Bruce K. Driver

Math 180C (Introduction to Probability) Notes

April 14, 2008 *File:180Notes.tex*

Contents

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:

$$ay_{n+1} + by_n + cy_{n-1} = 0. \tag{0.1}$$

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

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

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

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

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

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

1. Use first step analysis to show for $a < j < b$ that
$$u_j = \frac{1}{2}(u_{j+1} + u_{j-1}) + 1 \tag{0.5}$$
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

$$\mathbb{E}_j T_{a,b} = -ab + (b+a)\, j - j^2 = -a\,(b-j) + bj - j^2. \qquad (0.7)$$

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:

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

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 $T := \min(T_1, T_2, \dots, T_n)$ 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

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

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

Exercise 0.12. Consider the "pure birth" process with constant rates, $\lambda > 0$. In this case $S = \{0,1,2,\dots\}$ and if $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ 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 5.13. The second is to check that the given functions satisfy the differential equations.

1

Independence and Conditioning

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

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

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

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

for any finite subset, J_0, of J.

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

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

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

it follows that

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

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

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

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

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

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

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

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

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

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

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

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

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

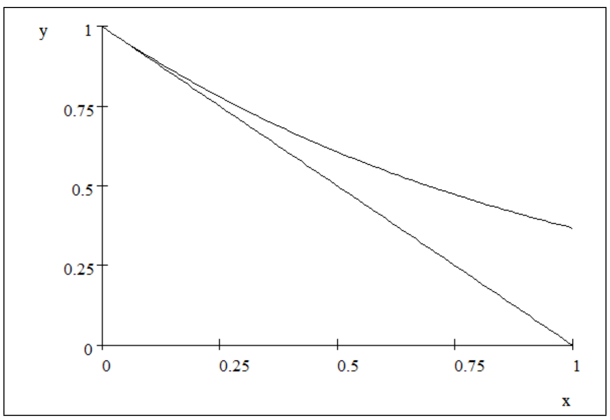


Fig. 1.1. Comparing e^{-x} and $1 - x$.

■

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

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*

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

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

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

Therefore,

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

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

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

while

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

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

Using this as motivation we make the following definition.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

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,

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

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

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

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*

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

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

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

■

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

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

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

2

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.

$$\sum_{y \in S} p(x,y) = 1 \text{ for all } i \in S. \tag{2.1}$$

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

$$\begin{aligned} &P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &\quad = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1}) \end{aligned} \tag{2.2}$$

where $\{x_j\}_{j=1}^{n+1}$ are allowed to range over S and n over $\mathbb{N}_0$. The identity in Eq. (2.2) 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. (2.2) may also be written as,

$$\mathbb{E}[f(X_{n+1}) \mid X_0, X_1, \dots, X_n] = \mathbb{E}[f(X_{n+1}) \mid X_n] = \sum_{y\in S} p(X_n, y) f(y) \tag{2.3}$$

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

---

[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.)

Proposition 2.2. *If P is a Markov probability as in Definition 2.1 and $\pi(x) := P(X_0 = x)$, then for all $n \in \mathbb{N}_0$ and $\{x_j\} \subset S$,*

$$P(X_0 = x_0, \dots, X_n = x_n) = \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n). \tag{2.4}$$

Conversely if $\pi : S \to [0,1]$ is a probability and $\{X_n\}_{n=0}^{\infty}$ is a sequence of random variables satisfying Eq. (2.4) for all n and $\{x_j\} \subset S$, then $(\{X_n\}, P, p)$ satisfies Definition 2.1.

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

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

($\Longleftarrow$) 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 2.3 *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. (2.4).*

Notation 2.4 *If*

$$\pi(y) = \delta_x(y) := \begin{cases} 1 \text{ if } x = y \\ 0 \text{ if } x \neq y \end{cases}, \tag{2.5}$$

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

$$P_\pi = \sum_{x \in S} \pi(x)\, P_x. \tag{2.6}$$

Notation 2.5 *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 2.6. Suppose that $S = \{1, 2, 3\}$, then

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

has the jump graph given by 2.1.

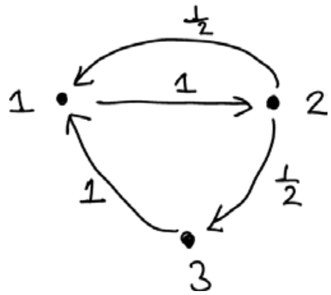


Fig. 2.1. A simple jump diagram.

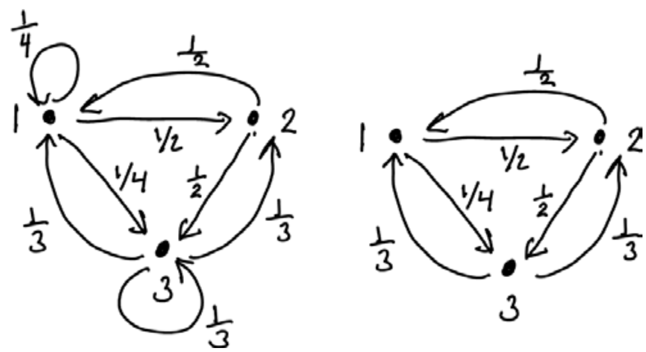


Fig. 2.2. The above diagrams contain the same information. In the one on the right we have dropped the jumps from a site back to itself since these can be deduced by conservation of probability.

Example 2.7. The transition matrix,

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

is represented by the jump diagram in Figure 2.2.

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

$$(p \cdot q)(x,y) := \sum_{z \in S} p(x,z)\, q(z,y). \quad \text{(Matrix Multiplication!)} \tag{2.7}$$

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

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

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 2.8. *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. (2.4) 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 2.9. *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.

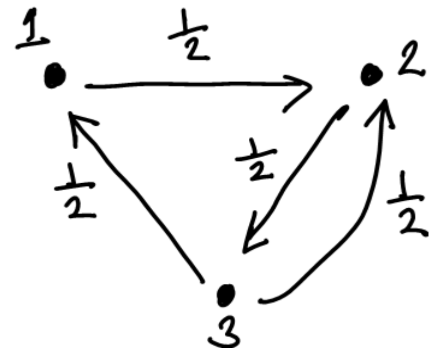
Example 2.10. Consider the following example,

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

with jump diagram given in Figure 2.10. We have

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

and



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

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

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

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

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

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

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

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

Example 2.11 (§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 $X_n = (\#\text{ of particles in } A)$ which we assume evolves by choosing a particle at random from $A \cup B$ and then replacing this particle in the opposite bin from which it was found. Suppose there are N total number of particles in the flask, then the transition probabilities are given by,

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

For example, if $N = 2$ we have

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

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

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

In the case $N = 2$,

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

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

and when $N = 3$,

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

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

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

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

:

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

We also have

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

and

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

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

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

2.1 First Step Analysis

We will need the following observation in the proof of Lemma 2.14 below. If T is a $\mathbb{N}_0 \cup \{\infty\}$ – valued random variable, then

$$\mathbb{E}_x T = \mathbb{E}_x \sum_{n=0}^{\infty} 1_{n<T} = \sum_{n=0}^{\infty} \mathbb{E}_x 1_{n<T} = \sum_{n=0}^{\infty} P_x\left(T > n\right). \tag{2.8}$$

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

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

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

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

Remark 2.12. 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 2.13. Suppose that $S = \{1,2,\dots,7\}$, $A = \{1,2,4,5,6\}$, $B = \{3,7\}$, and

$$P = \begin{array}{cc} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \end{array}.$$

Following the algorithm in Remark 2.12 leads to:

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

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

Lemma 2.14. *Keeping the notation above we have*

$$\mathbb{E}_x T = \sum_{n=0}^{\infty} \sum_{y \in A} Q^n(x,y) \text{ for all } x \in A, \tag{2.9}$$

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,

$$\begin{aligned} P_x(T > n) &= P_x(X_1, \dots, X_n \in A) \\ &= \sum_{x_1, \dots, x_n \in A} p(x, x_1) p(x_1, x_2) \dots p(x_{n-1}, x_n) \\ &= \sum_{y \in A} Q^n(x,y). \end{aligned} \tag{2.10}$$

Therefore Eq. (2.9) now follows from Eqs. (2.8) and (2.10). ■

Proposition 2.15. *Let us continue the notation above and let us further assume that A is a finite set and*

$$P_x(T < \infty) = P(X_n \in B \text{ for some } n) > 0 \ \forall\ x \in A. \tag{2.11}$$

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. (2.9) as

$$(\mathbb{E}_x T)_{x \in A} = (I - Q)^{-1} \mathbf{1} \tag{2.12}$$

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. (2.10) 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. (2.12). 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,

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

as claimed. ■

Remark 2.16. 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 $\mathbb{E}_i T = \infty$ 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 2.15.

