Homework 2

Benjamin Noland

Linear algebra example

Recall that row-equivalent matrices have the same rank. We have

$$P - I = \begin{pmatrix} -1/2 & 1/2 \\ 1/4 & -1/4 \end{pmatrix} \xrightarrow{\text{Row reduces to}} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} = R, \tag{1}$$

and therefore $\operatorname{rank}(P - I) = \operatorname{rank}(R) = 1$.

To find the subspace V, note that h solves $Ph = h$ if and only if $(P - I)h = 0$, and that the row reduction (1) implies that this latter system has the same solutions as the system $Rh = 0$, which is easily seen to have the solution space

$$V = \left\{ x \begin{pmatrix} 1 \\ 1 \end{pmatrix} : x \in \mathbb{R} \right\}.$$

To find the subspace W, note that μ solves $\mu P = \mu$ if and only if $\mu(P - I) = \mu$, which is equivalent to the system $(P - I)^t \mu^t = \mu^t$. Now,

$$(P - I)^t = \begin{pmatrix} -1/2 & 1/4 \\ 1/2 & -1/4 \end{pmatrix} \xrightarrow{\text{Row reduces to}} \begin{pmatrix} 1 & -1/2 \\ 0 & 0 \end{pmatrix} = R,$$

and thus $(P - I)^t \mu^t = \mu^t$ has the same solutions as the system $R\mu^t = 0$, which is easily seen to have the solution space

$$W = \left\{ x \begin{pmatrix} 1 \\ 2 \end{pmatrix} : x \in \mathbb{R} \right\}.$$

Basic properties of Markov chains

Exercise 1

For the first part of this problem, we have

$$\begin{aligned} \mathrm{P}(ABCDE) &= \mathrm{P}(E|ABCD)\mathrm{P}(ABCD) \\ &= \mathrm{P}(E|ABCD)\mathrm{P}(D|ABC)\mathrm{P}(ABC) \\ &= \mathrm{P}(E|ABCD)\mathrm{P}(D|ABC)\mathrm{P}(C|AB)\mathrm{P}(AB) \\ &= \mathrm{P}(E|ABCD)\mathrm{P}(D|ABC)\mathrm{P}(C|AB)\mathrm{P}(B|A)\mathrm{P}(A) \end{aligned}$$

through basic properties of conditional probability.

As for the generalization to events $A_1, \dots A_n$, I claim that

$$\mathrm{P}(A_1 \cdots A_n) = \mathrm{P}(A_n | A_{n-1} \cdots A_1)\mathrm{P}(A_{n-1}|A_{n-2} \cdots A_1) \cdots \mathrm{P}(A_2|A_1)\mathrm{P}(A_1). \tag{1}$$

To prove this, I will proceed by induction on n. When $n = 1$ the result is trivial, so let $n > 1$ and suppose that (1) holds for any events $A_1, \dots A_n$. Let A_{n+1} be an arbitrary event. Then

$$\begin{aligned}\mathrm{P}(A_1 \cdots A_n A_{n+1}) &= \mathrm{P}(A_{n+1}|A_n \cdots A_1)\mathrm{P}(A_1 \cdots A_n) \\ &= \mathrm{P}(A_{n+1}|A_n \cdots A_1)\mathrm{P}(A_n|A_{n-1} \cdots A_1)\mathrm{P}(A_{n-1}|A_{n-2} \cdots A_1) \cdots \mathrm{P}(A_2|A_1)\mathrm{P}(A_1)\end{aligned}$$

by the induction hypothesis. Hence (1) holds for every n by way of induction.

