Bruce K. Driver

Math 180C (Introduction to Probability) Notes

May 2, 2008 *File:180Notes.tex*

Contents

Part

Math 180C

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
$$T_b := \begin{cases} \min\{n : X_n = b\} & \text{if } \{X_n\} \text{ hits } b \\ \infty & \text{otherwise} \end{cases}$$
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:

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

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, $h_5 = P_5 (X_n \text{ hits } \{3, 4\})$ and $h_6 = P_6 (X_n \text{ hits } \{3, 4\})$.

Exercise 0.9. Find $\lim_{n\to\infty} P_6 (X_n = j)$ 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. $P (T_k > t) = e^{-q_k t}$ for all $t \geq 0$. Show that $T := \min (T_1, T_2, \ldots, T_n)$ 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, $P (T_k = T_j) = 0$ 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

$$P (T_1 < \min (T_2, \ldots, T_n)).$$

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

Exercise 0.12. Consider the "pure birth" process with constant rates, $\lambda > 0$. In this case $S = \{0, 1, 2, \ldots\}$ and if $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ 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

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 $\{A_j, A_j^c\}_{j \in J}$.*

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 $\{A_1, \ldots, A_n\} \subset \{A_j\}_{j \in J}$ we must show that

$$P(A_1^{\varepsilon_1} \cap \cdots \cap A_n^{\varepsilon_n}) = P(A_1^{\varepsilon_1}) \ldots P(A_n^{\varepsilon_n})$$

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

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

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

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

∎

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 $P(X_n = H \text{ i.o.}) = 1$.

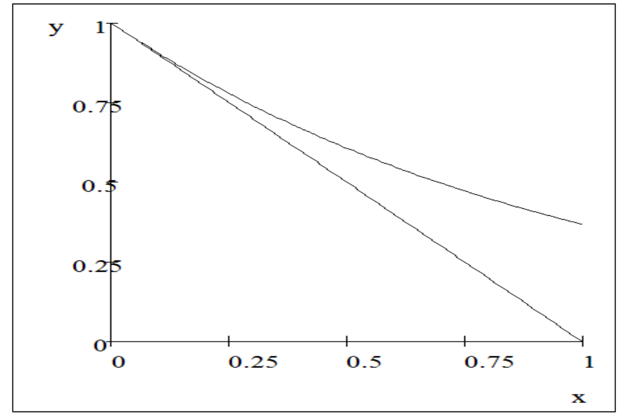


Fig. 1.1. Comparing e^{-x} and $1 - x$.

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 $(s_1, \dots, s_N)$ 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 $Y_k := \left(X_{kN+1}, \dots, X_{(k+1)N}\right)$. We then have

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

Therefore,

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

and so by the second Borel-Cantelli lemma,

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

1.2 Independent Random Variables

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

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

for all possible choices of $\{j_1, \dots, j_n\} \subset J$ 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*

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

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

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

Therefore,

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

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

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

while

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

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*

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

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*

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

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

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

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

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

The marginal density associated to Y is then given by

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

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

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

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).

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

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

Differentiating this equation in s implies,

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

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,

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

Hence it follows that

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

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 [3].

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

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

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

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

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

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

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,

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

Therefore,

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

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

2.3 Gamma Distributions

Lemma 2.7. *Suppose that $\{S_j\}_{j=1}^n$ are independent exponential random variables with parameter, θ. and $W_n = S_1 + \cdots + S_n$. 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 $W_k := S_1 + \cdots + S_k$. We then have,

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

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,

$$\begin{aligned}P(W_2 \le t) &= \int_0^t \left(1 - e^{-\theta(t-s)}\right) \theta e^{-\theta s} ds = \theta \int_0^t \left(e^{-\theta s} - e^{-\theta t}\right) ds \\ &= 1 - e^{-\theta t} - \theta t e^{-\theta t}.\end{aligned}$$

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

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

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,

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

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

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

■

2.4 Beta Distribution

Lemma 2.8. *Let*

$$B(x,y) := \int_0^1 t^{x-1} (1-t)^{y-1} dt \text{ for } \operatorname{Re} x, \operatorname{Re} y > 0. \tag{2.8}$$

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

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

■

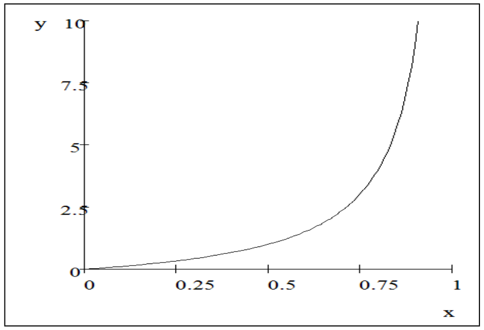


Fig. 2.1. Plot of $t/(1-t)$.

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

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

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

$$P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} & \\ & \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \end{matrix}$$

has the jump graph given by 3.1.

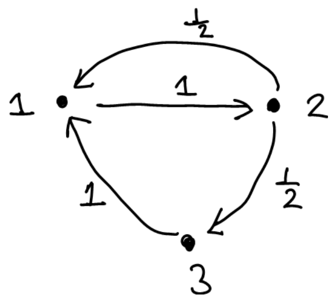


Fig. 3.1. A simple jump diagram.

Example 3.7. The transition matrix,

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

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

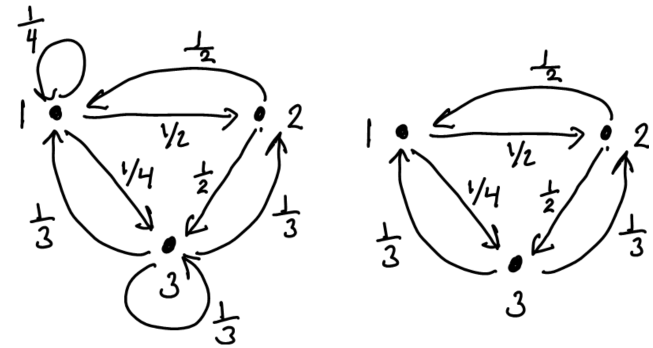


Fig. 3.2. The above diagrams contain the same information. In the one on the right we have dropped the jumps from a site back to itself since these can be deduced by conservation of probability.

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

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*

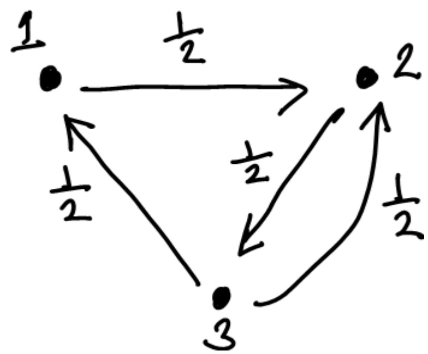
$$P_\pi(X_n = x) = \pi(x) \text{ for all } x \in S \text{ and } n \in \mathbb{N}.$$

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,

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

with jump diagram given in Figure 3.10. We have



$$P^2 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac14 & \frac12 & \frac14 \\ \frac14 & \frac12 & \frac14 \\ \frac14 & \frac12 & \frac14 \end{bmatrix}$$

and

$$P^3 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}^3 = \begin{bmatrix} \frac14 & \frac12 & \frac14 \\ \frac14 & \frac12 & \frac14 \\ \frac14 & \frac12 & \frac14 \end{bmatrix}.$$

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 $\pi_2 = \left[\frac14\ \frac12\ \frac14\right]$ which is the first row of P^2. It turns out in this case that this is the invariant distribution. Formally,

$$\left[\tfrac14\ \tfrac12\ \tfrac14\right] \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} = \left[\tfrac14\ \tfrac12\ \tfrac14\right].$$

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,

$$\pi = \pi P^2 = \left[\pi_1\ \pi_2\ \pi_3\right] \begin{bmatrix} \frac14 & \frac12 & \frac14 \\ \frac14 & \frac12 & \frac14 \\ \frac14 & \frac12 & \frac14 \end{bmatrix} = \left[\tfrac14\ \tfrac12\ \tfrac14\right].$$

Hence $\left[\frac14\ \frac12\ \frac14\right]$ 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

$$(p_{ij}) = \begin{array}{c} \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} 0 \\ 1 \\ 2 \end{matrix}$$

