Solutions 6

1. a) We have

$$\mathbb{P}(X \neq Y) = 1 - \mathbb{P}(X = Y) = 1 - \sum_{i \in S} \xi_i = 1 - \sum_{i \in S} \min(\mu_i, \nu_i)$$
$$= \sum_{i \in S \,:\, \mu_i > \nu_i} (\mu_i - \nu_i) = \sum_{i \in S \,:\, \nu_i > \mu_i} (\nu_i - \mu_i)$$

where the last two equalities hold because both μ and ν are distributions. Summing these last two equalities, we obtain

$$2\,\mathbb{P}(X \neq Y) = \sum_{i \in S} |\mu_i - \nu_i| = 2\ \|\mu - \nu\|_{\mathrm{TV}}$$

b) Observe first that if $\sum_{i \in S} \xi_i = 1$, then $X = Y$ with probability one. When $\sum_{i \in S} \xi_i < 1$, we obtain for $i \in \mathcal{S}$:

$$\mathbb{P}(X = i) = \mathbb{P}(X = Y = i) + \sum_{j \in S \setminus i} \mathbb{P}(X = i, Y = j) = \xi_i + \sum_{j \in S \setminus i} \frac{(\mu_i - \xi_i)\,(\nu_j - \xi_j)}{1 - \sum_{k \in S} \xi_k}$$
$$= \xi_i + \frac{\mu_i - \xi_i}{1 - \sum_{k \in S} \xi_k} \sum_{j \in S \setminus i} (\nu_j - \xi_j) = \xi_i + \frac{\mu_i - \xi_i}{1 - \sum_{k \in S} \xi_k} \left(1 - \nu_i - \sum_{j \in S} \xi_j + \xi_i \right)$$
$$= \xi_i + (\mu_i - \xi_i) - \frac{(\mu_i - \xi_i)\,(\nu_i - \xi_i)}{1 - \sum_{k \in S} \xi_k} = \mu_i$$

as $(\mu_i - \xi_i)\,(\nu_i - \xi_i) = 0$ necessarily. A similar reasoning shows that $\mathbb{P}(Y = j) = \nu_j$ for all $j \in S$.

c) We have

$$\widetilde{C}(P) = \sup_{\substack{\mu,\nu \\ \|\mu-\nu\|_{\mathrm{TV}} > 0}} \frac{\|\mu P - \nu P\|_{\mathrm{TV}}}{\|\mu - \nu\|_{\mathrm{TV}}} \geq \sup_{i \neq k \in \mathcal{S}} \frac{\|\delta_i P - \delta_k P\|_{\mathrm{TV}}}{\|\delta_i - \delta_k\|_{\mathrm{TV}}}$$
$$= \sup_{i \neq k \in \mathcal{S}} \|\delta_i P - \delta_k P\|_{\mathrm{TV}} = \sup_{i \neq k \in \mathcal{S}} \|P_i - P_k\|_{\mathrm{TV}} = C(P)$$

d) For any $i, k \in \mathcal{S}$ and $A \subset \mathcal{S}$, we have

$$P_i(A) - P_k(A) \leq \sup_{A \subset \mathcal{S}} P_i(A) - P_k(A) = \|P_i - P_k\|_{\mathrm{TV}}$$

where the right-hand side is either zero (for $i \neq k$) or less than $C(P)$ by definition. Therefore we have $P_i(A) - P_k(A) \leq C(P)\, 1_{\{i \neq k\}}$. Averaging this inequality over values of i and k with respect to a coupling between two random variables X and Y where X has the marginal distribution μ and Y has the marginal distribution ν leads to

$$\sum_{i,k} P(X = i, Y = k) P_i(A) - \sum_{i,k} P(X = i, Y = k) P_k(A) \leq C(P) \sum_{i,k} P(X = i, Y = k) 1_{\{i \neq k\}} \quad \Rightarrow$$
$$\sum_i \mu_i P_i(A) - \sum_k \nu_k P_k(A) \leq C(P) \sum_{i \neq k} P(X = i, Y = k) \quad \Rightarrow$$
$$\mu P(A) - \nu P(A) \leq C(P)\, P(X \neq Y)$$

