CPSC 536J Notes

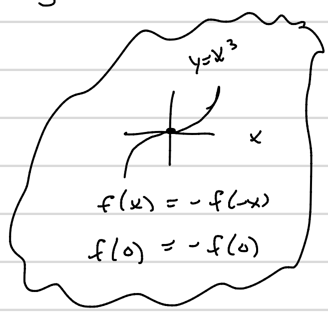
Starting Jan 25

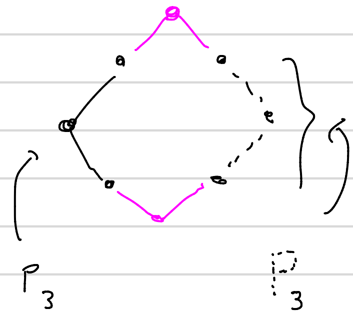
(Markov Chains & Reversibility)

Jan 25:

eigenpairs (adjacency)

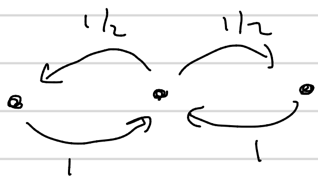
P_3





Reversible Markov Chains

Inner products, positive semi-definite matrices, – –



Markov chain $\longleftrightarrow$ Markov matrix $\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$

warning

$$\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

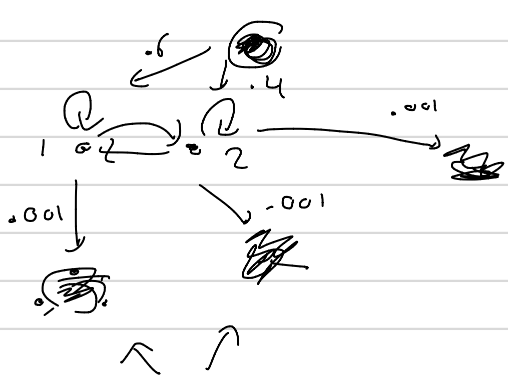
$$\begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \end{bmatrix}$$

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

P_{ij} = Prob from state i you move to state j

Markov chain/matrix is <u>irreducible</u> if from any state, you can get to any other state in some number of steps with non-zero probability.

Why?



Ans: Say i, j are <u>equivalent</u> if ① there is a path $i \to j$

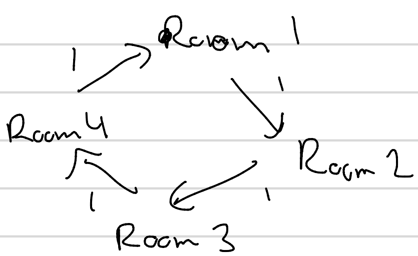
② " " " " $j \to i$

(in <u>digraphs</u> we say i is <u>strongly</u> connected to j)

An irreducible Markov chain/matrix is aperiodic if $\left\{\begin{matrix}\text{for some state} \\ \text{" any "}\end{matrix}\right\}$ i

$$\underbrace{GCD\{\text{lengths of paths from } i \text{ to } i\}}_{\text{called the period(icity) of the chain}} = 1$$

=



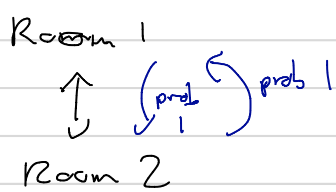
"Period" = 4

Thm: If P is an irreducible, aperiodic Markov matrix

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \pi_1 & \cdots & \pi_n \\ \vdots & & \\ \pi_1 & \cdots & \pi_n \end{bmatrix}$$ for some stochastic vector $[\pi_1 \cdots \pi_n]$

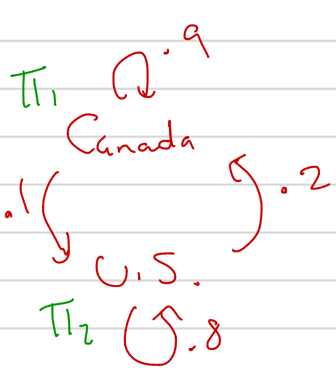
=

$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\quad P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ldots$ no limit ☹