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

$$(p_{ij}) = \begin{array}{c} \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix}.$$

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

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

$$= \begin{bmatrix} 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \\ 0.125 & 0.375 & 0.375 & 0.125 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \pi \\ \pi \end{bmatrix}.$$

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

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

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

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

Following the algorithm in Remark 3.12 leads to:

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

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

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*

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

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

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,

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

and

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

Combining the last two equations implies,

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

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

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

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 $P_\pi(\cdot|X_n=x)$, $\{X_{n+k}\}_{k\ge0}$ is independent of $\{X_0,\dots,X_n\}$ and $\{X_{n+k}\}_{k\ge0}$ has the same distribution as $\{X_k\}_{k=0}^{\infty}$ under P_x.*

Proof. According to Eq. (3.13),

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

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

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

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 $\{X_{n+k}\}_{k\ge0}$ 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 $\{X_{n+k}\}_{k\ge0}$ 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* $\mathbf{g} := (g(x))_{x\in A}$ *and* $\mathbf{h} := (h(y))_{y\in B}$ *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* $y \in B$,

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

Taking $g \equiv \mathbf{1}$ *(where* $\mathbf{1}(x) = 1$ *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 $g(x') = \delta_y(x')$ *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

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

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

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

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

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

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

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

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 $w(x) := \mathbb{E}_x\left[h(X_T)\right]$. Since X_T is the location of where $\{X_n\}_{n=0}^\infty$ first hits B if we are given $X_0 \in A$, then X_T is also the location where the sequence, $\{X_n\}_{n=1}^\infty$, 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,

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

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,

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

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

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

■

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

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

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 [3].

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,

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

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

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

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

$$h_1 = \frac{1}{3}h_2 + \frac{1}{3}h_3 + \frac{1}{3}h_4 = \frac{1}{3}h_2 + \frac{1}{3}$$
$$h_2 = \frac{3}{4}h_1 + \frac{1}{8}h_2 + \frac{1}{8}h_3 = \frac{3}{4}h_1 + \frac{1}{8}h_2 + \frac{1}{8}$$

i.e.

$$h_1 = \frac{1}{3}h_2 + \frac{1}{3}$$
$$h_2 = \frac{3}{4}h_1 + \frac{1}{8}h_2 + \frac{1}{8}$$

which have solutions,

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

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

$$h_1 = \frac{1}{3}h_2 + \frac{1}{3}$$
$$h_2 = \frac{3}{4}h_1 + \frac{1}{8}h_2$$

which has solutions,

$$P_1\,(X_n \text{ hits } 4) = h_1 = \frac{7}{15} \text{ and}$$
$$P_2\,(X_n \text{ hits } 4) = h_2 = \frac{2}{5}.$$

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

$$P_1\,(X_n \text{ hits } 3) + P_1\,(X_n \text{ hits } 4) = 1 \text{ and}$$
$$P_2\,(X_n \text{ hits } 3) + P_2\,(X_n \text{ hits } 4) = 1.$$

The output of one simulation is in Figure 3.3 below.

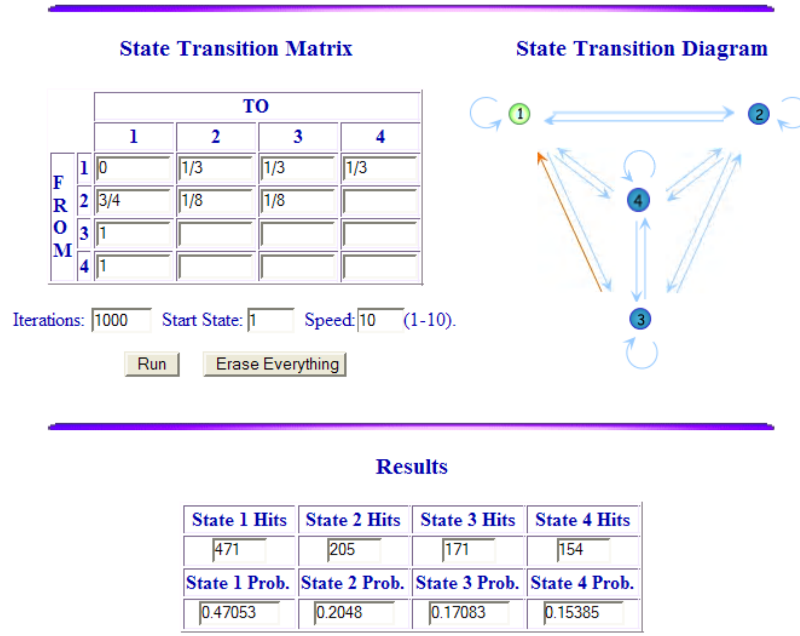


Fig. 3.3. In this run, rather than making sites 3 and 4 absorbing, we have made them transition back to 1. I claim now to get an approximate value for $P_1\,(X_n \text{ hits } 3)$ we should compute: (State 3 Hits)/(State 3 Hits + State 4 Hits). In this example we will get $171/(171+154) = 0.526\,15$ 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

$$\left[\begin{matrix} 1 & 2 & 3(\text{food}) \\ 4 & 5 & 6 \\ 7(\text{Shock}) & & \end{matrix}\right]$$

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,

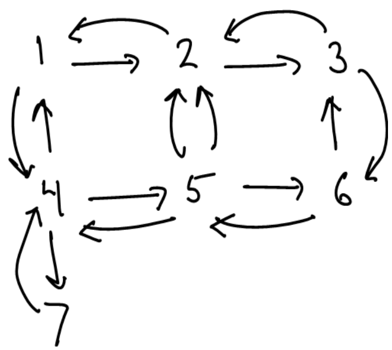


Fig. 3.4. Rat in a maze.

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

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

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

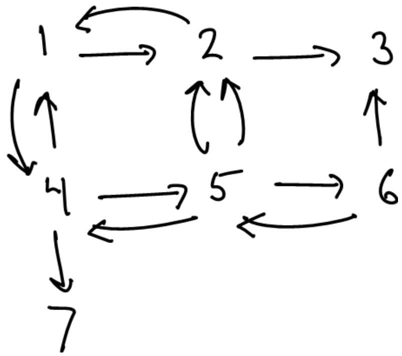
$$= \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \left[\begin{array}{cc} \frac{7}{12} & \frac{5}{12} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{5}{12} & \frac{7}{12} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{5}{6} & \frac{1}{6} \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\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

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

Let us now also work out the hitting probabilities,

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

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,



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

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

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

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:

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

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

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

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

[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.

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

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 $E_2T = E_4T = 5$.

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, $i = x_0, x_1, \ldots, x_n = j \in S$ such that $p(x_0, x_1)\, p(x_1, x_2) \ldots p(x_{n-1}, x_n) > 0$.*

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 4.45 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 $i = x_0, x_1, \ldots, x_n = j$ with all of the $\{x_j\}_{j=0}^n$ being distinct such that

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

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

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

$X_m \notin C$ 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.*

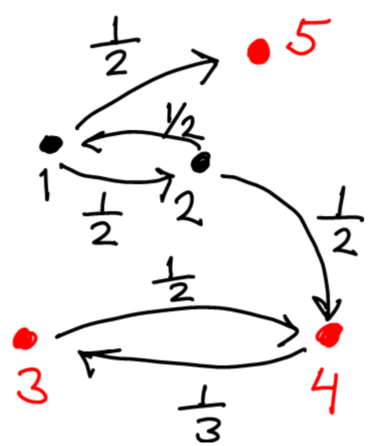
3. C is recurrent.

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 4.41.

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 4.42 below for the formal proof based on the strong Markov property in Corollary 4.41. ■

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),

$$\mathbb{E}_\nu M_j = \mathbb{E}_\nu \sum_{n\geq 1} 1_{X_n = j} = \sum_{n \geq 1} \mathbb{E}_\nu 1_{X_n = j}$$

$$= \sum_{n\geq 1} P_\nu(X_n = j) = \sum_{n=1}^{\infty} \sum_{j \in S} \nu(j) P_{jj}^n$$

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 $\sum_{n \geq 1} P_{jj}^n = \infty \implies \sum_{n \geq 1} P_{ii}^n = \infty$. Similarly $\sum_{n \geq 1} P_{ii}^n = \infty \implies \sum_{n \geq 1} P_{jj}^n = \infty$. 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 $P_i(R_j < \infty) = P_i(\{X_n\}_{n \geq 1} \text{ hits } j)$.)

