Mixing Times of Markov Chains, Michaelmas 2019. P.Sousi@statslab.cam.ac.uk

Example Sheet 2

1. Let X be a reversible Markov chain on a finite state space E with transition matrix P and invariant distribution π. Prove a generalisation of the Poincaré inequality, i.e. for all $f : E \to \mathbb{R}$ show that

$$\mathrm{Var}_\pi \left(P^t f\right) \le e^{-2t/t_{\mathrm{rel}}} \mathrm{Var}_\pi \left(f\right).$$

2. Let P be a reversible transition matrix on a finite state space with invariant distribution π. Define the total variation distance from stationarity from a typical point, i.e. for all t

$$d_{\mathrm{ave}}(t) = \sum_x \pi(x) \left\| P^t(x, \cdot) - \pi \right\|_{\mathrm{TV}}.$$

Suppose taht $1 = \lambda_1 \ge \ldots \ge \lambda_n \ge -1$ are the eigenvalues, then show that

$$4 d_{\mathrm{ave}}(t)^2 \le \sum_{j=2}^{n} \lambda_j^{2t}.$$

3. Let X be an irreducible Markov chain on the finite state space E with transition matrix P and invariant distribution π.

(i) Define the separation distance $s(t) = \max_{x,y} \left(1 - P^t(x,y)/\pi(y)\right)$. Show that $s(t)$ is decreasing as a function of t.

(ii) Define $t_{\mathrm{sep}}(\varepsilon) = \min\{t \ge 0 : s(t) \le \varepsilon\}$. Show that for all $\varepsilon \in (0,1]$ and all $k \in \mathbb{N}$ we have that

$$t_{\mathrm{sep}}(\varepsilon^k) \le k t_{\mathrm{sep}}(\varepsilon).$$

4. Consider two copies K_n and K_n' of the complete graph joined by a single edge. Find the order of the mixing time for a lazy simple random walk on the resulting graph.

5. Let X be an irreducible, lazy and reversible Markov chain on a finite state space with transition matrix P and stationary distribution π.

(i) Show that

$$\mathbb{E}_\pi[\tau_\pi] := \sum_{x,y} \pi(x)\pi(y)\mathbb{E}_x[\tau_y] = \sum_{i \ge 2} \frac{1}{1 - \lambda_i}.$$

where (λ_i) are all the eigenvalues. (*Hint:* Use question 12(b) from the first example sheet.)

(ii) Show that

$$\sum_{t=k}^{\infty} (P^t(x,x) - \pi(x)) \le e^{-k/t_{\mathrm{rel}}} \mathbb{E}_\pi[\tau_x].$$

5. Let X be a reversible Markov chain on the finite state space E with transition matrix P and invariant distribution π.

(ii) Prove that for all x, y

$$\frac{P^{2t}(x,y)}{\pi(y)} \ge \left(1 - \max_{z,w} \left\| P^t(z,\cdot) - P^t(w,\cdot) \right\|_{\mathrm{TV}}\right)^2.$$

Deduce that

$$P^{2t_{\text{mix}}}(x,y) \geq \frac{1}{4}\pi(y)$$

and that there exists a transition matrix $\widetilde{P}$ such that

$$P^{2t_{\text{mix}}}(x,y) = \frac{1}{4}\pi(y) + \frac{3}{4}\widetilde{P}(x,y)$$

(iii) Define

$$t_{\text{stop}} = \max_x \min\{\mathbb{E}_x[\Lambda_x] : \Lambda_x \text{ is a randomised stopping time s.t. } \mathbb{P}_x(X_{\Lambda_x} \in \cdot) = \pi(\cdot)\}.$$

(It is not clear by the definition that a stationary time achieving the minimum exists. One such example is the filling rule introduced by Baxter and Chacon.) By defining an appropriate stationary time, prove that

$$t_{\text{stop}} \leq 8t_{\text{mix}}.$$

We say that a randomised stopping time T starting from x has a halting state if there exists $z \in E$ such that $T \leq \tau_z$, where $\tau_z = \min\{t \geq 0 : X_t = z\}$.

(*Harder*) Show that if T has a halting state, then it is mean optimal, in the sense that

$$\mathbb{E}_x[T] = \min\{\mathbb{E}_x[\Lambda_x] : \Lambda_x \text{ is a randomised stopping time s.t. } \mathbb{P}_x(X_{\Lambda_x} \in \cdot) = \pi(\cdot)\}.$$

(*Hint*: For a stopping time S consider the exit frequencies from each state, i.e. $\nu(y) = \mathbb{E}_x\left[\sum_{k=0}^{T-1} \mathbf{1}(X_k = y)\right]$ for all y and compare them for different stopping times. Then use the uniqueness of the invariant measure up to multiplying by a constant.)

6. Let X be a reversible Markov chain with values in the finite space E, transition matrix P and invariant distribution π.

