Introduction to mixing times of random walks

Framework:

$X_t \in \mathcal{S}_n \quad (t=0,1,\dots)$

$X_0 = id$

Assumption: X_t is a Markov chain

$\mathbb{P}(X_t = x_t \mid X_0 = x_0, X_1 = x_1, \dots, X_{t-1} = x_{t-1}) = p(x_{t-1}, x_t)$

where $p(x,y) = \mathbb{P}$ (go from x to y in 1 shuffle)

Under mild conditions,

$\mathbb{P}(X_t = x) \xrightarrow[t\to\infty]{} \pi(x)$

Usually $\pi(x) = \frac{1}{n!}$

If μ, ν are prob. distr. on $\mathcal{S}_n$,

$\|\mu - \nu\|_{TV} = \sup_{A \subseteq \mathcal{S}_n} \|\mu(A) - \nu(A)\| = \frac{1}{2} \sum_{x \in \mathcal{S}_n} |\mu(x) - \nu(x)|$

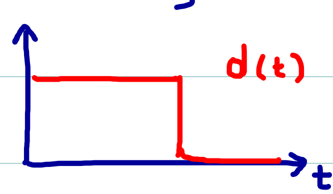
Introduce

$d(t) = \sup_{x \in \mathcal{S}_n} \|\mathcal{L}(X_t^x) - \pi\|_{TV}$

Observation

Diaconis - Shahshanam '81, Aldons '83

For "many" examples



Def

Def^n **Cutoff phenomenon** occurs at some time t_{mix} as $n \to \infty$ if

$$d\left(t_{mix}(1-\varepsilon)\right) \xrightarrow[n\to\infty]{} 1 \qquad \forall\, \varepsilon > 0$$

$$d\left(t_{mix}(1+\varepsilon)\right) \xrightarrow[n\to\infty]{} 0 \qquad \forall\, \varepsilon > 0$$

Can always define

$$t_{mix} = \inf\{t > 0 : d(t) \leq \tfrac{1}{4}\}$$

Examples

Random Transpositions

Pick two cards at random; switch them.

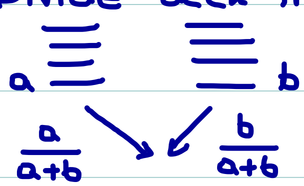
Thm (D-Sh. '81) Cutoff occurs as $n \to \infty$ for $t_{mix} = \frac{1}{2} n \log n$.

Random to top

Thm (Aldons) Cutoff at $t_{mix} = n \log n$

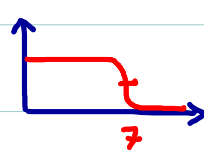
Riffle shuffle

Divide deck into 2 halves [using Binomial $(n, 1/2)$]



Thm (Aldons) Cutoff at $t_{mix} = \frac{3}{2} \log_2 n \simeq 8.55$

$n = 52$ **Thm** (Bayer-Diaconis 1992) $t_{mix}\left(\frac{1}{2}\right) = 7$ for $n = 52$



Maths techniques

- probabilistic tools
- representation theory
- analytic tools (functional ineq.)
- geometric tools

Probabilistic technique "Coupling"

Defn : If μ, ν on $\mathcal{S}_n$, then a coupling is a pair (X, Y) s.t. $X \sim \mu$ and $Y \sim \nu$. Say coupling is successful if $X = Y$.

Prop. We have $\displaystyle \|\mu - \nu\|_{TV} = \inf_{(X,Y) \text{ coupling}} \mathbb{P}(X \neq Y)$

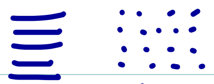
Rmk: In particular, if $\exists$ coupling (X, Y) s.t. $X = Y$ with high probability $\Rightarrow$ $\|\mu - \nu\|_{TV}$ small.

Proof of $\leq$: If A is any event and (X,Y) any coupling, then

$$
\begin{aligned}
|\mu(A) - \nu(A)| &= |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\
&= |\mathbb{P}(X \in A, X = Y) + \mathbb{P}(X \in A, X \neq Y) \\
&\quad - \mathbb{P}(Y \in A, X = Y) - \mathbb{P}(Y \in A, X \neq Y)| \leq \mathbb{P}(X \neq Y)
\end{aligned}
$$

Example of use of coupling

Random to top : cutoff occurs at $n \log n$



X_t $\quad$ Y_t = imaginary deck of cards started with $Y_0 \sim \pi$

Procedure :

At each step, pick $1 \leqslant i \leqslant n$ unif. at random

Choose card with label i, put it on top of deck.

Fact Once card with label i has been chosen, then this card is matched for both decks forever after.

Let τ = 1^{st} time all cards have been touched .

