1

LECTURE 19: Last remarks on Markov chains:

① Ergodicity: $V_x(n) := \sum_{i=0}^{n-1} \mathbb{1}\{Y_i = x\}$ = "time spent @ x".

If μ stationary prob. measure, expect $\frac{V_x(n)}{n} \longrightarrow \mu(x) = \frac{1}{\underbrace{E_x T_x}_{=: t_x}}$.

Thm: If (Y_n) irreducible + recurrent, $\forall x, y$,

$$\mathbb{P}_x\left[\frac{V_y(n)}{n} \longrightarrow \frac{1}{t_y}\right] = 1. \quad (\text{even if } t_y = \infty!)$$

If further positive recurrent, i.e. $t_y < \infty\ \forall y$, then $\forall f: S \to \mathbb{R}$ bdd,

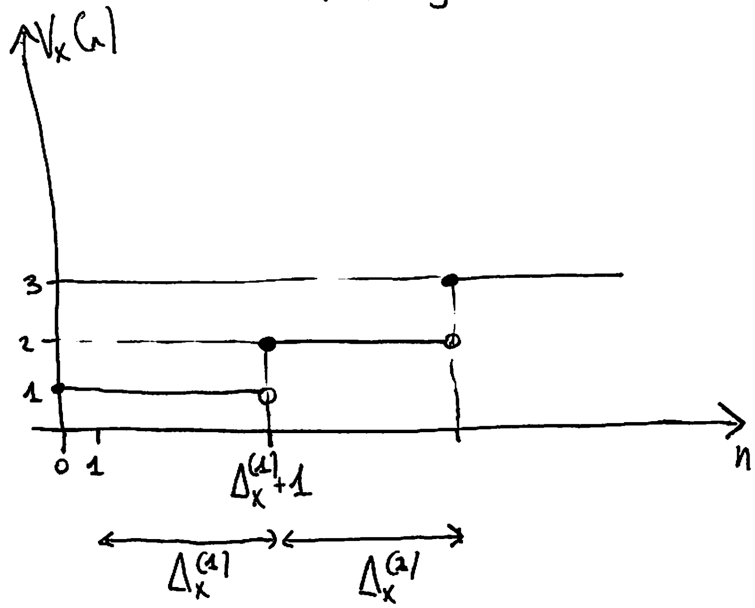
$$\mathbb{P}_x\left[\frac{1}{n}\sum_{i=0}^{n-1} f(Y_i) \longrightarrow \mathop{E}_{y\sim\mu} f(y)\right] = 1 \quad (\text{for } \mu \text{ unique stationary prob.})$$

(pos. rec. case.)

Pf: Exc: WLOG, $x = y$ (initial $x \rightsquigarrow y$ part insubstantial.)

$T_x^{(k)} :=$ time of k^{th} visit to x, $T_x^{(0)} = 0$, $\Delta_x^{(k)} = T_x^{(k)} - T_x^{(k-1)} \geq 0$

Strong Markov property $\Longrightarrow$ $\Delta_x^{(k)}$ iid, $E\Delta_x^{(k)} = t_x$.



$$1 + \Delta_x^{(1)} + \cdots + \Delta_x^{(V_x(n)-1)} \leq n$$
$$\leq \Delta_x^{(1)} + \cdots + \Delta_x^{(V_x(n))}$$

$$\Downarrow$$

$$\left|\frac{n}{V_x(n)} - \frac{\Delta_x^{(1)} + \cdots + \Delta_x^{(V_x(n))}}{V_x(n)}\right| \leq \frac{1 + \Delta_x^{(V_x(n))}}{V_x(n)}$$

Recurrence $\Longrightarrow$ $V_x(n) \longrightarrow \infty$ a.s. $\Longrightarrow$ this $\longrightarrow 0$ a.s.

②

SLLN $\Longrightarrow \dfrac{\Delta_x^{(1)} + \cdots + \Delta_x^{(V_x(n))}}{V_x(n)} \xrightarrow{a.s.} t_x \Longrightarrow \dfrac{n}{V_x(n)} \xrightarrow{a.s.} t_x$

$\Longrightarrow \dfrac{V_x(n)}{n} \xrightarrow{a.s.} \dfrac{1}{t_x} = \mu(x)$

For bdd functions $f: S \to \mathbb{R}$, WLOG $|f(y)| \le 1$

$$\Big| \underbrace{\frac{1}{n}\sum_{i=0}^{n-1} f(Y_i)}_{= \sum_y V_y(n) f(y)} - \sum_y \mu(y) f(y) \Big| = \Big| \sum_y \Big(\frac{V_y(n)}{n} - \mu(y) \Big) f(y) \Big|$$

$$\le \sum_y \Big| \frac{V_y(n)}{n} - \mu(y) \Big|$$

$$\le \sum_{y \in A} \Big| \frac{V_y(n)}{n} - \mu(y) \Big| + \sum_{y \notin A} \frac{V_y(n)}{n} + \mu(A^c)$$

$$\le 2 \sum_{y \in A} \Big| \frac{V_y(n)}{n} - \mu(y) \Big| + 2\mu(A^c).$$

Choose large finite $A \longrightarrow$ a.s. $\limsup \le \varepsilon$, $\forall \varepsilon > 0$.