(a) Let φ be an eigenfunction of P corresponding to eigenvalue $\lambda \neq 1$ and $\|\varphi\|_2 = 1$. Show that

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1} \varphi(X_s)\right)^2\right] \leq \frac{2t}{1-\lambda}.$$

(b) Let $f : E \to \mathbb{R}$ be a function with $\mathbb{E}_\pi[f] = 0$. Recall $\gamma = 1 - \lambda_2$ is the spectral gap. Show that

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1} f(X_s)\right)^2\right] \leq \frac{2t\mathbb{E}_\pi\left[f^2\right]}{\gamma}.$$

(c) Using coupling or otherwise, show that if $r \geq t_{\text{mix}}(\varepsilon/2)$ and $t \geq 4t_{\text{rel}}\text{Var}_\pi(f)/(\eta^2\varepsilon)$, then for all $x \in E$

$$\mathbb{P}_x\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_{r+s}) - \mathbb{E}_\pi[f]\right| \geq \eta\right) \leq \varepsilon.$$

7. Let P, π be a reversible finite Markov chain. Let $A \subsetneq \Omega$ and let $B = A^c$ with $k = |B|$. Suppose that the sub-stochastic matrix P_B (the restriction of P to B, i.e. $P_B(x,y) = P(x,y$ for $x, y \in B))$ is irreducible, in the sense that for all $x, y \in B$, there exists $n \geq 0$ such that $P_B^n(x,y) > 0$.

(i) By defining an appropriate inner product, show that P_B has k real eigenvalues

$$1 \geq \gamma_1 > \gamma_2 \geq \ldots \geq \gamma_k.$$

(ii) Show that there exist nonnegative numbers $a_1, \ldots, a_k$ satisfying $\sum_i a_i = 1$ such that for all $t \geq 0$ we have

$$\mathbb{P}_{\pi_B}(\tau_A > t) = \sum_{i=1}^{k} a_i \gamma_i^t$$

(iii) The Perron Frobenius theorem gives that $\gamma_1 > 0$ and $\gamma_1 \geq -\gamma_k$. Using the Courant-Fischer characterisation of eigenvalues establish that

$$\gamma_1 \leq 1 - \frac{\pi(A)}{t_{\mathrm{rel}}}.$$

(iv) Deduce that $\mathbb{P}_{\pi_B}(\tau_A > t) \leq \left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t \leq \exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right)$.

(v) By the Perron Frobenius theorem the eigenvector v corresponding to $\gamma_1 > 0$ is strictly positive. Let α be a probability distribution given by $\alpha = v / \sum_i v(i)$. Show that when the starting distribution is α, then the law of τ_A is geometric with parameter γ_1.

Prove that for all t and all y

$$\mathbb{P}_\alpha(X_t = y \mid \tau_A > t) = \alpha(y).$$

Finally show that for all $x \notin A$ we have

$$\mathbb{P}_x(X_t = y \mid \tau_A > t) \to \alpha(y) \text{ as } t \to \infty.$$

(The distribution α is called the quasi-stationary distribution.)

1. Let X be a reversible Markov chain on a finite state space E with transition matrix P and invariant distribution π. Prove a generalisation of the Poincaré inequality, i.e. for all $f : E \to \mathbb{R}$ show that

$$\mathrm{Var}_\pi\left(P^t f\right) \le e^{-2t/t_{\mathrm{rel}}} \mathrm{Var}_\pi\left(f\right).$$

WLOG f has $E_\pi[f] = 0$ else replace f by $f - \mathbb{E}_\pi[f]$ and its variance remains the same.

$\mathbb{E}_\pi[P^t f] = 0$

$$\mathbb{E}_\pi[P^t f] = \sum_x \sum_y P^t(x,y) f(y) \pi(x) = \sum_y f(y) \sum_x \pi(x) P^t(x,y)$$

$$\overset{\pi \text{ inv.}}{=} \sum_y f(y) \pi(y) = \mathbb{E}_\pi[f] = 0$$

Thus $f = \sum_{j=2}^{n} \langle f, f_j \rangle_\pi f_j$ and $P^t f = \sum_{j=2}^{n} \langle f, f_j \rangle_\pi P^t f_j = \sum_{j=2}^{n} \langle f, f_j \rangle_\pi \lambda_j^t f_j$

orthonormal basis

so
$$\mathrm{Var}_\pi(P^t f) = \|P^t f\|_2^2 = \sum_{i=2}^{n} \sum_{j=2}^{n} \langle f, f_j \rangle \langle f, f_i \rangle \lambda_i^t \lambda_j^t f_i f_j$$

$$= \sum_{j=2}^{n} \langle f, f_j \rangle^2 \lambda_j^{2t} f_j^2 \le \lambda_*^{2t} \sum_{j=2}^{n} \langle f, f_j \rangle^2 f_j^2 = \lambda_*^{2t} \|f\|_2^2$$

$$= \lambda_*^{2t} \mathrm{Var}_\pi(f)$$

$$\Rightarrow \mathrm{Var}_\pi(P^t f) \le (1 - \gamma_*)^{2t} \mathrm{Var}_\pi(f) = \left(1 - \frac{2t\gamma_*}{2t}\right)^{2t} \mathrm{Var}_\pi(f) \le e^{-2t\gamma_*} \mathrm{Var}_\pi[f]$$

$$= e^{-2t/t_{rel}} \mathrm{Var}_\pi[f].$$

2. Let P be a reversible transition matrix on a finite state space with invariant distribution π. Define the total variation distance from stationarity from a typical point, i.e. for all t

$$d_{\text{ave}}(t) = \sum_x \pi(x) \left\| P^t(x,\cdot) - \pi \right\|_{\text{TV}}.$$

