 P.Sousi@statslab.cam.ac.uk

Example Sheet 1

1. Let P be the transition matrix of a Markov chain with values in E and let μ and ν be two probability distributions on E. Show that

$$\|\mu P - \nu P\|_{\mathrm{TV}} \le \|\mu - \nu\|_{\mathrm{TV}} .$$

Deduce that $d(t) = \max_x \left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}}$ is decreasing as a function of t, where π is the invariant distribution.

2. Let $\Omega = \prod_{i=1}^n \Omega_i$, where Ω_i are finite sets. For each i, let μ_i and ν_i be probability distributions on Ω_i and set $\mu = \prod_{i=1}^n \mu_i$ and $\nu = \prod_{i=1}^n \nu_i$. Show that

$$\|\mu - \nu\|_{\mathrm{TV}} \le \sum_{i=1}^n \|\mu_i - \nu_i\|_{\mathrm{TV}} .$$

3. Let X and Y be Poisson random variables with parameters λ and μ respectively. Writing $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ for their laws, prove that

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\|_{\mathrm{TV}} \le |\lambda - \mu|.$$

4. Let Y be a random variable with values in $\mathbb{N}$ which satisfies

$$\mathbb{P}(Y = j) \le c, \text{ for all } j > 0 \text{ and } \mathbb{P}(Y = j) \text{ is decreasing in } j,$$

where c is a positive constant. Let Z be an independent random variable with values in $\mathbb{N}$. Prove that

$$\|\mathbb{P}(Y + Z = \cdot) - \mathbb{P}(Y = \cdot)\|_{\mathrm{TV}} \le c\mathbb{E}[Z] .$$

5. Let X be a Markov chain and let W and V be random variables taking values in $\mathbb{N}$ and suppose they are independent of X. Prove that

$$\|\mathbb{P}(X_W = \cdot) - \mathbb{P}(X_V = \cdot)\|_{\mathrm{TV}} \le \|\mathbb{P}(W = \cdot) - \mathbb{P}(V = \cdot)\|_{\mathrm{TV}}$$

6. Let $G = (V, E)$ be a finite connected graph with maximal distance between any two vertices equal to D. Suppose that X is a lazy simple random walk on G. Prove that for all $\varepsilon < 1/2$ we have

$$t_{\mathrm{mix}}(\varepsilon) \ge D/2.$$

7. Let X be a Markov chain in E with transition matrix P and invariant distribution π. Let $A \subseteq E$ be a subset with $\pi(A) \ge 1/8$. Let $\tau_A = \inf\{t \ge 0 : X_t \in A\}$. Prove that there exists a positive constant c so that

$$t_{\mathrm{mix}}(1/4) \ge c \max_x \mathbb{E}_x[\tau_A] .$$

1. Let P be the transition matrix of a Markov chain with values in E and let μ and ν be two probability distributions on E. Show that

$$\|\mu P - \nu P\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}}.$$

Deduce that $d(t) = \max_x \left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}}$ is decreasing as a function of t, where π is the invariant distribution.

Let $A = \{x : \mu P(x) \geq \nu P(x)\}$ $\quad B = \{x : \mu(x) \geq \nu(x)\}$

$\|\mu P - \nu P\|_{TV} = \frac{1}{2}\sum_x |\mu P(x) - \nu P(x)|$

$\mu P(A) - \nu P(A) \leq \mu P^{-1}(A) - \nu P^{-1}(A)$?

$= \sum_{x \in A}$

Let $t, s > 0$

π invariant

$d(t+s) = \max_x \|P^{t+s}(x,\cdot) - \pi\|_{TV} = \max_x \|P^t(x,\cdot)P^s - \pi P^s\|_{TV}$

$\leq \max_x \|P^t(x,\cdot) - \pi\|_{TV}$ from before

$= d(t)$

Hence $d(t+s) \leq d(t)$ $\forall t,s > 0$.

2. Let $\Omega = \prod_{i=1}^n \Omega_i$, where Ω_i are finite sets. For each i, let μ_i and ν_i be probability distributions on Ω_i and set $\mu = \prod_{i=1}^n \mu_i$ and $\nu = \prod_{i=1}^n \nu_i$. Show that

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sum_{i=1}^n \|\mu_i - \nu_i\|_{\mathrm{TV}}.$$

By induction it suffices to prove the result for $n=2$. Let (X_1, Y_1) be distributed with optimal coupling of (μ_1, ν_1) and (X_2, Y_2) with optimal coupling of (μ_2, ν_2). Let $X = (X_1, X_2)$ $Y = (Y_1, Y_2)$ then (X, Y) has law $(\mu_1 \times \mu_2, \nu_1 \times \nu_2)$.

$\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|_{TV} = \mathbb{P}(X \neq Y) \leq \mathbb{P}(X_1 \neq Y_1) + \mathbb{P}(X_2 \neq Y_2) = \|\mu_1 - \nu_1\|_{TV} + \|\mu_2 - \nu_2\|_{TV}.$

Needs checking

3. Let X and Y be Poisson random variables with parameters λ and μ respectively. Writing $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ for their laws, prove that

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\|_{\mathrm{TV}} \leq |\lambda - \mu|.$$

WLOG $\lambda \geqslant \mu$

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\|_{TV} = \frac{1}{2}\sum_n |f(n) - g(n)|$$

$$= \frac{1}{2}\sum_{n\geqslant 1}\left|\frac{e^{-\lambda}\lambda^n}{n!} - \frac{e^{-\mu}\mu^n}{n!}\right|$$

$$\leqslant \frac{1}{2}\sum_{n\geqslant 1}\frac{e^{-\lambda}}{n!}(\lambda^n - \mu^n)$$

$$= \frac{1}{2}(\lambda-\mu)\sum_{n\geqslant 1}\frac{e^{-\lambda}}{n!}(\lambda^{n-1} + \lambda^{n-2}\mu + \dots + \mu^{n-1})$$

$$\leqslant \frac{1}{2}(\lambda-\mu)\sum_{n\geqslant 1}\frac{e^{-\lambda}}{n!}(n\lambda^{n-1})$$

$$\leqslant \frac{1}{2}(\lambda-\mu)\left(1 + \sum_{n\geqslant 1} f(n)\right) = \frac{1}{2}\cdot 2(\lambda-\mu)$$

$$= \lambda - \mu \qquad \blacksquare$$

(From examples class)

Alternate proof: Take a coupling of (X,Y) by taking first $Y \sim \text{Poiss}(\mu)$ and then obtaining $Z \sim \text{Poiss}(\lambda-\mu)$ independently and set $X = Y + Z$.

