Lecture 12 (persistent = recurrent) ①

In this lecture <u>all Markov chains are finite</u> (if it is not said else). Transition matrix:

$\underline{P} = (p_{i,j})_{i,j=1}^{n} \in \mathbb{R}_{\geq 0}^{n \times n}$, $\forall i \in [n]: \sum_{j=1}^{n} p_{i,j} = 1$ (stochastic matrix) $\underline{D}$: $\bullet \xrightarrow{p_{i,j} > 0} \bullet$ – trans. digraph.

P is **irreducible**: $\forall i,j \in [n]\ \exists m \in \mathbb{N}$: $p_{i,j}^{(m)} > 0$. $\iff \forall i,j \in [n]\ \exists$ dir. path in D from i to j. (*)

For $i \in [n]$, let $C(i) = \{n \in \mathbb{N} \mid p_{i,i}^{(n)} > 0\}$ (*) $= \{$the lengths of dir. cycles from i to i in $D\}$.

i is <u>aperiodic</u>: $GCD(C(i)) = 1$

P, D is irred. $+ \ \exists$ aperiodic state $\Rightarrow \forall$ state is aperiodic. We say that the chain is <u>aperi.</u>

$\boxed{\text{Lemma (N.T.)}}$ $A = \{a_1, a_2, \ldots, a_q\} \subset \mathbb{N}$ s.t. $GCD(A) = 1 \Rightarrow \exists n_0 \in \mathbb{N}: \{n_0, n_0+1, n_0+2, \ldots\} \subset \{\sum_{i=1}^{q} d_i a_i \mid d_i \in \mathbb{N}_0\}$.

(*) $p_{i,j}^{(n)} = (P^n)_{i,j}$
$p_{i,i}^{(n)} = (P^n)_{i,i}$

Proof. By example, $A=\{3,5\} \rightsquigarrow S = (-3)\cdot 3 + 2\cdot 5 = 1$ ②

$6\cdot 3 = 18$, $6\cdot 3 + S = 3\cdot 3 + 2\cdot 5 = 19$, $6\cdot 3 + 2S = 0\cdot 3 + 4\cdot 5 = 20$,

$7\cdot 3 = 21$, $7\cdot 3 + S = 4\cdot 3 + 2\cdot 5 = 22$, $7\cdot 3 + 2S = 1\cdot 3 + 4\cdot 5 = 23, \dots$ ☒

$\bar{p} = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n_{\geq 0}$, $\sum_1^n p_i = 1$ — a distribution (on states)

$\bar{p}P = \left(\sum_{i=1}^n p_i p_{i,1}, \dots, \sum_{i=1}^n p_i p_{i,n}\right)$ — evolution of $\bar{p}$ by P in 1 unit of time.

$\boxed{\text{Def.}}$ A distrib. $\bar{p}$ is a stationary distrib. of the M. chain P if $\bar{p}P = \bar{p}$.

$\boxed{\text{Theorem } (\exists \text{ of stat. distrib.})}$ Every finite M. chain P that is irreducible and aperiodic has a unique stationary distribution $\bar{p}$. Moreover, $\forall i,j \in [n]: \lim_{k\to\infty} p^{(k)}_{i,j} = p_j$, $\bar{p} = (\dots, p_j, \dots)$

$p^{(k)}_{i,j} = (P^k)_{i,j}$

Proof – By the lectures notes of prof. J. Sgall.

$\boxed{\text{Lemma 1}}$ P aper. and irred. $\Rightarrow \exists k \in \mathbb{N}$ and $\exists \delta > 0$ s.t. $\forall i,j \in [n]: (P^k)_{i,j} > \delta$.

<u>Proof.</u> Since $\forall i \in [n]$ is aperiodic, by the Lemma we can find a $t_0 \in \mathbb{N}$ s.t. $\forall i \in [n]: t \geq t_0 \Rightarrow$
$\Rightarrow (P^t)_{i,i} > 0$. We claim that for $k := t_0 + n$ we have $\forall i,j \in [n]: (P^k)_{i,j} > 0$. Indeed, by irred. there $\exists\, \ell \leq n: (P^\ell)_{i,j} > 0$:

(shortest i-j path)

$$\rightsquigarrow (P^k)_{i,j} \geq \underbrace{(P^\ell)_{i,j}}_{>0} \underbrace{(P^{k-\ell})_{j,j}}_{>0,\text{ as } k-\ell \geq t_0} > 0. \quad \boxtimes$$

~~$\delta := \min_{i,j} (P^k)_{i,j}$.~~ We set $\delta := \min_{i,j} (P^k)_{i,j}$

$\bar{x} \in \mathbb{R}^n$, $|\bar{x}|_1 := |x_1| + |x_2| + \dots + |x_n|$.
$(x_1, x_2, \dots, x_n)$ $\quad \bar{x} = \bar{x}^+ - \bar{x}^-$ where $x_i^+ = \max(x_i, 0)$ and $x_i^- = \max(-x_i, 0)$, have disjoint supports, and $|\bar{x}|_1 = |\bar{x}^+|_1 + |\bar{x}^-|_1$.

