1 Two-State Chain With Linear Algebra

(a) $P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$

(b) We have $Pu = \lambda u$ for a vector u and scalar λ. This can be rewritten as $(\lambda I - P)u = 0$. Since u is in the nullspace of $\lambda I - P$, we can set $\det(\lambda I - P) = 0$ to find the characteristic polynomial of P and solve for the eigenvalues which are the roots of the characteristic polynomial. We have that

$$\begin{aligned}
\det(\lambda I - P) &= 0 \\
\begin{vmatrix} \lambda - 1 + \alpha & -\alpha \\ -\beta & \lambda - 1 + \beta \end{vmatrix} &= 0 \\
(\lambda - 1 + \alpha)(\lambda - 1 + \beta) - \alpha\beta &= 0 \\
\lambda^2 + (-1 + \alpha - 1 + \beta)\lambda + (-1 + \alpha)(-1 + \beta) - \alpha\beta &= 0 \\
\lambda^2 + (-2 + \alpha + \beta)\lambda + (1 - \alpha - \beta + \alpha\beta) - \alpha\beta &= 0 \\
\lambda^2 - 2\lambda + \lambda\alpha + \lambda\beta + 1 - \alpha - \beta &= 0 \\
\lambda(\lambda - 2) + \alpha(\lambda - 1) + \beta(\lambda - 1) + 1 &= 0 \\
(\lambda - 1)^2 + (\lambda - 1)(\alpha + \beta) &= 0 \\
(\lambda - 1)((\lambda - 1) + (\alpha + \beta)) &= 0.
\end{aligned}$$

Solving for the roots, we have $\lambda_1 = 1$ and $\lambda_2 = 1 - \alpha - \beta$. To find the eigenvectors u_1 and u_2 corresponding to these eigenvalues, we plug the eigenvalues into our equation $(\lambda I - P)u = 0$. For $\lambda = \lambda_1 = 1$ we get

$$\begin{aligned}
(\lambda_1 I - P)u_1 &= 0 \\
\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} u_1 &= 0.
\end{aligned}$$

From this, we can see that $u_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. For $\lambda = \lambda_2 = 1 - \alpha - \beta$ we get

$$\begin{aligned}
(\lambda_2 I - P)u_2 &= 0 \\
\begin{bmatrix} -\beta & -\alpha \\ -\beta & -\alpha \end{bmatrix} u_2 &= 0.
\end{aligned}$$

From this, we can see that $u_2 = \begin{bmatrix} 1 \\ -\frac{\beta}{\alpha} \end{bmatrix}$. From these results, we can see that $PU = U\Lambda$, where $U = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$ and $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{bmatrix}$. Since $PU = U\Lambda$, we can also write this as $P = U\Lambda U^{-1}$, where U and Λ are 2×2 matrices as required and Λ is the diagonal matrix of eigenvalues.

(c) $P^n = U\Lambda^n U^{-1}$, since $P = U\Lambda U^{-1}$.

(d) We have $\pi_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, so the PMF of X_n is given by

$$\begin{aligned}
\mathbb{P}[X_n = i] &= \pi_0 P^n(i) \\
&= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -\frac{\beta}{\alpha} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{bmatrix} \left(\frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \\ \alpha & -\alpha \end{bmatrix} \right) (i) \\
&= \frac{1}{\alpha+\beta} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{bmatrix} \begin{bmatrix} \beta & \alpha \\ \alpha & -\alpha \end{bmatrix} (i) \\
&= \frac{1}{\alpha+\beta} \begin{bmatrix} 1 & (1-\alpha-\beta)^n \end{bmatrix} \begin{bmatrix} \beta & \alpha \\ \alpha & -\alpha \end{bmatrix} (i) \\
&= \frac{1}{\alpha+\beta} \begin{bmatrix} \beta + \alpha(1-\alpha-\beta)^n & \alpha - \alpha(1-\alpha-\beta)^n \end{bmatrix} (i).
\end{aligned}$$

(e) $\lim_{n\to\infty} \mathbb{P}[X_n = 0] = \lim_{n\to\infty} \frac{\beta+\alpha(1-\alpha-\beta)^n}{\alpha+\beta} = \frac{\beta}{\alpha+\beta}$, since $|1-\alpha-\beta| < 1$.

2 Reducible Markov Chain

(a) See the Figures below.