If we take the coupling designed in the first step, we have $P(X \neq Y) = \|\mu - \nu\|_{\text{TV}}$. Then, by taking the supremum over $A \subset \mathcal{S}$ on both sides of the inequality, we obtain

$$\|\mu P - \nu P\|_{\text{TV}} = \sup_{A \subset \mathcal{S}} (\mu P(A) - \nu P(A)) \leq C(P)\, \|\mu - \nu\|_{\text{TV}} \quad \Rightarrow \quad \frac{\|\mu P - \nu P\|_{\text{TV}}}{\|\mu - \nu\|_{\text{TV}}} \leq C(P)$$

where we assume that $\mu \neq \nu$. Because the inequality holds for any distributions μ and ν such that $\mu \neq \nu$, we have

$$\widetilde{C}(P) = \sup_{\substack{\mu,\nu \\ \|\mu-\nu\|_{\text{TV}}>0}} \frac{\|\mu P - \nu P\|_{\text{TV}}}{\|\mu - \nu\|_{\text{TV}}} \leq C(P)$$

e) For any $i \in \mathcal{S}$ we have

$$\frac{\|P_i^n - \pi\|_{\text{TV}}}{\|P_i^{n-1} - \pi\|_{\text{TV}}} = \frac{\|P_i^{n-1} P - \pi P\|_{\text{TV}}}{\|P_i^{n-1} - \pi\|_{\text{TV}}} \leq \sup_{\substack{\mu,\nu \\ \|\mu-\nu\|_{\text{TV}}>0}} \frac{\|\mu P - \nu P\|_{\text{TV}}}{\|\mu - \nu\|_{\text{TV}}} = C(P)$$

which leads to (replacing $C(P)$ by C for convenience)

$$\|P_i^n - \pi\|_{\text{TV}} \leq C\, \|P_i^{n-1} - \pi\|_{\text{TV}} \leq ... \leq C^{n-1}\, \|P_i - \pi\|_{\text{TV}} \leq C^n\, \|\delta_i - \pi\|_{\text{TV}}$$

f) $C(P) = |1 - p - q| (= \lambda_*)$ which is 0 when $p + q = 1$ (i.e., when we have complete symmetry between two states) and is 1 when $p = q = 0$ (i.e. when there is no self-loop and the chain is periodic with period 2).

g) Given the transition matrix we have

$$C(P) = \max\left\{1 - p, \frac{1}{2}\big(\frac{1}{2} + p + |\frac{1}{2} - p|\big)\right\}$$

By observing that $\frac{1}{2}\big(\frac{1}{2} + p + |\frac{1}{2} - p|\big) = \max\{\frac{1}{2}, p\}$, we have

$$C(P) = \max\{1 - p, p\} (= \lambda_*)$$

which is 1 when $p = 0$ (i.e., when the chain not irreducible) or $p = 1$ (i.e. when there is no self-loop and the chain is periodic with period 2). Moreover, the minimum value of $C(P)$ is equal to 0.5 and happens at $p = 0.5$, when the transition probabilities from side states to the central state is the same as the probability of the opposite transitions.

h) For $N = 3$ we have $C(P) = \frac{p+q+|p-q|}{2} = \max\{p, 1-p\}$ which is 1 when $p = 0$ or $p = 1$ (i.e. when there is either only the clockwise path or only the counterclockwise path, and hence the chain is periodic with period 3) and has its minimum value equal to 0.5 and at $p = 0.5$ (i.e. when there is complete symmetry). For $N > 3$, $C(P)$ is always 1, which means that there always exists two different initial distributions that would not get closer after one transition – can you find those for $N = 5$?

2. a) In the case $p = 0$, the chain becomes a cyclic random walk with period 2. Since the chain is periodic, it cannot be ergodic in this case.

In the case $p = 1$, we obtain a chain with three equivalence classes, and the chain is thus not irreducible, so not ergodic.

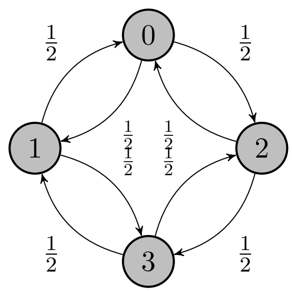

Figure 1: State diagram for the Markov chain for $p = 0$

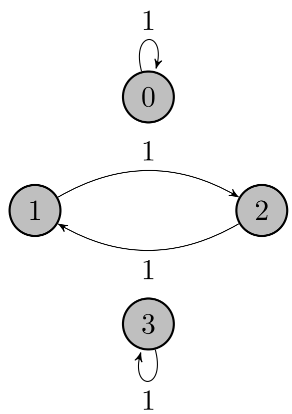

Figure 2: State diagram for the Markov chain for $p = 1$

In the case $0 < p < 1$, the chain is aperiodic since there are self-loops at states 0 and 3. It is irreducible since all states communicate. Finally, since the number of states is finite, the chain is positive-recurrent. Hence, with this choice of p, the chain is ergodic.

b) The transition matrix is given by

$$P = \begin{pmatrix} p & \frac{1-p}{2} & \frac{1-p}{2} & 0 \\ \frac{1-p}{2} & 0 & p & \frac{1-p}{2} \\ \frac{1-p}{2} & p & 0 & \frac{1-p}{2} \\ 0 & \frac{1-p}{2} & \frac{1-p}{2} & p \end{pmatrix}$$

Noting that the transition matrix is doubly stochastic (rows and columns of P sum to 1), we deduce that π is the uniform distribution on $\mathcal{S}$.