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\| \leqslant \mathbb{P}(X \neq Y) = \mathbb{P}(Z \neq 0) = 1 - e^{-(\lambda-\mu)} \leqslant |\lambda - \mu|$$

Neat!

4. Let Y be a random variable with values in $\mathbb{N}$ which satisfies

$$\mathbb{P}(Y = j) \le c, \text{ for all } j > 0 \text{ and } \mathbb{P}(Y = j) \text{ is decreasing in } j,$$

where c is a positive constant. Let Z be an independent random variable with values in $\mathbb{N}$. Prove that

$$\|\mathbb{P}(Y + Z = \cdot) - \mathbb{P}(Y = \cdot)\|_{\mathrm{TV}} \le c\mathbb{E}[Z] \,.$$

$$\begin{aligned}
\|\mathbb{P}(Y+Z=\cdot) - \mathbb{P}(Y=\cdot)\|_{TV} &= \frac{1}{2}\sum_m |\mathbb{P}(Y+Z=m) - \mathbb{P}(Y=m)| \\
&\overset{\text{indep}}{=} \frac{1}{2}\sum_m \Big|\sum_{n=1}^{m-1}\mathbb{P}(Y=n)\mathbb{P}(Z=m-n) - \mathbb{P}(Y=m)\Big| \\
&\le \frac{1}{2}\sum_m \mathbb{P}(Y=m')\Big|\sum_{n=1}^{m-1}\frac{\mathbb{P}(Y=n)}{\mathbb{P}(Y=m')}\mathbb{P}(Z=m-n) - 1\Big| \\
&\le \frac{1}{2}\sum_m c\,\Big|\sum_{n=1}^{m-1}\mathbb{P}(Z=m-n) - 1\Big| \\
&= \frac{1}{2}\sum_m c\,|1 - \mathbb{P}(Z\le m-1)| \\
&= \frac{1}{2}\sum_m c\,\mathbb{P}(Z\ge m) = \frac{1}{2}c\,\mathbb{E}[Z]
\end{aligned}$$

assuming $\exists\, m' \ge m$ s.t. $\mathbb{P}(Y=m') > 0$.

Suppose not, let $m = \max\{n > 0 : \mathbb{P}(Y=n) > 0\}$

Then

$$\begin{aligned}
\|\mathbb{P}(Y+Z=\cdot) - \mathbb{P}(Y=\cdot)\| &= \frac{1}{2}\sum_n |\mathbb{P}(Y+Z=n) - \mathbb{P}(Y=n)| \\
&\le \underbrace{\frac{1}{2}\sum_{n=1}^{m} c\,\mathbb{P}(Z\ge n)}_{\text{From before}} + \underbrace{\frac{1}{2}\sum_{n=m+1}^{\infty}\mathbb{P}(Y+Z=n)}_{\text{since } \mathbb{P}(Y+Z=n)=0 \text{ here}} \\
&\le \frac{1}{2}c\sum_{n=1}^{m}\mathbb{P}(Z\ge n) + \frac{1}{2}c\sum_{n=1}^{\infty}\mathbb{P}(Z\ge n) \\
&\le c\sum_{n=1}^{\infty}\mathbb{P}(Z\ge n) = c\,\mathbb{E}[Z]
\end{aligned}$$

Examples Class solution: $\eta(n) = \mathbb{P}(Y=n)$ $\quad \zeta(n) = \mathbb{P}(Z=n)$

$$\begin{aligned}
\mathbb{P}(Y=n) - \mathbb{P}(Y+Z=n) &= \eta(n) - \sum_{k=1}^{\infty}\zeta(n)\eta(n-k) \\
&= \sum_{k=1}^{\infty}\zeta(k)\eta(n) - \sum_{k=1}^{\infty}\zeta(k)\eta(n-k) \\
&= \sum_{k=1}^{\infty}\zeta(k)[\eta(n) - \eta(n-k)] \\
&= \sum_{k=n}^{\infty}\zeta(k)\eta(n) + \sum_{k=1}^{n-1}\zeta(k)\underbrace{[\eta(n) - \eta(n-k)]}_{\le 0 \text{ since } \eta \text{ non-negative}} \le \sum_{k=n}^{\infty}\zeta(k)\eta(n) = \mathbb{P}(Z\ge n)\mathbb{P}(Y=n) \\
&\le c\,\mathbb{P}(Z\ge n)
\end{aligned}$$

... Then done

5. Let X be a Markov chain and let W and V be random variables taking values in $\mathbb{N}$ and suppose they are independent of X. Prove that

$$\|\mathbb{P}(X_W = \cdot) - \mathbb{P}(X_V = \cdot)\|_{\text{TV}} \leq \|\mathbb{P}(W = \cdot) - \mathbb{P}(V = \cdot)\|_{\text{TV}}$$

$\exists$ a coupling of $\mu = \mathbb{P}(W = \cdot)$ and $\nu = \mathbb{P}(V = \cdot)$ such that $\mathbb{P}(W \neq V) = \|\mathbb{P}(W = \cdot) - \mathbb{P}(V = \cdot)\|_{TV}$.

Then we have

$$\|\mathbb{P}(X_W = \cdot) - \mathbb{P}(X_V = \cdot)\|_{TV} \overset{\text{From lectures}}{\leq} \mathbb{P}(X_W \neq X_V) \overset{\text{since } W=V \Rightarrow X_W = X_V}{\leq} \mathbb{P}(W \neq V)$$

$$= \|\mathbb{P}(W = \cdot) - \mathbb{P}(V = \cdot)\|_{TV}$$

6. Let $G = (V, E)$ be a finite connected graph with maximal distance between any two vertices equal to D. Suppose that X is a lazy simple random walk on G. Prove that for all $\varepsilon < 1/2$ we have

$$t_{\text{mix}}(\varepsilon) \geq D/2.$$

Examples Class: Idea: Consider $\bar{d}$ rather than d for lower bounds.

