Solutions to *Mixing Times of Markov Chains*, Example Sheet 3, 2019/20
Last compiled: 20th January 2020

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

Question 1 (i) Summing over paths from x to y which visit A only at the last step and using reversibility of the underlying Markov chain, one can check that

$$\pi(x)P_A(x,y) = \sum_{k=1}^{\infty} \pi(x)\mathbb{P}_x(X_k = y,\ \tau_A = k) = \sum_{k=1}^{\infty} \pi(y)\mathbb{P}_y(X_k = x,\ \tau_A = k) = \pi(y)P_A(y,x).$$

So P_A is reversible with respect to the conditional distribution $\pi_A(B) := \pi(A \cap B)/\pi(A)$.

(ii) Define $\tilde{\varphi}(x) := \mathbb{E}_x(\varphi(X_{\tau_A}))$. Then

$$\begin{aligned}(P_A\varphi)(x) &= \textstyle\sum_{y\in A} P_A(x,y)\mathbb{E}_y(\varphi(y)) = \mathbb{E}_x(\varphi(X_{\tau_A^+}))\\ &= \textstyle\sum_{y\in S} P(x,y)\mathbb{E}_y(\varphi(X_{\tau_A})) = \sum_{y\in S} P(x,y)\tilde{\varphi}(y) = (P\tilde{\varphi})(x).\end{aligned}$$

(iii) For $y \notin A$, we have $(P\tilde{\varphi})(y) = \varphi(y)$ by the Markov property; for $y \in A$, we have $\tilde{\varphi}(y) = \varphi(y)$. So

$$\mathcal{E}(\tilde{\varphi}) = \textstyle\sum_{x\in A}\big((I-P)\tilde{\varphi}(x)\big)\tilde{\varphi}(x)\pi(x) = \sum_{x\in A}\big((I-P_A)\varphi(x)\big)\varphi(x)\pi_A(x)\pi(A) = \pi(A)\mathcal{E}_A(\varphi).$$

Then, using that $\sum(f(x)-m)^2\pi(x)$ is minimised (over m) by $m := E_\pi(f)$, we obtain

$$\mathrm{Var}_\pi(\tilde{\varphi}) = \textstyle\sum_{x\in S}\big(\tilde{\varphi}(x) - E_\pi(\tilde{\varphi})\big)^2\pi(x) \geq \sum_{x\in A}\pi(A)\big(\varphi(x) - E_\pi(\varphi)\big)^2\pi_A(x) = \pi(A)\mathrm{Var}_{\pi_A}(\varphi).$$

Choosing φ to obtain the minimum in the variational form, the result now follows.

Question 2 Compare the chain with the SRW on the box observed at times when it visits the set A; denote that walk by using a tilde ($\sim$) as well as subscript A. By **Q1(iii)**, we have $\tilde{\gamma}_A \geq \gamma$, the spectral gap for SRW on the box. We know that $\gamma \gtrsim 1/n^2$. So it remains to show that $\gamma_A \geq \tilde{\gamma}_A$. To do this, we use the comparison technique.

By **Q1(i)**, the invariant distributions of P_A (ie the SRW on A) and $\tilde{P}_A$ (ie the SRW on the box observed only when in A) are the same. So, by Corollary 4.7, it suffices to show that $\mathcal{E}_A(f) \gtrsim \tilde{\mathcal{E}}_A(f)$ for all functions f, as then $\gamma_A \gtrsim \tilde{\gamma}_A$. From now on, drop the A-subscript—keep just the tildes ($\sim$).

Fix an arbitrary function $f : A \to \mathbb{R}$. Since P and $\tilde{P}$ have the same invariant distribution, it suffices to show that $P(x,y) \lesssim \tilde{P}(x,y)$. This, though, is straightforward. Note that there exist (x,y) with $P(x,y) = 0 \neq \tilde{P}(x,y)$, eg $x = (-1,0)$ and $y = (1,0)$. We need only study (x,y) with $P(x,y) \neq 0$, ie nearest neighbour (x,y). For such (x,y), we see that $P(x,y) = \frac{1}{4} = \tilde{P}(x,y)$. This completes the proof.

Question 3 (i) The rank of the matrix is 1 and 1 is an eigenvalue. Thus the remainder are 0, and so the spectral gap is 1.