Exercise 2

By Exercise 1 and the Markov property,

$$\begin{aligned}&\mathrm{P}_\mu\{X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0\} \\ &\quad= \mathrm{P}_\mu\{X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_0 = x_0\}\mathrm{P}_\mu\{X_{t-1} = x_{t-1}|X_{t-2} = x_{t-2}, \dots, X_0 = x_0\} \\ &\qquad\quad \times \cdots \times \mathrm{P}_\mu\{X_1 = x_1 | X_0 = x_0\}\mathrm{P}_\mu\{X_0 = x_0\} \\ &\quad= \mathrm{P}_\mu\{X_t = x_t|X_{t-1} = x_{t-1}\}\mathrm{P}_\mu\{X_{t-1} = x_{t-1}|X_{t-2} = x_{t-2}\} \\ &\qquad\quad \times \cdots \times \mathrm{P}_\mu\{X_1 = x_1|X_0 = x_0\}\mathrm{P}_\mu\{X_0 = x_0\} \\ &\quad= P(x_{t-1}, x_t)P(x_{t-2}, x_{t-1}) \cdots P(x_0, x_1)\mu(x_0).\end{aligned}$$

Exercise 3

$$\begin{aligned}\mathrm{P}_\mu\{f(X_0, \dots, X_k) = 1\} &= \sum_{x \in \mathcal{X}} \mathrm{P}_\mu\{f(X_0, \dots, X_k) = 1 | X_0 = x\}\mathrm{P}_\mu\{X_0 = x\} \\ &= \sum_{x \in \mathcal{X}} \mathrm{P}_x\{f(X_0, \dots, X_k) = 1\}\mu(x)\end{aligned}$$

Exercise 4

$$\mathrm{P}_\mu\{f(X_0, \dots, X_k) = 1\} = \mathrm{P}_\mu\{(X_0, \dots, X_k) \in \mathcal{P}\} = \sum_{(x_0, \dots, x_k) \in \mathcal{P}} \mathrm{P}_\mu\{X_0 = x_0, \dots, X_k = x_k\}$$

Exercise 5

In order to clarify the manipulations that follow, I will define the following sets:

$$\begin{aligned}\mathcal{P} &= \{(x_0, \dots, x_t) \in \mathcal{X}^{t+1} : p(x_0, \dots, x_t) = 1\} \\ \mathcal{F} &= \{(x_t, \dots, x_{t+k}) \in \mathcal{X}^{k+1} : f(x_t, \dots, x_{t+k}) = 1\}.\end{aligned}$$

Then we can write

$$
\begin{aligned}
&\mathrm{P}_\mu\{f(X_t,\ldots,X_{t+k})=1|X_t=x_t,p(X_0,\ldots,X_t)=1\}\\
&\quad=\mathrm{P}_\mu\{(X_t,\ldots,X_{t+k})\in\mathcal{F}|X_t=x_t,(X_0,\ldots,X_t)\in\mathcal{P}\}\\
&\quad=\mathrm{P}_\mu\{(X_t,\ldots,X_{t+k})\in\mathcal{F}|X_t=x_t\}\\
&\quad=\mathrm{P}_\mu\{f(X_t,\ldots,X_{t+k})=1|X_t=x_t\}
\end{aligned}
$$

by way of the Markov property. Moreover,

$$
\begin{aligned}
\mathrm{P}_\mu\{f(X_t,\ldots,X_{t+k})=1|X_t=x_t\}&=\mathrm{P}_\mu\{f(X_0,\ldots,X_k)=1|X_0=x_t\}\\
&=\mathrm{P}_{x_t}\{f(X_0,\ldots,X_k)=1\}
\end{aligned}
$$

since forcing the chain to be in state x_t at time t has the same effect on the state of the chain at time $t'>t$ that starting the chain in state x_t has on the state of the chain at time $t'-t>0$.

Exercise 6

By Exercise 5 and basic properties of conditional probability,

$$
\begin{aligned}
&\mathrm{P}_\mu\{f(X_t,\ldots,X_{t+k})=1,p(X_0,\ldots,X_t)=1|X_t=x_t\}\\
&\quad=\mathrm{P}_\mu\{f(X_t,\ldots,X_{t+k})=1|p(X_0,\ldots,X_t)=1,X_t=x_t\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|X_t=x_t\}\\
&\quad=\mathrm{P}_\mu\{f(X_t,\ldots,X_{t+k})=1|X_t=x_t\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|X_t=x_t\}\\
&\quad=\mathrm{P}_{x_t}\{f(X_0,\ldots,X_k)=1\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|X_t=x_t\}.
\end{aligned}
$$