Lemma 2

$\bar{v}, \bar{w} \in \mathbb{R}^n_{\geq 0}$ and with each entry $\geq \alpha \geq 0$. Then
$|\bar{v} - \bar{w}|_1 \leq |\bar{v}|_1 + |\bar{w}|_1 - 2\alpha n$.

<u>Proof.</u> $0 \leq \alpha \leq a \leq b \rightsquigarrow b - a \leq a + b - 2\alpha$
$(\Leftrightarrow 2\alpha \leq 2a \Leftrightarrow \alpha \leq \alpha\ \checkmark)$. $\boxtimes$

$\boxed{\text{Lemma 3}}$ $M \in [\delta, +\infty)^{n \times n}$, $\delta > 0$, and with all row sums $= 1$. Let $\bar{x} \in \mathbb{R}^n$, $\bar{x} = \bar{x}^+ - \bar{x}^-$ as described above. Then

(1) $|\bar{x} M|_1 \leq |\bar{x}|_1 - 2\delta n \min(|\bar{x}^+|_1, |\bar{x}^-|_1)$.

(2) If $\bar{x}$ has both > 0 and < 0 entries $\Rightarrow \bar{x} M \neq \bar{x}$

(3) If $\sum_1^n x_i = 0 \Rightarrow |\bar{x} M|_1 \leq (1 - \delta n) |\bar{x}|_1$.

Proof. (1) $\bar{v} := \bar{x}^+ M$, $\bar{w} := \bar{x}^- M$ are ≥ 0 vectors.

M has row sums $= 1 \rightsquigarrow |\bar{v}|_1 = |\bar{x}^+|_1$ and $|\bar{w}|_1 = |\bar{x}^-|_1 \rightsquigarrow |\bar{v}|_1 + |\bar{w}|_1 = |\bar{x}|_1$. Since $M \geq \delta$,

$\rightsquigarrow \forall i: v_i \geq \delta \sum_{i=1}^n x_i^+ = \delta |\bar{x}^+|_1$. Similarly, $\forall i: w_i \geq \delta |\bar{x}^-|_1$. By L2 with $d := \delta \min(|\bar{x}^+|_1, |\bar{x}^-|_1)$,

$|\bar{x} M|_1 = |\bar{v} - \bar{w}|_1$

$\leq |\bar{v}|_1 + |\bar{w}|_1 - 2\delta n \min(|\bar{x}^+|_1, |\bar{x}^-|_1) =$

$= |\bar{x}|_1 - 2\delta n \min(\text{—"—})$

(2) $\rightsquigarrow \min(|\bar{x}^+|_1, |\bar{x}^-|_1) > 0$. By (1) we

have that $|\bar{x}M|_1 < |\bar{x}|_1$ and $\bar{x}M \neq \bar{x}$. (5)

(3) $\leadsto |\bar{x}^+|_1 = |\bar{x}^-|_1 = \frac{|\bar{x}|_1}{2}$. So by (1),

$|\bar{x}M|_1 \leq |\bar{x}|_1 - \frac{\delta}{2}|\bar{x}|_1 = (1-\delta n)|\bar{x}|_1$. □

Let k, δ be as in L. 1.

$\boxed{\text{Proof of the theorem}}$ $\bar{x} = \bar{x}P$ is a system of n homog. lin. equations with n unknowns $x_1, \dots, x_n$. Row sums of $P = 1 \Rightarrow$ the matrix of the system has rank $\leq n-1 \leadsto \exists$ a nontrivial ($\neq \bar{0}$) solution $\bar{x}$. this $\bar{x} \geq 0$ or $\bar{x} \leq 0$ (else L 3.2 $\Rightarrow \bar{x}\underbrace{P^k}_{M} \neq \bar{x}$, in contrary with $\bar{x} = \bar{x}P$)

$\leadsto \bar{p} := \frac{1}{\sum_1^n x_i}\bar{x}$ is a distribution (on $[n]$).

$\leadsto \bar{p}P = \bar{p} \leadsto \bar{p}$ is a stationary distrib., which is moreover unique: if $\bar{q} \neq \bar{p}$ is another stat. distrib. $\leadsto$ ~~sols~~ $\bar{p}-\bar{q}$ is a nontriv. sol. of $\bar{y} = \bar{y}P$ with both >0 and <0 entries, $\lightning$ with L 3.2

Let $\bar{q}$ be any initial distrib. We show that (6)

$\boxed{\lim_{t\to\infty} \bar{q} P^t = \bar{p}}$. For $\bar{q} = (0,\dots,0,\underset{\substack{\uparrow\\ i}}{1},0,\dots,0)$ we

get that $(P^t)_{i,j} \to p_j$.