Suppose taht $1 = \lambda_1 \geq \ldots \geq \lambda_n \geq -1$ are the eigenvalues, then show that

$$4 d_{\text{ave}}(t)^2 \leq \sum_{j=2}^{n} \lambda_j^{2t}.$$

Let (f_i) be orthonormal basis of eigenfunctions

$$\begin{aligned}
4 d_{ave}(t)^2 &= 4 \left(\sum_x \pi(x) \| P^t(x,\cdot) - \pi \|_{TV} \right)^2 \\
&\leq \sum_x \sum_y \pi(x)\pi(y) \left(4 \| P^t(x,\cdot) - \pi \|_{TV}^2\right)^{1/2} \left(4 \| P^t(y,\cdot) - \pi \|_{TV}^2 \right)^{1/2} \\
&\leq \sum_x \sum_y \pi(x)\pi(y) \left(\sum_{i=2}^{n} f_i(x)^2 \lambda_i^{2t} \right)^{1/2} \left(\sum_{j=2}^{n} f_j(y)^2 \lambda_j^{2t} \right)^{1/2} \\
&= \left(\sum_x \left(\sum_{i=2}^{n} f_i(x)^2 \lambda_i^{2t} \right)^{1/2} \pi(x) \right) \left(\sum_y \left(\sum_{j=2}^{n} f_j(y)^2 \lambda_j^{2t} \right)^{1/2} \pi(y) \right) \\
&= \left(\sum_x \left(\sum_{i=2}^{n} f_i(x)^2 \lambda_i^{2t} \right)^{1/2} \pi(x) \right)^2 \\
&\overset{\text{cauchy-schwarz}}{\leq} \left(\sum_x \pi(x) \sum_{i=2}^{n} f_i(x) \lambda_i^{t} \right)^2
\end{aligned}$$

From lectures we have $4 \| P^t(x,\cdot) - \pi \|_{TV} \leq \sum_{j=2}^{n} \lambda_j^{2t} f(x)^2$

$$\begin{aligned}
4 d_{ave}(t)^2 &= 4 \left(\sum_x \pi(x) \| P^t(x,\cdot) - \pi \|_{TV} \right)^2 \\
&= \sum_x \sum_y \pi(x)\pi(y) \| P^t(x,\cdot) - \pi \|_{TV} \| P^t(y,\cdot) - \pi \|_{TV} \\
&\leq \sum_x \sum_y \pi(x)\pi(y) \left(\sum_{j=2}^{n} \lambda_j^{2t} \right) = \sum_{j=2}^{n} \lambda_j^{2t}
\end{aligned}$$

3. Let X be an irreducible Markov chain on the finite state space E with transition matrix P and invariant distribution π.

(i) Define the separation distance $s(t) = \max_{x,y}\left(1 - P^t(x,y)/\pi(y)\right)$. Show that $s(t)$ is decreasing as a function of t.

(ii) Define $t_{\mathrm{sep}}(\varepsilon) = \min\{t \geq 0 : s(t) \leq \varepsilon\}$. Show that for all $\varepsilon \in (0,1]$ and all $k \in \mathbb{N}$ we have that

$$t_{\mathrm{sep}}(\varepsilon^k) \leq k t_{\mathrm{sep}}(\varepsilon).$$

This is a neat trick to use, such a decomposition could work to show something other than $s(t)$

(i) Define Q_t by $P^t(x,y) = (1 - s_x(t))\pi(x) + Q_t(x,y)s_x(t)$

By def^n of $s_x(t)$ we have that $Q_t(x,y) \geq 0$ (Not immediate but does follow)

$$\sum_y P^t(x,y) = 1, \quad \sum_y \pi(y) = 1 \Rightarrow \sum_y Q_t(x,y) = 1$$ so $Q_t(x,y)$ is stochastic.

Using Chapman-Kolmogorov,

Taking max over y

$$P^{t+u}(x,y) \leq (1 - s_x(t))\pi(y) + s_x(t)s_x(u)\sum_z Q_t(x,z)Q_u(y,z)$$

Requires some algebra

Then $s_x(t+u) = \max_y \left(1 - \frac{P_{t+u}(x,y)}{\pi(y)}\right) \leq s_x(t)s_x(u)$ by plugging the above

and taking max over x gives $s(t+u) \leq s(t)s(u)$

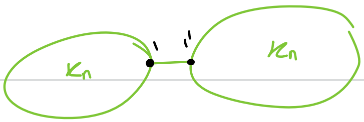
i.e. s is submultiplicative

Note: This proof comes about from seeing that $s_x(t) \geq \frac{\pi(y) - P^t(x,y)}{\pi(y)}$ $\forall y$

$\Rightarrow P^t(x,y) \geq (1 - s_x(t))\pi(y)$ then take up the "slack" with $Q_t(x,y)$.

General point: Such submultiplicative proofs are not generally very constructive

Q4



Lower bound Let τ be the first hitting time of the other half.

Assume the walk is not lazy (this speeds up the walk so reduces mixing time).

For $x \in K_n \setminus \{1\}$ $\quad \mathbb{P}_x(X_1 = 1) = \frac{1}{(n-1)}$, $\mathbb{P}_1(X_1 = 1') = \frac{1}{n}$ $\quad \tau \sim \text{Geom}\left(\frac{1}{n(n-1)}\right)$

$\Rightarrow \mathbb{E}[\tau] \approx n^2 \Rightarrow t_{mix}\left(\frac{1}{2} - \varepsilon\right) \geq n^2$

Upper bound

Pick $x \in X$, $y \in Y$.

