Proof. In view of the previous results,

$$\pi(x)\mathbb{E}_\pi(\tau_x) \geq \sum_{k=1}^{t}(P^k(x,x) - \pi(x)) \geq t(P^t(x,x) - \pi(x)).$$

Dividing by $t\pi(x)$,

$$\frac{\mathbb{E}_\pi(\tau_x)}{t} \geq \left|\frac{P^t(x,x)}{\pi(x)} - 1\right|.$$

Therefore, for $t \geq 2\max_{x\in V}\mathbb{E}_\pi(\tau_x)$ we have

$$\frac{1}{4} \geq \max_{x\in V}\frac{\mathbb{E}_\pi(\tau_x)}{2t} \geq \max_{x\in V}\left|\frac{P^{2t}(x,x)}{\pi(x)} - 1\right| \geq d_2(t)^2 \geq d_1(t)^2 = (2d(t))^2,$$

and so $d(t) \leq \frac{1}{4}$. □

This provides a sharp bound (up to a constant) for the lazy random walk on the cycle

Example 9.91. Label the states of $\mathbb{Z}_n$ with $\{0,\ldots,n-1\}$. By identifying the states 0 and n, we can see that $\mathbb{E}_k(\tau_0)$ for the lazy simple random walk on the cycle must be the same as the expected time to ruin or success in a lazy gambler's ruin on the path $\{0,1,\ldots,n\}$. Hence, for lazy simple random walk on the cycle we have

$$t_{\text{hit}} = \max_{x,y\in V}\mathbb{E}_y(\tau_x) = \max_{0\leq k\leq n} 2k(n-k) = \left\lfloor\frac{n^2}{2}\right\rfloor.$$

Therefore, the previous theorem gives $t_{\text{mix}} \leq n^2 + 1$.

Let us present a proof for Lemma 9.88.

Proof of Lemma 9.88. Consider the inner product

$$\langle f, g\rangle = \sum_{x\in V} f(x)g(x)\pi(x).$$

If P is the transition matrix of a reversible chain, then P is self-adjoint with respect to this inner product. Indeed, notice that

$$\langle Pf, g\rangle = \sum_{x\in V}\sum_{v\in V}\pi(v)P(v,x)f(x)g(v) = \sum_{x\in V}\sum_{v\in V}\pi(x)P(x,v)f(x)g(v) = \langle f, Pg\rangle.$$

Therefore, the spectral theorem guarantees that there is a orthonormal basis $\{f_j\}_{j=1}^{|V|}$ such that f_j are eigenfunctions of the real eigenvalues λ_j. Moreover, these eigenvalues are nonnegative by lazyness. Let δ_y be the function given by $\delta_y(x) = 1$ if $x = y$ and $\delta_y(x) = 0$ if $x \neq y$. We can express δ_y using the orthonormal basis $\{f_j\}_{j=1}^{|V|}$ as

$$\delta_y = \sum_{j=1}^{|V|}\langle\delta_y, f_j\rangle f_j = \sum_{j=1}^{|V|} f_j(y)\pi(y)f_j.$$

Since $P^t f_j = \lambda_j^t$ and $P^t(x,y) = (P^t\delta_y)(x)$, we obtain

$$P^t(x,y) = \sum_{j=1}^{|V|} f_j(y)\pi(y)\lambda_j^t f_j(x).$$

We can take $y = x$ to see that $P^t(x,x)$ is a decreasing function of t. □

Let (X_t) be a Markov chain on a state space V with n elements. The ***cover time variable*** τ_{cov} of (X_t) is the first time at which all the states have been visited, that is,

$$\tau_{\text{cov}} = \min\{t \geq 0 \colon V \subseteq \{X_0, X_1, \ldots, X_t\}\}.$$

We also define the ***cover time*** as the mean of τ_{cov} from the worst-case initial state:

$$t_{\text{cov}} = \max_{x \in V} \mathbb{E}_x \tau_{\text{cov}}.$$

Recall the definition of t_{hit}, and let x, $y \in V$ be states for which $t_{\text{hit}} = \mathbb{E}_x(\tau_y)$. Since any walk started at x must have visited y by the time all states are covered, we have

$$t_{\text{hit}} = \mathbb{E}_x(\tau_y) \leq \mathbb{E}_x(\tau_{\text{cov}}) \leq t_{\text{cov}}.$$

The next result gives an upper bound on cover times in terms of hitting times

Theorem 9.92 (Matthews)**.** *Let (X_t) be an irreducible finite Markov chain on $n > 1$ states. Then,*

$$t_{\text{cov}} \leq t_{\text{hit}} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1}\right).$$

Proof. Assume that our state space is $\{1, \ldots, n\}$ and our starting state is n. Let σ be a uniform random permutation on $\{1, \ldots, n-1\}$, chosen independently of the chain. Let T_k be the first time that the states $\sigma(1), \ldots, \sigma(k)$ have all been visited, and let $L_k = X_{T_k}$ be the last state among $\sigma(1), \ldots, \sigma(k)$ to be visited.

