1}^n$ are independent Gaussian random variables.*

Proof. The first item is clear since $B_\varepsilon(0) = 0$ for all $\varepsilon > 0$. The second follows from the central limit theorem. To prove the third, suppose that $0 \leq s < t < \infty$ are rational numbers. Then for $n \in \mathbb{N}$ sufficiently large chosen so that ns and nt are integers, we have $B_{n^{-1}}(t) - B_{n^{-1}}(s)$ is independent of $\left\{B_{n^{-1}}(\sigma) : \sigma \leq s\right\}$. This independence is preserved in the limit to learn that

$B(t)-B(s)$ is independent of $\{B(\sigma):\sigma\le s\}$. The continuity of B allows us to remove the restriction on s and t being rational. ■

Definition 12.3 (Brownian motion). *__Brownian motion__ refers to any continuous process, B, satisfying the properties in Proposition 12.2.*

In what follows, N will denote a standard normal random variable which is independent of B. We will make use of the fact that $B(t)-B(s)\overset{d}{=}\sqrt{t-s}N$ for all $0\le s<t<\infty$ and that

$$\mathbb{E}\left[f\left(x+\sqrt{\tau}N\right)\right]=\int_{\mathbb{R}}f\left(x+\sqrt{\tau}y\right)\frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}}dy=\int_{\mathbb{R}}f(y)\frac{e^{-\frac{1}{2\tau}(y-x)^2}}{\sqrt{2\pi\tau}}dy. \quad (12.1)$$

To simplify (and clarify) notation we will define,

$$p_t(x,y):=\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}|y-x|^2}\text{ for all }t>0\text{ and }x,y\in\mathbb{R}.$$

Therefore Eq. (12.1) may be written as

$$P_\tau f(x):=\mathbb{E}\left[f\left(x+\sqrt{\tau}N\right)\right]=\int_{\mathbb{R}}p_\tau(x,y)f(x)\,dy. \quad (12.2)$$

Corollary 12.4. *If $0=t_0<t_1<\dots<t_n<\infty$, $\Delta_i t:=t_i-t_{i-1}$, and $J_i=(a_i,b_i)\subset\mathbb{R}$ are given bounded intervals, then*

$$\begin{aligned}&P\left(B(t_i)\in J_i\text{ for }i=1,2,\dots,n\right)\\&\quad=\int\dots\int_{J_1\times\dots\times J_n}p_{\Delta_1 t}(0,x_1)\,p_{\Delta_2 t}(x_1,x_2)\dots p_{\Delta_n t}(x_{n-1},x_n)\,dx_1\dots dx_n.\end{aligned} \quad (12.3)$$

Proof. Let $x_0:=0$. We are going to prove by induction on n that

$$\begin{aligned}&\mathbb{E}F\left(B(t_1),\dots,B(t_n)\right)\\&\quad=\int F(x_1,\dots,x_n)\,p_{\Delta_1 t}(x_0,x_1)\dots p_{\Delta_n t}(x_{n-1},x_n)\,dx_1\dots dx_n.\end{aligned} \quad (12.4)$$

Eq. (12.3) will then follow by taking $F(x_1,\dots,x_n):=1_{J_1}(x_1)\dots 1_{J_n}(x_n)$.

For $n=1$, we have, using Eq. (12.2),

$$\mathbb{E}F\left(B(t_1)\right)=\mathbb{E}F\left(\sqrt{t_1}N\right)=\int_{\mathbb{R}}p_{t_1}(0,y)\,f(y)\,dy$$

which is Eq. (12.4) with $n=1$. For the induction step we begin with the following identity,

$$\begin{aligned}\mathbb{E}F\left(B(t_1),\dots,B(t_n)\right)&=\mathbb{E}F\left(B(t_1),\dots,B(t_{n-1}),B(t_{n-1})+B(t_n)-B(t_{n-1})\right)\\&=\mathbb{E}F\left(B(t_1),\dots,B(t_{n-1}),B(t_{n-1})+\sqrt{\Delta_n t}N\right)\\&=\mathbb{E}\int_{\mathbb{R}}F\left(B(t_1),\dots,B(t_{n-1}),y\right)p_{\Delta_n t}\left(B(t_{n-1}),y\right)dy\\&=\int_{\mathbb{R}}\mathbb{E}\left[F\left(B(t_1),\dots,B(t_{n-1}),y\right)p_{\Delta_n t}\left(B(t_{n-1}),y\right)\right]dy\end{aligned} \quad (12.5)$$