If $t \geqslant \tau$, then $X_t = Y_t$

$d(t) \leqslant \mathbb{P}(\tau > t)$

Fact 2. $\dfrac{\tau}{n \log n} \longrightarrow 1$ in probability

For instance,

$$\mathbb{E}(\tau) = \frac{n}{1} + \frac{n}{2} + \dots + \frac{n}{n} \sim n \log n$$

time to collect last card $\quad$ time to collect last but one $\quad$ time to collect first card

We deduce that

$d((1+\varepsilon) n \log n) \longrightarrow 0$

Lower bound

$d((1-\varepsilon) n \log n) \longrightarrow 1$?

Connection to rep^n theory

Pioneered by Diaconis-Shahshahami

Setup : finite group G_n

$p(x,y) = p(yx^{-1})$ for some probability p on G_n

Hence $X_t = X_0 g_1 g_2 \dots g_t$

where g_i are iid with distr. p.

Ex. Random transp.

$p(\tau) = \frac{2}{n^2}$ for any transposition τ

$p(id) = n \cdot \frac{1}{n^2} = \frac{1}{n}$

Fact : Then if $P^t(x) = \mathbb{P}(X_t = x \mid X_0 = id)$, then $P^t(x) = \underbrace{p^{*t}}_{\text{convolution}}(x)$

where $f * g(x) = \sum_{y \in G_n} f(xy^{-1}) g(y)$

E.g. $P(X_2 = x \mid X_0 = id) = \sum_y p(y) p(xy^{-1})$

Now, this suggests Fourier analysis.

Def^n If ρ is a representation and f is a function $G \to \mathbb{C}$

$\hat{f}(\rho) = \sum_{s \in G} f(s) \rho(s)$

$\hat{f}(\rho)$ is a matrix

ρ = frequency

$\hat{f}(\rho)$ = "amplitude" corresp. to ρ

<u>Fourier inversion thm</u>:

$$f(s) = \frac{1}{|G|} \sum_{\rho} d_\rho \operatorname{Tr}\left(\rho(s^{-1}) \hat{f}(\rho)\right), \text{ where } d_\rho = \dim \rho$$

$\hookrightarrow$ irred. repⁿ

<u>Parseval formula</u>:

$$\sum_{s\in G} f(s) h(s^{-1}) = \frac{1}{|G|} \sum_{\rho} d_\rho \operatorname{Tr}\left(\hat{f}(\rho) \hat{h}(\rho)\right)$$

For us, if $g(s) = h(s^{-1})$

$$\sum_{s\in G} f(s) g(s) = \frac{1}{|G|} \sum_{\rho} d_\rho \operatorname{Tr}\left(\hat{f}(\rho) \hat{g}(\rho)^*\right)$$

Indeed,

$$\hat{h}(\rho) = \sum_{s\in G} \rho(s) h(s) = \sum_{s\in G} \rho(s) g(s^{-1}) =$$

$$= \sum_{t\in G} \rho(t^{-1}) g(t) = \sum_{t\in G} \rho(t)^* g(t) = \left(\hat{g}(\rho)\right)^*$$

<u>Observation</u>: If $f, g : G \longrightarrow \mathbb{C}$, then $\widehat{f * g} = \hat{f} \cdot \hat{g}$

and in particular, $\widehat{p^{*t}(x)} = \hat{p}^{t}(x)$.

<u>Thm</u> (D-Sh "upper-bound lemma")

Let μ be any prob distr on G_n. Let π be uniform distr. on G_n

$$\|\mu - \pi\|_{TV}^2 \leq \frac{1}{4} \sum_{*} d_\rho \operatorname{Tr}\left(\hat{\mu}(\rho) \hat{\mu}(\rho)^*\right)$$

where $\sum_*$ is the sum over all irreducible, non trivial reps (ignoring the trivial Id rep^n)

Proof $\|\mu-\pi\|_{TV}^2 = \frac{1}{4}\left(\sum_{s\in G} |\mu(s)-\pi(s)|\right)^2$

$$\underset{\text{Cauchy-Schwarz}}{\leq} \frac{1}{4}|G| \sum_{s\in G} |\mu(s)-\pi(s)|^2$$

Use Parseval

$$\leq \frac{1}{4}|G| \frac{1}{|G|} \sum_{\rho \text{ irred.}} d_\rho \operatorname{Tr}(\hat{f}(\rho)\hat{f}(\rho)^*)$$

where $f(s) = \mu(s)-\pi(s)$.

But $\hat{f}(\rho) = \hat{\mu}(\rho)-\hat{\pi}(\rho)$

Observe that $\hat{\pi}(\rho)=0$ unless ρ is the trivial rep^n.