"steady state dist" $[.5 \quad .5]$

$[.5 \quad .5] P = [.5 \quad .5]$ (eigenvector, eigenvalue 1)

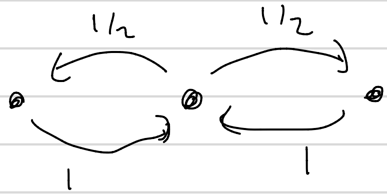


$P = \begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}$, $P^2 = \begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}\begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}$

$= \begin{bmatrix} .83 & .17 \\ .34 & .66 \end{bmatrix}$

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{bmatrix}$$

$\pi_1(.1) = \pi_2(.2)$



$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\neq P^T$$

P has real eigenvalues

$$\lim_{n\to\infty} P^n =$$

$$\det\left(\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} - \lambda I \right) = 0$$

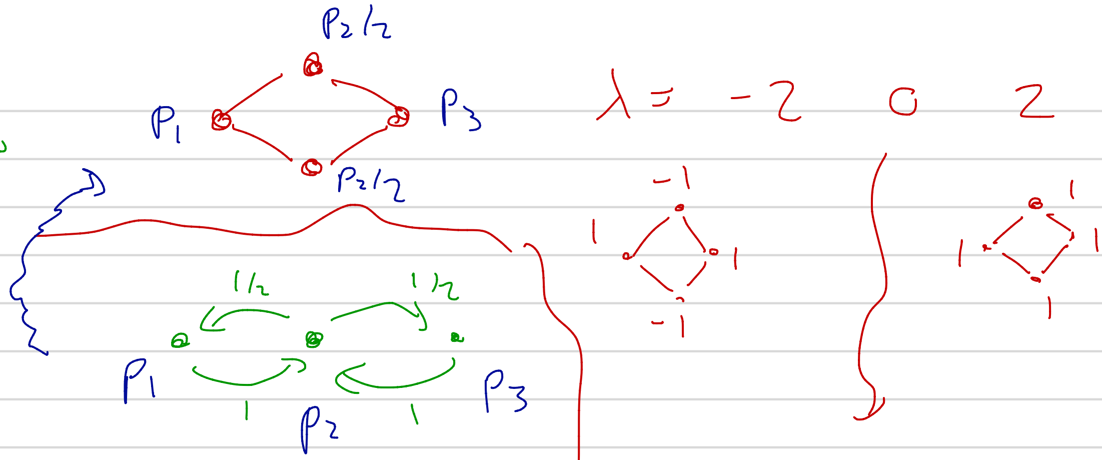
$$\begin{bmatrix} \lambda & -1 & 0 \\ -1/2 & \lambda & -1/2 \\ 0 & -1 & \lambda \end{bmatrix} = \lambda^3 - \lambda \qquad \lambda = \{1, -1, 0\}$$

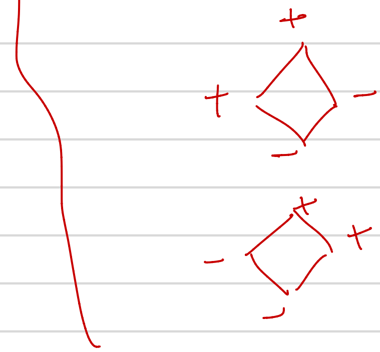
$$[\pi_1 \ \pi_2 \ \pi_3] \cdot I = [\pi_1 \ \pi_2 \ \pi_3] P$$

Periodic, period 2 $\quad \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$