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

Theorem 2.17 (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*

$$\mathbb{E}_\pi\left[f\left(X_n, X_{n+1}, \dots\right) g\left(X_0, \dots, X_n\right)\right] = \mathbb{E}_\pi\left[\left(\mathbb{E}_{X_n}\left[f\left(X_0, X_1, \dots\right)\right]\right) g\left(X_0, \dots, X_n\right)\right] \tag{2.13}$$

$$\mathbb{E}_\pi\left[f\left(X_n, X_{n+1}, \dots\right) | X_0 = x_0, \dots, X_n = x_n\right] = \mathbb{E}_{x_n} f\left(X_0, X_1, \dots\right) \tag{2.14}$$

for all $x_0, \dots, x_n \in S$ *such that* $P_\pi\left(X_0 = x_0, \dots, X_n = x_n\right) > 0$. *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 $F = f\left(X_0, X_1, \dots X_m\right)$, $\theta_n F := f\left(X_n, X_{n+1}, \dots X_{n+m}\right)$, and $G = g\left(X_0, \dots, X_n\right)$.

We then have,

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

and

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

Combining the last two equations implies,

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

as was to be proved.

Taking $g\left(y_0, \dots, y_n\right) = \delta_{x_0, y_0} \dots \delta_{x_n, y_n}$ is Eq. (2.13) implies that

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

which implies Eq. (2.14). 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 2.18. *Let $n \in \mathbb{N}_0$, $x \in S$ and π be any probability on S. Then relative to $P_\pi\left(\cdot|X_n = x\right)$, $\{X_{n+k}\}_{k\geq 0}$ is independent of $\{X_0, \dots, X_n\}$ and $\{X_{n+k}\}_{k\geq 0}$ has the same distribution as $\{X_k\}_{k=0}^\infty$ under P_x.*

Proof. According to Eq. (2.13),

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

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

$$\begin{aligned}
&\mathbb{E}_\pi\left[g\left(X_0, \dots, X_n\right) f\left(X_n, X_{n+1}, \dots\right) |X_n = x\right] \\
&\quad = \mathbb{E}_\pi\left[g\left(X_0, \dots, X_n\right) |X_n = x\right] \mathbb{E}_x\left[f\left(X_0, X_1, \dots\right)\right].
\end{aligned} \tag{2.15}$$

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

$$\mathbb{E}_\pi\left[f\left(X_n, X_{n+1}, \dots\right) |X_n = x\right] = \mathbb{E}_x\left[f\left(X_0, X_1, \dots\right)\right]. \tag{2.16}$$

This shows that $\{X_{n+k}\}_{k\geq 0}$ under $P_\pi\left(\cdot|X_n = x\right)$ has the same distribution as $\{X_k\}_{k=0}^\infty$ under P_x and, in combination, Eqs. (2.15) and (2.16) shows $\{X_{n+k}\}_{k\geq 0}$ and $\{X_0, \dots, X_n\}$ are conditionally independent on $\{X_n = x\}$. ■

Theorem 2.19. *Let us continue the notation and assumption in Proposition 2.15 and further let $g : A \to \mathbb{R}$ and $h : B \to \mathbb{R}$ be two functions. Let $\mathbf{g} := \left(g\left(x\right)\right)_{x\in A}$ and $\mathbf{h} := \left(h\left(y\right)\right)_{y\in B}$ to be thought of as column vectors. Then for all $x \in A$,*

$$\mathbb{E}_x\left[\sum_{n<T} g(X_n)\right] = x^{th} \text{ component of } (I - Q)^{-1}\mathbf{g} \tag{2.17}$$

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

$$P_x\left(X_T = y\right) = \left[(I - Q)^{-1}R\right]_{x,y}. \tag{2.18}$$

Taking $g \equiv \mathbf{1}$ (where $\mathbf{1}\left(x\right) = 1$ for all $x \in A$) in Eq. (2.17) shows that

$$\mathbb{E}_x T = \text{ the } x^{th} \text{ component of } (I - Q)^{-1}\mathbf{1} \tag{2.19}$$

in agreement with Eq. (2.12). If we take $g\left(x'\right) = \delta_y\left(x'\right)$ for some $x \in A$, then

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

and by Eq. (2.17) it follows that

$$\mathbb{E}_x\left[\text{number of visits to } y \text{ before hitting } B\right] = (I - Q)^{-1}_{xy}. \tag{2.20}$$

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 2.17 we have,

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

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

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

In matrix language this becomes, $\mathbf{u} = Q\mathbf{u} + \mathbf{g}$ and hence we have $\mathbf{u} = (I - Q)^{-1}\mathbf{g}$ which is precisely Eq. (2.17).

To prove Eq. (2.18), 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,

[2] In applying Theorem 2.17 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

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

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

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

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

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

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

∎

Remark 2.20. 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 2.21 (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*

$$h_i = P_i(H < \infty) = 1_{i\in A} + 1_{i\notin A}\sum_{n=0}^{\infty}[Q^n v]_i \tag{2.21}$$

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

$$\begin{aligned} &h_i = 1 \text{ if } i \in A \text{ and} \\ &h_i = \sum_{j\in S} p(i,j)\,h_j = \sum_{j\in A^c} Q(i,j)\,h_j + v_i \text{ for all } i \in A^c. \end{aligned} \tag{2.22}$$

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

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

and therefore,

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

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. (2.21). The remainder of the proof now follows from Lemma 2.22 below. Nevertheless, it is instructive to use the Markov property to show that Eq. (2.22) 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 2.22. *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 $x_i \geq 0$ for all i. Iterating this equation shows,

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

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

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

so that $h_i \leq x_i$ for all i. ∎

2.2 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 2.23. Consider the Markov chain determined by

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

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

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

i.e.

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

which have solutions,

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

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

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

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 2.3 below.

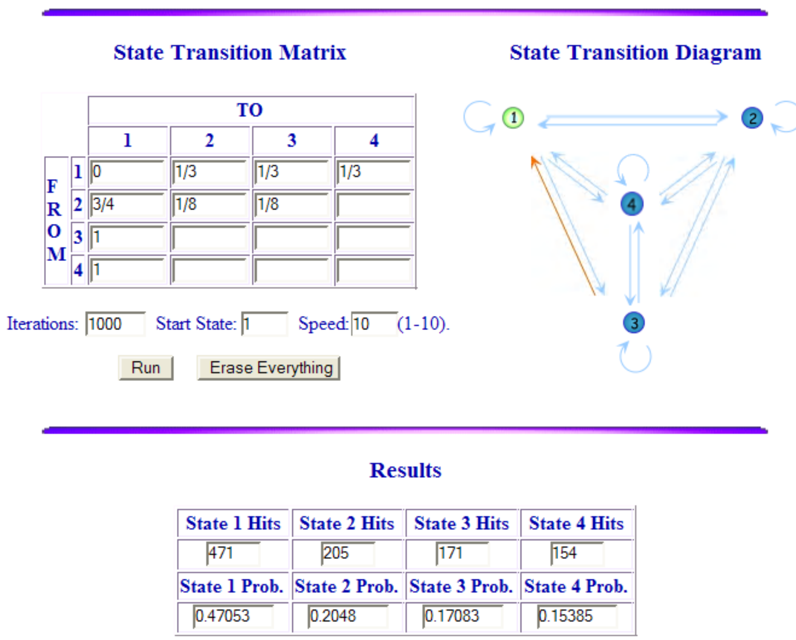


Fig. 2.3. In this run, rather than making sites 3 and 4 absorbing, we have made them transition back to 1. I claim now to get an approximate value for $P_1 (X_n \text{ hits } 3)$ we should compute: (State 3 Hits)/(State 3 Hits + State 4 Hits). In this example we will get $171/(171 + 154) = 0.526\,15$ which is a little lower than the predicted value of 0.533. You can try your own runs of this simulator.

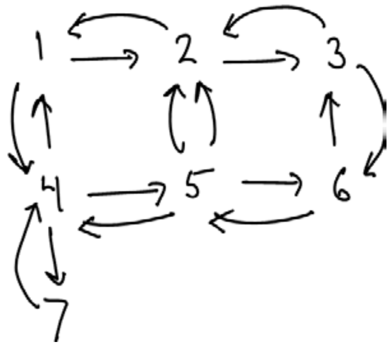


Fig. 2.4. Rat in a maze.