Indeed, $\hat{\pi}(\rho) = \sum_{s\in G} \frac{1}{|G|} \rho(s) = \frac{1}{|G|} \sum_{s\in G} \rho(s) =$

$$= \frac{1}{|G|} \sum_{s\in G} \rho(st) = \frac{1}{|G|} \left(\sum_{s\in G} \rho(s)\right) \rho(t)$$

true for all t, so $\sum_{s\in G} \rho(s) = 0$

When ρ = trivial rep^n

$\hat{\mu}(\rho) = \sum \mu(s)\cdot 1 = 1 \Rightarrow \hat{\mu}(\rho)-\hat{\pi}(\rho) = 0 \Rightarrow$

$\|\mu-\pi\|_{TV}^2 \leq \frac{1}{4}\sum_* d_\rho \operatorname{Tr}(\hat{\mu}(\rho)\hat{\mu}(\rho)^*)$, as desired.

Applying to RW on S_n, $X_t = \text{id}\, \underbrace{g_1 \cdots g_n}_{\text{iid}}$

Cor. $d(t)^2 \leqslant \frac{1}{4} \sum_{*} d_\rho^2 \lambda_\rho^{2t}$

where $\lambda_\rho = \left(\frac{1}{n} + \frac{n-1}{n} r(\rho)\right)$

and $r(\rho) = \frac{\chi_\rho(\tau)}{d_\rho}$ = character ratio, $\chi_\rho(\tau) = \chi_\rho$ evaluated at any transposition

Proof This is because $P(x)$ is itself a class function.

If ρ is a repⁿ and $s \in G$,

$$\rho(s)\hat{P}(\rho)\rho(s^{-1}) = \rho(s)\left(\sum_{t \in G} \rho(t)P(t)\right)\rho(s^{-1})$$
$$= \sum_{t \in G} \rho(sts^{-1})P(t)$$
$$= \sum_{t \in G} \rho(t)P(t) = \hat{P}(\rho)$$

$\Rightarrow \rho(s)\hat{P}(\rho) = \hat{P}(\rho)\rho(s) \quad \forall s \in G$

$\Rightarrow$ By Schur's lemma $\hat{P}(\rho) = \lambda I$

Here $\lambda d_\rho = \mathrm{Tr}(\hat{P}(\rho)) = \mathrm{Tr}\left(\sum_s \rho(s)P(s)\right) = \sum_{s \in G} P(s)\chi_\rho(s)$

Consequence:

$$\frac{1}{4}\sum_{*} d_\rho \,\mathrm{Tr}\Big(\underbrace{\hat{P}^t}_{\lambda_\rho^t I}\,\underbrace{(\hat{P}^t)^*}_{\overline{\lambda_\rho^t} I}\Big) = \frac{1}{4}\sum_{*} d_\rho^2 |\lambda_\rho|^{2t}$$

Spectral Gap

Let S finite set

$p(x,y)$ = trans. prob. on S

Prop (Perron-Frobenius)

1. If λ is an ($\mathbb{C}$) eigenvalue, then $|\lambda| \leq 1$.
2. If p is irreducible, then $\lambda = 1$ is an EV of dimension 1, generated by $(1, \_\_, 1)$
3. If p is aperiodic and reversible, then $\forall \lambda$ e.v. $\in \mathbb{R}$
4. $\exists f_j$ which is ON basis of eigenfunctions wrt $\langle , \rangle_\pi$

Reversibility: $\exists \pi(x) \geq 0 \quad \pi(x) P(x,y) = \pi(y) P(y,x)$

then π is the stat. distr. of the Markov chain.

Define $L^2(S,\pi)$ = space of functions $S \longrightarrow \mathbb{R}$

with $\langle f, g \rangle_\pi = \sum_{x \in S} f(x) g(x) \pi(x)$

We can always assume $\lambda \geq 0$ by considering $\tilde{P} = \frac{I+P}{2}$ [lazy chain]

Def^n. Let $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \geq 0$

$\gamma = 1 - \lambda_2$ = spectral gap

t_{rel} = relaxation time $= \frac{1}{\gamma}$

Observation: $P^t(x,y) = \mathbb{P}(X_t = y \mid X_0 = x) = (P^t \delta_y)(x)$

where $P^t f(x) = \sum P^t(x,y) f(y)$

and $\delta_y(s) = \mathbb{1}\{s=y\}$

Now $\delta_y = \sum_{j=1}^{n} \langle \delta_y, f_j \rangle_\pi f_j = \sum_{j=1}^{n} f_j(y) \pi(y) f_j$