wherein the second line we have again used Eq. (12.2). By the induction hypothesis,

$$\begin{aligned}&\mathbb{E}\left[F\left(B(t_1),\dots,B(t_{n-1}),y\right)p_{\Delta_n t}\left(B(t_{n-1}),y\right)\right]\\&\quad=\int F(x_1,\dots,x_{n-1},y)\,\rho(x_0,\dots,x_{n-1},y)\,dx_1\dots dx_{n-1},\end{aligned} \quad (12.6)$$

where

$$\rho(x_0,\dots,x_{n-1},y):=p_{\Delta_1 t}(x_0,x_1)\dots p_{\Delta_{n-1}t}(x_{n-2},x_{n-1})\,p_{\Delta_n t}(x_{n-1},y).$$

Combining Eqs. (12.5) and (12.6) and then replacing y by x_n verifies Eq. (12.4). ■

Theorem 12.5. *Let f be a C^2 – function which is bounded and has bounded first and second derivatives. As above, let*

$$P_t f(x):=\int_{\mathbb{R}}p_t(x,y)\,f(y)\,dy.$$

Then

$$\lim_{t\downarrow 0}P_t f(x)=f(x)\ \text{ for all }x\in\mathbb{R} \quad (12.7)$$

and

$$\frac{d}{dt}P_t f=\frac{1}{2}D^2P_t f=P_t\left(\frac{1}{2}D^2f\right). \quad (12.8)$$

Proof. From Eq. (12.2),

$$P_t f(x)=\mathbb{E}\left[f\left(x+\sqrt{t}N\right)\right]\to\mathbb{E}f(x)=f(x)\ \text{ as }t\downarrow 0$$

which proves Eq. (12.7). To prove Eq. (12.8), one show by an explicit computation that

$$\frac{d}{dt}p_t(x,y)=\frac{1}{2}\frac{\partial^2}{\partial x^2}p_t(x,y)=\frac{1}{2}\frac{\partial^2}{\partial y^2}p_t(x,y).$$

Therefore,

$$\begin{aligned}\frac{d}{dt}P_t f(x) &= \frac{d}{dt}\int_{\mathbb{R}} p_t(x,y) f(y)\,dy = \int_{\mathbb{R}} \frac{d}{dt} p_t(x,y) f(y)\,dy \\ &= \int_{\mathbb{R}} \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(x,y) f(y)\,dy = \frac{1}{2}\frac{\partial^2}{\partial x^2}\int_{\mathbb{R}} p_t(x,y) f(y)\,dy\end{aligned}$$

which combined with the integration by parts identity,

$$\begin{aligned}\int_{\mathbb{R}} \frac{\partial^2}{\partial x^2} p_t(x,y) f(y)\,dy &= \int_{\mathbb{R}} \frac{\partial^2}{\partial y^2} p_t(x,y) f(y)\,dy \\ &= \int_{\mathbb{R}} p_t(x,y)\,\frac{\partial^2}{\partial y^2} f(y)\,dy,\end{aligned}$$

verifies Eq. (12.8). ■

Remark 12.6. The last two results show that $\{B(t)\}_{t\geq 0}$ is a **Markov process**, P_t is the transition semigroup with infinitesimal generator being $Q := \frac{1}{2}D^2$, and $\{p_t(x,y)\}_{x,y\in\mathbb{R}}$ are the "matrix entries" of P_t.

12.1 Itô Calculus

Lemma 12.7. *Let N be a standard normal random variable. Then* $\operatorname{Var}(N^2) = 2$.

Proof. By integration by parts and the fact $\mathbb{E}N^2 = 1$, we find,

$$\begin{aligned}\mathbb{E}N^4 &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^4 e^{-x^2/2}dx = -\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^3 \frac{d}{dx} e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \frac{d}{dx}x^3 \cdot e^{-x^2/2}dx = 3\mathbb{E}N^2 = 3.\end{aligned}$$

Therefore,

$$\operatorname{Var}(N^2) = \mathbb{E}N^4 - \left(\mathbb{E}N^2\right)^2 = 3 - 1 = 2.$$

■

The next few results (see especially Corollary 12.9) are key ingredients in the Itô calculus and explains why it differs from ordinary calculus you learned in 20A-B.