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

Here is the maze

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

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

$$P = \begin{array}{cc} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \end{array}.$$

and the corresponding jump diagram is given in Figure 2.4

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 2.12. This gives,

$$Q = \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 4 & 5 & 6 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array} \end{array}$$

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,

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

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

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

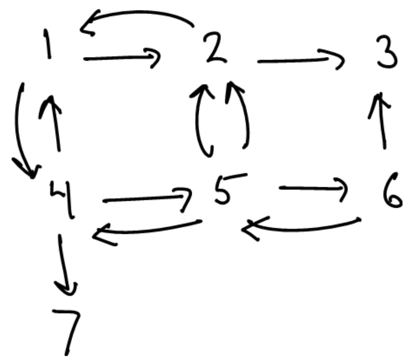
and

$$\begin{aligned} (I-Q)^{-1}R &= \begin{bmatrix} \frac{11}{6} & \frac{5}{4} & \frac{5}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{7}{4} & \frac{3}{4} & 1 & \frac{1}{3} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{4} & 1 & \frac{1}{3} \\ \frac{2}{3} & 1 & 1 & 2 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1/3 & 0 \\ 0 & 1/3 \\ 0 & 0 \\ 1/2 & 0 \end{bmatrix} \\ &= \begin{array}{c} \begin{array}{cc} 3 & 7 \end{array} \\ \left[\begin{array}{cc} \frac{7}{12} & \frac{5}{12} \\ \frac{3}{4} & \frac{1}{4} \\ \frac{5}{12} & \frac{7}{12} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{5}{6} & \frac{1}{6} \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 4 \\ 5 \\ 6 \end{array}. \end{aligned}$$

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,

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



in this example. To do this we make both 3 and 7 absorbing states so the jump diagram is in Figure 2.2.1. 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,

$$h_1 = \frac{4}{9},\ h_2 = \frac{2}{3},\ h_4 = \frac{2}{9},\ h_5 = \frac{5}{9},\ h_6 = \frac{7}{9}. \tag{2.23}$$

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

$$h_1 = \frac{5}{9},\ h_2 = \frac{1}{3},\ h_4 = \frac{7}{9},\ h_5 = \frac{4}{9},\ h_6 = \frac{2}{9}. \tag{2.24}$$

Notice that the sum of the hitting probabilities in Eqs. (2.23) and (2.24) add up to 1 as they should.

2.2.2 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:

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

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

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

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

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

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

So for example, $P_4(X_T = 3(\text{food})) = 1/3$, $E_4(\text{Number of visits to } 1) = 1$, $E_5(\text{Number of visits to } 2) = 3/2$ and $E_1 T = E_5 T = 6$ and $E_2 T = E_4 T = 5$.

3

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

$$R_i := \min\{n \geq 1 : X_n = i\} \tag{3.1}$$

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

$$M_i := \sum_{n \geq 1} 1_{X_n = i} \tag{3.2}$$

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

Definition 3.1. *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 3.2. *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 3.3. *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 3.4. *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.

3.1 The Main Results

Proposition 3.5 (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 3.13 for the assertion that being recurrent or transient is a class property. For the fact that positive and null recurrency is a class property, see Proposition 3.45 below. ■

Lemma 3.6. *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

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

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 3.7. *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:*

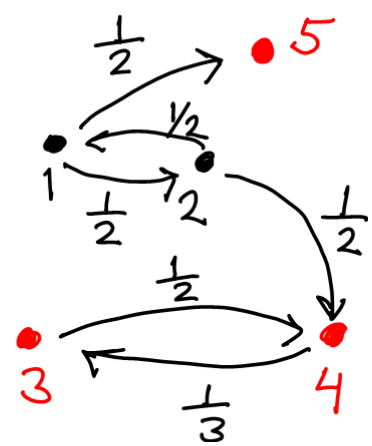
1. *C is closed.*
2. *C is positive recurrent.*
3. *C is recurrent.*

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 2.15. Also see Corollary 3.20 for another proof. ■

Remark 3.8. 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 3.7. 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 3.9. Let P be the Markov matrix with jump diagram given in Figure 3.9. 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 3.41.

Proposition 3.10. *If $j \in S$, $k \in \mathbb{N}$, and $\nu : S \to [0, 1]$ is any probability on S, then*

$$P_\nu (M_j \geq k) = P_\nu (R_j < \infty) \cdot P_j (R_j < \infty)^{k-1}. \tag{3.3}$$

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. (3.3). See Proposition 3.42 below for the formal proof based on the strong Markov property in Corollary 3.41. ∎

Corollary 3.11. *If $j \in S$ and $\nu : S \to [0, 1]$ is any probability on S, then*

$$P_\nu\left(M_j=\infty\right)=P_\nu\left(X_n=j\ i.o.\right)=P_\nu\left(R_j<\infty\right)1_{j\in S_r}, \tag{3.4}$$
$$P_j\left(M_j=\infty\right)=P_j\left(X_n=j\ i.o.\right)=1_{j\in S_r}, \tag{3.5}$$

$$\mathbb{E}_\nu M_j=\sum_{n=1}^{\infty}\sum_{i\in S}\nu\left(i\right)P_{ij}^n=\frac{P_\nu\left(R_j<\infty\right)}{1-P_j\left(R_j<\infty\right)}, \tag{3.6}$$

and

$$\mathbb{E}_i M_j=\sum_{n=1}^{\infty}P_{ij}^n=\frac{P_i\left(R_j<\infty\right)}{1-P_j\left(R_j<\infty\right)} \tag{3.7}$$

where the following conventions are used in interpreting the right hand side of Eqs. (3.6) and (3.7): $a/0:=\infty$ if $a>0$ while $0/0:=0$.

Proof. Since

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

it follows, using Eq. (3.3), that

$$P_\nu\left(X_n=j\text{ i.o. }n\right)=\lim_{k\to\infty}P_\nu(M_j\ge k)=P_\nu(R_j<\infty)\cdot\lim_{k\to\infty}P_j(R_j<\infty)^{k-1} \tag{3.8}$$

which gives Eq. (3.4). Equation (3.5) follows by taking $\nu=\delta_j$ in Eq. (3.4) and recalling that $j\in S_r$ iff $P_j\left(R_j<\infty\right)=1$. Similarly Eq. (3.7) is a special case of Eq. (3.6) with $\nu=\delta_i$. We now prove Eq. (3.6).

Using the definition of M_j in Eq. (3.2),

$$\begin{aligned}\mathbb{E}_\nu M_j&=\mathbb{E}_\nu\sum_{n\ge1}1_{X_n=j}=\sum_{n\ge1}\mathbb{E}_\nu 1_{X_n=j}\\&=\sum_{n\ge1}P_\nu(X_n=j)=\sum_{n=1}^{\infty}\sum_{j\in S}\nu\left(j\right)P_{jj}^n\end{aligned}$$

which is the first equality in Eq. (3.6). For the second, observe that

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

On the other hand using Eq. (3.3) we have

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

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

It is worth remarking that if $j\in S_t$, then Eq. (3.6) 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 3.12 (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. (3.5) and the equivalent of items 1. and 3. follows directly from Eq. (3.7) with $i = j$. ■

Proposition 3.13. *If $i \longleftrightarrow j$, then i is recurrent iff j is recurrent, i.e. the property of being recurrent or transient is a class property.*

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

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

which shows that $\sum_{n\geq 1} P_{jj}^n = \infty \Longrightarrow \sum_{n\geq 1} P_{ii}^n = \infty$. Similarly $\sum_{n\geq 1} P_{ii}^n = \infty \Longrightarrow \sum_{n\geq 1} P_{jj}^n = \infty$. Thus using item 3. of Theorem 3.12, it follows that i is recurrent iff j is recurrent. ■

Corollary 3.14. *If $C \subset S_r$ is a recurrent communication class, then*

$$P_i(R_j < \infty) = 1 \text{ for all } i, j \in C \tag{3.9}$$

and in fact

$$P_i(\cap_{j\in C}\{X_n = j \text{ i.o. } n\}) = 1 \text{ for all } i \in C. \tag{3.10}$$

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

$$P_\nu(\cap_{j\in C}\{X_n = j \text{ i.o. } n\}) = 1 \text{ for all } i \in C. \tag{3.11}$$