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

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

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

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

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 $\sum_{j\in S} \pi(j) \le 1$. 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

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

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 4.44 and Propositions 4.45 and 4.46 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

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

as in Theorem 4.21. Since C is closed,

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

1and therefore,

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

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 4.41) 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 4.41), 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

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

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 4.47 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 [3, Theorem 1.8.3] or Kallenberg [2, 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 $S = \{1, 2, \ldots, n\}$ 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 $\mathbf{1} := \left[1\ 1\ \ldots\ 1\right]^{\mathrm{tr}}$, 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 $\nu_i \geq 0$ for all i – but we don't. So let $\pi_i := |\nu_i|$ and observe that

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

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:

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

and therefore,

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

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

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

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

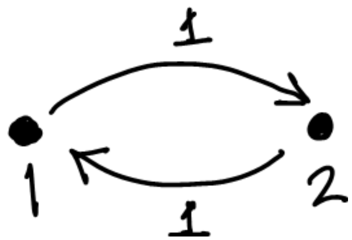
$$(\mathbb{E}_i R_j)_{i=1}^{n} = \left(I - P^{(j)}\right)^{-1} \mathbf{1}, \tag{4.25}$$

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

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

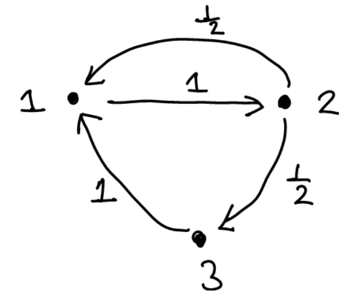


Fig. 4.1. A simple jump diagram.

has the jump graph given by 4.1. Notice that $P_{11}^2 > 0$ and $P_{11}^3 > 0$ 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,

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

is represented by the jump diagram in Figure 4.2. This chain is aperiodic. We

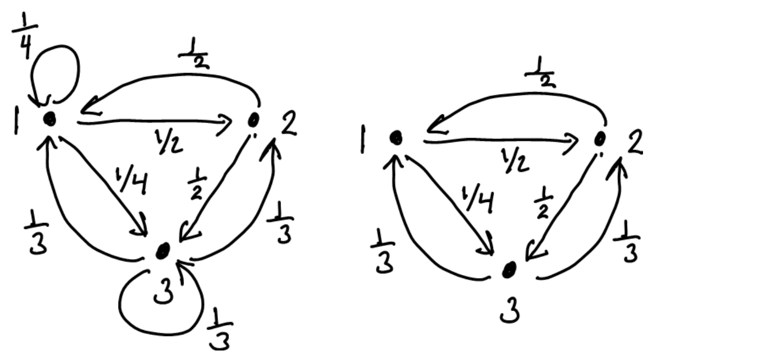


Fig. 4.2. The above diagrams contain the same information. In the one on the right we have dropped the jumps from a site back to itself since these can be deduced by conservation of probability.

find the invariant distribution as,

$$\mathrm{Nul}\,(P-I)^{\mathrm{tr}} = \mathrm{Nul}\left(\begin{bmatrix}1/4&1/2&1/4\\1/2&0&1/2\\1/3&1/3&1/3\end{bmatrix} - \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)^{\mathrm{tr}}$$

$$= \mathrm{Nul}\left(\begin{bmatrix}-\frac{3}{4}&\frac{1}{2}&\frac{1}{3}\\ \frac{1}{2}&-1&\frac{1}{3}\\ \frac{1}{4}&\frac{1}{2}&-\frac{2}{3}\end{bmatrix}\right) = \mathbb{R}\begin{bmatrix}1\\ \frac{5}{6}\\ 1\end{bmatrix} = \mathbb{R}\begin{bmatrix}6\\5\\6\end{bmatrix}$$

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

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,

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

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

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

$$= \begin{bmatrix}\frac{50}{17}\\ \frac{52}{17}\\ 4\\ \frac{30}{17}\end{bmatrix}$$

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

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

$$= \begin{bmatrix}\frac{54}{17}\\ 4\\ \frac{44}{17}\\ \frac{50}{17}\end{bmatrix}$$

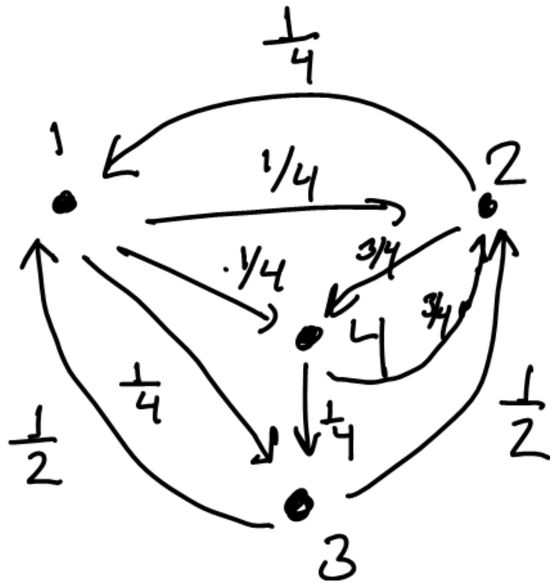


Fig. 4.3. The jump diagram for P.

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,

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

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,

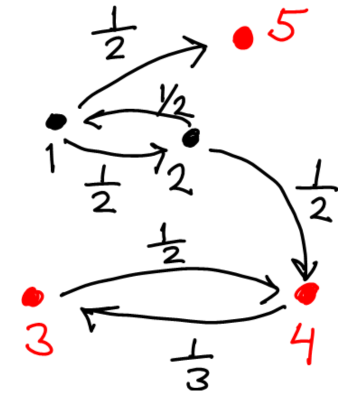


Fig. 4.4. The jump diagram for P above.

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

We do this in the usual way, namely

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

so that $\pi'_{\{3,4\}} = \frac{1}{5}\begin{bmatrix} 2 & 3 \end{bmatrix}$.

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

$$\pi_{\{3,4\}} = \begin{bmatrix} 0 & 0 & 2/5 & 3/5 & 0 \end{bmatrix}$$
$$\pi_{\{5\}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

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

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

and

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

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 $h_1 := P_1(X_n \text{ hits } 5)$. To do this let $h_i = P_i(X_n \text{ hits } 5)$ 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

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

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

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

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

[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 $h_3 = h_4 = 0$.

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

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

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

and

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

These answers already compare well with

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

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, $\tau = \tau(X_0, X_1, \dots)$. 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 $(X_0, \dots, X_n)$. Thus for each $n \in \mathbb{N}_0$ there should exist a function, σ_n such that $1_{\tau=n} = \sigma_n(X_0, \dots, X_n)$. In other words, the event $\{\tau = n\}$ may be described using only $(X_0, \dots, X_n)$ for all $n \in \mathbb{N}$.*

Remark 4.36. If τ is an $\{X_n\}_{n=0}^\infty$ - stopping time then

$$1_{\tau \ge n} = 1 - 1_{\tau < n} = 1 - \sum_{k<n} \sigma_k(X_0, \dots, X_k) =: u_n(X_0, \dots, X_{n-1}).$$

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 [4, p. 59-60].

Theorem 4.37 (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,

$$\begin{aligned}
\mathbb{E}\left[\sum_{n=0}^{\tau} f(X_n)\right] &= \mathbb{E}\left[\sum_{n=0}^{\infty} f(X_n) 1_{n\leq\tau}\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[f(X_n) 1_{n\leq\tau}\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[f(X_n) u_n(X_0, \dots, X_{n-1})\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[f(X_n)\right] \cdot \mathbb{E}\left[u_n(X_0, \dots, X_{n-1})\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[f(X_n)\right] \cdot \mathbb{E}\left[1_{n\leq\tau}\right] = \mathbb{E}f(X_0) \sum_{n=0}^{\infty} \mathbb{E}\left[1_{n\leq\tau}\right] \\
&= \mathbb{E}f(X_0) \cdot \mathbb{E}\left[\sum_{n=0}^{\infty} 1_{n\leq\tau}\right] = \mathbb{E}f(X_0) \cdot \mathbb{E}\tau.
\end{aligned}$$

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 4.38. 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 4.39 (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

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

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

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

Here is the analogue of

Theorem 4.40 (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*

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

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 4.41. *Let τ be a stopping time, $x \in S$ and π be any probability on S. Then relative to $P_\pi\left(\cdot|\tau<\infty,\ X_\tau = x\right)$, $\{X_{\tau+k}\}_{k\geq 0}$ 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),

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

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

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

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

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

This shows that $\{X_{\tau+k}\}_{k\geq 0}$ under $P_\pi\left(\cdot|\tau<\infty,\ X_\tau = x\right)$ 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 4.42. *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.