Proposition 12.8. *Let $T > 0$ and $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0,T]$,*

$$|\Pi| := \max\{t_i - t_{i-1} : 1 \leq i \leq n\} \text{ – the mesh size of } \Pi,$$

and

$$Q_\Pi := \sum_{i=1}^n (B_{t_i} - B_{t_i})^2.$$

Then $\mathbb{E}Q_\Pi = T$ and

$$\mathbb{E}\left[(Q_\Pi - T)^2\right] = \operatorname{Var}(Q_\Pi) = 2\sum_{i=1}^n (t_i - t_{i-1})^2 \leq 2T|\Pi|.$$

Therefore,

$$\mathbb{E}\left[(Q_\Pi - T)^2\right] \to 0 \text{ as } |\Pi| \to 0.$$

Proof. We have

$$\mathbb{E}Q_\Pi = \sum_{i=1}^n \mathbb{E}(B_{t_i} - B_{t_i})^2 = \sum_{i=1}^n (t_i - t_{i-1})\,\mathbb{E}N^2 = \sum_{i=1}^n (t_i - t_{i-1}) = T.$$

Since $\left\{(B_{t_i} - B_{t_i})^2\right\}_{i=1}^n$ are independent random variables,

$$\begin{aligned}\operatorname{Var} Q_\Pi &= \sum_{i=1}^n \operatorname{Var}\left[(B_{t_i} - B_{t_i})^2\right] = \sum_{i=1}^n (t_i - t_{i-1})^2 \operatorname{Var}\left[N^2\right] \\ &= 2\sum_{i=1}^n (t_i - t_{i-1})^2 \leq 2T|\Pi|.\end{aligned}$$

■

Corollary 12.9. *If $\{\Pi_n\}_{n=1}^\infty$ is a sequence of partitions of $[0,T]$ such that $\sum_{n=1}^\infty |\Pi_n| < \infty$, then $\lim_{n\to\infty} Q_{\Pi_n} = T$ a.s. We summarize this statement as $\int_0^T dB_t^2 = T$ or even more informally as $dB_t^2 = dt$.*

Proof. By Proposition 12.8,

$$\mathbb{E}\left[\sum_{n=1}^\infty (Q_{\Pi_n} - T)^2\right] = \sum_{n=1}^\infty \mathbb{E}(Q_{\Pi_n} - T)^2 \leq 2T\sum_{n=1}^\infty |\Pi_n| < \infty$$

and hence

$$\sum_{n=1}^\infty (Q_{\Pi_n} - T)^2 < \infty \text{ a.s.}$$

which implies $\lim_{n\to\infty}(Q_{\Pi_n} - T)^2 = 0$ a.s. ■

Definition 12.10. *The Itô integral of an adapted process*[1] $\{f_t\}_{t\geq 0}$, *is defined by*

$$\int_0^T f dB = \lim_{|\Pi|\to 0} \sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right) \tag{12.9}$$

when the limit exists.

Proposition 12.11. *Keeping the notation in Definition 12.10 and further assume* $\mathbb{E}f_t^2 < \infty$ *for all* t*. Then we have,*

$$\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] = 0$$

and

$$\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]^2 = \mathbb{E}\sum_{i=1}^n f_{t_{i-1}}^2 \left(t_i - t_{i-1}\right).$$

Proof. Since $\left(B_{t_i} - B_{t_{i-1}}\right)$ is independent of $f_{t_{i-1}}$ we have,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] &= \sum_{i=1}^n \mathbb{E}f_{t_{i-1}}\mathbb{E}\left(B_{t_i} - B_{t_{i-1}}\right)\\ &= \sum_{i=1}^n \mathbb{E}f_{t_{i-1}} \cdot 0 = 0.\end{aligned}$$

For the second assertion, we write,

$$\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]^2 = \sum_{i,j=1}^n f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right).$$

If $j < i$, then $f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}}$ is independent of $\left(B_{t_i} - B_{t_{i-1}}\right)$ and therefore,

$$\begin{aligned}&\mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]\\ &\quad = \mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}}\right] \cdot \mathbb{E}\left(B_{t_i} - B_{t_{i-1}}\right) = 0.\end{aligned}$$

Similarly, if $i < j$,

$$\mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right] = 0.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]^2 &= \sum_{i,j=1}^n \mathbb{E}\left[f_{t_{j-1}} \left(B_{t_j} - B_{t_{j-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]\\ &= \sum_{i=1}^n \mathbb{E}\left[f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right) f_{t_{i-1}} \left(B_{t_i} - B_{t_{i-1}}\right)\right]\\ &= \sum_{i=1}^n \mathbb{E}\left[f_{t_{i-1}}^2 \left(B_{t_i} - B_{t_{i-1}}\right)^2\right]\\ &= \sum_{i=1}^n \mathbb{E}f_{t_{i-1}}^2 \cdot \mathbb{E}\left(B_{t_i} - B_{t_{i-1}}\right)^2\\ &= \sum_{i=1}^n \mathbb{E}f_{t_{i-1}}^2 \left(t_i - t_{i-1}\right)\\ &= \mathbb{E}\sum_{i=1}^n f_{t_{i-1}}^2 \left(t_i - t_{i-1}\right),\end{aligned}$$

wherein the fourth equality we have used $B_{t_i} - B_{t_{i-1}}$ is independent of $f_{t_{i-1}}$. ■

This proposition motivates the following theorem which will not be proved here.