If $k < \frac{1}{2}D$ and $\text{dist}(x,y) = D$ then $\text{supp}(P^t(x,\cdot)) \cap \text{supp}(P^t(y,\cdot)) = \emptyset$ ← i.e. they never meet

Then $d(k) \geqslant \frac{1}{2}\underbrace{\bar{d}(k)}_{=1 \text{ since they never meet}} = \frac{1}{2}$ and the result follows

Reviewed

7. Let X be a Markov chain in E with transition matrix P and invariant distribution π. Let $A \subseteq E$ be a subset with $\pi(A) \geq 1/8$. Let $\tau_A = \inf\{t \geq 0 : X_t \in A\}$. Prove that there exists a positive constant c so that

$$t_{\mathrm{mix}}(1/4) \geq c \max_x \mathbb{E}_x[\tau_A]\,.$$

Idea: Haven't hit a set of size $1/8$, know you are lying in only $7/8$ so you can't be mixed (take as your distinguishing set the $1/8$th)

Let $t = t_{mix}(1/16)$ then $d(t) \leq 1/16$ and hence for all sets A we have $P^t(x, A) \geq \pi(A) - 1/16$

Let A be such that $\pi(A) = 1/8$, then $P^t(x,A) \geq \frac{1}{16}$.

Hence $\tau_A \overset{\text{stochastically dominated by}}{\lesssim} t\,\mathrm{Geom}(1/16)$ since $\tau_A \leq \underbrace{\inf\{k \geq 0 : X_{kt} \in A\}}_{\lesssim \mathrm{Geom}(1/16)}$ from the above.

Hence by submultiplicitive property of $\bar{d}$ and using $\bar{d}(t) \leq 2d(t)$, the claim follows.

Note: $\tau_A := \inf\{s \geq 0 : X_s \in A\}$ $\quad \tau_A' = \inf\{k \geq 0 \mid X_{tk} \in A\} \cdot t$

$\Rightarrow \tau_A \leq \tau_A'$

$\mathbb{P}(X_{(k+1)t} \in A \mid X_{kt} = x) \overset{\text{By Markov Property}}{=} P^t(x, A) \geq 1/16$

Define $I_k = \{X_{kt} \in A\}$ then $\mathbb{P}(I_{k+1} \mid I_k^c) \geq \frac{1}{16}$ by conditioning on each $x \in I_k^c$

This gives that $\tau_A \lesssim \mathrm{Geom}(1/16)$

Define $X^{(i)}, Y^{(i)}$, $i = 1, .., d$. At each step, toss fair coin indep. of all previous. If Heads X moves, Y stays same. Tails $\Rightarrow$ vice-versa.

Let $U_1, U_2, ...$ be iid uniform RVs on $\{1, ..., d\}$. Given $t \geq 0$ let $N_i(t) = \sum_{j=1}^{t} \mathbb{1}(U_j = i)$ and let $X_t = (X^{(1)}_{N_1(t)}, ..., X^{(d)}_{N_d(t)})$.

This is a fancy way of saying pick a component uniformly at random.

Let $\tilde{\tau}_i$ be coupling time of $(X^{(i)}, Y^{(i)})$ then this coupling is coalescent.

Let $\tau_i = \min\{ t \geq 0 : N_i(t) \geq \tilde{\tau}_i \}$ and $\tau = \max\{\tau_1, .., \tau_d\}$ (Coupling time for (X, Y))

$$d(t) \leq \max_{x,y \in} \mathbb{P}_{x,y}(X_t \neq Y_t) = \max_{x,y} \mathbb{P}_{x,y}(\tau > t)$$

$$\leq \max_{x,y} \frac{\mathbb{E}[\tau]}{t} \quad \text{by Chebyshev}$$

$$\leq \max_{x,y} \frac{\mathbb{E}[\tau_1 + \dots + \tau_d]}{t}$$

$\mathbb{E}[\tau_i] = \mathbb{E}[\underbrace{\mathbb{E}[\tau_i \mid \tilde{\tau}_i]}_{d \cdot \tilde{\tau}_i}] = d k (n-k)$ where k is clockwise distance between x_i, y_i

$\leq d \cdot \frac{n^2}{4}$ (from lectures)

$$\leq \frac{d^2 n^2}{4t} \leq \frac{1}{4} \quad \text{for } t = n^2 d^2$$

Looks fine

(a)

Note $n - n^\alpha \notin \mathbb{Z} \Rightarrow \mathbb{E}[T] = \infty$.

Else let τ_i denote the number of coupons acquired in order to get $i+1$ distinct coupons from i distinct ones. Then $\tau_i \sim \text{Geom}\left(\frac{n-i}{n}\right)$

Thus $\mathbb{E}[T] = \mathbb{E}[\tau_0 + \cdots + \tau_{n-n^\alpha-1}] = \sum_{i=0}^{n-n^\alpha-1} \frac{n}{n-i} = n \sum_{i=0}^{n-n^\alpha-1} \frac{1}{n-i} = n \sum_{i=n^\alpha+1}^{n} \frac{1}{i}$

(b) $T = \tau_0 + \cdots + \tau_{n-n^\alpha-1}$ where $\tau_i \sim \text{Geom}\left(\frac{n-i}{n}\right)$

