Mixing time for a random walk on a ring

Stephen Connor

Joint work with Michael Bate

THE UNIVERSITY of York

Aspects of Random Walks

Durham, April 2014

Random walks on groups

Let G be a finite group and let P be a probability distribution on G; that is, a function $P : G \to [0, 1]$ such that $\sum_{g \in G} P(g) = 1$.

Random walks on groups

Let G be a finite group and let P be a probability distribution on G; that is, a function $P : G \to [0, 1]$ such that $\sum_{g \in G} P(g) = 1$.

For example, we could have $G = S_n$, the symmetric group on $\{1, 2, \ldots, n\}$, and we could set

$$P(g) = \begin{cases} \frac{1}{n} & \text{if } g = 1 \text{ is the identity} \\ \frac{2}{n^2} & \text{if } g = (i, j) \text{ is a transposition} \\ 0 & \text{otherwise} \end{cases}$$

Random walks on groups

Let G be a finite group and let P be a probability distribution on G; that is, a function $P : G \to [0,1]$ such that $\sum_{g \in G} P(g) = 1$.

For example, we could have $G = S_n$, the symmetric group on $\{1, 2, \ldots, n\}$, and we could set

$$P(g) = \begin{cases} \frac{1}{n} & \text{if } g = 1 \text{ is the identity} \\ \frac{2}{n^2} & \text{if } g = (i,j) \text{ is a transposition} \\ 0 & \text{otherwise} \end{cases}$$

A *random walk on G* is then a Markov chain X with transitions governed by the distribution P. So we fix a starting point X_0, and then set

$$\mathbb{P}(X_{t+1} = hg \mid X_t = g) = P(h)$$

Distribution after repeated steps is given by *convolution*:

$$\mathbb{P}(X_2 = g | X_0 = 1) = P * P(g) = \sum_h P(gh^{-1})P(h)$$

As long as the probability distribution P isn't concentrated on a subgroup, the stationary distribution π for X is the uniform distribution; $\pi(g) = 1/|G|$ for all $g \in G$. When X is ergodic, we're interested in how long it takes for it to converge to equilibrium.

Distribution after repeated steps is given by *convolution*:

$$\mathbb{P}\left(X_2 = g | X_0 = 1\right) = P * P(g) = \sum_h P(gh^{-1})P(h)$$

As long as the probability distribution P isn't concentrated on a subgroup, the stationary distribution π for X is the uniform distribution; $\pi(g) = 1/|G|$ for all $g \in G$. When X is ergodic, we're interested in how long it takes for it to converge to equilibrium.

Definition

The **mixing time** of X is

$$\tau(\varepsilon) = \min\left\{t : \|\mathbb{P}\left(X_t \in \cdot\right) - \pi(\cdot)\|_{\mathrm{TV}} \leq \varepsilon\right\} .$$

Cutoff phenomenon

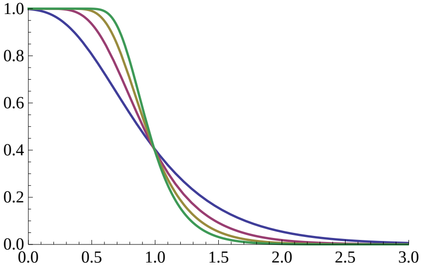
We're often interested in a natural sequence of processes $X^{(n)}$ on groups $G^{(n)}$ of increasing size: how does the mixing time $\tau^{(n)}$ scale with n?

Cutoff phenomenon

We're often interested in a natural sequence of processes $X^{(n)}$ on groups $G^{(n)}$ of increasing size: how does the mixing time $\tau^{(n)}$ scale with n?

In lots of nice examples a **cutoff phenomenon** is exhibited:

$$\text{for all } \varepsilon > 0, \quad \lim_{n\to\infty} \frac{\tau^{(n)}(\varepsilon)}{\tau^{(n)}(1-\varepsilon)} = 1$$



Random walk on a ring

What if we add some additional structure, and try to move from a walk on a *group* to a walk on a *ring*?

Random walk on a ring

What if we add some additional structure, and try to move from a walk on a *group* to a walk on a *ring*?

For example: random walk on $\mathbb{Z}_n$ (n odd) with

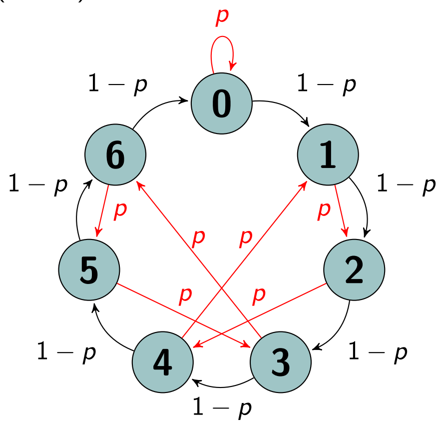
$$x \to \begin{cases} x+1 & \text{w.p. } 1-p_n \\ 2x & \text{w.p. } p_n \end{cases}$$

Random walk on a ring

What if we add some additional structure, and try to move from a walk on a *group* to a walk on a *ring*?