Define a coupling: (since $\| \quad \|_{TV} \leq \mathbb{P}(X \neq Y)$ for any coupling)

Pick $\tilde{x} \in \{1, \dots, n\} \setminus \{x\}$ move X to $\tilde{x}$ move Y to corresponding vertex in K_n'

If $x = 1$ and $y = 1'$, then sample $Z \sim \text{Bernoulli}(1/n)$. If $Z = 0$ then move indep. in their own halves.

If $Z = 1$ then move $X \to 1'$ and keep Y at $1'$ each wp $1/2$.

or $Y \to 1$ and keep X at 1

This coupling takes time order n^2 to coalesce.

Notes: Can use the bottleneck ratio for the lowerbound (proof should be approx 2 lines)

Q7

(a) In question $\|\varphi\|_2$ represents $\|\varphi\|_{2,\pi}$

$$\mathbb{E}_\pi[\varphi(X)^2] = \sum_x \pi(x)\varphi(x)^2 = \|\varphi\|_{2,\pi}^2 = 1$$

For $0 \le i, j$

$$\mathbb{E}_\pi[\varphi(X_i), \varphi(X_{i+j})] \overset{\text{since } \pi \text{ stat.}}{=} \mathbb{E}_\pi[\varphi(X_0)\varphi(X_j)]$$
$$= \sum_x \pi(x)\varphi(x) \sum_y P^j(x,y)\varphi(y)$$
$$= \sum_x \pi(x)\varphi(x)\lambda^j\varphi(x) = \lambda^j$$

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1}\varphi(X_s)\right)\right] = 2\sum_{0\le i<j\le t-1}\lambda^{j-i} + t$$
$$= \dots \le \frac{2t}{1-\lambda}$$

(b) Let e-vals be $1=\lambda_1 \ge \lambda_2 \dots \ge \lambda_n$ with corresponding e-functions $\varphi_1, \dots, \varphi_n$.

Let $f = \sum_{i=1}^n a_i\varphi_i$. Then $\mathbb{E}_\pi[f] = 0 \Rightarrow a_1 = 1$.

As before,

$$\mathbb{E}_\pi[\varphi_i(X_0)\varphi_j(X_0)] = \sum_x \pi(x)\lambda^t\varphi_i\varphi_j(x) = \lambda^t\delta_{ij}$$

$$\mathbb{E}_\pi\left[\left(\sum_{s=0}^{t-1}\sum_{i\ge2}a_i\varphi_i(X_i)\right)^2\right] = \mathbb{E}_\pi\left[\sum_{i\ge2}a_i^2\left(\sum_{s=0}^{t-1}\varphi_i(X_s)\right)^2\right] \quad \text{since cross terms 0}$$

$$\le \sum_{i\ge2}a_i^2\frac{2t}{1-\lambda_i} \le \sum_{i\ge2}a_i^2\frac{2t}{\gamma} = \frac{2t\,\mathbb{E}_\pi[f^2]}{\gamma}$$

(c) Replace $f \to f - \mathbb{E}_\pi[f]$. Assume $\mathbb{E}_\pi[f]=0$. $\|P^r(x,\cdot) - \pi(\cdot)\|_{TV} \le \varepsilon/2$.

Couple Y, Z with $Y \sim P^r(x,\cdot)$, $Z\sim\pi$ $\mathbb{P}(Y \ne Z) \le \varepsilon/2$.

$$\mathbb{E}_\pi\left[\frac{1}{t}\sum_{s=0}^{t-1}f(X_s)\right] = 0, \quad \mathrm{Var}_\pi\left[\frac{1}{t}\sum_{s=0}^{t-1}f(X_s)\right] \overset{(b)}{\le} \frac{2\mathbb{E}_\pi[f^2]}{t\gamma}$$

Q7 (cont.)

$$\mathbb{P}_x\left(\left| \frac{1}{t}\sum_{s=0}^{t-1} f(X_{r+s}) \right| > \eta \right) = \mathbb{P}_Y\left(\left| \frac{1}{t}\sum_{s=0}^{t-1} f(X_s) \right| > \eta \right) \leq P(Y \neq Z) + \mathbb{P}_\pi(\quad \cdots \quad)$$

$\leq \varepsilon/2$

chebyshev

$$\leq \frac{\mathrm{Var}_\pi(\ldots)}{\eta^2} \leq \frac{2\,\mathrm{Var}_\pi(t)}{t\gamma\eta^2} \leq \frac{\varepsilon}{2}$$

Solutions to *Mixing Times of Markov Chains*, Example Sheet 2, 2019/20

Last compiled: 5th December 2019

These *are not* model solutions. There may be be typos, and even logical errors. *Do not redistribute.*

Question 1 We first observe that

$$E_\pi(P^t f) = \pi P^t f = \pi f = E_\pi(f).$$

Since the first eigenfunction is constant, it follows from the spectral decomposition that

$$P^t f - E_\pi(P^t f) = \textstyle\sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t.$$

Since the $\{f_j\}$ form an orthonormal basis,

$$\mathrm{Var}_\pi(P^t f) = \|P^t f - E_\pi(P^t f)\|_2^2 = \textstyle\sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2 \lambda_j^{2t} \le (1 - 1/t_{\mathrm{rel}}) \sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2.$$

The result follows by upper bounding $1 - 1/t_{\mathrm{rel}} \le e^{-1/t_{\mathrm{rel}}}$ and observing that

$$\textstyle\sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2 = E_\pi(f^2) - E_\pi(f)^2 = \mathrm{Var}_\pi(f);$$

this follows by plugging in the spectral decomposition and using orthonormality of $\{f_j\}$.