Thus $\text{Var}(T) \overset{\text{iid}}{=} \sum_{i=0}^{n-n^\alpha-1} \text{Var}(\tau_i) = \sum_{i=0}^{n-n^\alpha-1} \frac{1-\left(\frac{n-i}{n}\right)}{\left(\frac{n-i}{n}\right)^2} = n \sum_{i=0}^{n-n^\alpha-1} \frac{i}{(n-i)^2}$

$$\Rightarrow \text{Var}\left(\frac{T}{\mathbb{E}[T]}\right) = \frac{\text{Var}(T)}{(\mathbb{E}[T])^2} = \frac{n \sum_{i=1}^{n-n^\alpha-1} \frac{i}{(n-i)^2}}{n^2 \left(\sum_{i=n^\alpha+1}^{n} \frac{1}{i}\right)^2} = \frac{1}{n} \cdot \underbrace{\left(\frac{\sum_{i=1}^{n-n^\alpha-1} \frac{i}{(n-i)^2}}{\left(\sum_{i=n^\alpha+1}^{n} \frac{1}{i}\right)^2} \right)}_{A}$$

$$\sum_{i=0}^{n-n^\alpha-1} \frac{i}{(n-i)^2} \overset{n-i \to i}{=} \sum_{i=n^\alpha+1}^{n} \frac{n-i}{i^2} \leq \sum_{i=n^\alpha+1}^{n} \frac{1}{i^2} \leq \sum_{i=n^\alpha+1}^{n} \frac{1}{i} \Rightarrow A \leq 1$$

so $\text{Var}\left(\frac{T}{\mathbb{E}[T]}\right) \to 0$ as $n \to \infty$ and $\mathbb{E}\left[\frac{T}{\mathbb{E}[T]}\right] \to 1$

and by Chebyshev

$$\mathbb{P}\left(\left|\frac{T}{\mathbb{E}[T]} - 1\right| > \varepsilon\right) \leq \varepsilon\left[\text{Var}\left(\frac{T}{\mathbb{E}[T]}\right)\right]^2 \leq \varepsilon \frac{1}{n^2} \to 0 \quad \text{as } n \to \infty$$

Examples Class (a): $X = \sum_{i=1}^{n} \mathbb{I}\{\sigma(i) = i\}$ then

$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(\sigma(i) = i) = n\mathbb{P}(\sigma(1) = 1)$
$= n \cdot \frac{1}{n} = 1$

$\mathbb{E}[X^2] = \mathbb{E}[\sum_{i=1}^{n} \mathbb{I}\{\sigma(i) = i\} + 2\sum_{i<j} \mathbb{I}\{\sigma(i) = i \cap \sigma(j) = j\}]$
$= 1 + 2 \cdot \binom{n}{2} \frac{(n-2)!}{n!} = 2$

(a) # permutations $= n!$

permutations with ≥ 1 FP $= (n-1)!$

Prove by induction on n that $\mathbb{E}[X] = 1$, $\mathbb{E}[X^2] = 2$

$\mathbb{E}[X] = 2 \times \frac{1}{2} = 1$ $\quad \mathbb{E}[X^2] = 4 \times \frac{1}{2} = 2$

$\Rightarrow \mathrm{Var}(X) = 2 - 1^2 = 1$

$\underline{n-1 \to n}$ $\quad \mathbb{E}[X] = \mathbb{E}[X \mid X \geq 1]\mathbb{P}(X \geq 1) = \frac{(n-1)!}{n!} \mathbb{E}[X \mid X \geq 1] = \frac{(n-1)!}{n!} \binom{n}{n-1} \cdot 1$
$= 1$

$\mathbb{E}[X^2] = \mathbb{E}[X^2 \mid X \geq 1]\, \mathbb{P}(X \geq 1) = \frac{(n-1)!}{n!} \binom{n}{n-1} \cdot 2 = 2$

Thus $\mathbb{E}[X] = 1$ and $\mathrm{Var}(X) = 2 - 1^2 = 1$ $\forall n$.

(b) Idea: Let A be the set of σ with atleast one FP. Then $\pi(A) \xrightarrow{n \to \infty} 1 - \frac{1}{e}$ (Exclusion - Inclusion)

Show $P^t(id, A) \to 1$ for $t = (\frac{1}{2} - \varepsilon)\log n$

Formally, let τ be the first time where all cards have been selected to go either left or right.

Then $P^t(id, A) \geq \mathbb{P}(\tau > t)$. Let τ' be the first time that all cards selected in time slowed down by $1/2$ (to account for moving left and right).

Then $\mathbb{P}(\tau > t) = \mathbb{P}(\tau' > 2t)$ but τ' is a coupon collector with

$\tau' = \mathrm{Geom}(\frac{n}{n}) + \mathrm{Geom}(\frac{n-1}{n}) + \dots$ $\quad \mathbb{E}[\tau'] = n(1 + \frac{1}{2} + \dots + \frac{1}{n}) = n\log(n+1) > n\log n$

$\mathrm{Var}(\tau') < n^2\pi^2/6$. Set $t = (\frac{1}{2} - \varepsilon)\log n$ then $\mathbb{P}(\tau' \leq 2t) \to 0$

$$\mathbb{P}(\tau' \leq 2t) = \mathbb{P}(\tau' - \mathbb{E}[\tau'] \leq 2t - \mathbb{E}[\tau']) \leq \mathbb{P}(\tau' - \mathbb{E}[\tau'] \leq 2t - n\log n)$$

$$= \mathbb{P}(\tau' - \mathbb{E}[\tau'] \leq -2\varepsilon n\log n)$$

$$\leq \frac{\mathrm{Var}(\tau')}{4\varepsilon^2 (n\log n)^2} < \frac{\pi^2}{24\varepsilon^2} \frac{1}{(\log n)^2} \xrightarrow[n\to\infty]{} 0$$

Hence $P^t(id, A) \geq 1 - \mathbb{P}(\tau' \leq 2t) \xrightarrow[n\to\infty]{} 1$

$$\Rightarrow \quad d(t) \geq |\underbrace{P^t(id, A)}_{\to 1} - \underbrace{\pi(A)}_{\to 1 - 1/e}| \to e^{-1} > 1/4$$

(a) let $\lambda \leftrightarrow v$ for P then $Pv = \lambda v \Rightarrow \sum_j p_{ij} v_j = \lambda v_j \quad \forall \; 1 \leq i \leq \dim P = n$

$$\left| \sum_j p_{ij} v_j \right| \leq \sum_j p_{ij} |v_j| \leq \max_i |v_i| = |v_k|$$

$$\Rightarrow |\lambda| |v_k| = \left| \sum_j p_{kj} |v_j| \right| \leq |v_k| \quad \Rightarrow |\lambda| \leq 1$$

(b) let $\lambda = -1 \longleftrightarrow v$. Then $Pv = -v \Rightarrow P^{2t-1} v = -v$ & $P^{2t} v = v \quad \forall \, t \geq 1$.