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

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

we have

$$\begin{aligned}
P_{ji}^m &= P_j(X_m = i) = P_j(M_j = \infty, X_m = i) \\
&\le P_j(X_m = i \text{ and } X_n = j \text{ for some } n > m) \\
&= \sum_{n>m} P_j(X_m = i, X_{m+1} \ne j, \dots, X_{n-1} \ne j, X_n = j) \\
&= \sum_{n>m} P_{ji}^m P_i(X_1 \ne j, \dots, X_{n-m-1} \ne j, X_{n-m} = j) \\
&= \sum_{n>m} P_{ji}^m P_i(R_j = n-m) = P_{ji}^m \sum_{k=1}^{\infty} P_i(R_j = k) \\
&= P_{ji}^m P_i(R_j < \infty). \qquad (3.12)
\end{aligned}$$

Because $P_{ji}^m > 0$, we may conclude from Eq. (3.12) that $1 \le P_i(R_j < \infty)$, i.e. that $P_i(R_j < \infty) = 1$ and Eq. (3.9) is proved. Feeding this result back into Eq. (3.4) 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. (3.10). Equation (3.11) follows by multiplying Eq. (3.10) by $\nu(i)$ and then summing on $i \in C$. ∎

Theorem 3.15 (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

$$\sum_{n=1}^{\infty} P_{ij}^n = \mathbb{E}_i M_j < \infty \implies \begin{cases} \lim_{n\to\infty} P_{ij}^n = 0 \\ P_i(X_n = j \text{ i.o. } n) = 0. \end{cases} \qquad (3.13)$$

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

$$\sum_{n=1}^{\infty} P_\nu(X_n = j) = \mathbb{E}_\nu M_j < \infty \implies \begin{cases} \lim_{n\to\infty} P_\nu(X_n = j) = 0 \\ P_\nu(X_n = j \text{ i.o. } n) = 0. \end{cases} \qquad (3.14)$$

Proof. The equivalence of the first two items follows directly from Eq. (3.5) and the equivalent of items 1. and 3. follows directly from Eq. (3.7) 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. (3.7) and (3.6) respectively. The remaining implication in Eqs. (3.13) and (3.6) 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 3.16. *1) If the state space, S, is a finite set, then $S_r \ne \emptyset$. 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 3.15, $\lim_{n\to\infty} P_{ij}^n = 0$ 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 3.17. *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 3.18. *We say a sub-probability,* $\pi : S \to [0,1]$, *is* ***invariant*** *if* $\pi P = \pi$, *i.e.*

$$\sum_{i\in S} \pi(i)\, p_{ij} = \pi(j) \text{ for all } j \in S. \tag{3.15}$$

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

Theorem 3.19. *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*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} 1_{X_n = j} = \pi_j \quad P_i - a.s. \tag{3.16}$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_i(X_n = j) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} P_{ij}^n = \pi_j. \tag{3.17}$$

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 3.44 and Propositions 3.45 and 3.46 below. ■

Using this result we can give another proof of Proposition 3.7.

Corollary 3.20. *If* C *is a closed finite communicating class then* C *is positive recurrent. (Recall that we already know that* C *is recurrent by Corollary 3.16.)*

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 3.21. 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 3.19 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 3.21 (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

$$\pi_i := \pi_i(\nu) = \frac{P_\nu (X_n \text{ hits } C)}{\mathbb{E}_i R_i}, \tag{3.18}$$

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

1. *P_ν – a.s.,*
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n = i} = \frac{1}{\mathbb{E}_i R_i} 1_{\{X_n \text{ hits } C\}}, \tag{3.19}$$
2. $$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{j\in S} \nu(j) P_{ji}^n = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\nu (X_n = i) = \pi_i, \tag{3.20}$$
3. *π is an invariant sub-probability for P, and*
4. *the mass of π is*
$$\sum_{i\in S} \pi_i = \sum_{C:\text{ pos. recurrent}} P_\nu (X_n \text{ hits } C) \le 1. \tag{3.21}$$

Proof. If $i \in S$ is a transient site, then according to Eq. (3.14), $P_\nu (M_i < \infty) = 1$ and therefore $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{X_n = i} = 0$ which agrees with Eq. (3.19) for $i \in S_t$.

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

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

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 3.41) and Corollary 3.14 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 3.41), 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 3.19 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. (3.19). Taking expectations of this equation, using the dominated convergence theorem, gives Eq. (3.20).

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

$$\sum_{i\in S} \pi_i P_{ij} = \sum_{i \in S_{\mathrm{pr}}} \pi_i P_{ij} = \sum_{C:\text{ pos-rec.}} P_\nu(X_n \text{ hits } C) \sum_{i\in C} \frac{1}{\mathbb{E}_i R_i} P_{ij}. \tag{3.22}$$

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 $j \notin S_{\mathrm{pr}}$ we have $\sum_{i\in S} \pi_i P_{ij} = 0 = \pi_j$ and if $j \in S_{\mathrm{pr}}$, then by Theorem 3.19,

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

Using this result in Eq. (3.22) shows that

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

so that π is an invariant distribution. Similarly, using Theorem 3.19 again,

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

■

Definition 3.22. *A state $i \in S$ is **aperiodic** if $P^n_{ii} > 0$ for all n sufficiently large.*

Lemma 3.23. *If $i \in S$ is aperiodic and $j \longleftrightarrow i$, then j is aperiodic. So being aperiodic is a class property.*

Proof. We have

$$P^{n+m+k}_{jj} = \sum_{w,z\in S} P^n_{j,w}P^m_{w,z}P^k_{z,j} \geq P^n_{j,i}P^m_{i,i}P^k_{i,j}.$$

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

Lemma 3.24. *A state $i \in S$ is aperiodic iff 1 is the greatest common divisor of the set,*

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

Proof. Use the number theory Lemma 3.47 below. ■

Theorem 3.25. *If P is an irreducible, aperiodic, and recurrent Markov chain, then*

$$\lim_{n\to\infty} P^n_{ij} = \pi_j = \frac{1}{\mathbb{E}_j(R_j)}. \tag{3.23}$$

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

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

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. ■

3.1.1 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 3.26. *The Markov matrix P has an invariant distribution.*

Proof. If $\mathbf{1} := \begin{bmatrix} 1 \; 1 \; \ldots \; 1 \end{bmatrix}^{\mathrm{tr}}$, 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 3.27. *Suppose that P is irreducible. (In this case we may use Proposition 2.15 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,

$$\mathbb{E}_i[R_j] = \sum_{k \neq j} P_{ik} \mathbb{E}_k[R_j] + 1. \tag{3.24}$$

Now suppose that π is any invariant distribution for P, then multiplying Eq. (3.24) by π_i and summing on i shows

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

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

We may use Eq. (3.24) 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. (3.24) 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

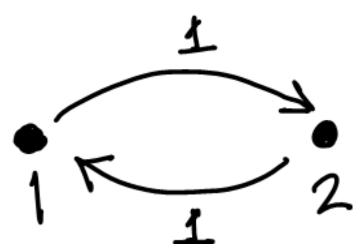
$$(\mathbb{E}_i R_j)_{i=1}^{n} = \left(I - P^{(j)}\right)^{-1} \mathbf{1}, \tag{3.25}$$

i.e.

$$\begin{bmatrix} \mathbb{E}_1 R_j \\ \vdots \\ \mathbb{E}_n R_j \end{bmatrix} = \left(I - P^{(j)}\right)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{3.26}$$

3.2 Examples

Example 3.28. Let $S = \{1, 2\}$ and $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with jump diagram in Figure 3.28. 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 3.29. Again let $S = \{1, 2\}$ and $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ with jump diagram in Figure 3.29. 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 3.30. Suppose that $S = \{1, 2, 3\}$, and

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

has the jump graph given by 3.1. Notice that $P_{11}^2 > 0$ and $P_{11}^3 > 0$ that P is

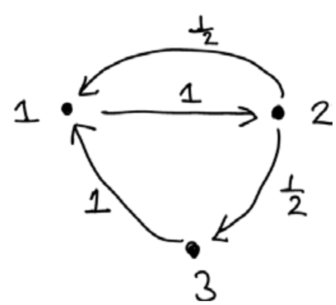


Fig. 3.1. A simple jump diagram.