Question 2 Recall the inequality given in Lemma 3.5:

$$4\,\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}}^2 \le \|P^t(x,\cdot) - \pi\|_2^2 = \textstyle\sum_{j=2}^{|\Omega|} f_j(x)^2 \lambda_j^{2t}.$$

(This is a consequence of Cauchy–Schwarz and Theorem 3.1.) Applying Cauchy–Schwarz to $d_{\mathrm{ave}}(t)$ gives

$$4\, d_{\mathrm{ave}}(t)^2 \le \textstyle\sum_x \pi(x) \|P^t(x,\cdot) - \pi\|_2^2 \le \sum_x \pi(x) \sum_{j \ge 2} f_j(x)^2 \lambda_j^{2t} = \sum_{j\ge 2} \lambda_j^{2t};$$

the final equality follows by swapping the summation order and using (ortho)normality of $\{f_j\}$.

Question 3 Write $s_x(t) := \max_y\{1 - P^t(x,y)/\pi(y)\}$. Define the matrix Q_t by

$$P^t(x,\cdot) = \big(1 - s_x(t)\big)\pi + s_x(t) Q_t(x,\cdot).$$

Since P^t has unit row-sums, so must Q_t; similarly, $\pi = \pi Q_t$. Rearrange the above display into

$$\pi(y) - P^t(x,y) - s_x(t)\pi(y) + s_x(t) Q_t(x,y) = 0.$$

Since $s_x(t)\pi(y) \ge \pi(y) - P^t(x,y)$ for all $x, y \in \Omega$, we see that $Q_t(x,y) \ge 0$ for all $x, y \in \Omega$. (Since $\sum_y P^t(x,y) = 1 = \sum_y \pi(y)$, we must have $s_x(t) \ge 0$ with equality if and only if $P^t(x,\cdot) = \pi$.)

We claim that, using this Q-representation above, we can write

$$P^{t+u}(x,y) = \big(1 - s_x(t)s(u)\big)\pi(y) + s_x(t)s(u) \textstyle\sum_z Q_t(x,z) Q_u(z,y).$$

This follows, after a little algebra, from Chapman–Kolmogorov. Rearranging this, we see that

$$s_x(t+u) \le s_x(t)s(u), \quad \text{and hence} \quad s(t+u) \le s(t)s(u).$$

Hence s is a decreasing, submultiplicative function, proving (i); given this, (ii) follows as for t_{mix}.

Question 4 We show that the mixing time is order n^2.

Lower Bound. Start the walk from the one of the complete graphs, but not at the distinguished vertex. To hit the other complete graph, the walk must get to the distinguished node, then cross. Each of these events has probability $1/n$ at each step, so to happen sequentially takes time order n^2 in expectation. Hence, by **S1Q7**, $t_{\mathrm{mix}} \gtrsim n^2$.

Upper Bound. Start two walks X and Y from x and y, respectively. Suppose first that x and y are in different copies and not at the distinguished vertices. Label the vertices $\{1, ..., n\}$ and $\{1', ..., n'\}$, with $1/1'$ the distinguished vertex. Move the walks together until they are both at their distinguished vertices. For the next step, take X: choose uniformly among $\{1', 2, ..., n\}$. If $1'$ is not chosen, then couple with Y as before. If $1'$ is chosen, then choose 1 for Y. Move one of X or Y with equal probability. This couples the walks. Similarly, this takes time order n^2.

Other cases are analysed similarly; the technical details are omitted here.

Please send comments/corrections to Sam Thomas at s.m.thomas@statslab.cam.ac.uk

Question 5 (i) Use **S1Q12(b)**, the spectral decomposition and orthonormality:

$$\begin{aligned}
\mathbb{E}_\pi(\tau_\pi) &:= \textstyle\sum_{x,y} \pi(x)\pi(y)\mathbb{E}_x(\tau_y) \\
&= \textstyle\sum_y \pi(y)\mathbb{E}_\pi(\tau_y) \\
&= \textstyle\sum_y \sum_{t=0}^\infty \big(P^t(y,y) - \pi(y)\big) \\
&= \textstyle\sum_y \sum_{t=0}^\infty \pi(y) \sum_{j=2}^{|\Omega|} f_j(y)^2 \lambda_j^t \\
&= \textstyle\sum_{j\ge 2} 1/(1-\lambda_j).
\end{aligned}$$