Theorem 12.12. *If* $\{f_t\}_{t\geq 0}$ *is an adapted process such that* $\mathbb{E}\int_0^T f_t^2 dt < \infty$*, then the Itô integral,* $\int_0^T f dB$*, exists and satisfies,*

$$\mathbb{E}\int_0^T f dB = 0 \text{ and}$$

$$\mathbb{E}\left(\int_0^T f dB\right)^2 = \mathbb{E}\int_0^T f_t^2 dt.$$

Corollary 12.13. *In particular if* τ *is a bounded stopping time (say* $\tau \leq T < \infty$*) then*

$$\mathbb{E}\int_0^\tau f dB = 0 \text{ and}$$

$$\mathbb{E}\left(\int_0^\tau f dB\right)^2 = \mathbb{E}\int_0^\tau f_t^2 dt.$$

Proof. The point is that, by the definition of a stopping time, $1_{0\leq t\leq\tau} f_t$ is still an adapted process. Therefore we have,

---

[1] To say f is adapted means that for each $t \geq 0$, f_t should only depend on $\{B_s\}_{s\leq t}$, i.e. $f_t = F_t\left(\{B_s\}_{s\leq t}\right)$.

$$\mathbb{E}\int_0^\tau f dB = \mathbb{E}\left[\int_0^T 1_{0\le t\le\tau} f_t dB_t\right] = 0$$

and

$$\begin{aligned}\mathbb{E}\left(\int_0^\tau f dB\right)^2 &= \mathbb{E}\left[\int_0^T 1_{0\le t\le\tau} f_t dB_t\right]^2 \\ &= \mathbb{E}\left[\int_0^T \left(1_{0\le t\le\tau} f_t\right)^2 dt\right] = \mathbb{E}\left[\int_0^\tau f_t^2 dt\right].\end{aligned}$$

■

Theorem 12.14 (Itô's Lemma). *If f is a C^2 – function, then*

$$\begin{aligned}df(B) &= f'(B)\,dB + \frac{1}{2}f''(B)\,dB^2 \\ &= f'(B)\,dB + \frac{1}{2}f''(B)\,dt.\end{aligned}$$

More precisely,

$$f(B_T) = f(B_0) + \int_0^T f'(B_t)\,dB_t + \frac{1}{2}\int_0^T f''(B_t)\,dt.$$

Roughly speaking, all differentials should be expanded out to second order using the multiplication rules,

$$dB^2 = dt \text{ and } dBdt = 0 = dt^2.$$

Proof. We do not give the proof here which is based in part on Taylor's theorem to order two and Corollary 12.9. ■

12.1.1 Examples of using Itô's formula

For this subsection, let $-\infty < a < 0 < b < \infty$,

$$\begin{aligned}\tau_b &:= \inf\{t > 0 : B(t) = b\}, \\ \tau_a &:= \inf\{t > 0 : B(t) = a\},\end{aligned}$$

and $\tau := \tau_a \wedge \tau_b$, , with the convention that $\inf\emptyset = \infty$.

Now let $f(x) = (x-a)(b-x)$, see Figure 12.2 below. By Itô's lemma we have, using $f'(x) = -2x + b + a$, and $f''(x) = -2$. that

$$df(B) = f'(B)\,dB - dt$$

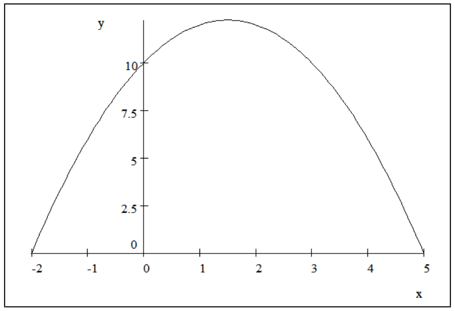


Fig. 12.2. A plot of f for $a = -2$ and $b = 5$.