For any $1 \leq s \leq n-1$, we have

$$\mathbb{E}_n(T_1 | \sigma(1) = s) = \mathbb{E}_n(\tau_s) \leq t_{\text{hit}}.$$

Averaging over s shows that $\mathbb{E}_x(T_1) \leq t_{\text{hit}}$. For any choice of distintct $1 \leq r \neq s \leq n-1$, we have

$$\mathbb{E}_n(T_k - T_{k-1} | L_{k-1} = r, \sigma(k) = s = L_k) = \mathbb{E}_r(\tau_s) \leq t_{\text{hit}}.$$

Averaging over r and s yields

$$\mathbb{E}_n(T_k - T_{k-1} | L_k = \sigma(k)) \leq t_{\text{hit}}.$$

Observe that, for any set S of k elementes, we have

$$\mathbb{P}_n\{L_k = \sigma(k) | \{\sigma(1), \ldots, \sigma(k)\} = S, \{X_t\}_t\} = \frac{1}{k}.$$

Consequently, since $\mathbb{E}_n(T_k - T_{k-1} | L_k \neq \sigma(k)) = 0$,

$$\mathbb{E}_n(T_k - T_{k-1}) \leq \mathbb{P}_n\{L_k = \sigma(k)\} \cdot t_{\text{hit}} = \frac{t_{\text{hit}}}{k}.$$

Therefore,

$$t_{\text{cov}} = \mathbb{E}_n(T_{n-1}) \leq t_{\text{hit}} \sum_{k=1}^{n-1} \frac{1}{k}.$$ □

A slightly modification of this technique can be used to prove lower bounds:

Proposition 9.93. *Let* $A \subseteq V$. *Set* $t_{\min}^A = \min_{a,b \in A, a \neq b} \mathbb{E}_a(\tau_b)$. *Then,*

$$t_{\text{cov}} \geq \max_{A \subseteq V} t_{\min}^A \left(1 + \frac{1}{2} + \cdots + \frac{1}{|A| - 1}\right).$$

These bounds can be used to study the simple random walk on $[0, n]^d$, where d is a natural number.

Markov Chains Mixing Times **November 11, 2022**

Lecture 10: Varopolous-Carne inequality

Lecturer: Yuval Peres Week: 5

Intuitively, mixing time is when the random walk is scattered quite like the stationary distribution. The stationary distribution for random walk on a graph is affected by the presence or absence of short paths between two vertices of the underlying graph. As an example, consider the case when the graph is a line graph and the case when the graph is the complete graph on the same number of vertices.

Let us begin with a definition.

The ***diameter*** of a simple graph G is the maximum among the shortest lengths between vertices of G; more precisely if for any two vertices x, y of G then we let $\rho_G(X, Y)$ denote the shortest path between X and Y in G, then the diameter of G is $\max_{x,y} \rho_G(x, y)$ where the maximum is taken over all vertices x, y of G. Note that the diameter of a graph which is not connected is infinity. So it is meaningful to speak about diameter of a graph G only if G is connected.

Setting: Consider a random walk on a connected graph G where one has $P(x, y) > 0 \iff$ the vertices x, y are adjacent in G.

We begin with the following preliminary bound.

Proposition 10.94. *In the above setting, one has*

$$t_{mix}(1/2 - \varepsilon) \geq \frac{diam(G)}{2} \quad \forall \varepsilon > 0.$$

In particular, the usual mixing time which is $t_{mix}(1/4)$ is lower bounded by $\frac{diam(G)}{2}$.

Proof. Suppose the smallest path between x_0, y_0 is of length $D(G) = \text{diam}(G)$.
Consider $A = \{x : \rho_G(x, x_0) < D(G)/2\}$; for $t < D(G)/2$ one has

$$\mathbf{P}_{x_0}\{X_t \in A^c\} = 0 = \mathbf{P}_{y_0}\{X_t \in A\}.$$

As a consequence $\overline{d}(t) \geq \mathbf{P}_{x_0}\{X_t \in A\} - \mathbf{P}_{y_0}\{X_t \in A\} = 1$ whence $d(t) \geq \overline{d}(t)/2 \geq 1/2$. □

Although this is sometimes sharp, it is usually the case that the mixing time is more like the square of the diameter.