For example: random walk on $\mathbb{Z}_n$ (n odd) with

$$x \to \begin{cases} x+1 & \text{w.p. } 1-p_n \\ 2x & \text{w.p. } p_n \end{cases}$$



Related results

Chung, Diaconis and Graham (1987) study the process (used in random number generation)

$$x \to \begin{cases} ax - 1 & \text{w.p. } \frac{1}{3} \\ ax & \text{w.p. } \frac{1}{3} \\ ax + 1 & \text{w.p. } \frac{1}{3} \end{cases}$$

Related results

Chung, Diaconis and Graham (1987) study the process (used in random number generation)

$$x \to \begin{cases} ax - 1 & \text{w.p. } \frac{1}{3} \\ ax & \text{w.p. } \frac{1}{3} \\ ax + 1 & \text{w.p. } \frac{1}{3} \end{cases}$$

- When $a = 1$ there exist constants C and C' such that:

$$e^{-Ct/n^2} < \left\| P_t^{(n)} - \pi^{(n)} \right\|_{\mathrm{TV}} < e^{-C't/n^2}$$

Related results

Chung, Diaconis and Graham (1987) study the process (used in random number generation)

$$x \to \begin{cases} ax - 1 & \text{w.p. } \frac{1}{3} \\ ax & \text{w.p. } \frac{1}{3} \\ ax + 1 & \text{w.p. } \frac{1}{3} \end{cases}$$

- When $a = 1$ there exist constants C and C' such that:

$$e^{-Ct/n^2} < \left\| P_t^{(n)} - \pi^{(n)} \right\|_{\mathrm{TV}} < e^{-C't/n^2}$$

- When $a = 2$ and $n = 2^m - 1$ there exist constants c and c' such that:

$$\text{for } t_n \geq c \log n \log\log n, \quad \left\| P_{t_n}^{(n)} - \pi^{(n)} \right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty$$

$$\text{for } t_n \leq c' \log n \log\log n, \quad \left\| P_{t_n}^{(n)} - \pi^{(n)} \right\|_{\mathrm{TV}} > \varepsilon \text{ as } n \to \infty$$

Back to our process...

General problem

The distribution of X_t isn't given by convolution.

$$X_t = 2I_tX_{t-1} + (1 - I_t)(X_{t-1} + 1)$$

where $I_t \sim \text{Bern}(p_n)$.

Back to our process...

General problem

The distribution of X_t isn't given by convolution.

$$X_t = 2I_tX_{t-1} + (1 - I_t)(X_{t-1} + 1)$$

where $I_t \sim \mathrm{Bern}(p_n)$.
But for this relatively simple walk, we can get around this by looking at the process **subsampled at jump times**. (Here we call a $+1$ move a *'step'* and a $\times 2$ move a *'jump'*.)

Back to our process...

General problem

The distribution of X_t isn't given by convolution.

$$X_t = 2I_t X_{t-1} + (1 - I_t)(X_{t-1} + 1)$$

where $I_t \sim \text{Bern}(p_n)$.

But for this relatively simple walk, we can get around this by looking at the process **subsampled at jump times**. (Here we call a $+1$ move a *'step'* and a $\times 2$ move a *'jump'*.)

So consider (with $X_0 = Y_0 = 0$)

$$Y_k = \sum_{j=1}^{k} 2^{k+1-j} S_j \pmod{n}, \quad S_j \overset{\text{i.i.d.}}{\sim} \text{Geom}(p_n) \pmod{n}$$

(i.e. Y_k is the position of X immediately following the k^{th} jump.)

Plan

1. Find lower bound for mixing time of Y;
2. Find upper bound for mixing time of Y;
3. Try to relate these back to the mixing time for X.

Plan

1. Find lower bound for mixing time of Y;
2. Find upper bound for mixing time of Y;
3. Try to relate these back to the mixing time for X.

Restrict attention to $p_n = \dfrac{1}{2n^\alpha}$, $\alpha \in (0,1)$.