Exercise 7

In order to clarify the manipulations that follow, I will define the following sets:

$$
\begin{aligned}
&\mathcal{P}=\{(x_0,\ldots,x_t)\in\mathcal{X}^{t+1}:p(x_0,\ldots,x_t)=1\}\\
&\mathcal{F}=\{(x_{t+m},\ldots,x_{t+m+k})\in\mathcal{X}^{k+1}:f(x_{t+m},\ldots,x_{t+m+k})=1\}.
\end{aligned}
$$

Then we can write

$$
\begin{aligned}
&\mathrm{P}_\mu\{f(X_{t+m},\ldots,X_{t+m+k})=1,p(X_0,\ldots,X_t)=1|E\}\\
&\quad=\mathrm{P}_\mu\{f(X_{t+m},\ldots,X_{t+m+k})=1|E\cap\{p(X_0,\ldots,X_t)=1\}\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|E\}\\
&\quad=\mathrm{P}_\mu\{(X_{t+m},\ldots,X_{t+m+k})\in\mathcal{F}|E\cap\{(X_0,\ldots,X_t)\in\mathcal{P}\}\}\mathrm{P}_\mu\{(X_0,\ldots,X_t)\in\mathcal{P}|E\}\\
&\quad=\mathrm{P}_\mu\{(X_{t+m},\ldots,X_{t+m+k})\in\mathcal{F}|E\}\mathrm{P}_\mu\{(X_0,\ldots,X_t)\in\mathcal{P}|E\}\\
&\quad=\mathrm{P}_\mu\{f(X_{t+m},\ldots,X_{t+m+k})=1|E\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|E\}
\end{aligned}
$$

by way of the Markov property and basic properties of conditional probability. Moreover, the Markov property and the result of Exercise 5 imply that

$$\begin{aligned}
&\mathrm{P}_\mu\{f(X_{t+m},\ldots,X_{t+m+k})=1|E\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|E\}\\
&\quad=\mathrm{P}_\mu\{(X_{t+m},\ldots,X_{t+m+k})\in\mathcal{F}|X_{t+m}=x_{t+m}\}\mathrm{P}_\mu\{(X_0,\ldots,X_t)\in\mathcal{P}|X_t=x_t\}\\
&\quad=\mathrm{P}_\mu\{f(X_{t+m},\ldots,X_{t+m+k})=1|X_{t+m}=x_{t+m}\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|X_t=x_t\}\\
&\quad=\mathrm{P}_{x_{t+m}}\{f(X_0,\ldots,X_k)=1\}\mathrm{P}_\mu\{p(X_0,\ldots,X_t)=1|X_t=x_t\}.
\end{aligned}$$

Note that the above equalities still hold if we condition on $\{X_t=x_t,X_{t+m}=x_{t+m}\}$ rather than E. This is because, due to the Markov property, conditioning on E in the above is equivalent to conditioning on either $X_t=x_t$ or $X_{t+m}=x_{t+m}$ depending on context; specifically,

$$\begin{aligned}
&\mathrm{P}_\mu\{(X_{t+m},\ldots,X_{t+m+k})\in\mathcal{F}|E\}=\mathrm{P}_\mu\{(X_{t+m},\ldots,X_{t+m+k})\in\mathcal{F}|X_{t+m}=x_{t+m}\}\\
&\mathrm{P}_\mu\{(X_0,\ldots,X_t)\in\mathcal{P}|E\}=\mathrm{P}_\mu\{(X_0,\ldots,X_t)\in\mathcal{P}|X_t=x_t\}.
\end{aligned}$$

Problems from the book

Exercise 1.2

Suppose G is connected. Let $x,y\in V$. Since G is connected there exists a sequence $x=x_0,x_1,\ldots,x_k=y$ of vertices leading from x to y. In particular, the definition of the random walk on G implies that there is a positive probability of reaching y from x in precisely k steps. Hence $P^k(x,y)>0$, so that the random walk is irreducible.

Conversely, suppose the random walk on G is irreducible. Let $x,y\in V$. Since the random walk is irreducible there exists an integer $k\geq 0$ with $P^k(x,y)>0$. The definition of the random walk on G therefore implies that it is possible to reach y from x in precisely k steps; that is, there exists a sequence $x=x_0,x_1,\ldots,x_k=y$ of vertices leading from x to y. Hence G is connected.