(ii) For each $i \in [n]$, define

$$\tilde{P}_i(\tilde{x},\tilde{y}) := P(\tilde{x}_i,\tilde{y}_i)\cdot \textstyle\prod_{j\neq i}\mathbf{1}(\tilde{x}_j = \tilde{y}_j);$$

this is the transition matrix for the chain on S^n which moves coordinate i according to P. Then the chain in question, with transition matrix given by $\tilde{P}$, satisfies

$$\tilde{P}(\tilde{x},\tilde{y}) = \tfrac{1}{n}\textstyle\sum_{i\in[n]}\tilde{P}_i(\tilde{x},\tilde{y}).$$

Define the *tensor product* $f : S^n \to \mathbb{R}$ of $(f_i)_{i\in[n]}$ by

$$f(\tilde{x}) := f_1(\tilde{x}_1)\cdots f_n(\tilde{x}_n).$$

Let $(f_i)_{i\in[n]}$ be a collection of eigenfunctions of P and $(\lambda_i)_{i\in[n]}$ the corresponding eigenvalues. It follows that f is an eigenvalue with eigenfunction $\frac{1}{n}\sum_{i\in[n]}\lambda_i$. The second largest eigenvalue corresponds to taking $\lambda_i := 1$ for $i \neq 1$ and $\lambda_1 := 1-\gamma$, ie the second largest eigenvalue. Hence $\tilde{\gamma} = \frac{1}{n}$.

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

(iii) From the variational form (Theorem 4.3), we know that

$$\tilde{\gamma} = \min_{f:\mathrm{Var}_{\tilde{\pi}}(f)\neq 0} \frac{\mathcal{E}(f)}{\mathrm{Var}_{\tilde{\pi}}(f)} \quad \text{where} \quad \mathcal{E}(f) := \tfrac{1}{2}\textstyle\sum_{\tilde{x},\tilde{y}} \big(f(\tilde{x}) - f(\tilde{y})\big)^2 \tilde{\pi}(\tilde{x})\tilde{P}(\tilde{x},\tilde{y}).$$

Rearranging, we see that, for any function f, we have

$$\begin{aligned}\mathrm{Var}(f(X)) = \mathrm{Var}_{\tilde{\pi}}(f) &\leq \tfrac{1}{2}\tilde{\gamma}^{-1} \textstyle\sum_{\tilde{x},\tilde{y}\in S^n} \big(f(\tilde{x}) - f(\tilde{y})\big)^2 \tilde{\pi}(\tilde{x})\tilde{P}(\tilde{x},\tilde{y}) \\ &= \tfrac{1}{2}\textstyle\sum_{i=1}^n \sum_{\tilde{x}\in S^n} \tilde{\pi}(\tilde{x}) \sum_{y\in S} P(\tilde{x},y)\big(f(\tilde{x}) - f(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n)\big)^2 \\ &= \tfrac{1}{2}\textstyle\sum_{i=1}^n \mathbb{E}\big((f(X_1,\dots,X_n) - f(X_1,\dots,X_{i-1},X_i',X_{i+1},\dots,X_n))^2\big).\end{aligned}$$

Question 4 We construct a path coupling. Note that the distance between two configurations is *half* the number vertices with different colours (since one change corrects two at once).

Consider two configurations x and y which differ only at two positions, say i and j, so $d(x,y) = 1$; wlog $i < j$. We couple a single step. First pick two positions L and R uar and interchange the the particles in positions L and R in x to create x'. We choose L' and R' as follows.

- Suppose that $L, R \notin \{i,j\}$. We set $L' := L$ and $R' := R$. Then $d(x',y') = d(x,y) = 1$.
- Suppose that $L \in \{i,j\}$ and $R \notin \{i,j\}$. Set $R' := R$ and let L' be the element of $\{i,j\} \setminus \{L\}$.
  - Suppose that $y_L = y_R$. Then $y' = y$ and $x_{R'} = y_R = \sigma_L = \sigma_{L'}$, using the fact that x and y differ only at $\{i,j\}$. Hence $x' = x$. So $d(x',y') = d(x,y) = 1$.
  - Suppose that $y_L \neq y_R$. Then, similarly, $x' = y'$. So $d(x',y') = 0$.
- Suppose that $L \notin \{i,j\}$ and $R \in \{i,j\}$. Do the analogue of the previous case.
  - Suppose that $y_R = y_L$. Then $d(x',y') = 1$.
  - Suppose that $y_R \neq y_L$. Then $d(x',y') = 0$.
- Suppose that $L, R \in \{i,j\}$. Set $L' := R$ and $R' := L$. Then $d(x',y') = d(x,y) = 1$.