So $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

c) For any two states $i, j \in \mathcal{S}$, one can see from the transition graph that we have $p_{ij} = p_{ji}$. Since the stationary distribution is the uniform one, detailed balance holds since $\pi_i p_{ij} = \pi_j p_{ij} = \pi_j p_{ji}$. The chain is thus reversible.

d) To find the eigenvalues of P, we find the characteristic polynomial of P, so $\det(P - \lambda I)$. We can compute this determinant using the cofactors formula to get

$$\begin{aligned}
\det(P - \lambda I) &= \begin{vmatrix} p-\lambda & \frac{1-p}{2} & \frac{1-p}{2} & 0 \\ \frac{1-p}{2} & -\lambda & p & \frac{1-p}{2} \\ \frac{1-p}{2} & p & -\lambda & \frac{1-p}{2} \\ 0 & \frac{1-p}{2} & \frac{1-p}{2} & p-\lambda \end{vmatrix} \\
&= (p-\lambda)\begin{vmatrix} -\lambda & p & \frac{1-p}{2} \\ p & -\lambda & \frac{1-p}{2} \\ \frac{1-p}{2} & \frac{1-p}{2} & p-\lambda \end{vmatrix} - \frac{1-p}{2}\begin{vmatrix} \frac{1-p}{2} & p & \frac{1-p}{2} \\ \frac{1-p}{2} & -\lambda & \frac{1-p}{2} \\ 0 & \frac{1-p}{2} & p-\lambda \end{vmatrix} + \frac{1-p}{2}\begin{vmatrix} \frac{1-p}{2} & -\lambda & \frac{1-p}{2} \\ \frac{1-p}{2} & p & \frac{1-p}{2} \\ 0 & \frac{1-p}{2} & p-\lambda \end{vmatrix} \\
&= \lambda^4 - \lambda^2 + 2\lambda p^3 - 2p^3 - \lambda^2 p^2 + p^2 - 2\lambda^3 p + 2\lambda^2 p.
\end{aligned}$$

Since we know $\lambda_0 = 1$, we know that there will be a factor $(\lambda - 1)$ in the characteristic polynomial. We have

$$\begin{aligned}
\det(P - \lambda I) &= \lambda^4 - \lambda^2 + 2\lambda p^3 - 2p^3 - \lambda^2 p^2 + p^2 - 2\lambda^3 p + 2\lambda^2 p \\
&= (\lambda - 1)(\lambda^3 + \lambda^2 - \lambda^2 2p - \lambda p^2 + 2p^3 - p^2).
\end{aligned}$$

Letting $f(\lambda) = \lambda^3 + \lambda^2(1-2p) - \lambda p^2 + 2p^3 - p^2$ and noting that $f(p) = 0$, we can further factor as

$$f(\lambda) = (\lambda - p)(\lambda^2 + \lambda(1-p) - 2p^2 + p) = (\lambda - p)(\lambda + p)(\lambda - (2p-1)).$$

Overall, we have

$$\det(P - \lambda I) = (\lambda - 1)f(\lambda) = (\lambda - 1)(\lambda - p)(\lambda + p)(\lambda - (2p-1)),$$

so that

$$\lambda_0 = 1, \quad \lambda_1 = -p, \quad \lambda_2 = p, \quad \lambda_3 = 2p - 1.$$

We find that $\lambda_* = \max\{p, |2p-1|\} = \max\{p, 2p-1, 1-2p\} = \max\{p, 1-2p\}$ since $p \geq 2p - 1$ for $0 \leq p \leq 1$. So $\gamma = 1 - \lambda_* = \min\{1-p, 2p\}$.

e) Using the bound seen in class together with the result from part b), we have

$$\max_{i \in S} \|P_i^n - \pi\|_{TV} \leq \frac{\lambda_*^n}{2\sqrt{\pi_i}} = \lambda_*^n = (1-\gamma)^n \leq e^{-\gamma n}.$$

To have $\max_{i \in S} \|P_i^n - \pi\|_{TV} \leq \epsilon$, we need

$$e^{-\gamma n} \leq \epsilon \iff n \geq \frac{1}{\gamma}\log(1/\epsilon) = \frac{1}{\min\{1-p, 2p\}}\log(1/\epsilon).$$

The convergence is fastest for the value of p which minimizes $\frac{1}{\min\{1-p,2p\}}$, or equivalently we want $\max_{0<p<1} \min\{1-p, 2p\}$. Since $2p$ is increasing and $1-p$ is decreasing, we can easily find graphically that the maximum is attained when both $1-p$ and $2p$ are equal, which happens at $p = \frac{1}{3}$.