The goal of this lecture will be to prove:

Theorem 10.95.

$$t_{mix} \geq \frac{diam(G)^2}{c \log n} \quad \textit{where } n = \textit{number of vertices in } G,$$

where c is some absolute constant.

Conjecture 10.96. For transitive graphs

$$t_{\text{mix}} \geq c_d \cdot \text{diam(G)}^2.$$

In particular, we can remove the logarithmic factor from the bound in the claim above for transitive graphs.

Now we shall prove the Varopolous-Carne inequality.

Theorem 10.97 (Varopolous-Carne inequality)**.** *Suppose (X_t) is a reversible Markov chain with transition matrix P and reversing measure π. Then for any two states x, y of the chain and any $t > 0$ one has*

$$P^t(x,y) \leq 2\sqrt{\frac{\pi(x)}{\pi(y)}}\ \mathbf{P}_0\{S_t \geq \rho(x,y)\} \leq 2\sqrt{\frac{\pi(x)}{\pi(y)}}\ e^{-\rho^2(x,y)/2t}.$$

Here $\rho(x,y)$ is the (unweighted) distance between the nodes x, y on the graph $\mathcal{G}$ for which there is a weighted random walk on G with transition matrix P, and (S_t) is the simple random walk on the integers.

Proof. First the easy part: the second $\leq$ can be proved as follows: notice that for all r we have

$$\mathbf{P}\{S_t \geq r\} = \mathbf{P}\{e^{\lambda S_t} \geq e^{\lambda r}\} \leq e^{-\lambda R} e^{t\lambda^2/2}$$

for all $\lambda > 0$; taking $\lambda = r/t$ gives the last inequality.

The other inequality is more tricky, but also has a very beautiful proof using Chebyshev polynomials.
We shall prove it in the special case when π is uniform, or equivalently P is symmetric.
We begin by recalling the Chebyshev polynomials. The Chebyshev polynomials (or more correctly the Chebyshev polynomials of the first kind) are a sequence of polynomials $(T_n(x))$ which satisfy the trigonometric relation $T_n(\cos\theta) = \cos(n\theta)$.
Thanks to the angle sum formulae of de Moivre, it follows that T_n is a polynomial of degree n. For instance $T_2(\cos\theta) = \cos(2\theta) = 2\cos^2(\theta) - 1$ implies that T_2 is the degree 2 polynomial $T_2(x) = 2x^2 - 1$.
Moreover, from the trigonometric identity

$$\cos((k+1)\theta) + \cos((k-1)\theta) = 2\cos(\theta)\cos(k\theta)$$

we see that

$$T_{k+1}(z) + T_{k-1}(z) = 2zT_k(z).$$

So T_k is a polynomial of degree k.
We also have $(\cos\theta)^t = \left(\frac{e^{i\theta}+e^{-i\theta}}{2}\right)^t = \sum_{k=-t}^{t} \mathbf{P}\{S_t = k\} e^{ik\theta}$.
Taking the real part of both sides we get

$$(\cos\theta)^t = \sum_{k=-t}^{t} \mathbf{P}_0\{S_t = k\} T_{|k|}(\cos\theta)$$

and since $\{\cos\theta : \theta \in \mathbf{R}\}$ is an infinite set so one has the polynomial identity

$$z^t = \sum_{k=-t}^{t} \mathbf{P}_0\{S_t = k\} T_{|k|}(z)$$

and since matrices preserve polynomial identity so we can plug in the transition matrix P in place of the indeterminate z to get

$$P^t = \sum_{k=-t}^{t} \mathbf{P}_0\{S_t = k\} T_{|k|}(P).$$

Now notice that if $\rho(x, y) > k$ then since T_k has degree k so $P^{\rho(x,y)}$ appears with a zero coefficient in $T_k(P)$. As a consequence, we conclude that $\rho(x, y) > k \implies T_k(P)(x, y) = 0$.
In general one has $T_k(P)(x, y) = \langle \delta_x, T_k(P)\delta_y \rangle$.
Now since P is assumed to be symmetric, so all the powers of P are symmetric, and hence the operator norm of P^m is the largest eigenvalue of P^m for each m, and since $|T_k(x)| \leq 1$ for all $x \in [-1, 1]$ so we get that

$$|T_k(P)(x, y)| = |\langle \delta_x, T_k(P)\delta_y \rangle| \leq 1$$

and that gives us the claim.

□