"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 3.31. The transition matrix,

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

is represented by the jump diagram in Figure 3.2. This chain is aperiodic. We find the invariant distribution as,

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

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

In this case

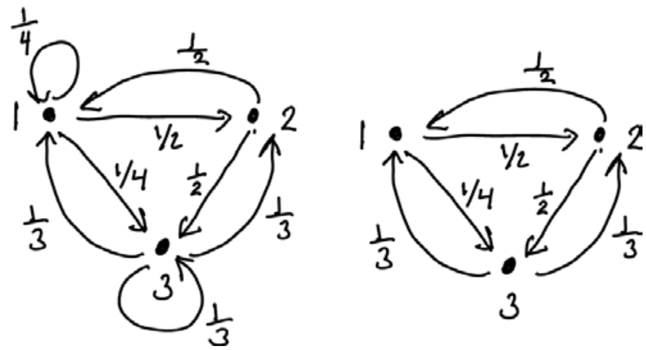


Fig. 3.2. The above diagrams contain the same information. In the one on the right we have dropped the jumps from a site back to itself since these can be deduced by conservation of probability.

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

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 3.32. Consider the following Markov matrix,

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

with jump diagram in Figure 3.3. 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

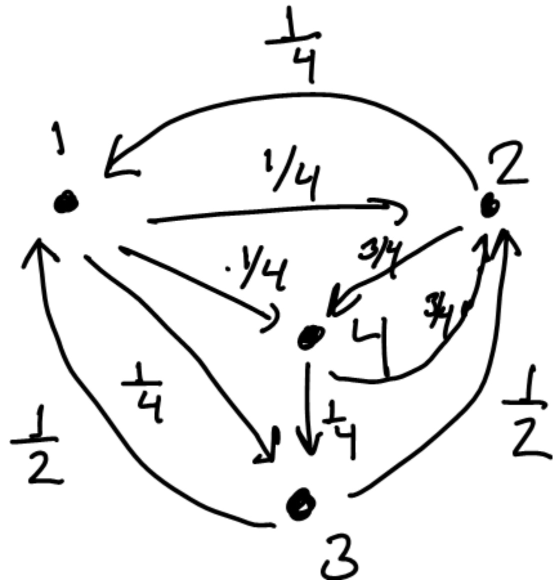


Fig. 3.3. The jump diagram for P.

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

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

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

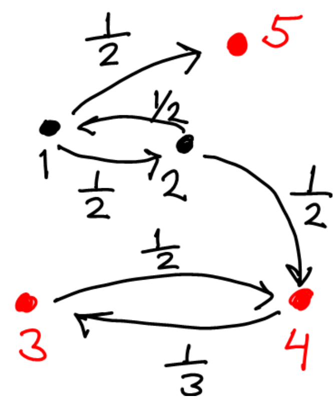
and again $\mathbb{E}_2 R_2 = 4 = 1/\pi_2$.

Example 3.33 (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,

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

Here are the steps to follow.

1. **Find the jump diagram for** P. In our case it is given in Figure 1.



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,

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

We do this in the usual way, namely

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

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

$$\text{first row of } \lim_{n\to\infty} P^n = P_1 (X_n \text{ hits } 5)\, \pi_{\{5\}} + P_1 (X_n \text{ hits } \{3,4\})\, \pi_{\{3,4\}} \tag{3.27}$$

and

$$\text{second row of } \lim_{n\to\infty} P^n = P_2 (X_n \text{ hits } 5)\, \pi_{\{5\}} + P_2 (X_n \text{ hits } \{3,4\})\, \pi_{\{3,4\}}. \tag{3.28}$$

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 $i = 1, 2, \ldots, 5$. 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]

---

[1]

Example 3.34. **Note:** If we were to make use of Theorem 2.21 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 $h_3 = h_4 = 0$.

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

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

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

7. Using these results in Eqs. (3.27) and (3.28) shows,

$$\begin{aligned} \text{first row of } \lim_{n\to\infty} P^n &= \frac{2}{3}\pi_{\{5\}} + \frac{1}{3}\pi_{\{3,4\}} \\ &= \left[0\ 0\ \tfrac{2}{15}\ \tfrac{1}{5}\ 2/3\right] \\ &= \left[0.0\ 0.0\ 0.133\,33\ 0.2\ 0.666\,67\right] \\ &\text{and} \\ \text{second row of } \lim_{n\to\infty} P^n &= \frac{1}{3}\pi_{\{5\}} + \frac{2}{3}\pi_{\{3,4\}} \\ &= \frac{1}{3}\left[0\ 0\ 0\ 0\ 1\right] + \frac{2}{3}\left[0\ 0\ 2/5\ 3/5\ 0\right] \\ &= \left[0\ 0\ \tfrac{4}{15}\ \tfrac{2}{5}\ \tfrac{1}{3}\right] \\ &= \left[0.0\ 0.0\ 0.266\,67\ 0.4\ 0.333\,33\right]. \end{aligned}$$

These answers already compare well with

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

3.3 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 2.17 continues to hold even when n is replaced by a random "stopping time."

Definition 3.35 (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 3.36. If τ is an $\{X_n\}_{n=0}^{\infty}$ - stopping time then

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

That is for a stopping time τ, $1_{\tau \geq n}$ is a function of $(X_0, \dots, X_{n-1})$ only for all $n \in \mathbb{N}_0$.

The following presentation of Wald's equation is taken from Ross [4, p. 59-60].

Theorem 3.37 (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*

$$\mathbb{E}\left[\sum_{n=0}^{\tau} f(X_n)\right] = \mathbb{E}f(X_0) \cdot \mathbb{E}\tau. \tag{3.29}$$

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. (3.29). ∎

Example 3.38. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. such that $P(X_n = 0) = P(X_n = 1) = 1/2$ and let

$$\tau := \min\{n : X_1 + \cdots + X_n = 10\}.$$

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 3.39 (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

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

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

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

which can not hold. Hence it must be that

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

Here is the analogue of

Theorem 3.40 (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 3.35. Then*

$$\begin{aligned}&\mathbb{E}_\pi\left[f(X_\tau, X_{\tau+1}, \dots)\, g_\tau(X_0, \dots, X_\tau)\, 1_{\tau<\infty}\right]\\&\quad= \mathbb{E}_\pi\left[\left[\mathbb{E}_{X_\tau} f(X_0, X_1, \dots)\right] g_\tau(X_0, \dots, X_\tau)\, 1_{\tau<\infty}\right].\end{aligned} \tag{3.30}$$

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 2.17. Indeed,

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

wherein we have used Theorem 2.17 in the third equality. ∎

The analogue of Corollary 2.18 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 3.41. *Let τ be a stopping time, $x \in S$ and π be any probability on S. Then relative to $P_\pi\left(\cdot | \tau < \infty,\ X_\tau = x\right)$, $\left\{X_{\tau+k}\right\}_{k\geq 0}$ is independent of $\left\{X_0, \dots, X_\tau\right\}$ and $\left\{X_{\tau+k}\right\}_{k\geq 0}$ has the same distribution as $\left\{X_k\right\}_{k=0}^{\infty}$ under P_x.*

Proof. According to Eq. (3.30),

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

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

$$\begin{aligned}
&\mathbb{E}_\pi \left[g\left(X_0, \dots, X_\tau\right) f\left(X_\tau, X_{\tau+1}, \dots\right) | \tau < \infty,\ X_\tau = x\right] \\
&\quad = \mathbb{E}_\pi \left[g\left(X_0, \dots, X_\tau\right) | \tau < \infty,\ X_\tau = x\right] \mathbb{E}_x\left[f\left(X_0, X_1, \dots\right)\right]. \qquad (3.31)
\end{aligned}$$

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

$$\mathbb{E}_\pi \left[f\left(X_\tau, X_{\tau+1}, \dots\right) | \tau < \infty,\ X_\tau = x\right] = \mathbb{E}_x\left[f\left(X_0, X_1, \dots\right)\right]. \qquad (3.32)$$

This shows that $\left\{X_{\tau+k}\right\}_{k\geq 0}$ under $P_\pi\left(\cdot | \tau < \infty,\ X_\tau = x\right)$ has the same distribution as $\left\{X_k\right\}_{k=0}^{\infty}$ under P_x and, in combination, Eqs. (3.31) and (3.32) shows $\left\{X_{\tau+k}\right\}_{k\geq 0}$ and $\left\{X_0, \dots, X_\tau\right\}$ are conditionally, on $\left\{\tau < \infty,\ X_\tau = x\right\}$, independent. ∎

To match notation in the book, let

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

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

Proposition 3.42. *Let $i \in S$ and $n \geq 1$. Then P_{ii}^n satisfies the "renewal equation,"*

$$P_{ii}^n = \sum_{k=1}^{n} P(R_i = k) P_{ii}^{n-k}. \tag{3.33}$$

Also if $j \in S$, $k \in \mathbb{N}$, and $\nu : S \to [0,1]$ is any probability on S, then Eq. (3.3) holds, i.e.