(ii) Use the spectral decomposition and then **S1Q12(b)**:

$$\begin{aligned}
\textstyle\sum_{t=k}^\infty \big(P^t(x,x) - \pi(x)\big) &= \textstyle\sum_{t=0}^\infty \big(P^{t+k}(x,x) - \pi(x)\big) \\
&= \textstyle\sum_{t=0}^\infty \sum_{i=2}^{|\Omega|} \lambda_i^{t+k} \pi(x) f_i(x)^2 \\
&\le \lambda_2^k \textstyle\sum_{t=0}^\infty \sum[|\Omega|]i = 2\lambda_i^t \pi(x) f_i(x)^2 \\
&= \lambda_2^k \textstyle\sum_{t=0}^\infty \big(P^t(x,x) - \pi(x)\big) \\
&= (1 - 1/t_{\mathrm{rel}})^k \pi(x)\mathbb{E}_\pi(\tau_x) \\
&\le \exp(-k/t_{\mathrm{rel}})\pi(x)\mathbb{E}_\pi(\tau_x).
\end{aligned}$$

Question 6 (i) Using Chapman–Kolmogorov and reversibility, we can write

$$\frac{P^{2t}(x,y)}{\pi(y)} = \sum_z \frac{P^t(x,z)P^t(z,y)}{\pi(y)} = \sum_z \pi(z) \frac{P^t(x,z)}{\pi(z)} \frac{P^t(y,z)}{\pi(z)}.$$

Applying Cauchy–Schwarz (in reverse) to the right-hand side, we obtain

$$\frac{P^{2t}(x,y)}{\pi(y)} \ge \Big(\textstyle\sum_z \sqrt{P^t(x,z)P^t(y,z)}\Big)^2 \ge \Big(\sum_z P^t(x,z) \wedge P^t(y,z)\Big)^2.$$

Using the version of total variation as the minimum, the first claim follows.

Taking $t := t_{\mathrm{mix}}(\frac{1}{4})$, we have $\bar{d}(t) \le 2d(t) \le \frac{1}{2}$, and hence

$$P^{2t}(x,y) \ge \pi(y)\big(1 - \bar{d}(t)\big)^2 \ge \tfrac{1}{4}\pi(y).$$

The final claim is immediate—the matrix $\tilde{P}$ must have unit row-sums and non-negative entries.

(ii) For $s \ge 0$, set $Y_s := X_{2st_{\mathrm{mix}}}$, so that $(Y_s)_{s\ge 0}$ is a Markov chain with transition matrix $P^{2t_{\mathrm{mix}}}$. Given that $Y_s = y$, we can obtain Y_{s+1} by the following procedure, justified by the above decomposition: sample $I_{s+1} \sim \mathrm{Bern}(\frac{1}{4})$; if $I_{s+1} = 1$, then select Y_{s+1} according to π and otherwise according to $Q(x,\cdot)$. Define the randomised stopping time

$$\sigma := \min\{s \ge 1 \mid I_s = 1\};$$

then σ is a strong stationary time, and $\mathbb{E}(\sigma) = 4$. The time $\tau := 2\sigma t_{\mathrm{mix}}$ is a randomised stopping time for $(X_t)_{t\ge 0}$ with $X_\tau \sim \pi$. Thus $t_{\mathrm{stop}} \le \mathbb{E}_x(\tau) = 8t_{\mathrm{mix}}$.

(iii) Let τ and σ be randomised stopping times for x; for $y \in \Omega$, define

$$\psi(y) := \mathbb{E}_x\big(\textstyle\sum_{k=0}^{\tau-1} \mathbf{1}(X_k = y)\big) \quad \text{and} \quad \varphi(y) := \mathbb{E}_x\big(\sum_{k=0}^{\sigma-1} \mathbf{1}(X_k = y)\big).$$

Using calculations similar to those of **S1Q12(a)**, one can check that

$$\psi P = \psi - \delta_x + \pi \quad \text{and} \quad \varphi P = \varphi - \delta_x + \pi.$$

(For the details, see Exercise 10.1 of [Levin/Peres/Wilmer], which has solution on page 404.) Thus

$$(\varphi - \psi)P = \varphi - \psi.$$

Hence, by uniqueness of the invariant distribution up to constant multiples, $\varphi - \psi = \alpha\pi$ for some $\alpha \in \mathbb{R}$.

Suppose now that τ has a halting state. Thus $\psi(z) = 0$ for some z. Then $\varphi(z) = \alpha\pi(z)$, and so $\alpha \ge 0$. Hence $\varphi(y) \ge \psi(y)$ for all y, and

$$\mathbb{E}_x(\sigma) = \textstyle\sum_y \varphi(y) \ge \sum_y \psi(y) = \mathbb{E}_x(\tau),$$

proving the mean-optimality of τ.

DO NOT REDISTRIBUTE WITHOUT PERMISSION

Question 7 (a) First consider the diagonal term: for s, by assumption, we have

$$\mathbb{E}_\pi\big(\varphi(X_s)\big) = 1.$$

Now consider the cross terms: for $r < s$, we have

$$\begin{aligned}\mathbb{E}_\pi\big(\varphi(X_r)\varphi(X_s)\big) &= \mathbb{E}_\pi\big(\mathbb{E}_\pi\big(\varphi(X_r)\varphi(X_s) \mid X_r\big)\big) \\ &= \mathbb{E}_\pi\big(\varphi(X_r)\mathbb{E}_{X_r}\big(\varphi(X_{s-r})\big)\big) = \mathbb{E}_\pi\big(\varphi(X_r)(P^{s-r}\varphi)(X_r)\big),\end{aligned}$$

by the Markov property. Since φ is an eigenfunction with $E_\pi(\varphi^2) = 1$, we have

$$\mathbb{E}_\pi\big(\varphi(X_r)\varphi(X_s)\big) = \lambda^{s-r}\mathbb{E}_\pi\big(\varphi(X_r)^2\big) = \lambda^{s-r}E_\pi(\varphi^2) = \lambda^{s-r}.$$

Combining the diagonal and cross terms when expanding the square, we have

$$\mathbb{E}_\pi\big(\big(\textstyle\sum_{s=0}^{t-1}\varphi(X_s)\big)^2\big) = t + 2\sum_{r=0}^{t-1}\sum_{s=1}^{t-1-r}\lambda^s = t + 2t\big(\lambda - \lambda t^{-1}(1-\lambda^t)/(1-\lambda)\big)/(1-\lambda).$$

Straightforward algebraic manipulations (considering $t = 1$ separately) prove the claim.