$P^{2t+1} v = -v$: P^{2t+1} stochastic $\Rightarrow$ 1 is eigenvalue w/ eigenvector $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$$\sum_j p_{ij} v_j = -v_i \quad \forall \, i$$

$$\Rightarrow \sum_i \sum_j p_{ij} v_j = -\sum v_i$$

$Pv = -v \Rightarrow P^2 v = v \Rightarrow (P^2 + P) v = 0 \quad \dots$

(a) We note that $\sum_y \mathbb{E}_a[\sum_{t=0}^{\tau-1} \mathbb{1}(X=Y) P_{yx}]$

$= \mathbb{E}_a[\sum_{t=0}^{\tau-1} \sum_y \mathbb{1}(X_t=y) P_{yx}]$

Markov Property

$= \mathbb{E}_a[\sum_{t=0}^{\tau-1} \sum_y \mathbb{1}(X_t=y)]\mathbb{E}[\mathbb{1}(X_{t+1}=x) \mid X_t=y]$

$= \mathbb{E}_a[\sum_{t=0}^{\tau-1} \mathbb{1}(X_{t+1}=x)]$

Shifting sum by 1 & using sum of exp. $= \sum_{t=1}^{\tau} \mathbb{E}_a[\mathbb{1}(X_t=x)]$

?? $\mathbb{1}(X_t=x) = \mathbb{1}(X_0=x) = \delta_{ax}$ by construction so done

(b) Let $\tau_x^m = \inf\{t \geq m : X_t = x\}$.

(a) $\Rightarrow$ $\mathbb{E}_x[\sum_{t=0}^{\tau_x^m-1} \mathbb{1}(X_t=x)] = \pi(x)\mathbb{E}_x[\tau_x^m]$

Consider $\mathbb{E}_\pi[\sum_{t=0}^{\tau_x^m-1} \mathbb{1}(X_t=x)]$.

If $X_0 \sim \pi$ then $X_t \sim \pi$ for all $t \geq 0$, hence:

$\mathbb{E}_\pi[\sum_{t=0}^{\tau_x^m-1} \mathbb{1}(X_t=x)] = \sum_{t=0}^{m-1} P_\pi(X_t=x) + \mathbb{E}_\pi[\sum_{t=0}^{\tau_x^m-1} \cancel{\mathbb{1}}(X_t=x)]$

$= \pi(x)m$ since π invariant.

This is the start of it, proof goes on for a while, c.f. solutions

From before

Thus $\mathbb{E}_x[\sum_{t=0}^{\tau_x^m-1} \mathbb{1}(X_t=x)] - \mathbb{E}_\pi[\sum_{t=0}^{\tau_x^m-1} \mathbb{1}(X_t=x)] = \pi(x)(\mathbb{E}_x[\tau_x^m] - m)$

$= \pi(x)(\mathbb{E}_{P^m(x,\cdot)}[\tau_x^m])$

by Markov Property

Check solution

For the second part:

Spectral Method:

$$\frac{P^{2t}(x,x)}{\pi(x)} = \sum_1^{|\Omega|} f_j^2(x)\lambda_j^{2t} \qquad |\lambda_j| \leq 1$$

$$\geq \sum_1^{|\Omega|} f_j^2(x)\lambda_j^{2t+2}$$

$$= \frac{P^{2t+2}(x,x)}{\pi(x)}$$

Solutions to *Mixing Times of Markov Chains*
Example Sheet 1, 2019/20
Last compiled: 7th November 2019

These *are not* model solutions. There will likely be typos, and maybe even logical errors. These *are not* for redistribution without my explicit permission.

Question 1 Since P is a stochastic matrix, any eigenvalue λ of P satisfies $|\lambda| \le 1$. Indeed, suppose that $Pv = \lambda v$. Suppose that $|v_k| = \max\{|v_1|, ..., |v_n|\}$; so $|v_k| > 0$. Then

$$|\lambda v_k| = \big| \textstyle\sum_1^n P_{k,j} v_j \big| \le |v_k|$$

since the row-sums of P are 1 and its entries are non-negative. Thus $|\lambda| \le 1$.

We hence know that P has matrix-norm $\|P\| \le 1$. The result follows from the inequality $\|\xi P\|_1 \le \|P\| \|\xi\|_1$ for any row-vector ξ. (The usual inequality is for column-vectors, but it holds for row-vectors too since the eigenvalues of P are the same as those of its transpose.)

The final claim simply uses the fact that $\pi P = P$.

Question 2 Use the coupling-representation of total variation. For each $i = 1, ..., n$, let (X_i, Y_i) be an optimal coupling of (μ_i, ν_i). Let $X := (X_1, ..., X_n)$ and $Y := (Y_1, ..., Y_n)$. Then (X, Y) is a coupling of (μ, ν). Thus

$$\|\mu - \nu\|_{\mathrm{TV}} \le \mathbb{P}(X \ne Y) \le \textstyle\sum_1^n \mathbb{P}(X_i \ne Y_i) = \sum_1^n \|\mu_i - \nu_i\|_{\mathrm{TV}}.$$

Question 3 We can express $\mathrm{Po}(\lambda)$ the number of arrivals in a Poisson process of rate λ, written $\mathrm{PP}(\lambda)$, in time 1. Assume that $\lambda \ge \mu$. Couple two Poisson processes: sample a $\mathrm{PP}(\lambda)$; create a $\mathrm{PP}(\mu)$ by filtering this process using acceptance probability μ/λ.