$$P_\nu \left(M_j \geq k\right) = P_\nu \left(R_j < \infty\right) \cdot P_j \left(R_j < \infty\right)^{k-1}. \tag{3.34}$$

Proof. To prove Eq. (3.33) 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],

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

For Eq. (3.34) 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 3.41) that;

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

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

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

By the last two displayed equations,

$$P_i(M_j \geq k) = P_j(M_j \geq k-1) P_i(R_j < \infty) \tag{3.35}$$

Taking $i = j$ in this equation shows,

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

and so by induction,

$$P_j(M_j \geq k) = P_j(R_j < \infty)^k. \tag{3.36}$$

Equation (3.34) now follows from Eqs. (3.35) and (3.36). ■

3.4 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 3.14 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 3.43. *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 3.14) and $X_{T_n} = j$, we may apply the strong Markov property in the form of Corollary 3.41 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 3.44. *Suppose that X_n is a irreducible recurrent Markov chain, and let $j \in S$ be a fixed state. Define*

$$\pi_j := \frac{1}{\mathbb{E}_j(R_j)}, \tag{3.37}$$

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

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} 1_{X_n=j} = \pi_j \quad P_i - a.s. \tag{3.38}$$

for all $i \in S$ and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} P_{ij}^n = \pi_j. \tag{3.39}$$

Proof. Let us first note that Eq. (3.39) follows by taking expectations of Eq. (3.38). So we must prove Eq. (3.38).

By Lemma 3.43, 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

$$\lim_{N\to\infty} \frac{\tau_1 + \tau_2 + \cdots + \tau_N}{N} = \mathbb{E}_i\tau_2 = \mathbb{E}_j\tau_1 = \mathbb{E}_j R_j \qquad (P_i\text{– a.s.}). \tag{3.40}$$

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

$$\lim_{N\to\infty} \frac{R_j^{(N)}}{N} = \mathbb{E}_j R_j \qquad (P_i\text{– a.s.}) \tag{3.41}$$

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. $R_j^{(\nu_N)} \le N$. 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. (3.41) shows that

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

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. (3.38). ■

Proposition 3.45. *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. (3.37). 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*

$$\sum_{i\in S} \pi_i = 1 \text{ and} \tag{3.42}$$

$$\sum_{i\in S} \pi_i P_{ij} = \pi_j \text{ for all } j \in S. \tag{3.43}$$

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

Proof. Let us define

$$T_{ki}^n := \frac{1}{n} \sum_{l=1}^{n} P_{ki}^l \tag{3.44}$$

which, according to Theorem 3.44, 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

$$\alpha = \sum_{i\in S} \pi_i = \sum_{i\in S} \lim\inf_{n\to\infty} T^n_{ki} \le \lim\inf_{n\to\infty} \sum_{i\in S} T^n_{ki} = 1$$

and

$$\sum_{i\in S} \pi_i P_{ij} = \sum_{i\in S} \lim_{n\to\infty} T^n_{li} P_{ij} \le \lim\inf_{n\to\infty} \sum_{i\in S} T^n_{li} P_{ij} = \lim\inf_{n\to\infty} T^{n+1}_{lj} = \pi_j$$

where $l \in S$ is arbitrary. Thus

$$\sum_{i\in S} \pi_i =: \alpha \le 1 \text{ and } \sum_{i\in S} \pi_i P_{ij} \le \pi_j \text{ for all } j \in S. \tag{3.45}$$

By induction it also follows that

$$\sum_{i\in S} \pi_i P^k_{ij} \le \pi_j \text{ for all } j \in S. \tag{3.46}$$

So if $\pi_j = 0$ for some $j \in S$, then given any $i \in S$, there is a integer k such that $P^k_{ij} > 0$, and by Eq. (3.46) 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. (3.45) 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. (3.43) is proved.

From Eq. (3.43) and induction we also have

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

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

$$\sum_{i\in S} \pi_i T^k_{ij} = \pi_j \text{ for all } j \in S. \tag{3.47}$$

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. (3.47) to find

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

Since $\pi_j > 0$, this implies that $\alpha = 1$ and hence Eq. (3.42) is now verified. ■

Proposition 3.46. *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

$$\mu_j = \sum_{i\in S}\mu_i T_{ij}^k \text{ for all } k \in \mathbb{N} \tag{3.48}$$

where T_{ij}^k is defined above in Eq. (3.44). As in the proof of Proposition 3.45, 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. (3.48) allows us to conclude that $\mu_j = 0$ for all j which contradicts the fact that μ was assumed to be a distribution. ■

Lemma 3.47 (A number theory lemma). *Suppose that* 1 *is the greatest common denominator of a set of positive integers,* $\Gamma := \{n_1, \dots, n_k\}$. *Then there exists* $N \in \mathbb{N}$ *such that the set,*

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

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,

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

■

4

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 4.1 and 4.2.

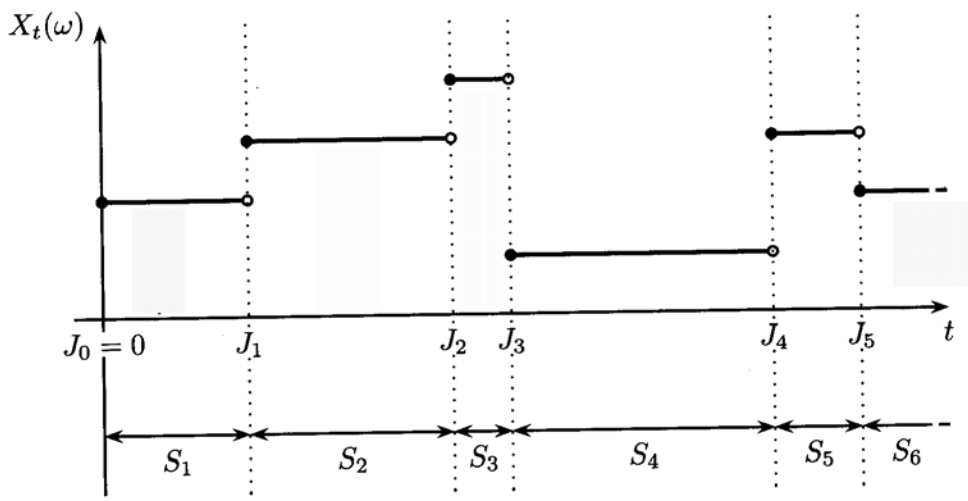


Fig. 4.1. Typical sample paths of a continuous time Markov chain in a discrete state space.

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

$$P_{ij}(t) := P_i(X(t) = j) \tag{4.1}$$

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

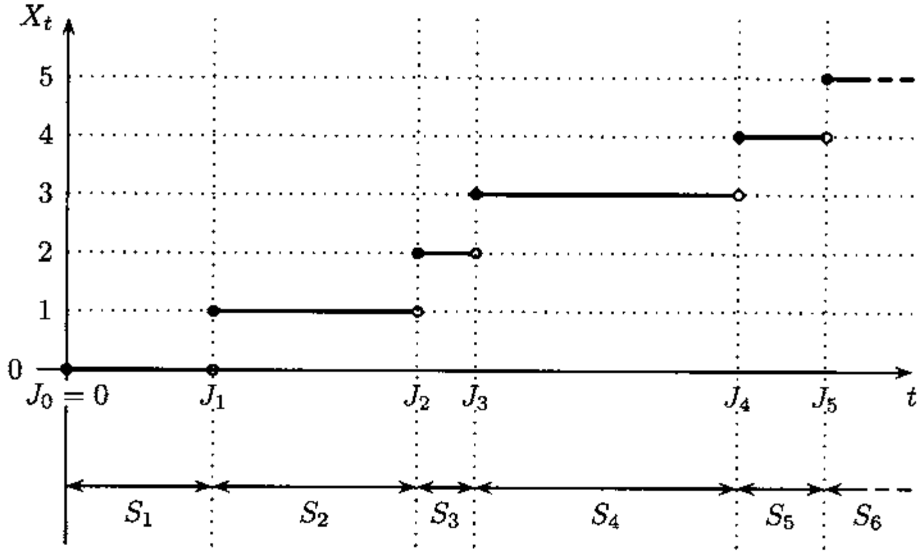


Fig. 4.2. A sample path of a birth process. Here the state space is $\{0, 1, 2, \dots\}$ to be thought of the possible population size.