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

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],

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

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

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

Since, on $R_j < \infty$, $X_{R_j} = j$, it follows by the strong Markov property (Corollary 4.41) that;

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

By the last two displayed equations,

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

---

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

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

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 4.43. *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 4.41 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. ■

Theorem 4.44. *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 4.43, the sequence $\{\tau_n\}_{n\geq 2}$ is i.i.d. relative to P_i and $\mathbb{E}_i\tau_n = \mathbb{E}_j\tau_j = \mathbb{E}_j R_j$ for all $i \in S$. We may now use the strong law of large numbers (Theorem 1.14) to conclude that

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

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

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

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 $R_j^{(\nu_N)} \le N \le R_j^{(\nu_N+1)}$. Putting these facts together along with Eq. (4.41) shows that

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

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

Proposition 4.45. *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*

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

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

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 4.44, satisfies,

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

Observe that,

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

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

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

and

$$\sum_{i\in S}\pi_i P_{ij} = \sum_{i\in S}\lim_{n\to\infty} T_{li}^n P_{ij} \le \liminf_{n\to\infty}\sum_{i\in S} T_{li}^n P_{ij} = \liminf_{n\to\infty} T_{lj}^{n+1} = \pi_j$$

where $l \in S$ is arbitrary. Thus

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

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 4.46. *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 4.45, 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 4.47 (A number theory lemma). *Suppose that 1 is the greatest common denominator of a set of positive integers, $\Gamma := \{n_1, \ldots, n_k\}$. Then there exists $N \in \mathbb{N}$ such that the set,*

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

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

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

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

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 $m \geq 0$ and $1 \leq r < k$. 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 $n \geq N$, then $n - a^2 = ma + r$ for some $m \geq 0$ and $0 \leq r < a$. Therefore,

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

■

5

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.

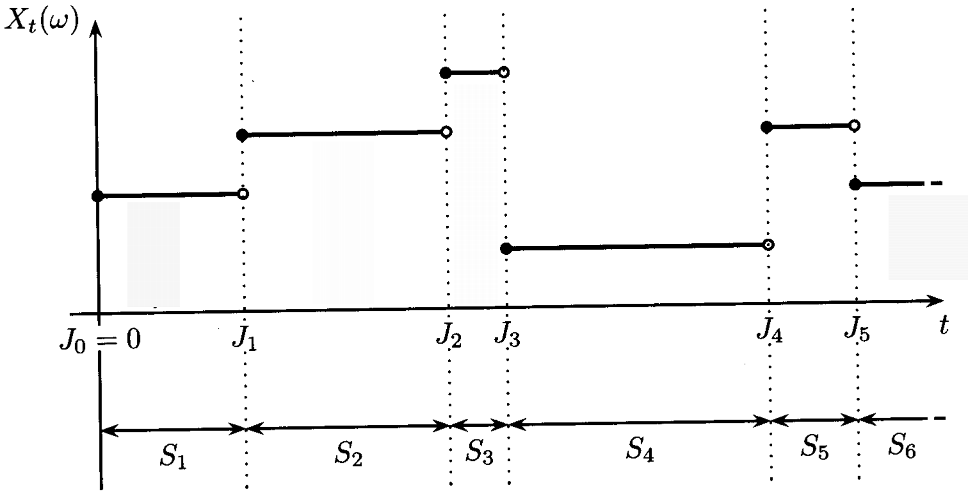


Fig. 5.1. Typical sample paths of a continuous time Markov chain in a discrete state space.

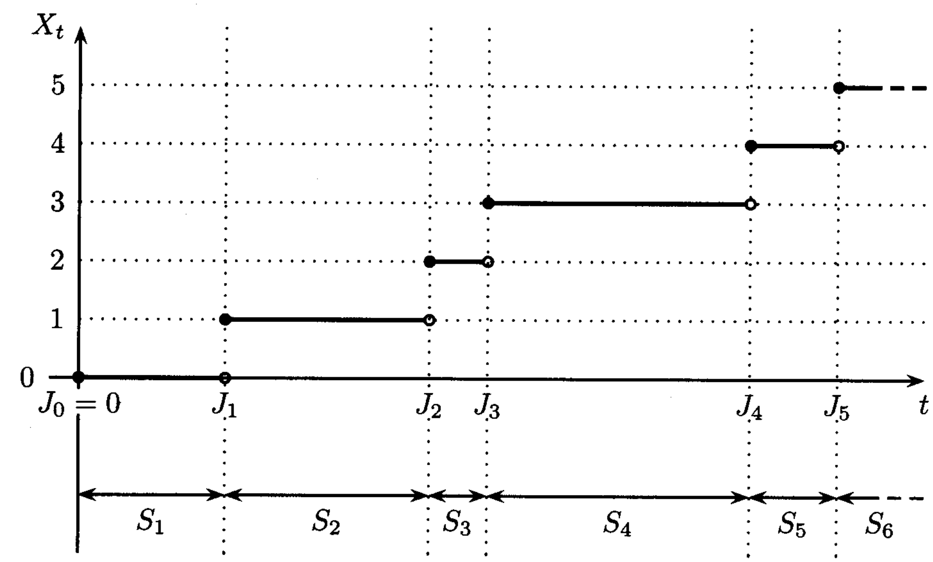


Fig. 5.2. A sample path of a birth process. Here the state space is $\{0, 1, 2, \dots\}$ to be thought of the possible population size.

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.

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 $t\geq s$ 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 $\{T\leq t\}$ is determined from the knowledge of $\{X(s):0\leq s\leq t\}$. Alternatively put, for each $t\geq 0$, 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 $\{X(t)\}_{t\geq 0}$ is a Markov chain, T is a stopping time, and $j\in S$, then, conditioned on $\{T<\infty \text{ and } X_T=j\}$,*

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

are $\{X(t+T):t\geq 0\}$ has the same distribution as $\{X(t)\}_{t\geq 0}$ 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*

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

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 $t\geq 0$, $P(t)$ 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 $\{P(t)\}_{t\geq 0}$ 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. $P(t)\neq I$ 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 $\lambda>0$,

$$P(t)=e^{-\lambda t}\begin{array}{c}\begin{matrix}0&1&2&3&4&5&6&\dots\end{matrix}\\\begin{bmatrix}1&\lambda t&\frac{(\lambda t)^2}{2!}&\frac{(\lambda t)^3}{3!}&\frac{(\lambda t)^4}{4!}&\frac{(\lambda t)^5}{5!}&\dots\\0&1&\lambda t&\frac{(\lambda t)^2}{2!}&\frac{(\lambda t)^3}{3!}&\frac{(\lambda t)^4}{4!}&\dots\\0&0&1&\lambda t&\frac{(\lambda t)^2}{2!}&\frac{(\lambda t)^3}{3!}&\dots\\0&0&0&1&\lambda t&\frac{(\lambda t)^2}{2!}&\dots\\\vdots&\vdots&\vdots&\ddots&\ddots&\ddots&\ddots\end{bmatrix}\end{array}\begin{matrix}0\\1\\2\\3\\\vdots\end{matrix}.$$

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

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

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 $k = 0, 1, 2, \dots, n$, then

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

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

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

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:

$$Q_{ij} \le 0 \text{ for all } i \ne j, \text{ and} \qquad (5.8)$$

$$\sum_j Q_{ij} = 0, \text{ i.e. } -Q_{ii} = \sum_{j \ne i} Q_{ij} \text{ for all } i. \qquad (5.9)$$

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 $\#(S) = \infty$.