Suppose that there are N arrivals (by time 1) in $\mathrm{PP}(\lambda)$. The probability that the two processes differ is, by the union bound, at most $N(1 - \mu/\lambda)$. Taking expectation over this, using the fact that $\mathbb{E}(N) = \lambda$, gives

$$\mathbb{P}\big(\mathrm{PP}(\lambda) = \mathrm{PP}(\mu)\big) \le \lambda(1 - \mu/\lambda) = \lambda - \mu.$$

Using the coupling-representation of total variation completes the proof.

Question 4 Since the law of Y is decreasing, for all $k \ge 0$ we have $\mathbb{P}(Y = j) \ge \mathbb{P}(Y + k = j)$ for all $j \ge 0$. Thus

$$\begin{aligned}
\big\|\mathbb{P}(Y + k = \cdot) - \mathbb{P}(Y = \cdot)\big\|_{\mathrm{TV}} &= \textstyle\sum_{j : \mathbb{P}(Y=j) \ge \mathbb{P}(Y+k=j)} \big(\mathbb{P}(Y = j) - \mathbb{P}(Y + k = j)\big) \\
&= \textstyle\sum_j \big(\mathbb{P}(Y = j) - \mathbb{P}(Y + k = j)\big) \\
&= 1 - \mathbb{P}(Y \ge k) = \mathbb{P}(Y \le k) \le ck.
\end{aligned}$$

The claim now follows from the fact that Z is independent of Y.

Question 5 Sample the infinite path $(X_k)_{k \ge 0}$, and sample V and W (independently of X). Using the independence, we have

$$\mathbb{P}\big(X_V = X_W\big) = \textstyle\sum_{v,w} \mathbb{P}\big(X_v = X_w\big) \mathbb{P}\big(V = v, W = w\big).$$

We lower bound $\mathbb{P}(X_v = X_w) \ge 0$ when $v \ne w$, but $\mathbb{P}(X_v = X_w) = 1$ when $v = w$. Hence

$$\mathbb{P}\big(X_V = X_W\big) \ge \textstyle\sum_u \mathbb{P}\big(V = u = W\big) = \mathbb{P}\big(V = W\big).$$

Using the coupling-representation of total variation completes the proof.

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

Question 6 Consider $\bar{d}$, rather than d. Note that if $k < D/2$ and x and y are at (maximal) graph distance D, then $P^k(x, \cdot)$ and $P^k(y, \cdot)$ are positive on disjoint sets. Hence $\bar{d}(k) = 1$ for $k < \frac{1}{2}D$. Thus $d(k) \ge \frac{1}{2}\bar{d}(k) = \frac{1}{2}$ for $k < \frac{1}{2}D$. Hence $t_{\text{mix}}(\varepsilon) \ge \frac{1}{2}D$ for all $\varepsilon < \frac{1}{2}$.

Question 7 Let $t := t_{\text{mix}}(\frac{1}{16})$. Then $d(t) \le \frac{1}{16}$, so for all sets A and starting points x we have

$$P^t(x, A) \ge \pi(A) - \tfrac{1}{16}.$$

Let A be a set with $\pi(A) \ge \frac{1}{8}$. Then, for all x, we have

$$P^t(x, A) \ge \tfrac{1}{16}.$$

Hence this shows that $\tau_A \lesssim t \cdot \text{Geom}(\frac{1}{16})$. Hence $\mathbb{E}_x(\tau_A) \lesssim t$.

Finally, from the submultiplicativity of $\bar{d}$ and the inequality $d(t) \le 2\bar{d}(t)$, we deduce the claim.

Question 8 Define a coupling in the following way:

- Choose a coordinate, from $\{1, ..., d\}$ uniformly at random.
- If the positions in the chosen coordinates of the two walks agree, then move the walks together in this coordinate.
- If they disagree, choose one of the chains uniformly to move and keep the other fixed.

Let $X_t := (X_t^1, ..., X_t^d)$ and $Y_t := (Y_t^1, ..., Y_t^d)$, and let

$$\tau_i := \inf\{t \ge 0 \mid X_t^i = Y_t^i\} \quad \text{for each} \quad i = 1, ..., d.$$

Then the coupling time $\tau_c = \max_i \tau_i$.

For each coordinate i, the difference $|X_t^i - Y_t^i|$, viewed at times when coordinate i is selected, behaves as a (non-lazy) SRW on $\mathbb{Z}_n$ with 0 absorbing. It is known that the expected time for a SRW to traverse a distance n is $\Theta(n^2)$—in fact, it is upper bounded by $\frac{1}{4}n^2$; see [Levin–Peres–Wilmer, Proposition 2.1]. Since each coordinate is selected with probability $1/d$, we have

$$\mathbb{E}(\tau_i) \le \tfrac{1}{4}dn^2 \quad \text{for any starting state.}$$

By the union bound and Markov's inequality, we obtain

$$\mathbb{P}(\tau_c > t) \le \textstyle\sum_1^d \mathbb{P}(\tau_i > t) \le \frac{1}{4}d^2n^2/t.$$

Hence $t_{\text{mix}}(\frac{1}{4}) \le d^2n^2$. (So take $c := d^2$.)

Extension. Prove that we may take $c \asymp d \log d$; in fact, $c := d\lceil \log_4(d/\varepsilon)\rceil$ works for $t_{\text{mix}}(\varepsilon)$.

Question 9 (a) The transition rates for X are given by

$$p(x, x+1) = 1 - x/n \quad \text{and} \quad p(x, x) = x/n \quad \text{for} \quad x = 0, ..., n.$$

Write $G_\alpha \sim \text{Geo}_1(\alpha)$ for a geometric random variable (with support $\{1, 2, ...\}$) with success probability α; then $\mathbb{E}(G_\alpha) = 1/\alpha$. Write

$$\tau_k := \inf\{t \ge 0 \mid X_t = k\}$$

for the time it takes to collect k coupons. Then, by the strong Markov property, we may write

$$\tau_k = \textstyle\sum_{i=0}^{k-1} G_{1-i/n}, \quad \text{and hence} \quad \mathbb{E}(\tau_k) = \sum_{i=0}^{k-1} 1/(1 - i/n) = n \sum_{i=0}^{k-1} 1/(n - i).$$