Definition 4.1. *The* ***time homogeneous Markov property*** *states for every* $0 \le s < t < \infty$ *and any choices of* $0 = t_0 < t_1 < \cdots < t_n = s < t$ *and* $i_1, \dots, i_n \in S$ *that*

$$P_i\left(X\left(t\right) = j | X\left(t_1\right) = i_1, \dots, X\left(t_n\right) = i_n\right) = P_{i_n,j}\left(t - s\right), \tag{4.2}$$

and consequently,

$$P_i\left(X\left(t\right) = j | X\left(s\right) = i_n\right) = P_{i_n,j}\left(t - s\right). \tag{4.3}$$

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

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

So again a continuous time Markov process is forgetful in the sense what the chain does for $t \ge s$ depend only on where the chain is located, $X(s)$, at time s and not how it got there. See Fact 4.3 below for a more general statement of this property.

Definition 4.2 (Informal). *A stopping time,* T, *for* $\{X(t)\}$, *is a random variable with the property that the event* $\{T \le t\}$ *is determined from the knowledge of* $\{X(s) : 0 \le s \le t\}$. *Alternatively put, for each* $t \ge 0$, *there is a functional,* f_t, *such that*

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

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

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

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

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

Theorem 4.4 (Finite dimensional distributions). *Let $0 < t_1 < t_2 < \cdots < t_n$ and $i_0, i_1, i_2, \ldots, i_n \in S$. Then*

$$\begin{aligned} P_{i_0}(X_{t_1} = i_1, X_{t_2} = i_2, \ldots, X_{t_n} = i_n) \\ = P_{i_0,i_1}(t_1) P_{i_1,i_2}(t_2 - t_1) \ldots P_{i_{n-1},i_n}(t_n - t_{n-1}). \end{aligned} \tag{4.4}$$

Proof. The proof is similar to that of Proposition 2.2. 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 4.5 (Properties of P). *Let $P_{ij}(t) := P_i(X(t) = j)$ be as above. Then:*

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

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

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

.

2. $\lim_{t\downarrow 0} P_{ij}(t) = \delta_{ij}$ *for all $i, j \in S$.*

*3. The **Chapman – Kolmogorov equation** holds:*

$$P(t+s) = P(t)P(s) \text{ for all } s,t \geq 0, \tag{4.5}$$

i.e.

$$P_{ij}(t+s) = \sum_{k\in S} P_{ik}(s)\,P_{kj}(t) \text{ for all } s,t \geq 0. \tag{4.6}$$

*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 (4.4) 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 4.6. In this example we will take $S = \{0, 1, 2, \ldots\}$ and then define, for $\lambda > 0$,

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

In components this may be expressed as,

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

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 $i \leq k \leq j$, therefore we have

$$\sum_{k \in S} P_{ik}(t) P_{kj}(s) = 1_{i \le j} \sum_{i \le k \le j} P_{ik}(t) P_{kj}(s)$$
$$= 1_{i \le j} e^{-\lambda(t+s)} \sum_{i \le k \le j} \frac{(\lambda t)^{k-i}}{(k-i)!} \frac{(\lambda s)^{j-k}}{(j-k)!}. \tag{4.7}$$

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

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

and hence by the Binomial formula we find,

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

Combining this with Eq. (4.7) shows that

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

Proposition 4.7. *Let $\{X_t\}_{t \ge 0}$ is the Markov chain determined by $P(t)$ of Example 4.6. 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 $0 = s_0 < s_1 < s_2 < \dots < s_n = s$ and $i_k \in S$ for $k = 0, 1, 2, \dots, n$, then

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

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

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

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 4.8. 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 $\{N(t)\}_{t\geq 0}$ 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. (4.2) 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:

$$Q_{ij} \leq 0 \text{ for all } i \neq j, \text{ and} \tag{4.8}$$

$$\sum_j Q_{ij} = 0, \text{ i.e. } -Q_{ii} = \sum_{j\neq i} Q_{ij} \text{ for all } i. \tag{4.9}$$

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.

$$\dot{\pi}_j(t) = \sum_{i\neq j}\pi_i(t) Qij - q_j\pi_j(t), \tag{4.10}$$

where $q_j = -Q_{j\,j}$. (See Example 5.12 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. 4.10 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 $\#(S) = \infty$.

5

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 5.1. *An $N \times N$ matrix function $P(t)$ for $t \geq 0$ is **Markov semigroup** if*

1. *$P(t)$ is Markov matrix for all $t \geq 0$, i.e. $P_{ij}(t) \geq 0$ for all i, j and*
$$\sum_{j \in S} P_{ij}(t) = 1 \text{ for all } i \in S. \tag{5.1}$$
The condition in Eq. (5.1) may be written in matrix notation as,
$$P(t)\mathbf{1} = \mathbf{1} \text{ for all } t \geq 0. \tag{5.2}$$
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 5.2. *An $N \times N$ matrix, Q, is an **infinitesimal generator** if $Q_{ij} \geq 0$ for all $i \neq j$ and*

$$\sum_{j \in S} Q_{ij} = 0 \text{ for all } i \in S. \tag{5.3}$$

The condition in Eq. (5.3) may be written in matrix notation as,

$$Q\mathbf{1} = 0. \tag{5.4}$$

In this section we are going to make use of the following facts from the theory of linear ordinary differential equations.

Theorem 5.3. *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,*

$$\dot{P}(t) = AP(t) \text{ with } P(0) = B \tag{5.5}$$

which is in fact given by

$$P(t) = e^{tA}B \tag{5.6}$$

where

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \dots \tag{5.7}$$

The matrix function e^{tA} may be characterized as the unique solution Eq. (5.5) 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),

$$e^{(t+s)A} = e^{tA}e^{sA} \text{ for all } s, t \geq 0. \tag{5.8}$$

Proof. We will only prove Eq. (5.8) 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 5.4 (Thanks to Mike Gao!). If $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $A^n = 0$ for $n \geq 2$, so that

$$e^{tA} = I + tA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Similarly if $B = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$, then $B^n = 0$ for $n \geq 2$ and

$$e^{tB} = I + tB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -t & 1 \end{bmatrix}.$$

Now let $C = A + B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. In this case $C^2 = -I$, $C^3 = -C$, $C^4 = I$, $C^5 = C$ etc., so that

$$C^{2n} = (-1)^n I \text{ and } C^{2n+1} = (-1)^n C.$$

Therefore,

$$\begin{aligned} e^{tC} &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} C^{2n} + \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!} C^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} (-1)^n I + \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)!} (-1)^n C \\ &= \cos(t) I + \sin(t) C = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix} \end{aligned}$$

which is the matrix representing rotation in the plan by t degrees.

Here is another way to compute e^{tC} in this example. Since $C^2 = -I$, we find

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

It is now easy to verify the solution to this second order equation is given by,

$$e^{tC} = \cos t \cdot I + \sin t \cdot C$$

which agrees with our previous answer.

Remark 5.5. **Warning:** if A and B are two $N \times N$ matrices it is not generally true that

$$e^{(A+B)} = e^A e^B \tag{5.9}$$

as can be seen from Example 5.4.

However we have the following lemma.

Lemma 5.6. *If A and B commute, i.e. $AB = BA$, then Eq. (5.9) holds. In particular, taking $B = -A$, shows that $e^{-A} = \left[e^A\right]^{-1}$.*

Proof. First proof. Simply verify Eq. (5.9) 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,

$$\begin{aligned}
e^{(A+B)} &= \sum_{n=0}^{\infty} \frac{1}{n!} (A+B)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} \\
&= \sum_{0 \le k \le n < \infty} \frac{1}{k!} \frac{1}{(n-k)!} A^k B^{n-k} \qquad (\text{let } n-k=l) \\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!} \frac{1}{l!} A^k B^l = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \cdot \sum_{l=0}^{\infty} \frac{1}{l!} B^l = e^A e^B.
\end{aligned}$$

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 5.3 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 5.3, 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
$$e^{-B} e^{-A} e^{(A+B)} = I. \tag{5.10}$$
Multiplying Eq. (5.10) on the left by $e^A e^B$ gives Eq. (5.9).

■

The following is the main theorem of this chapter.

Theorem 5.7. *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 5.8 – 5.11 below. (You might look at Example 5.4 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 5.7. The first result is technical and you may safely skip its proof.

