Stochastic Models — Second HW problem set

Gábor Pete
http://www.math.bme.hu/~gabor

April 5, 2021

The number of dots $^\bullet$ is the value of an exercise. **Hand in solutions for 12 points by April 19 Monday 11:59 am.** If you have seriously tried to solve some problem, but got stuck, I will be happy to help. Also, if your final solution to a problem has some mistake but has some potential to work, then I will give it back and you can try and correct the mistake.

The first exercise would have been better in the first set, but I forgot to put it there:

$\triangleright$ **Exercise 1.**$^{\bullet\bullet}$ Recall that a discrete P-harmonic function $f : V \longrightarrow \mathbb{R}$ on the state space of a Markov chain P is one that satisfies $Pf(x) = f(x)$ for every $x \in V$.

(a) Show that if P is irreducible on a finite set V, then every harmonic function is constant.

(b) Let $(X_n)_{n\geq 0}$ be simple random walk on the 3-regular infinite tree, $\mathbb{T}_3$, with Markov operator P. Take any vertex $o \in V(\mathbb{T}_3)$, and let A be one of the three connected components of $\mathbb{T}_3 \setminus \{o\}$. Show that $f(x) := \mathbf{P}_x[\,\exists n_0 : X_n \in A\ \forall n \geq n_0\,]$ is a non-constant bounded harmonic function, where, remember, $\mathbf{P}_x[\,\cdot\,]$ means that $X_0 = x$.

(c) Consider the lamplighter graph $G = \mathbb{Z}_2 \wr \mathbb{Z}^3$, with the standard 7 generators (six for moving the marker in the city $\mathbb{Z}^3$, one for switching the lamp where the marker is). Give an example of a non-constant bounded harmonic function for SRW on G. (Hint: follow the strategy of the previous part, but with a different notion of "what happens eventually". Namely, note that the marker visits the origin of $\mathbb{Z}^3$ only a finite number of times, hence there is a "final" state of the lamp there.)

Remark: If you recall the "amazing theorem" mentioned in class, that the existence of bounded harmonic functions on a transitive graph is equivalent to the speed of the SRW being positive, then this $\mathbb{Z}_2 \wr \mathbb{Z}^3$ is an example where the graph is amenable, but the speed is positive.

Now, finite Markov chains.

$\triangleright$ **Exercise 2.**$^{\bullet\bullet}$ This is a linear algebra reminder on why there is an orthonormal basis of real-valued eigenvectors for any reversible Markov chain on a finite set. We have partly done this in class, and you should have seen it in some linear algebra course anyway, so please hand it in only if it is new for you.

When the state space V has n elements, and π is a reversible distribution, recall that for $u, v \in \mathbb{C}^n$ we defined the inner product $(u, v) = (u, v)_\pi := \sum_{x\in V} u(x)\overline{v(x)}\pi(x)$.

(a) Show that $(v, u) = \overline{(u, v)}$, and $(Pu, v) = (u, Pv)$. Deduce that if $v \in \mathbb{C}^n$ is an eigenvector of P with eigenvalue λ, then $\lambda \in \mathbb{R}$.

(b) From the fundamental theorem of algebra we know that $\det(P - \lambda I)$ has a root $\lambda \in \mathbb{C}$. Recall that this implies that there exists a nonzero $v_\lambda \in \mathbb{C}^n$ in the kernel of $P - \lambda I$, hence λ is an eigenvalue, with eigenvector v_λ.

(c) Show that $v^\perp := \{u \in \mathbb{C}^n : (u, v) = 0\}$ is a linear subspace for any v, and $Pv_\lambda^\perp \subseteq v_\lambda^\perp$, when v_λ is the eigenvector found in the previous part.

(d) Prove by induction that P has an orthonormal basis of eigenvectors $v_1, \dots, v_n \in \mathbb{C}^n$, with all real eigenvalues $\lambda_1, \dots, \lambda_n$.