(b) Decompose f as $f = \sum_{j=1}^{|\Omega|} a_j f_j$ with $a_j := \langle f, f_j\rangle_\pi$. By Parseval's identity, $E_\pi(f^2) = \sum_{j=1}^{|\Omega|} a_j^2$. Observe that $a_1 = \langle f, f_1\rangle_\pi = \langle f, 1\rangle_\pi = E_\pi(f) = 0$. Writing $G_j := \sum_{s=0}^{t-1} f_j(X_s)$, we have

$$\textstyle\sum_{s=0}^{t-1} f(X_s) = \sum_{j=1}^{|\Omega|} a_j G_j.$$

Consider now the cross terms: for $r \le s$ and $j \ne k$, similarly to above, we have

$$\mathbb{E}_\pi\big(f_j(X_s)f_k(X_r)\big) = \lambda_j^{s-r}\mathbb{E}_\pi\big(f_k(X_r)f_j(X_r)\big) = \lambda_j^{s-r}E_\pi(f_k f_j) = 0,$$

using orthonormality; hence $\mathbb{E}_\pi(G_j G_k) = 0$, and thus

$$\mathbb{E}_\pi\big(\big(\textstyle\sum_{s=0}^{t-1} f(X_s)\big)^2\big) = \sum_{i=2}^{|\Omega|} a_i^2\mathbb{E}_\pi(G_i^2).$$

Applying (a) to G_j (noting its summation definition), we deduce that

$$\mathbb{E}_\pi\big(\big(\textstyle\sum_{s=0}^{t-1} f(X_s)\big)^2\big) \le \sum_{i=2}^{|\Omega|} 2ta_i^2/(1-\lambda_i) \le 2tE_\pi(f^2)/\gamma.$$

(c) Assume wlog that $E_\pi(f) = 0$. Let μ_r be the optimal coupling of $P^r(x,\cdot)$ with π. We define a process $(Y_t, Z_t)_{t\ge0}$ as follows. Let (Y_0, Z_0) have distribution μ_r. Given (Y_0, Z_0), let $(Y_t)_{t\ge1}$ and $(Z_t)_{t\ge1}$ evolve independently with transition matrix P, until the first time that they meet; after they meet, evolve them together according to P. The chain $(Y_t, Z_t)_{t\ge0}$ has transition matrix Q given by

$$Q\big((y,z),(u,v)\big) := \begin{cases} P(y,u) & \text{if } \ y = z \text{ and } u = v, \\ P(y,u)P(z,v) & \text{if } \ y \ne z, \\ 0 & \text{otherwise.}\end{cases}$$

The sequences $(Y_t)_{t\ge0}$ and $(Z_t)_{t\ge0}$ are each Markov chains with transition matrix P, started with distributions $P^r(x,\cdot)$ and with μ, respectively. By definition of the optimal coupling,

$$\mathbb{P}(Y_0 \ne Z_0) = \big\|P^r(x,\cdot) - \pi\big\|_{\mathrm{TV}}.$$

Since $(Y_t)_{t\ge0}$ has the same distribution as $(X_{r+t})_{t\ge0}$, we can rewrite the probability in question as

$$\begin{aligned}\mathbb{P}_x\big(\big|\tfrac{1}{t}\textstyle\sum_{s=0}^{t-1} f(X_{r+s})\big| \ge \eta\big) &= \mathbb{P}\big(\big|\tfrac{1}{t}\textstyle\sum_{s=0}^{t-1} f(Y_s)\big| \ge \eta\big) \\ &\le \mathbb{P}(Y_0 \ne Z_0) + \mathbb{P}\big(\big|\tfrac{1}{t}\textstyle\sum_{s=0}^{t-1} f(Z_s)\big| \ge \eta\big).\end{aligned}$$

By definition of $t_{\mathrm{mix}}(\varepsilon)$, if $r \ge t_{\mathrm{mix}}(\varepsilon/2)$ then the first term $\mathbb{P}(Y_0 \ne Z_0) \le \varepsilon/2$. By (b),

$$\mathrm{Var}_\pi\big(\tfrac{1}{t}\textstyle\sum_{s=0}^{t-1} f(Z_s)\big) \le 2\mathrm{Var}_\pi(f)/(t\gamma).$$

Applying Chebyshev, the second term is bounded by $\varepsilon/2$ if $t \ge \lfloor 4\mathrm{Var}_\pi(f)/(\eta^2\varepsilon)\rfloor\gamma^{-1}$.