Taking $k := n - n^\alpha$ gives $\mathbb{E}(T) = \mathbb{E}(\tau_k)$. For general k, we may write

$$\tfrac{1}{n}\mathbb{E}(\tau_k) = \textstyle\sum_{j=1}^{n} 1/j - \sum_{j=1}^{n-k} 1/j.$$

Using expression for harmonic sums, we obtain

$$\tfrac{1}{n}\mathbb{E}(\tau_k) = \log\left(\tfrac{n}{n-k}\right) + \mathcal{O}\left(\tfrac{1}{n-k}\right).$$

If $k = n - n^\alpha$, then this becomes

$$\mathbb{E}(T) = n\big(\log(n^{1-\alpha}) + \mathcal{O}(n^{-\alpha})\big) = (1-\alpha)n \log n + \mathcal{O}(n^{1-\alpha}).$$

(b) Write $I_i(t)$ for the indicator that coupon i *has not* been collected by time t. Let $Y_t := \sum_1^n I_i(t)$ be the number *not* collected by time t; so $X_t + Y_t = n$. It is straightforward to show, by direct calculation, that $\{I_j(t)\}_1^n$ are negatively correlated and, writing $p_t := (1-1/n)^t$, we have

$$\mathbb{E}(Y_t) = np_t \quad \text{and} \quad \mathbb{V}\mathrm{ar}(Y_t) \le np_t(1-p_t) \le np_t.$$

(See [Levin–Peres–Wilmer, Lemma 7.13].) Take

$$t := c(1-\alpha)n\log n; \quad \text{then} \quad p_t = \exp\big(-c(1-\alpha)\log n + o(1)\big) = n^{-c(1-\alpha)+o(1)}.$$

Since $\mathbb{V}\mathrm{ar}(Y_t) \ll \mathbb{E}(Y_t)^2$, the random variable Y_t concentrates. This implies that

$$\mathbb{P}(Y_t > n^{1-\alpha}) = \begin{cases} 1 - o(1) & \text{when} \quad c > 1, \\ o(1) & \text{when} \quad c < 1, \end{cases}$$

and that the same inequality holds for $\mathbb{P}(Y_t < n^{1-\alpha})$ with $c > 1$ and $c < 1$ reversed. This implies concentration, in probability, of T around its expectation.

Question 10 (a) Consider $\sigma(1)$. By uniformity, it is uniformly distributed over $\{1, ..., n\}$. Hence, by linearity of expectation, the expected number of fixed points is

$$\mathbb{E}(X) = \mathbb{E}\big(\textstyle\sum_1^n \mathbf{1}(\sigma(i) = i)\big) = n \cdot 1/n = 1.$$

To calculate the variance, we use the representation

$$\mathbb{E}(X^2) = \mathbb{E}\big(\textstyle\sum_{i,j=1}^n \mathbf{1}(\sigma(i) = i,\ \sigma(j) = j)\big).$$

Now, if $\sigma(i) = i$, then $\sigma_{\setminus i}$, which is σ but with i removed, is a uniform permutation on $\{1, ..., n\} \setminus \{i\}$. Hence, for $i \ne j$, we have

$$\mathbb{P}(\sigma(i) = i,\ \sigma(j) = j) = \mathbb{P}(\sigma(j) = j \mid \sigma(i) = i)\mathbb{P}(\sigma(i) = i) = \tfrac{1}{n-1}\tfrac{1}{n}.$$

If $i = j$, then $\mathbb{P}(\sigma(i) = i,\ \sigma(j) = j) = 1/n$. Hence we obtain

$$\mathbb{E}(X^2) = n \cdot \tfrac{1}{n} + 2\tbinom{n}{2} \cdot \tfrac{1}{n-1}\tfrac{1}{n} = n \cdot \tfrac{1}{n} + 2 \cdot \tfrac{1}{2}n(n-1) \cdot \tfrac{1}{n-1}\tfrac{1}{n} = 1 + 1.$$

Hence $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = (1+1) - 1^2 = 1$.

(b) Start from the identity permutation. We use as our distinguishing statistic the number of fixed points; even more restrictively, we look only at which cards have not been touched, as they must be fixed points. After t steps, a total of $2t$ (not necessarily distinct) cards have been touched.

The same argument as used in **Question 9(b)** applies here.

Question 11 (a) We showed this at the start of **Question 1**.

(b) Suppose that -1 is an eigenvalue of P and let f be the corresponding eigenfunction. Then $Pf = -f$, which means that, for all x, we have

$$f(x) = -\textstyle\sum_y P(x,y)f(y).$$

Now choose x so that $|f(x)| = \max_y |f(y)|$. Taking absolute values above, we obtain

$$|f(x)| \le \textstyle\sum_y P(x,y)|f(y)| \le |f(x)|, \quad \text{and hence} \quad |f(x)| = \sum_y P(x,y)|f(y)|.$$

So if $P(x,y) > 0$ then $|f(x)| = |f(y)|$, since $|f(x)|$ is maximal. By irreducibility, this implies that $|f(x)| = |f(y)|$ for all y. By rescaling, suppose that $f(x) = 1$.

For every y with $P(x,y) > 0$, we then have $f(y) = -1$, since $f(x) = -\sum_y P(x,y)f(y)$ and $\sum_y P(x,y) = 1$. Since $P^2 f = f$, it follows that if x and y are such that $P^2(x,y) > 0$ then $f(x) = f(y)$, by the same reasoning. So we can partition the space into two sets:

$$A := \{y \mid f(y) = +1\} = \{y \mid P^t(x,y) > 0 \text{ for some even } t\};$$
$$B := \{y \mid f(y) = -1\} = \{y \mid P^t(x,y) > 0 \text{ for some odd } t\}.$$

This proves that the chain is periodic with period 2, ie $T(x) \subseteq \mathbb{Z}$, since if w and z belong to the same set of the partition then $P^t(w,z) = 0$ for all t odd.