(e) Show that P also has an orthonormal basis of eigenvectors $u_1, \dots, u_n \in \mathbb{R}^n$, with the same eigenvalues.

▷ **Exercise 3.**$^\bullet$ Let P be a reversible Markov chain on n states; that is, the random walk on a finite graph G with symmetric edge-weights. We have seen that P has eigenvalues $-1 \le \lambda_n \le \cdots \le \lambda_1 = 1$. Show that $\lambda_n > -1$ if and only if G is not bipartite.

▷ **Exercise 4.**$^\bullet$ Let P be any reversible finite Markov chain. Let $\bar{P}$ be its 1/2-lazy version: in each step, with probability 1/2 we stay put, while with probability 1/2 we take a step according to P. This is a usual way to get rid of periodicity. Show that the spectrum of $\bar{P}$ is contained in the interval $[0, 1]$.

▷ **Exercise 5.**$^{\bullet\bullet}$ For simple random walk on any finite or infinite d-regular graph, show that after any even number of steps the most likely position is the starting vertex.

▷ **Exercise 6.**$^{\bullet\bullet}$ Let P be a reversible Markov chain on a finite state space V, with reversible distribution π. Recall that the chain is then just the random walk w.r.t. the symmetric edge-weights $c(x, y) := \pi(x)p(x, y)$. There is the following version of the Courant-Fisher-Rayleigh theorem (which you don't have to prove):

$$\lambda_2 = \sup \left\{ \frac{(Pf, f)_\pi}{\|f\|_\pi} : \mathbf{E}_\pi[\, f\,] := \sum_{x \in V} f(x)\pi(x) = 0 \right\}.$$

Using this, show that the **spectral gap** has the following formula:

$$1 - \lambda_2 = \inf \left\{ \frac{\frac{1}{2}\sum_{x,y} \big(f(x) - f(y)\big)^2 c(x, y)}{\mathrm{Var}_\pi[f]} : \mathrm{Var}_\pi[f] := \mathbf{E}_\pi[f^2] - (\mathbf{E}_\pi f)^2 \neq 0 \right\}.$$

Show that the numerator can be written as $\mathbf{E}_{X_0 \sim \pi}\Big[\mathrm{Var}\big[f(X_1) \,\big|\, X_0\big]\Big]$. Thus, this formula is the infimum ratio of the local variance to the global one.

▷ **Exercise 7.** Finding good functions in the formula of the previous exercise will give you upper bounds on the spectral gap (hence lower bounds on the relaxation time, see the next exercise) of reversible Markov chains. Using this strategy, show:

(a)$^\bullet$ On the cycle C_n, the gap is at most $O(1/n^2)$.

(b)$^\bullet$ On the hypercube $\{0, 1\}^k$, the gap is at most $O(1/k)$.

(c)$^\bullet$ On the dumbbell graph (two complete graphs K_n joined by a single edge), the gap is at most $O(1/n^2)$.

(d)$^\bullet$ What bound can you give on the following lollipop graph: a complete graph K_n, with a length n^2 path emanating from it?

▷ **Exercise 8.**$^{\bullet\bullet}$ Consider a reversible Markov chain P on a finite state space V with reversible distribution π and absolute spectral gap g_{abs}. This exercise explains why $T_{\mathrm{relax}} = 1/g_{\mathrm{abs}}$ is called the **relaxation time**.

Show that $g_{\mathrm{abs}} > 0$ implies that $\lim_{t\to\infty} P^t f(x) = \mathbf{E}_\pi f$ for all $x \in V$. Moreover,

$$\mathrm{Var}_\pi[P^t f] \le (1 - g_{\mathrm{abs}})^{2t}\, \mathrm{Var}_\pi[f]\,,$$

with equality at the eigenfunction corresponding to the λ_i giving $g_{\mathrm{abs}} = 1 - |\lambda_i|$. Hence T_{relax} is the time needed to reduce the standard deviation of any function to $1/e$ of its original standard deviation.