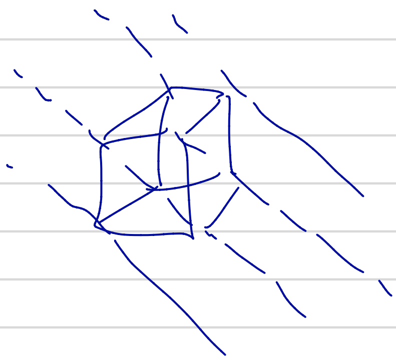
=

New





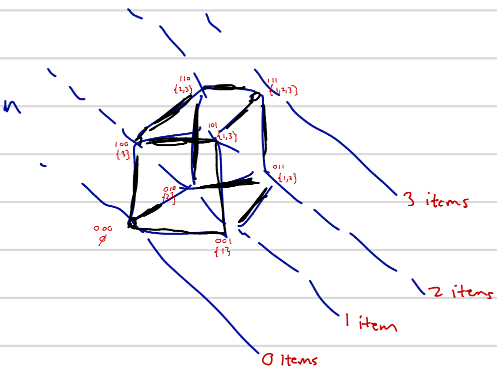
= $\mathbb{B}^n$

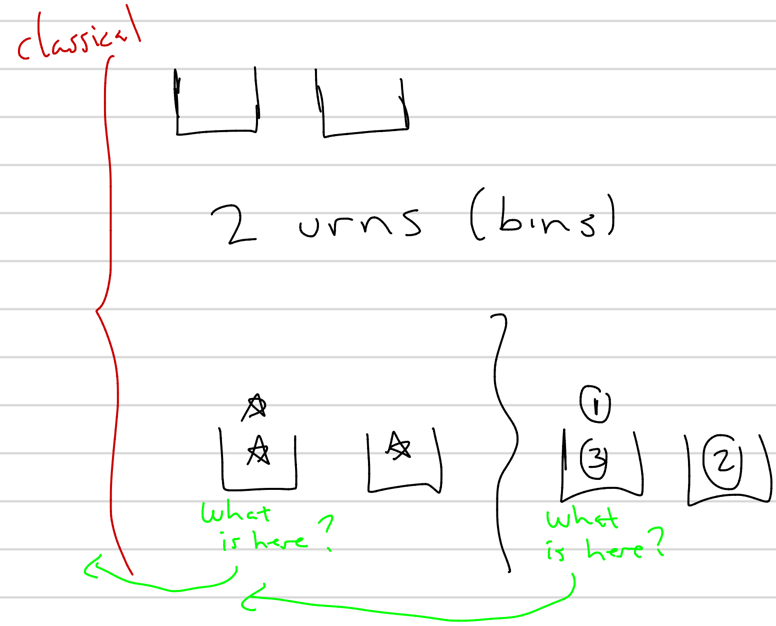


n distinct balls
n indistinguishable "

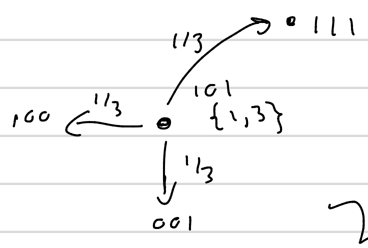
Jan 28

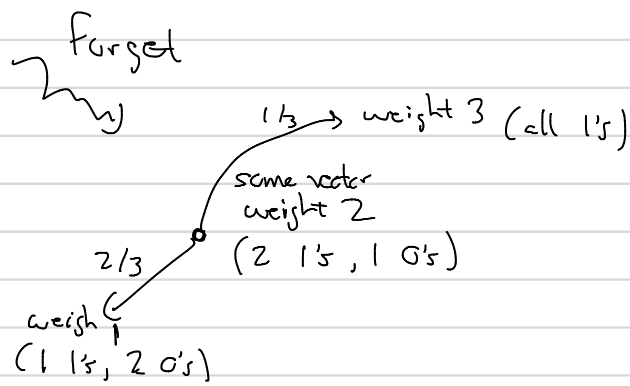
$\mathbb{B}^n =$



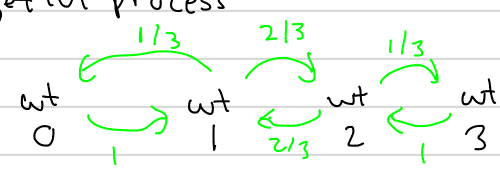


$\mathbb{B}^n$





Forgetful process



$$
P = \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline \text{wt } 0 & 0 & 1 & 0 & 0 \\ \text{wt } 1 & 1/3 & 0 & 2/3 & 0 \\ 2 & 0 & 2/3 & 0 & 1/3 \\ 3 & 0 & 0 & 1 & 0 \end{array}
$$

Given a finite dimension vector space $\vec{V} = \begin{cases} \mathbb{R}^{\{\text{states of our Markov chain}\}} \\ \mathbb{C}^{\{\text{''}\}} \end{cases}$

$\mathbb{R}^n, \mathbb{C}^n$, n = # states in Markov chain.