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

$$\begin{aligned} \frac{d^2}{dt^2} e^{tC} &= C^2 e^{tC} = -e^{tC} \text{ with} \\ e^{0C} &= I \text{ and } \frac{d}{dt} e^{tC}|_{t=0} = C. \end{aligned}$$

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,*
$$\dot{P}(t) = P(t)Q \quad \text{(Kolmogorov's forward Eq.)}$$
and
$$\dot{P}(t) = QP(t) \quad \text{(Kolmogorov's backwards Eq.)}$$
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,

$$\dot{\pi}(t) = \lambda e^{\lambda t}\pi(0) + \lambda e^{\lambda t}\int_0^t e^{-\lambda s} h(s)\, ds + e^{\lambda t} e^{-\lambda t} h(t)$$
$$= \lambda\pi(t) + h(t).$$

■

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

$$Q = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

with rate diagram being given in Figure 6.1. We now find e^{tQ} using Theorem

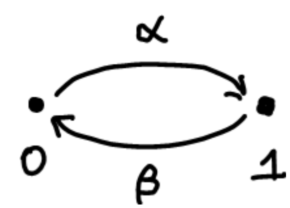


Fig. 6.1. Two state Markov chain rate diagram.

6.8. To do this we start by observing that

$$\det(Q-\lambda I)=\det\left(\begin{bmatrix}-\alpha-\lambda & \alpha\\ \beta & -\beta-\lambda\end{bmatrix}\right)=(\alpha+\lambda)(\beta+\lambda)-\alpha\beta$$
$$=\lambda^2+\tau\lambda=\lambda(\lambda+\tau).$$

Thus the eigenvalues of Q are $\{0,-\tau\}$. The eigenvector for 0 is $\begin{bmatrix}1 & 1\end{bmatrix}^{\mathrm{tr}}$. Moreover,

$$Q-(-\tau)I=\begin{bmatrix}\beta & \alpha\\ \beta & \alpha\end{bmatrix}$$

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 $P(t)=e^{tQ}$ and $\pi(t)=\pi(0)P(t)$, then

$$\dot\pi(t)=\pi(t)Q=[\pi_0(t),\pi_1(t)]\begin{bmatrix}-\alpha & \alpha\\ \beta & -\beta\end{bmatrix}$$
$$\begin{bmatrix}-\alpha\pi_0(t)+\beta\pi_1(t) & \alpha\pi_0(t)-\beta\pi_1(t)\end{bmatrix},$$

i.e

$$\dot\pi_0(t)=-\alpha\pi_0(t)+\beta\pi_1(t) \tag{6.18}$$
$$\dot\pi_1(t)=\alpha\pi_0(t)-\beta\pi_1(t). \tag{6.19}$$

The latter pair of equations is easy to write down using the jump diagram and the movement of sand interpretation. If we assume that $\pi_0(0)+\pi_1(0)=1$ then we know $\pi_0(t)+\pi_1(t)=1$ for all later times and therefore we may rewrite Eq. (6.18) as

$$\begin{aligned}\dot\pi_0(t)&=-\alpha\pi_0(t)+\beta(1-\pi_0(t))\\ &=-\tau\pi_0(t)+\beta\end{aligned}$$

where $\tau:=\alpha+\beta$. We may use Lemma 6.14 below to find

$$\begin{aligned}\pi_0(t)&=e^{-\tau t}\pi_0(0)+\int_0^t e^{-\tau(t-s)}\beta ds\\ &=e^{-\tau t}\pi_0(0)+\frac{\beta}{\tau}\left(1-e^{-\tau t}\right).\end{aligned}$$

We may also conclude that

$$\begin{aligned}\pi_1(t)=1-\pi_0(t)&=1-e^{-\tau t}\pi_0(0)-\frac{\beta}{\tau}\left(1-e^{-\tau t}\right)\\ &=1-e^{-\tau t}(1-\pi_1(0))-\frac{\beta}{\tau}\left(1-e^{-\tau t}\right)\\ &=e^{-\tau t}\pi_1(0)+\left(1-e^{-\tau t}\right)-\frac{\beta}{\tau}\left(1-e^{-\tau t}\right)\\ &=e^{-\tau t}\pi_1(0)+\frac{\alpha}{\tau}\left(1-e^{-\tau t}\right).\end{aligned}$$

By taking $\pi_0(0)=1$ and $\pi_1(0)=0$ we get the first row of $P(t)$ is equal to

$$\begin{bmatrix}e^{-\tau t}1+\frac{\beta}{\tau}\left(1-e^{-\tau t}\right) & \frac{\alpha}{\tau}\left(1-e^{-\tau t}\right)\end{bmatrix}=\frac{1}{\tau}\begin{bmatrix}e^{-\tau t}\alpha+\beta & \alpha\left(1-e^{-\tau t}\right)\end{bmatrix}$$

and similarly the second row of $P(t)$ is found by taking $\pi_0(0)=0$ and $\pi_1(0)=1$ to find

$$\begin{bmatrix}\frac{\beta}{\tau}\left(1-e^{-\tau t}\right) & e^{-\tau t}+\frac{\alpha}{\tau}\left(1-e^{-\tau t}\right)\end{bmatrix}=\frac{1}{\tau}\begin{bmatrix}\beta\left(1-e^{-\tau t}\right) & \beta e^{-\tau t}+\alpha\end{bmatrix}.$$

Hence we have found

$$\begin{aligned}P(t)&=\frac{1}{\tau}\begin{bmatrix}e^{-\tau t}\alpha+\beta & \alpha\left(1-e^{-\tau t}\right)\\ \beta\left(1-e^{-\tau t}\right) & \beta e^{-\tau t}+\alpha\end{bmatrix}\\ &=\frac{1}{\tau}\begin{bmatrix}\left(e^{-\tau t}-1\right)\alpha+\beta+\alpha & \alpha\left(1-e^{-\tau t}\right)\\ \beta\left(1-e^{-\tau t}\right) & \beta\left(e^{-\tau t}-1\right)+\alpha+\beta\end{bmatrix}\\ &=I+\frac{1}{\tau}\left(1-e^{-\tau t}\right)\begin{bmatrix}-\alpha & \alpha\\ \beta & -\beta\end{bmatrix}\\ &=I+\frac{1}{\tau}\left(1-e^{-\tau t}\right)Q.\end{aligned}$$

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

$$\pi = \lim_{t\to\infty} \pi P(t) = \frac{1}{\tau}\left[\beta \ \alpha\right] = \left[\tfrac{\beta}{\alpha+\beta} \ \tfrac{\alpha}{\alpha+\beta}\right] \tag{6.20}$$

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

$$\operatorname{Nul} Q^{\mathrm{tr}} = \operatorname{Nul}\begin{bmatrix} -\alpha & \beta \\ \alpha & -\beta \end{bmatrix} = \mathbb{R}\cdot\begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

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;

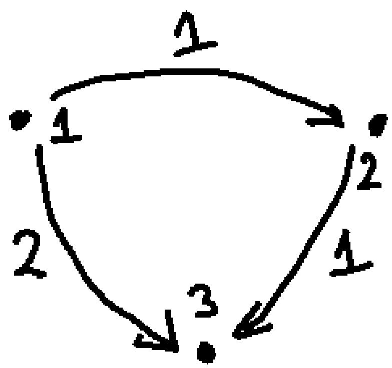
$$\dot{P}(t) = e^{-\tau t} Q \text{ while}$$

$$P(t)\, Q = Q - \frac{1}{\tau}\left(e^{-\tau t} - 1\right)(-\tau)\, Q = Q + \left(e^{-\tau t} - 1\right) Q = \dot{P}(t).$$

Example 6.19. Let $S = \{1, 2, 3\}$ and

$$Q = \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} -3 & 1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

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

$$A_n := \left\{X\left(\frac{i}{2^n}t\right) = x \text{ for } i = 1, 2, \ldots, 2^n - 1, 2^n\right\}.$$