The distance either stays at 1 or decreases to 0; the latter happens only when one of L and R is chosen in $\{i,j\}$ and $y_L \neq y_R$. By counting edges connecting red-blue, this happens with probability

$$2 \cdot \left(\frac{1}{n}\frac{n-k-1}{n-1} + \frac{1}{n}\frac{k-1}{n-1}\right) = \frac{2}{n}\frac{n-2}{n-1} \geq \frac{1}{n}.$$

Observe that the diameter is at most k. To see this, simply 'correct' vertices one at a time. Hence, by the path coupling technique, the claim follows.

Question 5 Consider the following variation on the question: instead of flipping δn coordinates in one go, choose δn in one go and resample each independently. (This is a little like viewing the usual SRW at multiples of δn, but not quite the same—in essence, some sampling without replacement is going on.)

Just as with the usual SRW, once all coordinates have been resampled, by the independence, the chain is at equilibrium. In t steps, the probability that a given coordinate hasn't been sampled is $(1-\delta)^t$. Setting this equal to ε/n and solving for t gives $t = -\log(n/\varepsilon)/\log(1-\delta)$. By the union bound, the probability that all coordinates have been resampled at least once is at least $1-\varepsilon$. Hence

$$d_{\mathrm{TV}}\big((\log n + C)/(-\log(1-\delta))\big) \to 0 \quad \text{as } C \to \infty.$$

Question 6 Suppose that f and g are distinct colourings and let x be a vertex with $f(x) \neq g(x)$; let $c := g(x)$. Since $q \geq \Delta + 2$, for each neighbour y of x satisfying $f(y) = c$, we can find a different colour c_y' so that changing $f(y)$ to c_y' will result in a proper colouring. After making these changes, no neighbour of x has f-colour c, so changing $f(x)$ to c will again result in a proper colouring.

We have produced a proper colouring f' such that $d(f',g) < d(f,g)$, along with a path in the space of proper colourings from f to f'. By induction on the distance, we can produce a path from f to g.

Question 7 Using the explicit form of the transition matrix, it is straightforward to check that

$$P^t(\mathbf{0},\mathbf{1}) \le P^t(x,y) \le P^t(\mathbf{0},\mathbf{0}) \quad \text{for all} \quad x,y \in \{0,1\}^n.$$

Using this and the explicit form of $P^t(\mathbf{0},\mathbf{1})$, the claim is relatively easy to deduce.

Due to the laziness, once all have been picked, the system is exactly in the (full) invariant distribution. To get from $\mathbf{0}$ to $\mathbf{1}$, we need all to have been picked. Write A_t for the event that all coordinates have been picked in t steps. Thus

$$P^t(\mathbf{0},\mathbf{1}) = \mathbb{P}(A_t)\pi(\mathbf{1}), \quad \text{and hence} \quad 1 - P^t(\mathbf{0},\mathbf{1})/\pi(\mathbf{1}) = \mathbb{P}(A_t^c).$$

Estimating A_t is the coupon-collector problem. We have shown before that it concentrates at $n \log n$.

Question 8 We say that a random variable Z, or more precisely a sequence $(Z_n)_{n\in\mathbb{N}}$, *concentrates* if $\mathrm{Var}(Z) = o(\mathbb{E}(Z)^2)$. Assuming this, for all $\varepsilon \in (0,1)$, by Chebeshev, we have

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > \varepsilon\mathbb{E}(Z)) \le \varepsilon^{-2}\mathbb{E}(Z)^2/\mathrm{Var}(Z) = o(1).$$

Let X and Y evolve independently until they coalesce, and evolve together from then on. Clearly the coalescence time is upper bounded by the hitting time of 0 from n. Since the walk is biased with drift $-\frac{1}{6}$, this takes a time concentrated around $6n$. This set-up is that of the Gambler's ruin. The claim can be proved in this context. Alternatively remove the upper limit n to get a biased SRW on $\mathbb{Z}$; one can then apply the usual strong Markov property arguments to find the time to move from n to 0. This proves the upper bound.