We expect things to happen (for Y) sometime around

$$T_n := \log_2(np_n) \sim (1-\alpha)\log_2 n$$

In fact:

Theorem

Y exhibits a **cutoff** *at time T_n, with window size $O(1)$.*

In fact:

Theorem

Y exhibits a **cutoff** *at time T_n, with window size $O(1)$.*

To prove this, we need to show that (for n odd)

$$\liminf_{n\to\infty} \left\| \mathbb{P}\left(X^{(n)}_{T_n-\theta} \in \cdot \right) - \frac{1}{n} \right\|_{\mathrm{TV}} \geq 1 - \varepsilon(\theta)$$

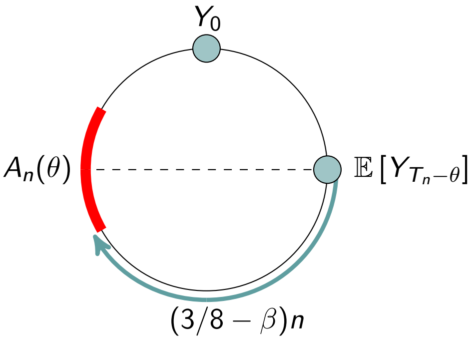
and

$$\limsup_{n\to\infty} \left\| \mathbb{P}\left(X^{(n)}_{T_n+\theta} \in \cdot \right) - \frac{1}{n} \right\|_{\mathrm{TV}} \leq \varepsilon(\theta)\,,$$

where $\varepsilon(\theta) \to 0$ as $\theta \to \infty$.

Lower bound

For a lower bound, we simply find a set $A_n(\theta)$ (of considerable size) that Y has very little chance of hitting before time $T_n - \theta$, for large $\theta \in \mathbb{N}$. (Recall the definition of total variation distance!)



- If $A_n(\theta)$ is chosen as above then $\pi(A_n(\theta)) = 1/4 + 2\beta$.

- If $A_n(\theta)$ is chosen as above then $\pi(A_n(\theta)) = 1/4 + 2\beta$.
- Using Chebychev's inequality:

$$\mathbb{P}\left(Y_{T_{n-\theta}} \in A_n(\theta)\right) \leq \frac{4^{1-\theta}}{3(3/8-\beta)^2}\,.$$

- If $A_n(\theta)$ is chosen as above then $\pi(A_n(\theta)) = 1/4 + 2\beta$.
- Using Chebychev's inequality:

$$\mathbb{P}\left(Y_{T_n - \theta} \in A_n(\theta)\right) \leq \frac{4^{1-\theta}}{3(3/8 - \beta)^2}.$$

- Now choose $\beta = \beta(\theta)$ to make the difference between these large. . .

- If $A_n(\theta)$ is chosen as above then $\pi(A_n(\theta)) = 1/4 + 2\beta$.
- Using Chebychev's inequality:

$$\mathbb{P}\left(Y_{T_{n-\theta}} \in A_n(\theta)\right) \leq \frac{4^{1-\theta}}{3(3/8-\beta)^2}\,.$$

- Now choose $\beta = \beta(\theta)$ to make the difference between these large. . .

Lemma (Lower bound for Y)

For $\theta \geq 3$,

$$\left\|\mathbb{P}\left(Y_{T_{n-\theta}} \in \cdot\right) - \pi_n(\cdot)\right\|_{\mathrm{TV}} \geq 1 - 4^{1-\theta/3}\,.$$

Upper bound

Let P be a probability on a group G. A (complex) representation ρ is a group homomorphism $\rho : G \to \mathrm{GL}_d(\mathbb{C})$, where $\mathrm{GL}_d(\mathbb{C})$ denotes the group of $d \times d$ invertible complex matrices. We write

$$\boxed{\hat{P}(\rho) = \sum_{g \in G} P(g)\rho(g)}$$

for the *Fourier transform* of P at ρ.

Upper bound

Let P be a probability on a group G. A (complex) representation ρ is a group homomorphism $\rho : G \to \mathrm{GL}_d(\mathbb{C})$, where $\mathrm{GL}_d(\mathbb{C})$ denotes the group of $d \times d$ invertible complex matrices. We write

$$\boxed{\hat{P}(\rho) = \sum_{g \in G} P(g)\rho(g)}$$

for the *Fourier transform* of P at ρ.