Exercise 1.6

Fix a state $x_0\in\mathcal{X}$, and for each $1\leq i\leq b-1$ define a set

$$C_i=\{x\in\mathcal{X}:P^{mb+i}(x_0,x)>0\text{ for some }m\geq 0\}.$$

I claim that the C_i's form a partition of $\mathcal{X}$. Let $x\in\mathcal{X}$. Then since the chain is irreducible there exists an integer $t\geq 0$ with $P^t(x_0,x)>0$. However, $t=mb+i$ for some choice of integers $m\geq 0$ and $0\leq i\leq b-1$. Hence $x\in C_i$. Now suppose $x\in C_j$ as well for some other index $0\leq j\leq b-1$. Since the chain is irreducible there exists an integer $r\geq 0$ with $P^r(x,x_0)>0$. Hence

$$P^{mb+i+r}(x_0,x_0)\geq P^{mb+i}(x_0,x)P^r(x,x_0)>0.$$

Thus, since the chain has period b, we have that b divides $(mb+i+r)$, and hence divides $(i+r)$. Similarly, b divides $(j+r)$. If $i \geq j$, then b divides $(i+r)-(j+r) = i-j \geq 0$, and since $0 \leq i, j \leq b-1$, we must have $i-j=0$, so that $i=j$. If $i \leq j$, then we reach the same conclusion. Thus $i=j$ in any case, so that x belongs to one and only one of the C_i's. So the C_i's partition $\mathcal{X}$, as claimed.

Now let $x, y \in \mathcal{X}$ satisfy $P(x,y) > 0$. By the above argument, $x \in C_i$ for some $0 \leq i \leq b-1$. Thus $P^{mb+i}(x_0, x) > 0$ for some integer $m \geq 0$, so that

$$P^{mb+(i+1)}(x_0, y) \geq P^{mb+i}(x_0, x)P(x,y) > 0.$$

Hence $y \in C_{i+1}$ (where the addition $(i+1)$ is modulo b).

Exercise 1.7

Note that the stationary distribution π on $\mathcal{X}$ is given by

$$\pi(x) = \frac{1}{|\mathcal{X}|} \quad \text{for any } x \in \mathcal{X}.$$

In particular, $\pi(x) = \pi(y)$ for any $x, y \in \mathcal{X}$. Thus for any $y \in \mathcal{X}$, the symmetry of P implies that

$$\sum_{x \in \mathcal{X}} \pi(x)P(x,y) = \sum_{x \in \mathcal{X}} \pi(y)P(y,x) = \pi(y) \sum_{x \in \mathcal{X}} P(y,x) = \pi(y),$$

so that π is stationary for P.

Exercise 1.8

For any $x, y \in \mathcal{X}$,

$$\begin{aligned}
\pi(x)P^2(x,y) = \pi(x) \sum_{z \in \mathcal{X}} P(x,z)P(z,y) &= \sum_{z \in \mathcal{X}} \pi(x)P(x,z)P(z,y) \\
&= \sum_{z \in \mathcal{X}} \pi(z)P(z,x)P(z,y) = \sum_{z \in \mathcal{X}} \pi(z)P(z,y)P(z,x) \\
&= \sum_{z \in \mathcal{X}} \pi(y)P(y,z)P(z,x) = \pi(y) \sum_{z \in \mathcal{X}} P(y,z)P(z,x) \\
&= \pi(y)P^2(y,x).
\end{aligned}$$

Thus P^2 is also reversible with respect to π.

Exercise 1.9

Let $z \in \mathcal{X}$ be the starting state of the chain. For any $y \in \mathcal{X}$, define a random variable N_y to be the number of visits the chain makes to y before returning to z. Then the distribution $\tilde{\pi}$ is defined by

$$\tilde{\pi}(y) = \mathrm{E}_z(N_y) \quad \text{for any } y \in \mathcal{X}.$$

For any $y \in \mathcal{X}$ and integer $t \geq 0$, define a random variable $N_{y,t}$ to be the number of visits the chain makes to y at time t before returning to z (hence $N_{y,t} = 0$ or $N_{y,t} = 1$). Note that the condition $N_{y,t} = 1$ is equivalent to the condition that the chain be in state y at time t, and that $\tau_z^+ > t$. Therefore,

$$E(N_{y,t}) = 0 \cdot \mathrm{P}\{N_{y,t} = 0\} + 1 \cdot \mathrm{P}\{N_{y,t} = 1\} = \mathrm{P}\{N_{y,t} = 1\} = \mathrm{P}\{X_t = y, \tau_z^+ > t\}.$$