For the lower bound, for every $\varepsilon \in (0,1)$, the set $A_\varepsilon := \{0, ..., \varepsilon n\}$ has $\pi(A_\varepsilon) \to 1$ as $n \to \infty$. Start the walk from n. It takes time concentrated around $6n(1-\varepsilon)$ to hit A, and hence $6n(1-2\varepsilon)$ is a lower bound on the mixing time. Similar arguments can be used to show this claim. For the SRW approach, start at $(1-\varepsilon)n$ and first show that with probability tending to 1 the SRW does not reach n before 0 (eg using optional stopping arguments). This proves the matching lower bound.

Question 9 Let $f : \{\pm 1\}^n \to \{\pm 1\} : x \mapsto x_1$. By the variation form (Theorem 4.3),

$$\gamma \le \mathcal{E}(f)/\mathrm{Var}_\pi(f) \quad \text{recalling that} \quad \mathcal{E}(f) = \tfrac{1}{2}\textstyle\sum_{x,y}\big(f(x) - f(y)\big)^2\pi(x)P(x,y).$$

Note that $E_\pi(f) = 0$, by symmetry, and so $\mathrm{Var}_\pi(f) = E_\pi(f^2) = 1$ since f takes values in ± 1. Hence if we can show that $\mathcal{E}(f) \le \frac{2}{n}$ then we deduce that $\gamma \le \frac{2}{n}$, and hence $t_{\mathrm{rel}} \ge \frac{1}{2}n$.

To show this bound on $\mathcal{E}(f)$, first note that $f(x) = f(y)$ unless $x_1 \ne y_1$. So to have $f(x) \ne f(y)$ and $P(x,y) \ne 0$, we need $x_1 \ne y_1$ and $x_j = y_j$ for $j \ne 1$. So we can replace the sum over $(x,y) \in \Omega^2$ with $x \in \Omega$ and y with $y_1 = -x_1$ and $y_j = x_j$ for $j \ne 1$.

If vertex i is selected for updating, then a positive spin is placed at i with probability

$$\tfrac{1}{2}\big(1 + \tanh(\beta S_1(x))\big) \quad \text{where} \quad S_1(x) := \textstyle\sum_{j\sim i} x_j.$$

Since vertices for updating are chosen uniform at random, for (x,y) as above, we have

$$P(x,y) = \begin{cases} \frac{1}{n}\cdot\frac{1}{2}\big(1 + \tanh(\beta S_1(x))\big) & \text{if} \quad x_1 = -1, \\ \frac{1}{n}\cdot\frac{1}{2}\big(1 - \tanh(\beta S_1(x))\big) & \text{if} \quad x_1 = +1. \end{cases}$$

Combining all these, we deduce that

$$\begin{aligned} \mathcal{E}(f) &= \tfrac{1}{2}\textstyle\sum_{x,y}\big(f(x) - f(y)\big)^2\pi(x)P(x,y) \\ &= \tfrac{1}{n}\textstyle\sum_{x:x_1=+1}\pi(x)\cdot\big(1 + \tanh(\beta S_1(x))\big) + \tfrac{1}{n}\textstyle\sum_{x:x_1=-1}\pi(x)\cdot\big(1 - \tanh(\beta S_1(x))\big) \\ &= \tfrac{1}{n} + \tfrac{1}{n}\textstyle\sum_{x:x_1=+1}\tanh(\beta S_1(x)) + \tfrac{1}{n}\textstyle\sum_{x:x_1=-1}\tanh(\beta S_1(x)). \end{aligned}$$

Now observe that $\pi(x) = \pi(-x)$. Hence, replacing $x \to -x$ in the second sum, we have

$$\mathcal{E}(f) = \tfrac{1}{n} + \tfrac{1}{n}\textstyle\sum_{x:x_1=+1}\pi(x)\big(\tanh(\beta S_1(x)) - \tanh(\beta S_1(-x))\big).$$

Note that $\tanh\theta \in (-1,1)$ for all $\theta \in \mathbb{R}$ and $\sum_{x:x_1=+1}\pi(x) = \frac{1}{2}$ by symmetry. As such, we obtain

$$\mathcal{E}(f) \le \tfrac{1}{n} + \tfrac{1}{n}\textstyle\sum_{x:x_1=+1} 2\pi(x) = \tfrac{1}{n} + \tfrac{1}{n} = \tfrac{2}{n} \quad \text{as required.}$$

In fact, $S_1(-x) = -S_1(x)$ and $\tanh(-\theta) = -\tanh(\theta)$, but this is not needed.