We have accepted that the total variation distance between probability measures can be written as

$$d_{\mathrm{TV}}(\mu, \nu) = \min \big\{ \mathbf{P}[\, X \neq Y\,] : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu \big\}. \tag{1}$$

Consider now any Markov chain with a unique stationary measure π, and define

$$d(t) := \sup_{x \in V} d_{\mathrm{TV}}\big(p_t(x, \cdot), \pi(\cdot)\big) \quad \text{and} \quad \bar{d}(t) := \sup_{x,y \in V} d_{\mathrm{TV}}\big(p_t(x, \cdot), p_t(y, \cdot)\big)\,.$$

Furthermore, define the **total variation mixing time** by

$$T_{\mathrm{mix}}(\epsilon) := \inf\big\{t : d(t) \le \epsilon\big\} \qquad \text{and} \qquad T_{\mathrm{mix}} := T_{\mathrm{mix}}(1/4).$$

The following exercise explains why we introduced $\bar{d}(t)$ and why this 1/4 definition is a good one.

▷ **Exercise 9.**

(a)• Show that $d(t) \le \bar{d}(t) \le 2d(t)$.

(b)• Using (1), show that $\bar{d}(t+s) \le \bar{d}(t)\,\bar{d}(s)$.

(c)• Conclude from the previous two parts that $T_{\mathrm{mix}}(2^{-\ell}) \le \ell\, T_{\mathrm{mix}}(1/4)$.

▷ **Exercise 10.**•• Consider simple random walk on the dumbbell graph: take two copies of the complete graph K_n, add a loop at each vertex (so that the degrees become n), except at one distinguished vertex in each copy, which will be connected to each other by an edge. Show that $d(1) = 1/2$, but $T_{\mathrm{mix}} \ge cn^2$ for some uniform $c > 0$. That is, in the definition of T_{mix}, the 1/4 should not be replaced by 1/2.

▷ **Exercise 11.**•• Consider lazy SRW on the cycle C_n. Using the Central Limit Theorem, show that for any $t > 0$ there exists $\delta_0(t), \delta_1(t) > 0$, such that, for any n, we have $\delta_0(t) < d(tn^2) < 1 - \delta_1(t)$. Moreover, show that one can achieve $\lim_{t\to 0} \delta_0(t) = 1$. This proves the lower bound $T_{\mathrm{mix}} \ge cn^2$ for some uniform $c > 0$.

Note that (1) implies that, for any coupling of two copies of the Markov chain, $x = X_0, X_1, \dots$ and $y = Y_0, Y_1, \dots$ on V,

$$d_{\mathrm{TV}}\big(p_t(x,\cdot), p_t(y,\cdot)\big) \le \mathbf{P}[\,X_t \ne Y_t \mid X_0 = x,\, Y_0 = y\,]\,.$$

If we choose $y = Y_0$ according to the stationary measure, then Y_t is also stationary, hence, in any coupling,

$$d(t) \le \max_x \mathbf{P}[\,X_t \ne Y_t \mid X_0 = x,\, Y_0 \sim \pi\,]\,.$$

This results in the **coupling method** to give upper bounds on the TV-mixing time, as in the next exercise.

▷ **Exercise 12.**•• For lazy SRW on the cycle C_n, show by coupling that $\lim_{t\to\infty} \delta_1(t) = 1$ can also be achieved in the previous exercise, hence the mixing time is $T_{\mathrm{mix}} \le O(n^2)$, so altogether $\asymp n^2$.

The following three exercises together prove that the total variation mixing time of the 1/2-lazy random walk $X_0, X_1, \dots$ on the hypercube $\{0,1\}^k$ is $\sim \frac{1}{2} k \log k$.