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,

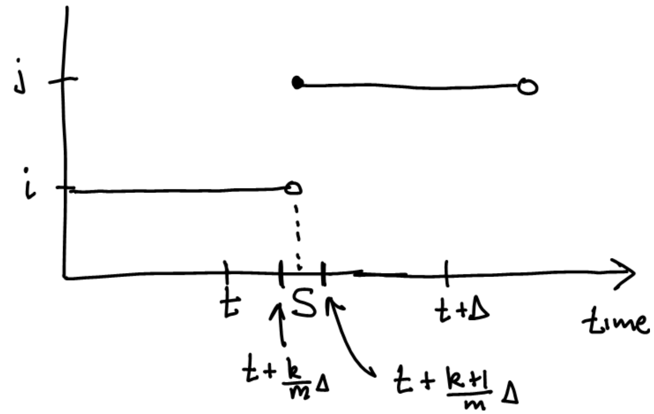
$$\begin{aligned} P_x(X_S = y, t < S \leq t + \Delta) &= P_x(X_{t+\Delta} = y, S > t) + o(\Delta) \\ &= P_x(X_{t+\Delta} = y, X_t = x, S > t) + o(\Delta) \\ &= \lim_{n\to\infty}\left[1 - \frac{tQ_x}{n} + O(n^{-2})\right]^n P_{x,y}(\Delta) + o(\Delta) \\ &= e^{-tQ_x}P_{x,y}(\Delta) + o(\Delta). \end{aligned}$$

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

$$P_x(S > t, X_{t+\delta} = y) = \lim_{n\to\infty} P_x\left(X_{t+\delta} = y \text{ and } X\left(\frac{i}{2^n}t\right) = x \text{ for } i = 1, 2, \dots, 2^n\right)$$
$$= \lim_{n\to\infty} \left[P_{x,x}(t/2^n)\right]^{2^n} P_{xy}(\delta)$$
$$= P_{xy}(\delta) \lim_{n\to\infty}\left[1 - \frac{tQ_x}{2^n} + O\left(2^{-2n}\right)\right]^{2^n} = P_{xy}(\delta)e^{-tQ_x}.$$

With this computation in hand, we may now compute $P_x(X_S = y, t < S \leq t + \Delta)$ using the Figure 6.5 as our guide

So according Figure 6.5, we must have $X_S = y \;\&\; t < S \leq t + \Delta$ iff for all large n there exists $0 \leq k < n$ such that $S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y$ and therefore

$$\begin{aligned}
&P_x(X_S = y \;\&\; t < S \leq t + \Delta) \\
&\quad= \lim_{n\to\infty} P_x\begin{pmatrix} S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y \\ \text{for some } 0 \leq k < n\end{pmatrix} \\
&\quad= \lim_{n\to\infty} \sum_{k=0}^{n-1} P_x(S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y) \\
&\quad= \lim_{n\to\infty} \sum_{k=0}^{n-1} P_{xy}(\Delta/n)e^{-(t+k\Delta/n)Q_x} \\
&\quad= \lim_{n\to\infty} \sum_{k=0}^{n-1} e^{-(t+k\Delta/n)Q_x}\left(\frac{\Delta}{n}Q_{xy} + o(n^{-1})\right) \\
&\quad= Q_{xy}\int_t^{t+\Delta} e^{-Q_x s}ds = \frac{Q_{x,y}}{Q_x}\int_t^{t+\Delta} Q_x e^{-Q_x s}ds \\
&\quad= \frac{Q_{x,y}}{Q_x}P_x(t < S \leq t + \Delta).
\end{aligned}$$

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

$$P_x(X_S = y, t < S \leq t + \Delta) = P_x(X_S = y) \cdot P_x(t < S \leq t + \Delta).$$

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,

$$\begin{aligned}
&\{X_S = y \;\&\; t < S \leq t + \Delta\} \\
&\quad= \cup_{N=1}^{\infty} \cap_{n\geq N} \cup_{0\leq m<n} \left\{S > t + k\Delta/n \;\&\; X_{t+(k+1)\Delta/n} = y\right\}.
\end{aligned}$$

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

$$\tilde{P}_{xy} := \begin{cases} \frac{Q_{x,y}}{-Q_{x,x}} & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases} \text{ for all } x, y \in S. \tag{6.24}$$

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 $\{Y_j = x_j : j \leq n\}$, $S_j \stackrel{d}{=} \exp\left(Q_{x_{j-1}}\right)$ 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 $X_t := Y_{Nt}$. Then $\{X_t\}_t \geq 0$ is the Markov process starting at x with Markov semi-group, $P(t) = e^tQ$.*

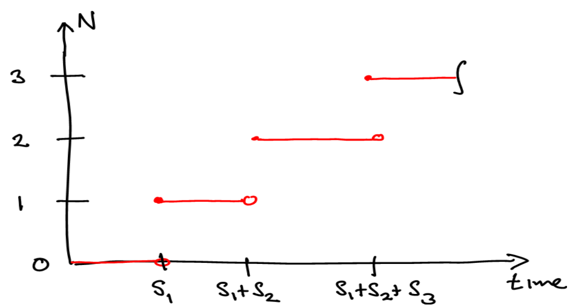


Fig. 6.2. Defining N_t.

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 [3, 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

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{array} \left[\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \dots \\ -\lambda_0 & \lambda_0 & & & & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & & \\ & & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \\ & & & \ddots & \ddots & \ddots \\ & & & & & \end{array} \right].$$

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,

$$0 \underset{\frac{\mu_1}{\lambda_1+\mu_1}}{\overset{1}{\rightleftarrows}} 1 \underset{\frac{\mu_2}{\lambda_2+\mu_2}}{\overset{\frac{\lambda_1}{\lambda_1+\mu_1}}{\rightleftarrows}} 2 \underset{\frac{\mu_3}{\lambda_3+\mu_3}}{\overset{\frac{\lambda_2}{\lambda_2+\mu_2}}{\rightleftarrows}} 3 \dots \rightleftarrows (n-1) \underset{\frac{\mu_n}{\lambda_n+\mu_n}}{\overset{\frac{\lambda_{n-1}}{\lambda_{n-1}+\mu_{n-1}}}{\rightleftarrows}} n \underset{\frac{\mu_{n+1}}{\lambda_{n+1}+\mu_{n+1}}}{\overset{\frac{\lambda_n}{\lambda_n+\mu_n}}{\rightleftarrows}} \dots$$

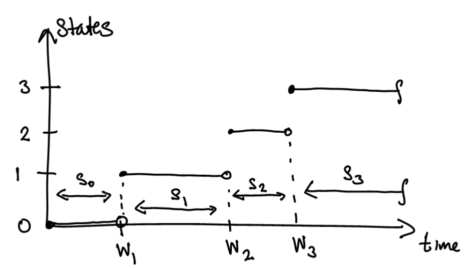
In the jump hold description, a particle follows this discrete time chain. When it arrives at a site, say n, it stays there for an $\exp(\lambda_n + \mu_\nu)$ – time and then jumps to either $n-1$ or n with probability $\frac{\mu_n}{\lambda_n+\mu_n}$ or $\frac{\lambda_n}{\lambda_n+\mu_n}$ respectively. 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

$$Q_{i,i+1} = \lambda_i \text{ and } Q_{ii} = -\lambda_i \text{ for all } i = 0, 1, 2, \ldots$$

with all other entries being zero. Thus we have

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \ldots \end{array} \\ \begin{bmatrix} -\lambda_0 & \lambda_0 & & & & \\ & -\lambda_1 & \lambda_1 & & & \\ & & -\lambda_2 & \lambda_2 & & \\ & & & -\lambda_3 & \lambda_3 & \\ & & & & \ddots & \ddots \end{bmatrix} \end{array}.$$

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

$$\pi_n(t) = P(X(t) = n | X(0) = 0) = \lambda_0 \ldots \lambda_{n-1} \left[\sum_{k=0}^n B_{k,n} e^{-\lambda_k t} \right]$$

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 $(\pi_0(t), \ldots, \pi_3(t)) = (1, 0, 0, 0) e^{tQ'}$. 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

$$(\pi_0(t), \ldots, \pi_3(t)) = (1, 0, 0, 0) S \begin{bmatrix} e^{-t\lambda_0} & & & \\ & e^{-t\lambda_1} & & \\ & & e^{-t\lambda_2} & \\ & & & e^{-t\lambda_3} \end{bmatrix} S^{-1}$$

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

$$\begin{aligned}\frac{d}{dt}(\pi_0(t), \dots, \pi_3(t)) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} &= \frac{d}{dt}(1,0,0,0)\, e^{tQ'} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= (1,0,0,0)\, e^{tQ'} Q' \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\end{aligned}$$