<u>Inner</u> <u>product</u> on $\vec{V}$ is $\vec{V} \times \vec{V} \to \begin{cases} \mathbb{R} \\ \mathbb{C} \end{cases}$

(1) $\forall \vec{v} \in \vec{V}$, $\langle \vec{v}, \vec{v} \rangle \in \mathbb{R}_{\geq 0}$ and $\langle \vec{v}, \vec{v} \rangle = 0 \iff \vec{v} = 0$

(2) $\langle\ ,\ \rangle$ bilinear : conj bilinear, sesqu "

① $\langle \vec{v}_1, \vec{v}_2 + \vec{v}_3 \rangle$ = what you think $\langle \vec{v}_1, \vec{v}_2 \rangle + \langle \vec{v}_1, \vec{v}_3 \rangle$

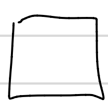
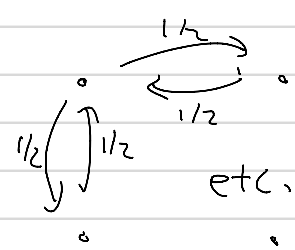
② $\langle \vec{v}_1, \vec{v}_2 \rangle = \overline{\langle \vec{v}_2, \vec{v}_1 \rangle}$ ← complex conj

③ $\langle \vec{v}_1, \alpha \vec{v}_2 \rangle = \alpha \langle \vec{v}_1, \vec{v}_2 \rangle$

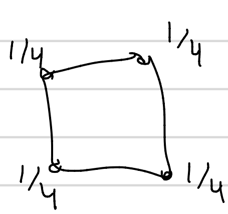
$(\vec{v}_i \in \vec{V},\ \alpha \in \mathbb{C})$

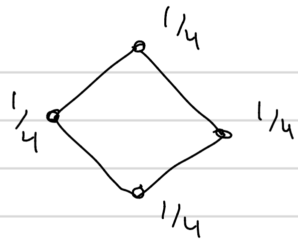
$\langle \alpha \vec{v}_1, \vec{v}_2 \rangle = \bar{\alpha} \langle \vec{v}_1, \vec{v}_2 \rangle \Leftarrow$ ②, ③

Compromise: For graphs, reversible Markov chains, can work in $\mathbb{R}$. (Sometimes better over $\mathbb{R}$)

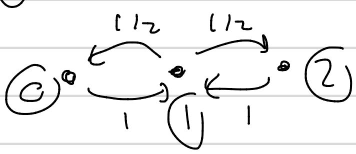
Example:  Markov Chain $\rightsquigarrow$ 

etc.

Steady-state stationary: 



Forgot



$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$ not symmetric

$f, g : \{0, 1, 2\} \to \mathbb{R}$:

$$\langle f, g \rangle_\pi = \frac{1}{4} f(0)g(0) + \frac{1}{2} f(1)g(1) + \frac{1}{4} f(2)g(2)$$

$\pi = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$

$$\langle f, g \rangle_{\pi^{-1}} = 4\, f(0)g(0) + 2\, f(1)g(1) + 4\, f(2)g(2)$$

Rem: $f \cdot g$ or (f, g) denotes : $\sum_i f(i) g(i)$,

then any $\langle \, , \rangle$ given as $f \cdot (Mg)$ where M is symmetric & pos def

and conversely

For any $\mathcal{L} : V \to W$ ~~and~~ any inner product $\langle \, , \rangle_V$ on V, $\langle \, , \rangle_W$ on W

we ~~define~~ $\mathcal{L}^*$ by

$\langle v$

Given $\mathcal{L} : \bar{V} \longrightarrow \bar{W}$ lin trans vect sp $\bar{V}, \bar{W}$ over $\left\{ \begin{array}{l} \mathbb{R} \\ \mathbb{C} \\ \mathbb{F} \text{ field} \end{array} \right.$

there is a unique $\mathcal{L}^* : \bar{W}^* \longrightarrow \bar{V}^*$ ~~that~~ satisfies