Question 12 (a) Write $\rho_x := \mathbb{E}_a(\sum_{t=0}^{\tau} \mathbf{1}(X_t = x))$. We show that ρ is an invariant measure:

$$\textstyle\sum_j \rho_x p_{x,y} = \rho_y \quad \forall y.$$

Then, by uniqueness of the invariant measure up to constants,

$$\rho(\cdot) = c \cdot \pi(\cdot) \quad \text{for} \quad c = \textstyle\sum_y \rho_y = \mathbb{E}_a(\tau).$$

We use direct calculation:

$$\begin{aligned}
\rho_a &= \textstyle\sum_{t=0}^{\infty} \mathbb{P}_a\big(X_t = a,\, \tau > t\big) \\
&= \textstyle\sum_{t=0}^{\infty} \mathbb{P}_a\big(X_{t+1} = a,\, \tau > t\big) \quad \text{since } \mathbb{P}_a\big(X_\tau = a\big) = 1 \\
&= \textstyle\sum_{t=0}^{\infty} \sum_x \mathbb{P}_a\big(X_t = x,\, \tau > t,\, X_{t+1} = a\big) \\
&= \textstyle\sum_{t=0}^{\infty} \sum_j p_{x,a} \mathbb{P}_a\big(X_t = x,\, \tau > t\big) \quad \text{by the Markov propery} \\
&= \textstyle\sum_x \rho_x p_{x,a}.
\end{aligned}$$

Question 13 (a) Using reversibility and Cauchy–Schwarz, we obtain

$$\begin{aligned}
\frac{P^{2t}(x,y)}{\pi(y)} &= \sum_z P^t(x,z)\frac{P^t(z,y)}{\pi(y)} = \sum_z \frac{P^t(x,z)}{\pi(z)}\frac{P^t(y,z)}{\pi(z)}\pi(z) \\
&\le \sqrt{\sum_z \pi(z)\frac{P^t(x,z)}{\pi(z)}\frac{P^t(z,x)}{\pi(x)}} \cdot \sqrt{\sum_z \pi(z)\frac{P^t(y,z)}{\pi(z)}\frac{P^t(z,y)}{\pi(y)}} \\
&= \sqrt{\frac{P^{2t}(x,x)}{\pi(x)} \cdot \frac{P^{2t}(y,y)}{\pi(y)}}.
\end{aligned}$$

(b) Since $P^{2t+2}(x,x) = \sum_{y,z} P^t(x,y)P^2(y,z)P^t(z,x)$, using reversibility we have

$$\pi(x)P^{2t+2}(x,x) = \textstyle\sum_{y,z} P^t(y,x)\pi(y)P^2(y,z)P^t(z,x) = \sum_{y,z}\psi(y,z)\psi(z,y)$$
$$\text{where} \quad \psi(y,z) := P^t(y,x)\sqrt{\pi(y)P^2(y,z)}.$$

Applying Cauchy-Schwarz and using reversibility, we obtain

$$\pi(x)P^{2t+2}(x,x) \le \textstyle\sum_{y,z}\psi(y,z)^2 = \sum_y \big(P^t(y,x)\big)^2\pi(y) = P^{2t}(x,x).$$

8. Let X be a lazy simple random walk on the d-dimensional discrete torus $\mathbb{Z}_n^d$. Show that there exists a positive constant c (depending on the dimension d) so that

$$t_{\text{mix}}(1/4) \le cn^2.$$

9. A company issues n different coupons. In order to win the prize, a collector needs all n coupons. We suppose that each coupon he acquires is equally likely to be each of the n types. Let X_t denote the number of different types represented among the collector's first t coupons. For $\alpha \in (0,1)$, define $T = \min\{t \ge 0 : X_t = n - n^\alpha\}$.

(a) What is $\mathbb{E}[T]$?

(b) Show that $T/\mathbb{E}[T] \to 1$ in probability as $n \to \infty$.

10. (a) Let S_n be the symmetric group and let $\sigma \in S_n$ be a uniform random permutation. Let X denote the number of fixed points of σ, i.e. the number of $1 \le i \le n$ such that $\sigma(i) = i$. Show that $\mathbb{E}[X] = 1$ and $\mathrm{Var}\,(X) = 1$.

(b) Consider the random transposition shuffle as a method of shuffling a deck of n cards. At each step, the shuffler chooses two cards, L_t and R_t, independently and uniformly at random. If L_t and R_t are different, then transpose them. Otherwise, do nothing. Prove that for any $\varepsilon > 0$ and all n sufficiently large we have

$$t_{\text{mix}}(1/4) \ge \left(\frac{1}{2} - \varepsilon\right) n \log n.$$

11. (a) Let P be a transition matrix. Show that if λ is an eigenvalue, then $|\lambda| \le 1$.

(b) Suppose that P is irreducible and for every x consider the set $T(x) = \{t : P^t(x,x) > 0\}$. Show that $T(x) \subseteq 2\mathbb{Z}$ if and only if -1 is an eigenvalue of P.

12. (a) Let τ be a stopping time for a finite and irreducible Markov chain satisfying $\mathbb{E}[\tau] < \infty$ and $\mathbb{P}_a(X_\tau = a) = 1$. Show that for all x

$$\mathbb{E}_a\left[\sum_{t=0}^{\tau} \mathbf{1}(X_t = x)\right] = \pi(x)\mathbb{E}_a[\tau]\,.$$

(b) Consider a finite, irreducible and aperiodic Markov chain. Prove that for all x

$$\pi(x)\mathbb{E}_\pi[\tau_x] = \sum_{t=0}^{\infty} \left(P^t(x,x) - \pi(x)\right).$$

Hint: Count the number of visits to x up until $\tau_x^m = \inf\{t \ge m : X_t = x\}$ in two different ways: using part (a) and also using the convergence to equilibrium theorem.

13. Let P be the transition matrix of a finite reversible chain with invariant distribution π.

Using the Cauchy-Schwarz inequality or otherwise prove that for all x, y and all t

$$\frac{P^{2t}(x,y)}{\pi(y)} \le \sqrt{\frac{P^{2t}(x,x)}{\pi(x)} \cdot \frac{P^{2t}(y,y)}{\pi(y)}} \quad \text{and} \quad P^{2t+2}(x,x) \le P^{2t}(x,x).$$