or in its integrated form,

$$f(B_t) = f(0) + \int_0^t f'(B)\,db - t.$$

integrated form. Taking $t = T\wedge\tau$ in this formula and then taking expectations gives,

$$\mathbb{E}f(B_{T\wedge\tau}) = -ab - \mathbb{E}[T\wedge\tau],$$

i.e.

$$\mathbb{E}[T\wedge\tau] = -ab - \mathbb{E}f(B_{T\wedge\tau}) \le -ab.$$

By MCT we may let $T\uparrow\infty$, to discover, $\mathbb{E}\tau \le -ab < \infty$ and in particular $P(\tau<\infty) = 1$. An application of DCT now implies that $\lim_{T\uparrow\infty}\mathbb{E}f(B_{T\wedge\tau}) = \mathbb{E}f(B_\tau) = 0$ and therefore we have shown

$$\mathbb{E}[\tau_a\wedge\tau_b] = -ab.$$

This is the same formula we had for simple random walks and in fact formally follows from the random walk formula by our construction of Brownian motion as a limit of scaled random walks.

For our next application of Itô's formula, let $f(x) = x - a$. Since $f'(x) = 1$ and $f''(x) = 0$, it follows by Itô's formula that

$$df(B) = dB,$$

i.e.

$$-a + B_t = -a + \int_0^t dB.$$

Evaluating this equation at $t = \tau\wedge T$ and then taking expectations implies, $\mathbb{E}[-a + B_{T\wedge\tau}] = -a$. By the MCT we may now let $T\uparrow\infty$ to find,

$$-a = \mathbb{E}\left[-a + B_\tau\right] = 0P\left(\tau_a < \tau_b\right) + (-a+b)\,P\left(\tau_b < \tau_a\right).$$

Thus we have shown

$$P\left(\tau_b < \tau_a\right) = \frac{-a}{b-a}$$

which again should be compared with our random walk results. Since $\{\tau_b < \tau_a\} \subset \{\tau_b < \infty\}$ for all a, it follows that

$$P\left(\tau_b < \infty\right) \geq \frac{-a}{b-a} \to 1 \text{ as } a \downarrow -\infty$$

and therefore, $P\left(\tau_b < \infty\right) = 1$. This shows that Brownian motion hits every point in $\mathbb{R}$ and by the Markov property is therefore, recurrent. Again these results agree with what we found for simple random walks.

12.2 Option Pricing

In this section we are going to try to explain the Black–Scholes formula for option pricing. The following excerpt is taken from *http://en.wikipedia.org/wiki/Black-Scholes.*

> Robert C. Merton was the first to publish a paper expanding the mathematical understanding of the options pricing model and coined the term "Black-Scholes" options pricing model, by enhancing work that was published by Fischer Black and Myron Scholes. The paper was first published in 1973. The foundation for their research relied on work developed by scholars such as Louis Bachelier, A. James Boness, Sheen T. Kassouf, Edward O. Thorp, and Paul Samuelson. The fundamental insight of Black-Scholes is that the option is implicitly priced if the stock is traded.
>
> Merton and Scholes received the 1997 Nobel Prize in Economics for this and related work. Though ineligible for the prize because of his death in 1995, Black was mentioned as a contributor by the Swedish academy.

Definition 12.15. *A **European stock option at time** T **with strike price** K is a ticket that you would buy from a trader for the right to buy a particular stock at time T at a price K. If the stock price, S_T, at time T is greater that K you could then buy the stock at price K and then instantly resell it for a profit of $(S_T - K)$ dollars. If the $S_T < K$, you would not turn in your ticket but would loose whatever you paid for the ticket. So the pay off of the option at time T is $(S_T - K)_+$.*

12.2.1 The question and the general setup

Question: What should be the price (q) at time zero of such a stock option?

To answer this question, we will use a simplified version of a financial market which consists of only two assets; a no risk bond worth $\beta_t = \beta_0 e^{rt}$ (for some $r > 0$) dollars per share at time t and a risky stock worth S_t dollars per share. We are going to model S_t via a geometric "Brownian motion."