$\forall \bar{v} \in \bar{V}$, $\forall\, l : \bar{W}^*$

$$l(\mathcal{L} v) = (\mathcal{L}^* l)\, v$$

(in W, W^*) (in $\bar{V}, \bar{V}^*$)

(here $\bar{V}^* = \mathrm{Hom}(\bar{V}, \mathbb{F}) =$ (lin trans $\bar{V} \to \mathbb{F}$))

=

Special case: If $\langle\, , \rangle$ on $\bar{V}$ (then $\bar{V} \overset{\langle , \rangle}{\longleftrightarrow} \bar{V}^*$)

and $\langle\, , \rangle$ on $\bar{W}$, $\mathcal{L} : \bar{V} \to \bar{W}$

$$\langle \mathcal{L} v, w \rangle_{\bar{W}} = \langle v, \mathcal{L}^* w \rangle_{\bar{V}}$$

=

Special case: $\bar{V} = \bar{W}$, $\mathcal{L} : \bar{V} \to \bar{V}$, then

$$\langle \mathcal{L} v, w \rangle_{\bar{V}} = \langle v, \mathcal{L}^* w \rangle_{\bar{V}} \qquad v, w \in \bar{V}$$

=

Special case: $\bar{V} = \mathbb{R}^n$, $\langle\, , \rangle$ ~~is~~ dot product

$$(A v, w) = (v, A^T w)$$

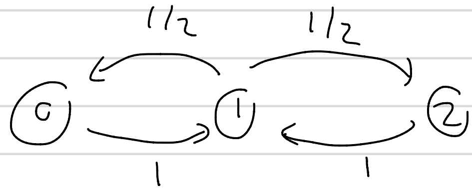
where $A : \mathbb{R}^n \to \mathbb{R}^n$ think of it as a matrix over $\vec{e}_1, \ldots, \vec{e}_n$

or $A : \mathbb{C}^n \to \mathbb{C}^n$, $A^T \rightsquigarrow A^H = (\bar{A})^T$

Jan 30

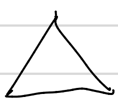
Detailed Balance, Entropy:

One more reason:



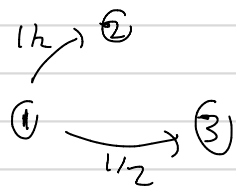
need $\left\{\begin{array}{l}\text{steady-state} \\ \text{stationary}\end{array}\right\}$ distribution

=

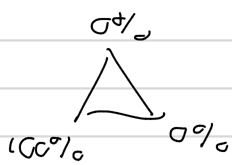


graph : Markov chain

2-regular graph



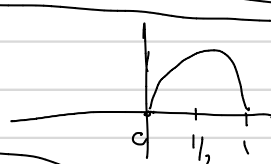
Start $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$
$\quad$ 1 $\quad$ 2 $\quad$ 3



$1-p_1-p_2$

$$\text{Entropy}_2(p_1, p_2, p_3) = \sum p_i \log_2 \frac{1}{p_i}$$
$$= -\sum p_i \log_2 p_i$$

$$-(1-p)\log_2(1-p) - p\log_2 p$$



$$\text{Ent}_2(1,0,0) = 0$$

$$\binom{n}{pn} \overset{\text{within } 2^{o(n)}}{\sim} 2^{H_2(p)\cdot n}$$

Computer Scientists

$$H_2\left(\tfrac{1}{2}\right)$$

$$H_2(p) = -p\log p - (1-p)\log(1-p)$$

$$\begin{bmatrix}1\\0\\0\end{bmatrix} \xrightarrow[\text{~~Markov~~}]{\text{op by}} \begin{bmatrix}0\\1/2\\1/2\end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix}1/3 +\\1/3 +\\1/3 +\end{bmatrix}$$

small

$$H_2\left(\tfrac{1}{2}\right) = -\left(1-\tfrac{1}{2}\right)\log_2\left(1-\tfrac{1}{2}\right) - \tfrac{1}{2}\log_2\tfrac{1}{2}$$

$$\text{''} = H_2(p,q) \qquad = \left(-\tfrac{1}{2}\log_2\tfrac{1}{2}\right)2$$

$$q = 1-p \qquad \text{''} \qquad = 1$$

$$\lim \longrightarrow \begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix}$$

$$\begin{bmatrix}1\\0\\0\end{bmatrix} \longrightarrow \begin{bmatrix}0\\1/2\\1/2\end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix}$$