▷ **Exercise 13.**•• Let Y_t be the number of missing coupons at time t in the coupon collector's problem with k coupons. Show that, for $\alpha \in (0,1)$ fixed,

$$\mathbf{E}\, Y_{\alpha\, k \log k} \sim k^{1-\alpha} \quad \text{and} \quad \mathbb{D}\, Y_{\alpha\, k\log k} = o(k^{1-\alpha}).$$

Using Markov's and Chebyshev's inequalities, deduce that $Y_{\alpha\, k\log k}/\sqrt{k} \to 0$ or ∞ in probability, for $\alpha > 1/2$ and $< 1/2$, respectively.

▷ **Exercise 14.**•• Let $\mathsf{N}(\mu,\sigma^2)$ denote the normal distribution. Show that, for any sequence $\sigma_k \to \sigma \in (0,\infty)$, we have that $d_{\mathrm{TV}}\big(\mathsf{N}(0,\sigma^2),\, \mathsf{N}(\mu_k,\sigma_k^2)\big) \to 0$ or 1, for $\mu_k \to 0$ and $\mu_k \to \infty$, respectively. Using this and the local version of the de Moivre–Laplace theorem, prove that

$$d_{\mathrm{TV}}\big(\mathsf{Binom}(k,1/2),\, \mathsf{Binom}(k-k^\beta,1/2)+k^\beta\big) \to \begin{cases} 0 & \text{if } \beta < 1/2\,, \\ 1 & \text{if } \beta > 1/2\,. \end{cases}$$

▷ **Exercise 15.**

(a)• For $X_0 = (0,0,\dots,0) \in \{0,1\}^k$, let the distribution of X_t be μ_t. What is it, conditioned on $\|X_t\|_1 = \ell$?

(b)• What is the distribution of $\|Z\|_1$, where Z has distribution π, uniform on $\{0,1\}^k$?

(c)$^{\bullet\bullet}$ Let Y_t be the number of coordinates that have not been rerandomized by time t in X_t. Compare the distribution of $k - \|X_t\|_1$, conditioned on $Y_t \geq y$, to $\mathsf{Binom}(k-y, 1/2) + y$. Deduce from the previous parts and the previous exercises that $d_{\mathrm{TV}}\big(\mu_{\alpha\, n \log n}, \pi\big) \to 0$ or 1, for $\alpha > 1/2$ and $< 1/2$, respectively.

The L^∞**- or uniform mixing time** of a Markov chain is usually defined as

$$T_{\mathrm{mix}}^{\infty} := \inf\left\{ t : \sup_{x,y} \left| \frac{p_t(x,y)}{\pi(y)} - 1 \right| < \frac{1}{e} \right\}.$$

▷ **Exercise 16.**$^{\bullet\bullet}$ Using Exercise 13, show that the uniform mixing time of the hypercube $\{0,1\}^k$ is $\sim k \log k$.

▷ **Exercise 17.** This exercise explains why it is hard to construct large expander graphs. A *covering map* $\varphi : G' \longrightarrow G$ between graphs is a surjective graph homomorphism that is locally an isomorphism: denoting by $N_G(v)$ the subgraph induced by $v \in G$ and all its neighbours, we require that each connected component of the subgraph of G' induced by the full inverse image $\varphi^{-1}(N_G(v))$ be isomorphic to $N_G(v)$.

(a)$^{\bullet}$ If $G' \longrightarrow G$ is a covering map of infinite graphs, then the spectral radii satisfy $\rho(G') \leq \rho(G)$, i.e., the larger graph is more non-amenable. In particular, if G is an infinite k-regular graph, then $\rho(G) \geq \rho(\mathbb{T}_k) = \frac{2\sqrt{k-1}}{k}$. (Hint: use the return probability definition of $\rho(G)$.)

(b)$^{\bullet}$ If $G' \longrightarrow G$ is a covering map of finite graphs, then $\lambda_2(G') \geq \lambda_2(G)$, i.e., the larger graph is a worse expander. (Hint: eigenfunctions on G can be "lifted" to G'.