Definition 12.16 (Geometric Brownian Motion). *Let $\sigma, \mu > 0$ be given parameters. We say that the solution to the "**stochastic differential equation,**"*

$$\frac{dS_t}{S_t} = \sigma dB_t + \mu dt \tag{12.10}$$

*with S_0 being non-random is a **geometric Brownian motion**. More precisely, S_t, is a solution to*

$$S_t = S_0 + \sigma \int_0^t S dB + \mu \int_0^t S_s ds. \tag{12.11}$$

Notice that $\frac{dS}{S}$ is the relative change of S and formally, $\mathbb{E}\left(\frac{dS}{S}\right) = \mu dt$ and $\operatorname{Var}\left(\frac{dS}{S}\right) = \sigma^2 dt$. Taking expectation of Eq. (12.11) gives,

$$\mathbb{E}S_t = S_0 + \mu \int_0^t \mathbb{E}S_s ds.$$

Differentiating this equation then implies,

$$\frac{d}{dt}\mathbb{E}S_t = \mu \mathbb{E}S_t \text{ with } \mathbb{E}S_0 = S_0,$$

which yields, $\mathbb{E}S_t = S_0 e^{\mu t}$. So on average, S_t is growing or decaying exponentially depending on the sign of μ.

Proposition 12.17 (Geometric Brownian motion). *The stochastic differential Equation (12.11) has a unique solution given by*

$$S_t = S_0 \exp\left(\sigma B_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right).$$

Proof. We do not bother to give the proof of uniqueness here. To prove existence, let us look for a solution to Eq. (12.10) of the form;

$$S_t = S_0 \exp\left(aB_t + bt\right),$$

for some constants a and b. By Itô's lemma, using $\frac{d}{dx}e^x = \frac{d^2}{dx^2}e^x = e^x$ and the multiplication rules, $dB^2 = dt$ and $dt^2 = dB \cdot dt = 0$, we find that

$$\begin{aligned} dS &= S\left(adB + bdt\right) + \frac{1}{2}S\left(adB + bdt\right)^2 \\ &= S\left(adB + bdt\right) + \frac{1}{2}Sa^2dt, \end{aligned}$$

i.e.

$$\frac{dS}{S} = adB + \left(b + \frac{1}{2}a^2\right)dt.$$

Comparing this with Eq. (12.10) shows that we should take $a = \sigma$ and $b = \mu - \frac{1}{2}\sigma^2$ to get a solution. ■

Definition 12.18 (Holdings and Value Processes). *Let (a_t, b_t) be the* ***holdings process*** *which denotes the number of shares of stock and bonds respectively that are held in the portfolio at time t. The* ***value process****, V_t, of the portfolio, is*

$$V_t = a_tS_t + b_t\beta_t. \tag{12.12}$$

Suppose time is partitioned as,

$$\varPi = \{0 = t_0 < t_1 < t_2 < \cdots < t_n = T\}$$

for some time T in the future. Let us suppose that (a_t, b_t) is constant on the intervals, $[0, t_1]$, $(t_1, t_2]$, $\ldots, (t_{n-1}, t_n]$. Let us write $(a_t, b_t) = (a_{i-1}, b_{i-1})$ for $t_{i-1} < t \le t_i$, see Figure 12.3.

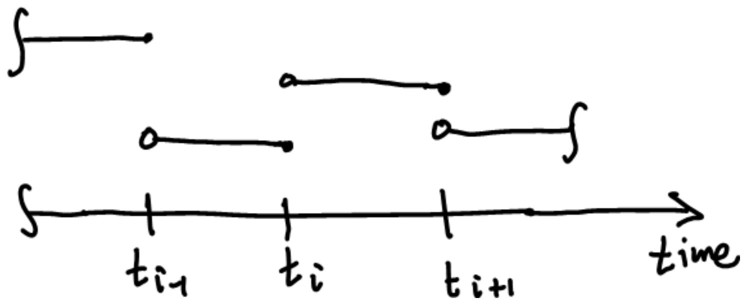


Fig. 12.3. A possible graph of either a_t or b_t.

Therefore the value of the portfolio is given by

$$V_t = a_{i-1}S_t + b_{i-1}\beta_t \text{ for } t_{i-1} < t \le t_i.$$

If our holding process is said to be **self financing** (i.e. we do not add any external money to portfolio other than what was invested, $V_0 = a_0S_0 + b_0\beta_0$, at the initial time $t = 0$), then we must have[2]

$$a_{i-1}S_{t_i} + b_{i-1}\beta_{t_i} = V_{t_i} = a_iS_{t_i} + b_i\beta_{t_i} \text{ for all } i. \tag{12.13}$$