Moreover, note that $N_y = \sum_{t=0}^{\infty} N_{y,t}$. Thus, for any $y \in \mathcal{X}$,

$$\tilde{\pi}(y) = \mathrm{E}_z(N_y) = \mathrm{E}_z\left(\sum_{t=0}^{\infty} N_{y,t}\right) = \sum_{t=0}^{\infty} \mathrm{E}_z(N_{y,t}) = \sum_{t=0}^{\infty} \mathrm{P}\{X_t = y, \tau_z^+ > t\}.$$

Exercise 1.11

First, note that if π is a stationary distribution for an irreducible Markov chain, then $\pi(x) > 0$ for every $x \in \mathcal{X}$. To see this, let $x \in \mathcal{X}$. Since π is a probability distribution, there exists a state $z \in \mathcal{X}$ with $\pi(z) > 0$. Then since the chain is irreducible there exists an integer $t \geq 0$ with $P^t(z, x) > 0$. Therefore, since π is stationary,

$$\pi(x) = \sum_{y \in \mathcal{X}} \pi(y) P^t(y, x) \geq \pi(z) P^t(z, x) > 0.$$

In particular, this means that the ratio $\pi_1(x)/\pi_2(x)$ is defined for any $x \in \mathcal{X}$. Thus, since the state space $\mathcal{X}$ is finite, there exists a state $x \in \mathcal{X}$ that minimizes the ratio $\pi_1(x)/\pi_2(x)$. The fact that π_1 and π_2 are stationary for this chain implies that for any integer $t \geq 0$,

$$\pi_1(x) = \sum_{y \in \mathcal{X}} \pi_1(y) P^t(y, x) \tag{1}$$

and

$$\pi_1(x) = \frac{\pi_1(x)}{\pi_2(x)} \pi_2(x) = \sum_{y \in \mathcal{X}} \frac{\pi_1(x)}{\pi_2(x)} \pi_2(y) P^t(y, x). \tag{2}$$

Subtracting (2) from (1), we get

$$\sum_{y \in \mathcal{X}} \left[\pi_1(y) - \frac{\pi_1(x)}{\pi_2(x)} \pi_2(y)\right] P^t(y, x) = 0. \tag{3}$$

Note that for any $y \in \mathcal{X}$,

$$\frac{\pi_1(x)}{\pi_2(x)} \pi_2(y) \leq \frac{\pi_1(y)}{\pi_2(y)} \pi_2(y) = \pi_1(y)$$

since x minimizes the ratio $\pi_1(x)/\pi_2(x)$ by construction. Thus each term in the sum (3) is zero, and hence for any $y \in \mathcal{X}$ with $P^t(y, x) > 0$,

$$\pi_1(y) = \frac{\pi_1(x)}{\pi_2(x)} \pi_2(y). \tag{4}$$

However, since the chain is irreducible, we can always choose $t \geq 0$ such that $P^t(y, x) > 0$ for a given state $y \in \mathcal{X}$, and hence by rearranging (4) we see that

$$\frac{\pi_1(y)}{\pi_2(y)} = \frac{\pi_1(x)}{\pi_2(x)} \quad \text{for any } y \in \mathcal{X}.$$

In particular, $\pi_1(y)/\pi_2(y) = \pi_1(z)/\pi_2(z)$ for any $y, z \in \mathcal{X}$, and therefore,

$$1 = \sum_{y \in \mathcal{X}} \pi_1(y) = \sum_{y \in \mathcal{X}} \frac{\pi_1(z)}{\pi_2(z)} \pi_2(y) = \frac{\pi_1(z)}{\pi_2(z)},$$

so that $\pi_1(z) = \pi_2(z)$. Thus $\pi_1 = \pi_2$, so that the chain has a unique stationary distribution.