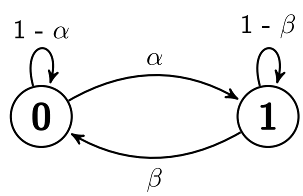

Figure 1: Recurrent class

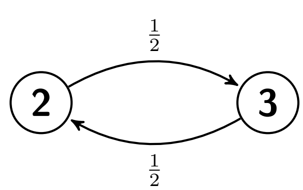

Figure 2: Transient class

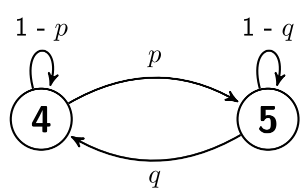

Figure 3: Recurrent class

(b) $\mathbb{P}[T_0 < T_5 \mid X_0 = 2] = \alpha(2)$. Using the recurrence relations, we can get the following equations:

$$\begin{aligned}
\alpha(2) &= \frac{1}{2}\alpha(1) + \frac{1}{2}\alpha(3) \\
\alpha(1) &= (1-\beta)\alpha(1) + \beta\alpha(0) = 1 \\
\alpha(0) &= 1 \\
\alpha(3) &= \frac{1}{2}\alpha(4) + \frac{1}{2}\alpha(2) \\
\alpha(4) &= (1-p)\alpha(4) + p\alpha(5) = 0 \\
\alpha(5) &= 0.
\end{aligned}$$

Solving for $\alpha(2)$, we get that

$$\begin{aligned}
\alpha(2) &= \frac{1}{2}\alpha(1) + \frac{1}{2}\alpha(3) \\
&= \frac{1}{2}(1) + \frac{1}{2}\left(\frac{1}{2}\alpha(4) + \frac{1}{2}\alpha(2)\right) \\
&= \frac{1}{2} + \frac{1}{4}(0 + \alpha(2)) \\
&= \frac{2}{3}.
\end{aligned}$$

(c) The stationary distributions must of of the form $\pi = \left[c\frac{\beta}{\alpha+\beta} \quad c\frac{\alpha}{\alpha+\beta} \quad 0 \quad 0 \quad (1-c)\frac{q}{p+q} \quad (1-c)\frac{p}{p+q}\right]$, where $c \in [0,1]$. Since $\pi(i)$ represents the long-term fraction of time spent in state i, $\pi(2) = \pi(3) = 0$ as 2 & 3 are in the transient class. Since the chain ultimately ends up in one of the recurrent classes, we can use our result from 1(e) to find $\pi(0), \pi(1), \pi(4), \pi(5)$, where they are symmetrically of the form $\frac{\beta}{\alpha+\beta}$. However, we must also account for a constant factor c (and symmetrically, $1-c$), which represents the probability of ending up in the recurrent class $\{0,1\}$ (and symmetrically, $\{4,5\}$), based on the initial distribution.

(d) Yes, the distribution of the chain converges to the stationary distribution in part (c). In particular, we have that $c = \gamma\alpha(2) + (1-\gamma)\alpha(3) = \frac{2\gamma}{3} + (1-\gamma)\frac{1}{3} = \frac{\gamma}{3} + \frac{1}{3}$. Since the recurrent classes themselves are irreducible & aperiodic, their respective distributions will converge a.s. as $n \to \infty$ to their respective stationary distributions, scaled by the corresponding probabilities of ending up in the respective classes, as in part (c).

3 Product of Rolls of a Die

We can effectively model this with a Markov chain as follows. Let S represent the start state (where no dice have been rolled yet), and A represent rolling a 1 or a 5 (both of which do not multiply up to 12 with any other roll), and let B represent rolling a 2, 3, 4, or 6 (where one of the rolls $\in B$ multiplies to 12 with another roll). Finally, let C denote the exit state of obtaining a product of 12 from the last 2 rolls. Then, we have the following state transition diagram:

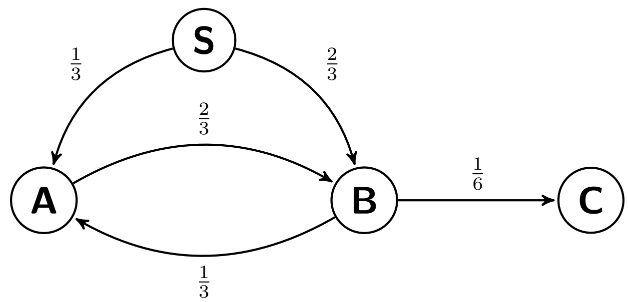

Figure 4: State transition diagram for rolling a product of 12 from the last 2 rolls

Now we can just calculate the mean hitting time to C starting from S, $\beta(S)$ from the recurrence relations and equations as follows:

$$\begin{aligned}
\beta(S) &= 1 + \frac{1}{3}\beta(A) + \frac{2}{3}\beta(B) \\
\beta(A) &= 1 + \frac{1}{3}\beta(A) + \frac{2}{3}\beta(B) \\
\beta(B) &= 1 + \frac{1}{3}\beta(A) + \frac{1}{2}\beta(B) + \frac{1}{6}\beta(C) \\
\beta(C) &= 0.
\end{aligned}$$

Solving for $\beta(S)$, we get that $\beta(S) = 10.5$.

4 Metropolis-Hastings Algorithm

(a) Simulating the Markov chain can be done efficiently as the ratio $\frac{\pi(y)}{\pi(x)}$ can be efficiently computed as $\frac{\tilde{\pi}(y)}{\tilde{\pi}(x)}$ since the normalizing constant will cancel out.