That is to say, when we rebalance our portfolio at time t_i, we are only use the money, V_{t_i}, dollars in the portfolio at time t_i. Using Eq. (12.13) at i and $i-1$ allows us to conclude,

$$\begin{aligned} V_{t_i} - V_{t_{i-1}} &= a_{i-1}S_{t_i} + b_{i-1}\beta_{t_i} - \left(a_{i-1}S_{t_{i-1}} + b_{i-1}\beta_{t_{i-1}}\right) \\ &= a_{i-1}\left(S_{t_i} - S_{t_{i-1}}\right) + b_{i-1}\left(\beta_{t_i} - \beta_{t_{i-1}}\right) \text{ for all } i, \qquad (12.14) \end{aligned}$$

which states the change of the portfolio balance over the time interval, $(t_{i-1}, t_i]$ is due solely to the gain or loss made by the investments in the portfolio. The Equations (12.13) and (12.14) are equivalent. Summing Eq. (12.14) then gives,

$$\begin{aligned} V_{t_j} - V_0 &= \sum_{i=1}^{j} a_{i-1}\left(S_{t_i} - S_{t_{i-1}}\right) + \sum_{i=1}^{j} b_{i-1}\left(\beta_{t_i} - \beta_{t_{i-1}}\right) \qquad (12.15) \\ &= \int_0^{t_j} a_r dS_r + \int_0^{t_j} b_r d\beta_r \text{ for all } j. \qquad (12.16) \end{aligned}$$

More generally, if we throw any arbitrary point, $t \in [0, T]$, into our partition we may conclude that

$$V_t = V_0 + \int_0^t adS + \int_0^t bd\beta \text{ for all } 0 \le t \le T. \tag{12.17}$$

The interpretation of this equation is that $V_t - V_0$ is equal to the gains or losses due to trading which is given by

$$\int_0^t adS + \int_0^t bd\beta.$$

Equation (12.17) now makes sense even if we allow for continuous trading. The previous arguments show that the integrals appearing in Eq. (12.17) should be taken to be Itô – integrals as defined in Definition 12.10. Moreover, if the investor does not have psychic abilities, we should assume that holding process is adapted.

---

[2] Equation (12.13) may be written as

$$\left(a_i - a_{i-1}\right)S_{t_i} + \left(b_i - b_{i-1}\right)\beta_{t_i} = 0.$$

This explains why the continuum limit of this equation is not $S_tda_t + \beta_tdb_t = 0$ but rather must be interpreted as $S_{t+dt}da_t + \beta_{t+dt}db_t = 0$. It is also useful to observe that

$$\begin{aligned} d\left(XY\right)_t &= X_{t+dt}Y_{t+dt} - X_tY_t \\ &= \left(X_{t+dt} - X_t\right)Y_{t+dt} + X_t\left(Y_{t+dt} - Y_t\right), \end{aligned}$$

and hence there is not quadratic differential term when $d\left(XY\right)$ is written out this way.

12.2.2 Pricing the Option

Now that we have set the stage we can now try to price the option. (We will closely follow [1, p. 255-264.] here.) The guiding principle is:

Fundamental Principle: The price of the option, $q := f(S_0, T, K, r)$, should be equal to the amount of money, V_0, that an investor would have to put into the bond-stock market at time $t = 0$ so as there exists a self-financing holding process (a_t, b_t), such that

$$V_T = a_T S_T + b_T \beta_T = (S_T - K)_+ .$$

Remark 12.19 (Free Money). If we price the option higher than V_0, i.e. $q > V_0$, we could make risk free money by selling one of these options at q dollars, investing $V_0 < q$ of this money using the holding process (a_t, b_t) to cover the payoff at time T and then pocket the different, $q - V_0$.

If the price of the option was less than V_0, i.e. $q < V_0$, the investor should buy the option and then pursue the trading strategy, $(-a, -b)$. At time zero the investor has invested $q + (-a_0 S_0 - b_0 \beta_0) = q - V_0 < 0$ dollars, i.e. he is holding $V_0 - q$ dollars in hand at time $t = 0$. The value of his portfolio at time T is now $-V_T = -(S_T - K)_+$. If $S_T > K$, the investor then exercises her option to pay off the debt she as accrued in the portfolio and if $S_T \le K$, she does nothing since his portfolio is worth zero dollars. Either way, she still has the $V_0 - q$ dollars in hand from the start of the transactions at $t = 0$.