Let $\bar{a}$ be any distrib. For $s = 0,1,2,\dots$ let

$\bar{v}^{(s)} := \bar{a} P^{sk} - \bar{p}$. $\bar{p}$ is stat. distr. $\Rightarrow \bar{v}^{(s)} =$

$= \dots = (\bar{a} - \bar{p}) P^{sk}$ and $\bar{v}^{(s+1)} = \bar{v}^{(s)} P^k$. The

coordinates of each $\bar{v}^{(s)}$ sum to 0 and $(P^k)_{i,j} > 0$

by L1. By L3.3:

$$|\bar{v}^{(s+1)}|_1 = |\bar{v}^{(s)} P^k|_1 \leq$$

$\leq (1-\delta n)|\bar{v}^{(s)}|_1 \leadsto |\bar{v}^{(s)}|_1 \leq \underbrace{(1-\delta n)^s}_{\in[0,1)} |\bar{v}^{(0)}|_1$

$\leadsto |\bar{v}^{(s)}|_1 \to 0$ for $s \to \infty$ $\leadsto \bar{v}^{(s)} \to \bar{0}$ and

$\boxed{\lim_{s\to\infty} \bar{v}^{(s)} = \bar{0}}$ and $\boxed{\lim_{s\to\infty} \bar{a} P^{sk} = \bar{p}}$.

Now let $\bar{a} := \bar{q} P^l$ for $\underbrace{l = 0,1,\dots,k-1}_{\text{each fixed}}$. By

$\lim_{s\to\infty} \bar{q} P^{sk+l} = \bar{p}$. Hence $\lim_{t\to\infty} \bar{q} P^t = \bar{p}$. ☒

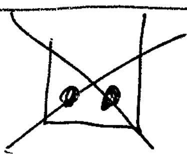
Balls and Bins

m balls ○ ○ ⋯ ○

↓ randomly

⊔ ⊔ ⋯ ⊔ n bins.

Properties of the distribution of balls in bins.

$m \le n$, the probab. that , i.e. ≤ 1 ball in each bin, is

$$\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right) \overset{m=o(n)}{\approx} \prod_{j=1}^{m-1} e^{-j/n} = e^{-\frac{m(m-1)}{2n}} \approx$$

$$\approx e^{-m^2/2n}$$ ← The probab. that for m people and n (e.g. $n=365$) possible birthdays, all have different birthdays.

For example, for $m=\sqrt{2(\log 2)n}$ we get this probab. $\doteq \frac{1}{2}$.

Proposition

$m=n$, $\Pr\Big(\underbrace{\max_{\text{bins}} |\square|}_{\text{maximum load}} > \frac{3\log n}{\log\log n}\Big) \le \frac{1}{n}$ for $n \ge n_0$.

$\Pr(\overset{\geq M}{\text{bin } 1}) \leq \binom{n}{M}\left(\frac{1}{n}\right)^M$ (union bound) ⑧

$\leq \frac{1}{M!} \leq \left(\frac{e}{M}\right)^M.$

Again by the union bound, for $M \geq \frac{3\log n}{\log\log n}$ we have that $\Pr(\overset{\geq M}{\text{any bin}}) \leq n \cdot \left(\frac{e \log\log n}{3\log n}\right)^{\frac{3\log n}{\log\log n}} \leq$

$\leq n \cdot \left(\frac{\log\log n}{\log n}\right)^{\frac{3\log n}{\log\log n}} = \exp(\log n) \cdot \exp\left((\log_{(3)} n - \log_{(2)} n)\frac{3\log n}{\log_{(2)} n}\right) = \exp\left(-2\log n + \underbrace{\frac{3\log n \cdot \log_{(3)} n}{\log_{(2)} n}}_{= o(\log n)}\right) \leq$

$\leq \frac{1}{n}$ for $n \geq n_0$ $\boxtimes$

Bucket sort $\overbrace{n = 2^m \text{ random integers}}^{\text{elements}}$ from $[0, 2^k)$ where $k \geq m$. Using BS we sort the numbers in expected time $O(n)$. The algor. is deterministic.

1st stage: elements $\longrightarrow$ $\overbrace{\sqcup_1 \ \sqcup_2 \ \cdots \ \sqcup_n}^{n \text{ buckets}}$

In the j-th bucket: numbers with first m bin. digits $= j$. We get $\sqcup_1 < \sqcup_2 < \cdots < \sqcup_n$ and need only $O(n)$ time, assuming that $\forall$ element can be placed in the appropr. bucket in $O(1)$ time. (?)

elements in buckets: binomial distribution $B(n, 1/n)$.

2nd stage: $\forall$ bucket (i.e. the elements in it) is sorted in quadratic time.

$X_j :=$ # elements in the j-th bucket, time to sort it: $\leq c X_j^2$. The expected time for the 2nd stage:

$$\mathbb{E}\left(\sum_{j=1}^{n} c X_j^2\right) = c \sum_{j=1}^{n} \mathbb{E} X_j^2 = c n \, \mathbb{E} X_1^2.$$

As we know or will know (in the last lecture), $\mathbb{E} X_1^2 = \frac{n(n-1)}{n^2} + 1 = 2 - \frac{1}{n} < 2$.

[X_1 has $B(n, \frac{1}{n})$ distribution. $p \in [0,1]$, $n \in \mathbb{N}$, $X \in \{0, 1, \ldots, n\}$, $B(n,p)$: $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$]. So the 2nd stage takes $O(n)$ time too.

Thank you!