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,

$$\frac{d}{dt}(\pi_0(t), \dots, \pi_3(t)) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \le 0.$$

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 $\left\{t \ge \sum_{n=1}^{\infty} S_n\right\}$. With this definition, $\{X(t)\}_{t\ge 0}$ 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,

$$0 \underset{\mu_1}{\longleftarrow} 1 \underset{\mu_2}{\longleftarrow} 2 \underset{\mu_3}{\longleftarrow} 3 \ldots \underset{\mu_{N-1}}{\longleftarrow} (N-1) \underset{\mu_N}{\longleftarrow} N.$$

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 $j = 1, 2, \ldots, N$ by using the solutions for the pure Birth process with $0 \to N$, $1 \to (N-1)$, $2 \to (N-2), \ldots,$ 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

$$\begin{aligned}\pi_n(t) &= P(X(t) = n | X(0) = N) \\ &= \binom{N}{n} e^{-n\alpha t} \left(1 - e^{-\alpha t}\right)^{N-n} \text{ for } n = 0, 1, 2, \dots, N. \end{aligned} \tag{7.5}$$

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*

$$P\left(B | \cup_{j=1}^N A_j\right) = P(B|A_1). \tag{7.6}$$

We also use the identity, that

$$P(B|A \cap C) = P(B|A) \tag{7.7}$$

whenever C is independent of $\{A, B\}$.

Proof. The proof is the following simple computation,

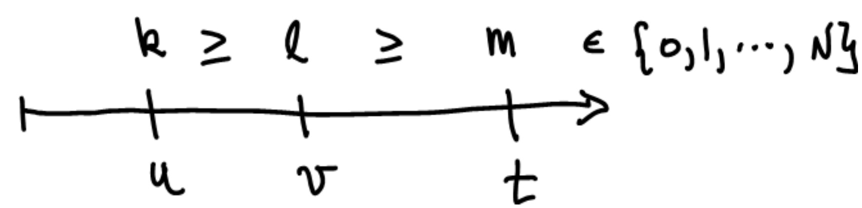
$$\begin{aligned} P\left(B | \cup_{j=1}^n A_j\right) &= \frac{P\left(B \cap \left(\cup_{j=1}^n A_j\right)\right)}{P\left(\cup_{j=1}^n A_j\right)} = \frac{P\left(\cup_{j=1}^n B \cap A_j\right)}{P\left(\cup_{j=1}^n A_j\right)} \\ &= \frac{\sum_{j=1}^n P(B \cap A_j)}{\sum_{j=1}^n P(A_j)} = \frac{nP(B \cap A_1)}{nP(A_1)} = P(B|A_1). \end{aligned}$$

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 $V = \{1, 2, \ldots, l\} \subset U = \{1, 2, \ldots, k\}$, that

$$\begin{aligned}
&P(X_t = m | X_u = k, X_v = l) = P(X_t = m | A_{U,V}) \\
&\quad = P(\text{Exactly } m \text{ of } \xi_1, \ldots, \xi_l \ > t | \xi_1 > v, \ldots, \xi_l > v, v \geq \xi_{l+1} > u, \ldots, v \geq \xi_k > u) \\
&\quad = P(\text{Exactly } m \text{ of } \xi_1, \ldots, \xi_l \ > t | \xi_1 > v, \ldots, \xi_l > v) \\
&\quad = \binom{l}{m} P(\xi_1 > t, \ldots, \xi_m > t, \xi_{m+1} \leq t, \ldots \xi_l \leq t | \xi_1 > v, \ldots, \xi_l > v) \\
&\quad = \binom{l}{m} \frac{P(\xi_1 > t)^m \cdot P(v < \xi_1 \leq t)^{l-m}}{P(v < \xi_1)^l} \\
&\quad = \binom{l}{m} \frac{(e^{-\alpha t})^m \cdot (e^{-vt} - e^{-\alpha t})^{l-m}}{e^{-\alpha v l}} = \binom{l}{m} e^{-\alpha m (t-v)} \left(1 - e^{-\alpha(t-v)}\right)^{l-m}.
\end{aligned}$$

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 $S_1 = W_1$, $S_2 = W_2 - W_1, \ldots, S_N = W_N - W_{N-1}$ 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 $W_1 = S_1 = \min(\xi_1, \ldots, \xi_N)$, 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

$$\begin{aligned}
&P(W_3 - W_2 > t | W_1 = a, W_2 = a + b) \\
&= P(\min(\xi_3, \ldots, \xi_N) - \xi_2 > t | \xi_1 = a,\ \xi_2 = a + b,\ \min(\xi_3, \ldots, \xi_N) > \xi_2) \\
&= \frac{P(\min(\xi_3, \ldots, \xi_N) > t + a + b,\ \xi_1 = a,\ \xi_2 = a + b,\ \min(\xi_3, \ldots, \xi_N) > \xi_2)}{P(\xi_1 = a,\ \xi_2 = a + b,\ \min(\xi_3, \ldots, \xi_N) > a + b)} \\
&= \frac{P(\min(\xi_3, \ldots, \xi_N) > t + a + b)}{P(\min(\xi_3, \ldots, \xi_N) > a + b)} = e^{-(N-2)\alpha t}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[1_{W_2 - W_1 > t}\ g(W_1)\right] &= \sum_k \mathbb{E}\left[1_{W_2 - W_1 > t}\ g(W_1) : \xi_k < T_k\right] \\
&= \sum_k \mathbb{E}\left[1_{T_k - \xi_k > t}\ g(\xi_k) : \xi_k < T_k\right] \\
&= \sum_k \mathbb{E}\left[1_{T_k - \xi_k > t}\ g(\xi_k)\right] \\
&= \sum_k \mathbb{E}\left[\exp(-(N-1)\alpha(t + \xi_k))\, g(\xi_k)\right] \\
&= \exp(-(N-1)\alpha t) \sum_k \mathbb{E}\left[\exp(-(N-1)\alpha \xi_k)\, g(\xi_k)\right] \\
&= \exp(-(N-1)\alpha t) \sum_k \mathbb{E}\left[1_{T_k - \xi_k > 0}\ g(\xi_k)\right] \\
&= \exp(-(N-1)\alpha t) \sum_k \mathbb{E}\left[1_{T_k - \xi_k > 0}\ g(W_1)\right] \\
&= \exp(-(N-1)\alpha t) \cdot \mathbb{E}\left[g(W_1)\right].
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[1_{W_3 - W_2 > t}\ g(W_1, W_2)\right] &= \sum_{j \neq k} \mathbb{E}\left[1_{W_3 - W_2 > t}\ g(W_1, W_2) : \xi_j < \xi_k < T_{jk}\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > t}\ g(\xi_j, \xi_k) : \xi_j < \xi_k < T_{jk}\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > t}\ g(\xi_j, \xi_k) : \xi_j < \xi_k\right] \\
&= \sum_{j \neq k} \mathbb{E}\left[\exp(-(N-2)\alpha(t + \xi_k))\, g(\xi_j, \xi_k) : \xi_j < \xi_k\right] \\
&= \exp(-(N-2)\alpha t) \sum_{j \neq k} \mathbb{E}\left[\exp(-(N-2)\alpha \xi_k)\, g(\xi_j, \xi_k) : \xi_j <\right. \\
&= \exp(-(N-2)\alpha t) \sum_{j \neq k} \mathbb{E}\left[1_{T_{jk} - \xi_k > 0}\ g(\xi_j, \xi_k) : \xi_j < \xi_k\right] \\
&= \exp(-(N-2)\alpha t) \sum_{j \neq k} \mathbb{E}\left[g(W_1, W_2) : \xi_j < \xi_k < T_{jk}\right] \\
&= \exp(-(N-2)\alpha t) \cdot \mathbb{E}\left[g(W_1, W_2)\right].
\end{aligned}$$

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. ■

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

$$\pi_i = \frac{1}{q_i\cdot\mathbb{E}_i R_i} \text{ where } q_i := -Q_{ii}.$$

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.*
$$\sum_{i\in S}\pi_i Q_{ij}=0 \text{ for all } j\in S,$$
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*
$$\sum_{i\in S}\nu_i Q_{ij}=0 \text{ for all } j\in S \text{ with } \sum_{i\in S}\nu_i=1.$$
If such a solution exists it is unique and $\nu=\pi$.