If we have such a self-financing holding process (a_t, b_t), then $\{(a_s, b_s)\}_{t \le s \le T}$ is a self-financing holding process on $[t, T]$ such that $V_T = a_T S_T + b_T \beta_T = (S_T - K)_+$, therefore if the stock price is S_t at time t, the option price at this time, $f(S_t, T - t, K)$, should be V_t, i.e. we have

$$V_t = f(S_t, T - t, K) . \tag{12.18}$$

By Itô's lemma (dropping K from the notation),

$$\begin{aligned} dV_t &= f_x(S_t, T-t)\, dS_t + \frac{1}{2} f_{xx}(S_t, T-t)\, dS_t^2 - f_t(S_t, T-t)\, dt \\ &= f_x(S_t, T-t)\, S_t (\sigma dB_t + \mu dt) + \left[\frac{1}{2} f_{xx}(S_t, T-t)\, S_t^2 \sigma^2 - f_t(S_t, T-t)\right] dt \\ &= f_x(S_t, T-t)\, S_t \sigma dB_t \\ &\quad + \left[f_x(S_t, T-t)\, S_t \mu + \frac{1}{2} f_{xx}(S_t, T-t)\, S_t^2 \sigma^2 - f_t(S_t, T-t)\right] dt \end{aligned}$$

On the other hand from Eqs. (12.17) and (12.10), we know that

$$\begin{aligned} dV_t &= a_t dS + b_t \beta_0 r e^{rt} dt \\ &= a_t S_t (\sigma dB_t + \mu dt) + b_t \beta_0 r e^{rt} dt \\ &= a_t S_t \sigma dB_t + \left[a_t S_t \mu + b_t \beta_0 r e^{rt}\right] dt. \end{aligned}$$

Comparing these two equations implies,

$$a_t = f_x(S_t, T - t) \tag{12.19}$$

and

$$\begin{aligned} & a_t S_t \mu + b_t \beta_0 r e^{rt} \\ & \quad = f_x(S_t, T-t)\, S_t \mu + \frac{1}{2} f_{xx}(S_t, T-t)\, S_t^2 \sigma^2 - f_t(S_t, T-t) . \end{aligned} \tag{12.20}$$

Using Eq. (12.19) and

$$\begin{aligned} f(S_t, T-t) = V_t &= a_t S_t + b_t \beta_0 e^{rt} \\ &= f_x(S_t, T-t)\, S_t + b_t \beta_0 e^{rt} \end{aligned}$$

in Eq. (12.20) allows us to conclude,

$$\begin{aligned} \frac{1}{2} f_{xx}(S_t, T-t)\, S_t^2 \sigma^2 - f_t(S_t, T-t) &= r b_t \beta_0 e^{rt} \\ &= r f(S_t, T-t) - r f_x(S_t, T-t)\, S_t. \end{aligned}$$

Thus we see that the unknown function f should solve the partial differential equation,

$$\begin{gathered} \frac{1}{2} \sigma^2 x^2 f_{xx}(x, T-t) - f_t(x, T-t) = r f(x, T-t) - r x f_x(x, T-t) \\ \text{with } f(x, 0) = (x - K)_+ , \end{gathered}$$

i.e.

$$f_t(x, T-t) = \frac{1}{2} \sigma^2 x^2 f_{xx}(x, T-t) + r x f_x(x, T-t) - r f(x, T-t) \tag{12.21}$$

$$\text{with } f(x, 0) = (x - K)_+ . \tag{12.22}$$

Fact 12.20 *Let N be a standard normal random variable and $\Phi(x) := P(N \le x)$. The solution to Eqs. (12.21) and (12.22) is given by;*

$$f(x, t) = x \Phi(g(x, t)) - K e^{-rt} \Phi(h(x, t)) , \tag{12.23}$$

where,

$$\begin{aligned} g(x, t) &= \frac{\ln(x/K) + \left(r + \frac{1}{2}\sigma^2\right) t}{\sigma \sqrt{t}}, \\ h(x, t) &= g(x, t) - \sigma \sqrt{t}. \end{aligned}$$

Theorem 12.21 (Option Price). *Given the above setup, the "rational" price" of the European call option is $q = f(S_0, T)$ where f is given as in Eq. (12.23).*

References

1. K. L. Chung and R. J. Williams, *Introduction to stochastic integration*, second ed., Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 1990. MR MR1102676 (92d:60057)
2. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
3. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
4. Gregory F. Lawler, *Introduction to stochastic processes*, Chapman & Hall Probability Series, Chapman & Hall, New York, 1995. MR MR1372946 (97a:60001)
5. J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144)
6. Philip Protter, *Stochastic integration and differential equations*, Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 1990, A new approach. MR MR1037262 (91i:60148)
7. Sheldon M. Ross, *Stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1983, Lectures in Mathematics, 14. MR MR683455 (84m:60001)