Proposition 5.8 (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

$$\frac{d}{dt}\left[P\left(t\right)\Pi\right]=\frac{1}{\varepsilon}\left(P\left(t+\varepsilon\right)-P\left(t\right)\right).$$

As Π is invertible, we may conclude that $P\left(t\right)$ is differentiable and that

$$\dot{P}\left(t\right):=\frac{1}{\varepsilon}\left(P\left(t+\varepsilon\right)-P\left(t\right)\right)\Pi^{-1}.$$

Since the right hand side of this equation is continuous in t it follows that $\dot{P}\left(t\right)$ is continuous as well. ■

Proposition 5.9. *If* $\{P\left(t\right)\}_{t\geq0}$ *is a Markov semi-group and* $Q:=\frac{d}{dt}|_{0+}P\left(t\right)$, *then*

1. $P\left(t\right)$ *satisfies* $P\left(0\right)=I$ *and both,*
$$\dot{P}\left(t\right)=P\left(t\right)Q\qquad\text{(Kolmogorov's forward Eq.)}$$
and
$$\dot{P}\left(t\right)=QP\left(t\right)\qquad\text{(Kolmogorov's backwards Eq.)}$$
hold.
2. $P\left(t\right)=e^{tQ}$.
3. Q *is an infinitesimal generator.*

Proof. 1.-2. We may compute $\dot{P}\left(t\right)$ using

$$\dot{P}\left(t\right)=\frac{d}{ds}|_{0}P\left(t+s\right).$$

We then may write $P\left(t+s\right)$ as $P\left(t\right)P\left(s\right)$ or as $P\left(s\right)P\left(t\right)$ and hence

$$\dot{P}\left(t\right)=\frac{d}{ds}|_{0}\left[P\left(t\right)P\left(s\right)\right]=P\left(t\right)Q\text{ and}$$
$$\dot{P}\left(t\right)=\frac{d}{ds}|_{0}\left[P\left(s\right)P\left(t\right)\right]=QP\left(t\right).$$

This proves Item 1. and Item 2. now follows from Theorem 5.3.

3. Since $P\left(t\right)$ is continuously differentiable, $P\left(t\right)=I+tQ+O\left(t^{2}\right)$, and so for $i\neq j$,

$$0\leq P_{ij}\left(t\right)=\delta_{ij}+tQ_{ij}+O\left(t^{2}\right)=tQ_{ij}+O\left(t^{2}\right).$$

Dividing this inequality by t and then letting $t\downarrow0$ shows $Q_{ij}\geq0$. Differentiating the Eq. (5.2), $P\left(t\right)\mathbf{1}=\mathbf{1}$, at $t=0_{+}$ to show $Q\mathbf{1}=0$. ■

Proposition 5.10. *Let* Q *be any matrix such that* $Q_{ij}\geq0$ *for all* $i\neq j$. *Then* $\left(e^{tQ}\right)_{ij}\geq0$ *for all* $t\geq0$ *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 5.6 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. ∎

Proposition 5.11. *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

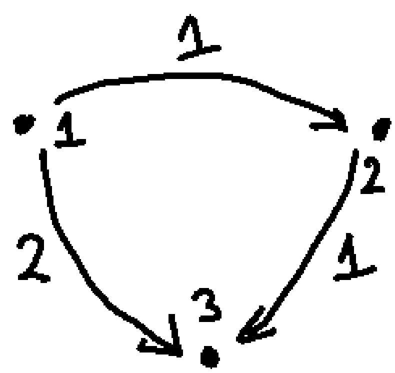
$$\frac{d}{dt}e^{tQ}\mathbf{1} = e^{tQ}Q\mathbf{1} = 0,$$

it follows that $e^{tQ}\mathbf{1} = e^{tQ}\mathbf{1}|_{t=0} = \mathbf{1}$. ∎

Example 5.12. 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 5.12. 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

[1] Actually if you do not want to use Lemma 5.6, 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.

$$\dot{\pi}(t) = \pi e^{tQ} Q = \pi(t) Q.$$

In this particular example this gives,

$$\begin{aligned}\begin{bmatrix} \dot{\pi}_1 & \dot{\pi}_2 & \dot{\pi}_3 \end{bmatrix} &= \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} \begin{bmatrix} -3 & 1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -3\pi_1 & \pi_1 - \pi_2 & 2\pi_1 + \pi_2 \end{bmatrix},\end{aligned}$$

or equivalently,

$$\dot{\pi}_1 = -3\pi_1 \tag{5.11}$$

$$\dot{\pi}_2 = \pi_1 - \pi_2 \tag{5.12}$$

$$\dot{\pi}_3 = 2\pi_1 + \pi_2. \tag{5.13}$$

Notice that these equations are easy to read off from Figure 5.12. 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. (5.11) gives,

$$\pi_1(t) = e^{-3t}\pi_1(0)$$

and therefore Eq. (5.12) becomes

$$\dot{\pi}_2 = e^{-3t}\pi_1(0) - \pi_2$$

which, by Lemma 5.13 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. (5.13) 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

$$\begin{bmatrix}\pi_1\\ \pi_2\\ \pi_3\end{bmatrix} = \begin{bmatrix} e^{-3t}\pi_1(0) \\ \frac{1}{2}\left(e^{-t}-e^{-3t}\right)\pi_1(0) + e^{-t}\pi_2(0) \\ \left(1-\frac{1}{2}\left[e^{-t}+e^{-3t}\right]\right)\pi_1(0) + \left(1-e^{-t}\right)\pi_2(0) + \pi_3(0)\end{bmatrix}$$

$$= \begin{bmatrix} e^{-3t} & 0 & 0 \\ \frac{1}{2}\left(e^{-t}-e^{-3t}\right) & e^{-t} & 0 \\ 1-\frac{1}{2}\left[e^{-t}+e^{-3t}\right] & 1-e^{-t} & 1\end{bmatrix} \begin{bmatrix}\pi_1(t)\\ \pi_2(t)\\ \pi_3(t)\end{bmatrix}.$$

From this we may conclude that

$$P(t) = e^{tQ} = \begin{bmatrix} e^{-3t} & 0 & 0 \\ \frac{1}{2}\left(e^{-t}-e^{-3t}\right) & e^{-t} & 0 \\ 1-\frac{1}{2}\left[e^{-t}+e^{-3t}\right] & 1-e^{-t} & 1\end{bmatrix}^{\mathrm{tr}}$$

$$= \begin{bmatrix} e^{-3t} & \left(\frac{1}{2}e^{-t}-\frac{1}{2}e^{-3t}\right) & \left(1-\frac{1}{2}e^{-t}-\frac{1}{2}e^{-3t}\right) \\ 0 & e^{-t} & -e^{-t}+1 \\ 0 & 0 & 1\end{bmatrix}.$$

Lemma 5.13 (ODE Lemma). *If $h(t)$ is a given function and $\lambda \in \mathbb{R}$, then the solution to the differential equation,*

$$\dot{\pi}(t) = \lambda\pi(t) + h(t) \tag{5.14}$$

is

$$\pi(t) = e^{\lambda t}\left(\pi(0) + \int_0^t e^{-\lambda s}h(s)\,ds\right) \tag{5.15}$$

$$= e^{\lambda t}\pi(0) + \int_0^t e^{\lambda(t-s)}h(s)\,ds. \tag{5.16}$$

Proof. If $\pi(t)$ satisfies Eq. (5.14), 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

$$\pi(t) = e^{\lambda t}\pi(0) + e^{\lambda t}\int_0^t e^{-\lambda s}h(s)\,ds \tag{5.17}$$

which is the same as Eq. (5.15). A direct check shows that $\pi(t)$ so defined solves Eq. (5.14). Indeed using Eq. (5.17) and the fundamental theorem of calculus shows,

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

■

Corollary 5.14. *Suppose $\lambda \in \mathbb{R}$ and $\pi(t)$ is a function which satisfies, $\dot{\pi}(t) \geq \lambda\pi(t)$, then*

$$\pi(t) \geq e^{\lambda t}\pi(0) \ \text{ for all } t \geq 0. \tag{5.18}$$

In particular if $\pi(0) \geq 0$ then $\pi(t) \geq 0$ for all t.

Proof. Let $h(t) := \dot{\pi}(t) - \lambda\pi(t) \geq 0$ and then apply Lemma 5.13 to conclude that

$$\pi(t) = e^{\lambda t}\pi(0) + \int_0^t e^{\lambda(t-s)}h(s)\,ds. \tag{5.19}$$

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. (5.19) leads to the estimate in Eq. (5.18). ■

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)