Proof. We refer the reader to [3, 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),

$$\nu_j=\lim_{t\to\infty}\sum_i \nu_i P_{ij}(t)=\sum_i\lim_{t\to\infty}\nu_i P_{ij}(t)=\left(\sum_i\nu_i\right)\pi_j=\pi_j.$$

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:

$$\begin{aligned}\sum_{j\in S}\pi_j P_{jk}(s)&=\sum_{j\in S}\lim_{t\to\infty}P_{ij}(t)P_{jk}(s)=\lim_{t\to\infty}\sum_{j\in S}P_{ij}(t)P_{jk}(s)\\&=\lim_{t\to\infty}P_{ik}(t+s)=\pi_k.\end{aligned}$$

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 (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,

$$\theta = 1,\ \theta_1 = \frac{\lambda}{\mu},\ \theta_2 = \frac{\lambda^2}{2\mu^2},\ \theta_3 = \frac{\lambda^3}{3! \cdot \mu^3}, \ldots, \theta_K = \frac{\lambda^K}{K!\mu^K}$$

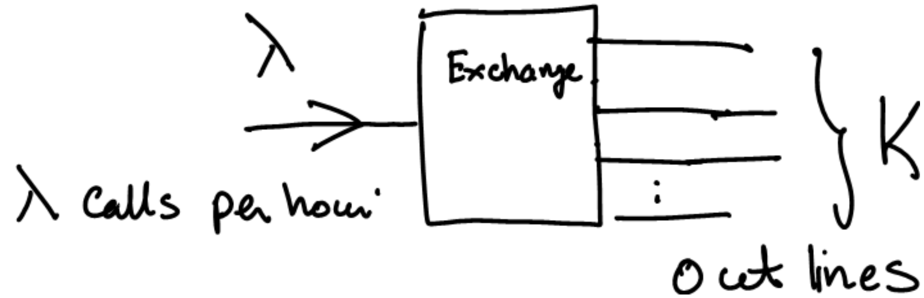


Fig. 8.1. Schematic of a telephone exchange.

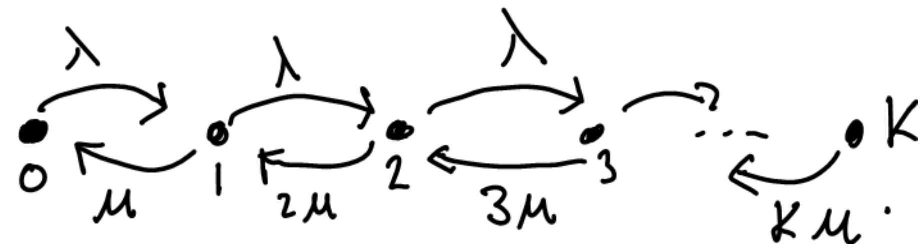


Fig. 8.2. Rate diagram for the telephone exchange.

so that

$$\Theta := \sum_{k=0}^{K} \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \cong e^{\lambda/\mu} \text{ for large } K.$$

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

$$\pi_{25} = \frac{\frac{1}{25!}(25)^{25}}{\sum_{k=0}^{25}\frac{1}{k!}(25)^k} \cong 0.144.$$

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,

$$\pi_{30} = \frac{\frac{1}{30!}(25)^{30}}{\sum_{k=0}^{30}\frac{1}{k!}(25)^k} \cong 0.053$$

and

$$\pi_{35} = \frac{\frac{1}{35!}(25)^{35}}{\sum_{k=0}^{35} \frac{1}{k!}(25)^k} \cong .012$$

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,

$$\begin{aligned}\theta_n &= \frac{a(a+\lambda)(a+2\lambda)\dots(a+(n-1)\lambda)}{n!\mu^n} \\ &= \left(\frac{\lambda}{\mu}\right)^n \frac{\frac{a}{\lambda}\left(\frac{a}{\lambda}+1\right)\left(\frac{a}{\lambda}+2\right)\dots\left(\frac{a}{\lambda}+(n-1)\right)}{n!}.\end{aligned}$$

Using Lemma 8.7 with $\alpha = a/\lambda$ and $x = \lambda/\mu$ which we need to assume is less than 1, we find

$$\Theta := \sum_{n=0}^{\infty} \theta_n = \left(1 - \frac{\lambda}{\mu}\right)^{-a/\lambda}$$

and therefore,

$$\pi_n = \left(1 - \frac{\lambda}{\mu}\right)^{a/\lambda} \frac{\frac{a}{\lambda}\left(\frac{a}{\lambda}+1\right)\left(\frac{a}{\lambda}+2\right)\dots\left(\frac{a}{\lambda}+(n-1)\right)}{n!}\left(\frac{\lambda}{\mu}\right)^n$$

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 8.11. *If*

$$M(t) := \mathbb{E}[X(t)] = \sum_{n=1}^{\infty} nP(X(t) = n) = \sum_{n=1}^{\infty} n\pi_n(t)$$

be the expected population size for our linear birth and death process with immigration, then

$$M(t) = \frac{a}{\lambda - \mu}\left(e^{t(\lambda-\mu)} - 1\right) + M(0)e^{t(\lambda-\mu)}$$

which when $\lambda = \mu$ should be interpreted as

$$M(t) = at + M(0).$$

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,

$$\begin{aligned}\dot{M}(t) &= \sum_{n=1}^{\infty} n\dot{\pi}_n(t) \\ &= \sum_{n=1}^{\infty} n\begin{pmatrix} (a+\lambda(n-1))\pi_{n-1}(t) \\ -(a+\lambda n+\mu n)\pi_n(t) + \mu(n+1)\pi_{n+1}(t)\end{pmatrix} \\ &= \sum_{n=1}^{\infty} n(a+\lambda(n-1))\pi_{n-1}(t) \\ &\quad - \sum_{n=1}^{\infty} n(a+\lambda n+\mu n)\pi_n(t) + \sum_{n=1}^{\infty} \mu n(n+1)\pi_{n+1}(t) \\ &= \sum_{n=0}^{\infty} (n+1)(a+\lambda n)\pi_n(t) \\ &\quad - \sum_{n=1}^{\infty} n(a+\lambda n+\mu n)\pi_n(t) + \sum_{n=2}^{\infty} \mu(n-1)n\pi_n(t) \\ &= a\pi_0(t) + [2(a+\lambda) - (a+\lambda+\mu)]\pi_1(t) \\ &\quad + \sum_{n=2}^{\infty} [(n+1)(a+\lambda n) + \mu(n-1)n - n(a+\lambda n+\mu n)]\pi_n(t) \\ &= a\pi_0(t) + [a+\lambda-\mu]\pi_1(t) + \sum_{n=2}^{\infty} [(a+\lambda n) - \mu n]\pi_n(t) \\ &= a\pi_0(t) + \sum_{n=1}^{\infty} [a+\lambda n-\mu n]\pi_n(t) \\ &= \sum_{n=0}^{\infty} [a+\lambda n-\mu n]\pi_n(t) = a + (\lambda-\mu)M(t).\end{aligned}$$

Thus we have shown that

$$\dot{M}(t) = a + (\lambda-\mu)M(t) \text{ with } M(0) = \sum_{n=1}^{\infty} n\pi_n(0),$$

where $M(0)$ 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 $P(X_0 = x_0, \dots, X_n = x_n)$ 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 $P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2, \dots, X_{t_n} = i_n)$ 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

   $$P(t) = \begin{array}{c} \begin{matrix} \quad 1 & \qquad 2 & \qquad 3 \end{matrix} \\ \begin{bmatrix} \frac{1}{3}e^{-3t} + \frac{2}{3} & 0 & -\frac{1}{3}e^{-3t} + \frac{1}{3} \\ -e^{-2t} + \frac{1}{3}e^{-3t} + \frac{2}{3} & e^{-2t} & -\frac{1}{3}e^{-3t} + \frac{1}{3} \\ -\frac{2}{3}e^{-3t} + \frac{2}{3} & 0 & \frac{2}{3}e^{-3t} + \frac{1}{3} \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \end{array}$$

   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.

References

1. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
2. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
3. 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)
4. 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)