Question 8 (i) Define an inner product $\langle\cdot,\cdot\rangle$ by

$$\langle f, g\rangle := \textstyle\sum_x f(x)g(x)\pi_B(x) \quad \text{where} \quad \pi_B(x) := \pi(x)/\pi(B).$$

As in the proof of the spectral decomposition,

$$M_B(x,y) := P(x,y)\sqrt{\pi_B(x)/\pi_B(y)}$$

defines a symmetric matrix (over $x, y \in \Omega$), and so has real eigenvalues. It is substochastic, so the same proof as in **S1Q1** shows that the eigenvalues have modulus at most 1.

(ii) Let (g_i) be the corresponding eigenfunctions. Then we can write

$$\begin{aligned}\mathbb{P}_{\pi_B}(\tau_A > t) &= \textstyle\sum_{x,y\notin A}\pi_B(x)P_B^t(x,y) \\ &= \textstyle\sum_{x,y\notin A}\pi_B(x)\sum_{i=1}^k g_i(x)g_i(y)\pi_B(y)\gamma_i^t \\ &= \textstyle\sum_{i=1}^k \gamma_i^t\big(\sum_{x\notin A}\pi_B(x)g_i(x)\big)^2 =: \sum_{i=1}^k a_i\gamma_i^t.\end{aligned}$$

Considering $t := 0$ proves that $\sum_1^k a_i = 1$.

(iii) The Courant–Fischer characterisation of eigenvalues gives

$$\gamma_1 = \sup_{f \text{ not constant}} \frac{\sum_{x,y}\pi(x)f(x)P_B(x,y)f(y)}{\sum_x \pi(x)f(x)^2}$$

Clearly the supremum is attained by non-negative f, and since the sum in the denominator is non-zero when $x, y \notin A$, we can restrict to f which are 0 on A. Hence we get

$$\gamma_1 = \sup_{\substack{f\ge0,\, f=0 \text{ on } A \\ f \text{ not constant}}} \frac{\sum_{x,y}\pi(x)f(x)P_B(x,y)f(y)}{\sum_x \pi(x)f(x)^2}, \quad \text{and equivalently} \quad 1 - \gamma_1 = \inf_{\substack{f\ge0,\, f=0 \text{ on } A \\ f \text{ not constant}}} \frac{\mathbb{E}(f)}{\langle f, f\rangle}.$$

By Cauchy–Schwarz, we have $\mathbb{E}_{\pi_B}(f^2) \ge \mathbb{E}_{\pi_B}(f)^2$, which gives

$$\mathrm{Var}_\pi(f)\cdot\langle f - \mathbb{E}_\pi(f), f - \mathbb{E}_\pi(f)\rangle \ge \pi(A)\cdot\langle f, f\rangle.$$

This combined with the variational characterisation of the relaxation time implies that

$$1 - \gamma_1 \ge \pi(A)/t_{\mathrm{rel}}.$$

(iv) This follows easily from (iii) and (iv): $|\gamma_i| \le \gamma_1 \le 1 - \pi(A)/t_{\mathrm{rel}}$ for all i, and hence

$$\mathbb{P}_{\pi_B}(\tau_A > t) = \textstyle\sum_1^k a_i\gamma_i^t \le \sum_1^k a_i\gamma_1^t = \gamma_1^t \le \big(1 - \pi(A)/t_{\mathrm{rel}}\big)^t \le \exp\big(-t\pi(A)/t_{\mathrm{rel}}\big).$$

(v) The eigenvector v is taken to be the left eigenvector corresponding to the eigenvalue γ_1. By reversibility, $g_1(x) = C\alpha(x)/\pi(x)$, where C is a positive constant. Since α is a left eigenvector, we have

$$\mathbb{P}_\alpha(\tau_A > t) = \textstyle\sum_{x,y\notin A}\alpha(x)P_B^t(x,y) = \sum_{y\notin A}\alpha(y)\gamma_1^t = \gamma_1^t.$$

using $\sum_{y\notin A}\alpha(y) = 1$. For the second part, we have

$$\mathbb{P}_\alpha\big(X_t = y \mid \tau_A > t\big) = \mathbb{P}_\alpha\big(X_t = y,\, \tau_A > t\big)/\mathbb{P}_\alpha\big(\tau_A > t\big) = \textstyle\sum_{x\notin A}\alpha(x)P_B^t(x,y)/\gamma_1^t = \alpha(y).$$

For the final part, we have

$$\mathbb{P}_x(X_t = y \mid \tau_A > t) = \frac{\mathbb{P}_x(X_t = y,\, \tau_A > t)}{\mathbb{P}_x(\tau_A > t)} = \frac{\sum_1^k \gamma_i^t g_i(x)g_i(y)\pi_B(y)}{\sum_1^k \gamma_i^t g_i(x)\sum_{z\notin A} g_i(z)\pi_B(z)}.$$

By the aperiodicity assumption and the Perron–Frobenius theorem, we have $|\gamma_i| < \gamma_1$ for all $i > 1$. So taking the limit as $t \to \infty$, we get

$$\mathbb{P}_x(X_t = y \mid \tau_A > t) \to \frac{g_1(y)\pi_B(y)}{\sum_{z\notin A} g_1(z)\pi_B(z)} = \alpha(y),$$

where the last equality holds since $g_1(x) \propto \alpha(x)/\pi(x) \propto \alpha(x)/\pi_B(x)$.