Rk: Previously to estimate $\mathbb{E}_{X\sim\mu} f(X)$ with MCMC, would run $Y_n^{(1)}, \ldots, Y_n^{(k)} \overset{iid}{\sim} \mu_n \approx \mu$, form estimate $\frac{1}{k}\sum_{i=1}^{k} f(Y_n^{(i)})$.

Ergodicity $\Longrightarrow$ enough to take $\frac{1}{n}\sum_{i=1}^{n} f(Y_i)$ – "one "chain".

② Convergence estimates: S finite ($|S| = N$), irreducible, recurrent, μ stationary prob. measure, $P_{ij} := p(i,j)$, $P \in \mathbb{R}^{N\times N}$, $\mu_i := \mu(i)$, $\mu \in \mathbb{R}^N$.

Def: Reversible if $\mu_i P_{ij} = \mu_j P_{ji}$ $\forall\, i,j \in [N]$.

$\mu_i P_{ij} P_{jk} P_{kl} = P_{ji} P_{kj} P_{lk} \mu_l = \mu_l P_{lk} P_{kj} P_{ji}$ — equal prob. to see forward vs. reverse sequence @ stationarity.

③

$\mu \rightsquigarrow (D_\mu)_{ii} = \mu_i$ diagonal mx.

Reversibility $\longleftrightarrow D_\mu P = P^T D_\mu \longleftrightarrow D_\mu P$ <u>symmetric</u>.

$\Longrightarrow D_\mu^{1/2} P D_\mu^{-1/2} =: M$ also symmetric.

$(\sqrt{\mu})_i := \sqrt{\mu_i} \longrightarrow M\sqrt{\mu} = D_\mu^{1/2} P \mathbb{1} = D_\mu^{1/2}\mathbb{1} = \sqrt{\mu}$.

$\longrightarrow \sqrt{\mu}$ eigvec with $\lambda = 1$.

<u>Prop</u>: $\|M\| \leq 1$.

<u>Pf</u>: $$\|M\| = \max_{\substack{f\in\mathbb{R}^N\\ \neq 0}} \frac{|f^T M f|}{\|f\|^2} = \max_{f\neq 0} \frac{|f^T D_\mu^{1/2} P D_\mu^{-1/2} f|}{(f^T f)}$$

Cauchy-Schwarz.

$$(f := D_\mu^{1/2} g) \quad = \max_{g\neq 0} \frac{|g^T D_\mu P g|}{(g^T D_\mu g)} = \max_{g\neq 0} \frac{|\mathbb{E}_\mu\, g(Y_0)\, g(Y_1)|}{\mathbb{E}_\mu\, g(Y_0)^2} \leq 1$$

So, if $\lambda_1 \geq \dots \geq \lambda_N$ sorted eigvals, $\lambda_1 = 1$.

Suppose $|\lambda_2|, |\lambda_N| < 1 \Longrightarrow M^k = D_\mu^{1/2} P^k D_\mu^{-1/2} \approx \sqrt{\mu}\sqrt{\mu}^T$

$\Longrightarrow P^k \approx \mathbb{1}\mu^T$: same as convergence!

$\boxed{\text{Rk}: \|\sqrt{\mu}\| = 1}$

Concretely,

$$\max\{|\lambda_2|, |\lambda_N|\}^k \geq \|M^k - \sqrt{\mu}\sqrt{\mu}^T\| = \|D_\mu^{1/2}(P^k - \mathbb{1}\mu^T) D_\mu^{-1/2}\|$$

$$\geq \sqrt{\frac{\min \mu_i}{\max \mu_i}}\, \|P^k - \mathbb{1}\mu^T\|$$

<u>Thm</u>: $$\frac{\max_{x\in S} |\mathbb{E}_x f(Y_k) - \mathbb{E}_{y\sim\mu} f(y)|}{\sqrt{\operatorname{Var}_{y\sim\mu} f(y)}} \leq \sqrt{\frac{\max \mu_i}{\min \mu_i}} \cdot \left(\max\{|\lambda_2|, |\lambda_N|\}\right)^k.$$

<u>Ex</u>: G d-regular graph (all degrees $= d$), p given by SRW.

A = adjacency mx $\rightsquigarrow$ $P = \frac{1}{d} A$.

N vertices $\longrightarrow$ $\mu_i \equiv \frac{1}{N}$ $\longrightarrow$ $M = P$, $M\mathbb{1} = \mathbb{1}$.

If $\underset{\frac{1}{d}\lambda_2(A)}{\underset{\shortparallel}{|\lambda_2(P)|}}, \underset{\frac{1}{d}\lambda_N(A)}{\underset{\shortparallel}{|\lambda_N(P)|}} \leq (1-\varepsilon) \Longrightarrow$ exponentially fast mixing!

<u>Prop</u>: $\lambda_2(A) = d \longleftrightarrow G$ disconnected $\rightsquigarrow$ MC reducible

$\lambda_N(A) = -d \longleftrightarrow G$ bipartite $\longrightarrow$ MC periodic

Otherwise, exp. fast convergence.

<u>Rk</u>: <u>Expanders</u> = graphs w/ $|\lambda_2|, |\lambda_N| \leq 1 - \varepsilon$ for ε, d <u>const</u> as $N \to \infty$. Many CS applications!