This behaves very nicely with respect to convolution:

$$\widehat{P * P}(\rho) = \hat{P}(\rho)\hat{P}(\rho)$$

A basic but extremely useful result is the following:

Lemma (Diaconis and Shahshahani, 1981)

$$\|P-\pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4} \sum_{\substack{\mathrm{non-triv} \\ \mathrm{irr}\ \rho}} d_\rho \mathrm{tr}\left(\hat{P}(\rho)\hat{P}(\rho)^*\right)$$

(Here $A^* = (\overline{a_{ji}})$ denotes the complex conjugate transpose of the matrix $A = (a_{ij})$, and tr denotes the trace function on square matrices)

A basic but extremely useful result is the following:

Lemma (Diaconis and Shahshahani, 1981)

$$\|P-\pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4} \sum_{\substack{\mathrm{non-triv} \\ \mathrm{irr}\ \rho}} d_\rho \mathrm{tr}\left(\hat{P}(\rho)\hat{P}(\rho)^*\right)$$

(Here $A^* = (\overline{a_{ji}})$ denotes the complex conjugate transpose of the matrix $A = (a_{ij})$, and tr denotes the trace function on square matrices)

Our subsampled walk Y **is** a random walk on the group $(\mathbb{Z}_n, +)$, whose n irreducible (one-dimensional) representations are determined by

$$\rho_s(1) := e^{i\frac{2\pi s}{n}} \qquad \text{for } 0 \leq s \leq n-1$$

The Upper Bound Lemma becomes

$$\|P - \pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4} \sum_{s=1}^{n-1} |\hat{P}(\rho_s)|^2$$

where $\hat{P}(\rho_s)$ is now just a complex number.

The Upper Bound Lemma becomes

$$\|P-\pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4}\sum_{s=1}^{n-1} |\hat{P}(\rho_s)|^2$$

where $\hat{P}(\rho_s)$ is now just a complex number.

Substituting the correct distribution for Y_t leads us to the following upper bound:

$$\|\delta_0 P_t-\pi\|_{\mathrm{TV}}^2 \leq \frac{1}{4}\sum_{s=1}^{n-1}\prod_{k=1}^{t} \frac{p_n^2}{1-2(1-p_n)\cos(\frac{2\pi}{n}2^k s)+(1-p_n)^2}$$

Lemma (Upper bound for Y)

Let $p_n = 1/2n^\alpha$, *with* $\alpha \in (0,1]$. *For* $\theta \in \mathbb{N}$,

$$\limsup_{n\to\infty} \|\mathbb{P}\left(Y_{T_n+\theta} \in \cdot\right) - \pi_n(\cdot)\|_{\mathrm{TV}} = O(4^{-\theta})\,.$$

Lemma (Upper bound for Y)

Let $p_n = 1/2n^\alpha$, with $\alpha \in (0,1]$. For $\theta \in \mathbb{N}$,

$$\limsup_{n\to\infty} \left\| \mathbb{P}\left(Y_{T_n+\theta} \in \cdot\right) - \pi_n(\cdot) \right\|_{\mathrm{TV}} = O(4^{-\theta})\,.$$

Proof.

Careful analysis of the right-hand side!
(Identify which terms really contribute to the sum ($s = (n \pm 1)/2$ accounts for nearly everything), deal with these, and show that nothing else really matters.) □

This completes our proof of a cutoff for Y.

Moving from Y to X

We've seen that Y mixes in an interval of length $O(1)$ around $T_n = \log_2(np_n)$: what does this tell us about the mixing time for X?

Moving from Y to X

We've seen that Y mixes in an interval of length $O(1)$ around $T_n = \log_2(np_n)$: what does this tell us about the mixing time for X?

Corollary

For $p_n = 1/2n^\alpha$, with $\alpha \in (0,1)$, X exhibits a cutoff at time

$$T_n^X = T_n/p_n = 2(1-\alpha)n^\alpha \log_2 n$$

with window size $\sqrt{T_n}/p_n$.

Proof.

Essentially follows from the observation that the number of jumps by time $T_n^X + c\sqrt{T_n}/p_n$ concentrates (in an interval of order $\sqrt{T_n}$) around $T_n + c\sqrt{T_n}$. □

And finally: open problems

1. We can deal with more interesting steps in our walk, but not yet with more interesting jumps, e.g. consider

$$x \to \begin{cases} x+1 & \text{w.p. } 1-p_n \\ 2x & \text{w.p. } p_n/2 \\ \left(\frac{n+1}{2}\right)x & \text{w.p. } p_n/2 \end{cases}$$

or more general rules, such as $x \to x^2 \ldots$

And finally: open problems

1. We can deal with more interesting steps in our walk, but not yet with more interesting jumps, e.g. consider

$$x \to \begin{cases} x+1 & \text{w.p. } 1-p_n \\ 2x & \text{w.p. } p_n/2 \\ \left(\frac{n+1}{2}\right)x & \text{w.p. } p_n/2 \end{cases}$$

or more general rules, such as $x \to x^2 \ldots$

2. Or how about this process?

$$x \to \begin{cases} x+1 & \text{w.p. } 1-p_n \\ ax & \text{w.p. } p_n \end{cases}$$

where multiplication by a is not invertible? (Stationary distribution won't even be uniform...)