(b) Since $\pi_k(x)P(x,y) = \pi_k(y)P(y,x)$, we have that

$$\begin{aligned}
\pi_{k+1}(y) &= \sum_{x \in \mathcal{X}} \pi_k(x)P(x,y) \\
&= \sum_{x \in X} \pi_k(y)P(y,x) \\
&= \pi_k(y) \sum_{x \in \mathcal{X}} P(y,x) \\
&= \pi_k(y).
\end{aligned}$$

Since $\pi_{k+1}(y) = \pi_k(y)\ \forall y$, we have that $\pi P = \pi$, and thus π is the stationary distribution of the chain.

(c) For the Metropolis-Hastings chain, to get the term $\pi(x)P(x,y)$, we note that $P(x,y)$ can be represented as $f(x,y)A(x,y) = f(x,y)\min\left(1, \frac{\pi(y)f(y,x)}{\pi(x)f(x,y)}\right)$. Similarly, we have $P(y,x) = f(y,x)A(y,x) = f(y,x)\min\left(1, \frac{\pi(x)f(x,y)}{\pi(y)f(y,x)}\right)$. Now we consider the different cases we can arrive at. For $\pi(y)f(y,x) > \pi(x)f(x,y)$, we have that $P(x,y) = f(x,y)\min\left(1, \frac{\pi(y)f(y,x)}{\pi(x)f(x,y)}\right) = f(x,y)$. Correspondingly, $P(y,x) = f(y,x)\frac{\pi(x)f(x,y)}{\pi(y)f(y,x)} = \frac{\pi(x)f(x,y)}{\pi(y)}$. As a result, we have $\pi(x)P(x,y) = \pi(x)f(x,y) = \pi(y)P(y,x)$, satisfying detailed balance. For $\pi(y)f(y,x) = \pi(x)f(x,y)$, we get $P(x,y) = f(x,y)$ and $P(y,x) = f(y,x)$, so that $\pi(x)P(x,y) = \pi(x)f(x,y) = \pi(y)f(y,x) = \pi(y)P(y,x)$, also satisfying detailed balance. Finally, for $\pi(y)f(y,x) < \pi(x)f(x,y)$, we have a complementary symmetric case to $\pi(y)f(y,x) > \pi(x)f(x,y)$, and can conclude that detailed balance also holds under this condition as well. From part (b), we can thus conclude that π is the stationary distribution of the chain.

(d) The lazy chain is aperiodic because of the forced self-loop from a state to itself. This ensures that it is possible to reach the same state again in 1 step, or that $P_{ii} \geq \frac{1}{2} > 0$, such that $d(i) = \gcd\left(n \geq 1 \mid P_{ii}^n > 0\right) = 1\ \forall i$ and so the lazy chain is aperiodic. The stationary distribution is the same as before since only a constant factor is introduced on both sides of the detailed balance equation. More concretely, we now have that $P(x,y) = \frac{1}{2}f(x,y)A(x,y)$ and $P(y,x) = \frac{1}{2}f(y,x)A(y,x)$. Thus our analysis in part (c) still holds as the constant factor of $\frac{1}{2}$ is introduced on both sides, and $f(x,y)A(x,y)$ and $f(y,x)A(y,x)$ are not modified. Thus, the stationary distribution is the same as before.

5 Reversible Markov Chains

(a) Since x is a leaf node in the tree from the graph of an irreducible Markov chain, then this must mean that x only had non-zero probability transitions to itself or y, since it would not be a leaf node in the graph of the Markov chain otherwise. Furthermore, since π is the stationary distribution of the Markov chain, we have that $\sum_{j\neq i}\pi(j)P_{ji} = \pi(i)\sum_{i\neq j}P_{ij}$, or that flow in = flow out. Since x only flows out to y and only receives flow back in from y (when discluding self loops), we thus have that $\pi(y)P(y,x) = \pi(x)P(x,y)$, satisfying detailed balance.

(b) After removing the leaf x from the Markov chain, this does not affect the stationary distribution of the chain for the rest of the states that are not x or y, since x only had non-zero probability transitions to itself or y. As such, the balance equation still hold for these states. Since the probability of a self-transition at y is increased by $P(y,x)$, this also balances the balance equations for y. In particular, removing x removes the transitions $P(y,x)$ and $P(x,y)$ (as well as $P(x,x)$, but this is not part of y's balance equations). Moreover, by increasing the self-transition at y by $P(y,x)$ we account for this since by detailed balance from part (a), $\pi(y)P(y,x) = \pi(x)P(x,y)$, and a self-loop occurs exactly on both sides of the balance equations for y, so that $\pi(y)P(y,x) = \pi(x)P(x,y)$ is the flow from the self-loop that exits and also enters into y, as desired. Thus the balance equations also hold for y, and the stationary distribution of the original chain restricted to $\mathcal{X}\setminus\{x\}$ is the same as that for the new chain. By induction, we see that every state $i \in \mathcal{X}$ satisfies the detailed balance equations on the correspondingly restricted states, and thus the Markov chain is reversible.

6 [Bonus] Entropy Rate of a Markov Chain

(a)

(b)