$\Rightarrow P^t(x,y) = \sum_{j=1}^{n} f_j(y) \pi(y) \lambda_j^t f_j(x)$

Hence $\dfrac{P^t(x,y)}{\pi(y)} = \sum_{j=1}^{n} f_j(x) f_y(y) \lambda_j^t$

<u>Thm</u> Assume (irred., aperiodicity, reversibility)

Let $t_{mix}(\varepsilon) = \inf\{t \geq 0 : d(t) \leq \varepsilon\}$

Then $\forall \varepsilon \in (0,1)$ $\boxed{t_{rel} \log\left(\frac{1}{\varepsilon}\right) \leq t_{mix}(\varepsilon) \leq t_{rel} \log\left(\frac{1}{2\varepsilon\sqrt{\pi_{min}}}\right)}$

where

$\pi_{min} = \min\{\pi(x) : x \in S\}$

<u>Proof</u> Upper-bound:

Recall $\|P_t(x,\cdot) - \pi(\cdot)\|_{TV} = \frac{1}{2} \sum_{y \in S} |P_t(x,y) - \pi(y)|$

$$= \frac{1}{2} \sum_{y \in S} \pi(y) \left| \frac{P_t(x,y)}{\pi(y)} - 1 \right|$$

$$= \frac{1}{2} \left\| \frac{P_t(x,\cdot)}{\pi(\cdot)} - 1 \right\|_1 \leq \frac{1}{2} \left\| \frac{P_t(x,\cdot)}{\pi(\cdot)} - 1 \right\|_2$$

(Jensen)

$\Rightarrow 4d(t)^2 \leq \sum_{j=2}^{n} f_j(x)^2 \lambda_j^{2t} \leq (1-\gamma)^{2t} \sum_{j=1}^{n} f_j(x)^2$

Recall $\delta_x = \sum_{j=1}^{n} \langle \delta_x, f_j \rangle f_j = \sum_{j=1}^{n} f_j(x) \pi(x) f_j$

So $\langle \delta_x, \delta_x \rangle_\pi = \pi(x) = \sum_{j=1}^{n} f_j(x)^2 \pi(x)^2$

$\Rightarrow \dfrac{1}{\pi(x)} = \sum_{j=1}^{n} f_j(x)^2$

$\Rightarrow 4d(t)^2 \leq \dfrac{1}{\pi(x)} (1-\gamma)^{2t}$

$\Rightarrow d(t) \leq \dfrac{1}{2\sqrt{\pi_{min}}} e^{-t\gamma}$

Hence $\boxed{t_{mix}(\varepsilon) \leq t_{rel} \log\left(\frac{1}{2\varepsilon\sqrt{\pi_{min}}}\right)}$

Example: $\mathbb{Z}/n\mathbb{Z}$

$$\lambda_2 = \cos\left(\frac{2\pi}{n}\right) \Rightarrow \gamma = 1 - \cos\left(\frac{2\pi}{n}\right) \simeq \frac{2\pi^2}{n^2}$$

$$\Rightarrow t_{rel} = O(n^2)$$

Dirichlet form & path method

Suppose (P, π) reversible.

Define $\mathcal{E}(f,g) = \frac{1}{2} \sum_{x,y \in S} \pi(x) P(x,y) \big(f(y) - f(x)\big)\big(g(y) - g(x)\big)$

for $f, g : S \to \mathbb{R}$

$$\text{"}\approx\text{"} \frac{1}{2} \int \nabla_f \nabla_g = -\frac{1}{2} \int f \Delta g$$

$$\mathcal{E}(f,g) \underset{\text{computation}}{=} \langle (I-P)f, g \rangle_\pi$$

Thm $\gamma = \min\limits_{\substack{f: S \to \mathbb{R} \\ \mathbb{E}_\pi(f) = 0}} \frac{\mathcal{E}(f,f)}{\|f\|_2^2} = \min\limits_{\substack{f: S \to \mathbb{R} \\ \mathbb{E}_\pi(f) = 0 \\ \|f\|_2 = 1}} \mathcal{E}(f,f)$

Def^n: π Poincaré inequality if $\forall f : S \longrightarrow \mathbb{R}$

$$\mathrm{var}_\pi(f) \leq C\, \mathcal{E}(f,f) \qquad (*)$$

This always holds with $C = 1/\gamma$, and this is optimal.

Any inequality such as $(*)$ gives $t_{rel} \leq C$.

Pointers:

- Diaconis - Saloff - Coste 1996 :

  Comparison ineq.

  $\tilde{\mathcal{E}}(f,f) \leq A\, \mathcal{E}(f,f)$

- **New** Y. Ollivier

  Discrete notion of Ricci curvature

  $K(x,y) \geq K_0 > 0$

  $\Rightarrow d(t) \leq (1-K_0)^t$ diameter (S).