$\mathrm{Entr}(1,0,0) = 0$ $\qquad$ $\mathrm{Entropy} \begin{cases} > 0 \\ < \log_2 3 \end{cases}$ $\qquad$ $\mathrm{Entropy}_2(1/3, 1/3, 1/3)$

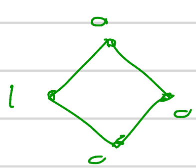
$$= 3\left(\tfrac{1}{3}\log_2 3\right)$$

$$= \log_2 3$$

= (connected) d-reg graph on n vertices

$$\begin{bmatrix}0\\1\\0\\0\\0\\0\end{bmatrix} \longrightarrow \qquad \text{poss! } \lim = \begin{bmatrix}1/n\\1/n\\\vdots\\1/n\end{bmatrix}$$

Entropy = 0 $\qquad$ (aperiodic) (irreducible)

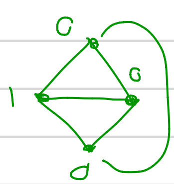


$\longrightarrow \cdots \longrightarrow$ $\begin{bmatrix}1/2\\0\\0\\1/2\end{bmatrix} \longrightarrow \begin{bmatrix}0\\1/2\\1/2\\0\end{bmatrix} \longrightarrow \cdots$

periodic
period = 2

+ small $\qquad$ + small

Entropy $\log_2 2 = 1$

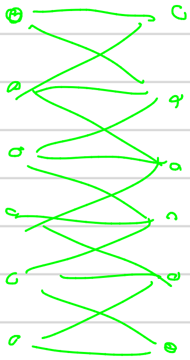


$\longrightarrow \cdots \longrightarrow$ $\begin{bmatrix}1/4\\1/4\\1/4\\1/4\end{bmatrix}$

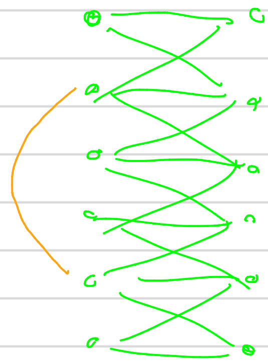
Entropy 0 $\qquad$ $\mathrm{Ent}(p_1, \dots, p_n)$ $\qquad$ Entropy $\log_2 4 = 2$

$$= \sum p_i \log \frac{1}{p_i}$$

Rem: $\mathrm{Entropy}(p_1, \dots, p_n) \le \log_2 n$, equality iff $p_i = \frac{1}{n}$



period 2



period 2

without extra edge,

← Cayley on $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$

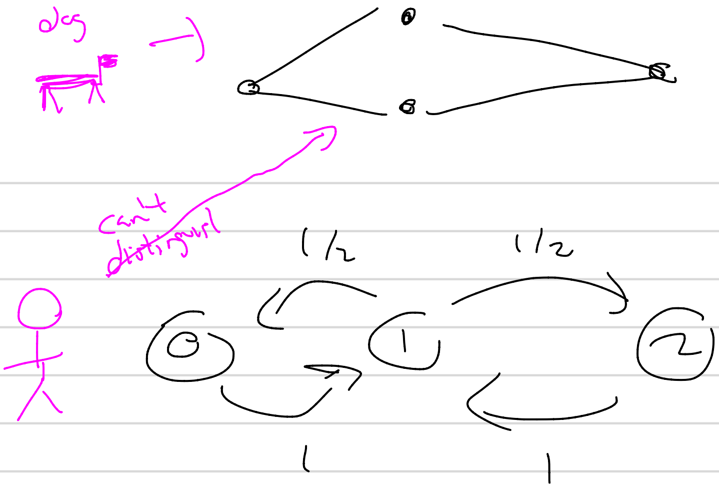
or

$\mathrm{Cycle}_m \times \mathrm{Cycle}_2$

lim Entropy:

$$\log_2(n/2) = \log_2 n - 1$$

n = # vertices



2 balls in 2 urns/bins

① 2 balls

② ① ② " " distinguished

$\frac{1}{4}$ $\frac{1}{2}$ $\frac{1}{4}$

1/3 1/3 1/3

Entropy $= \log_2 3$

relative entropy

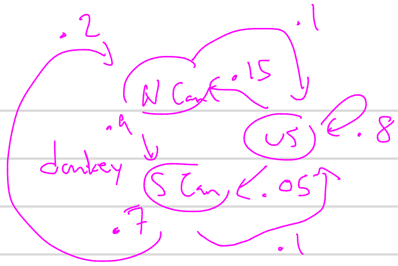
limit

Decreased entropy?

1/4 1/2 1/4

Entropy $< \log_2 3$

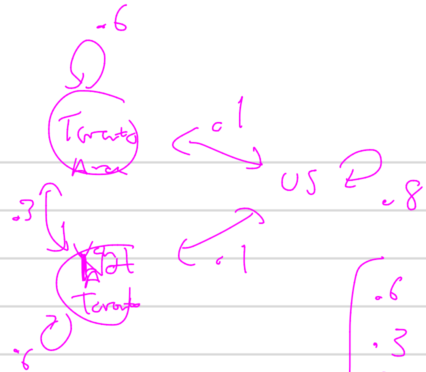
Feb 1



Dolphin

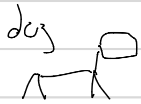
U US

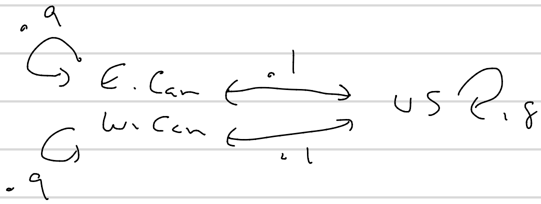
cat



$$\begin{bmatrix} .6 & .3 & .1 \\ .3 & .6 & .1 \\ .1 & .1 & .8 \end{bmatrix}$$

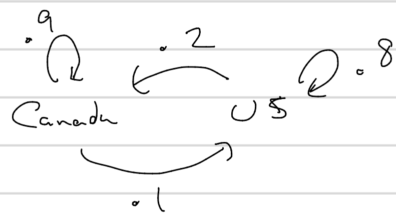
dog





$$\begin{bmatrix} .9 & 0 & .1 \\ 0 & .9 & .1 \\ .1 & .1 & .8 \end{bmatrix}$$

Refinement



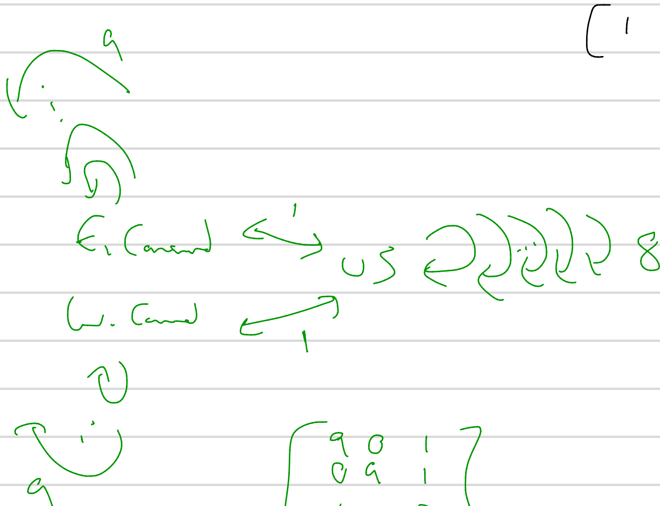
$$\begin{array}{c|cc} & C & U \\ \hline \text{Can} & .9 & .1 \\ \text{US} & .2 & .8 \end{array}$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix} = \begin{bmatrix} .9x_1 + .2x_2, & .1x_1 + .8x_2 \end{bmatrix}$$

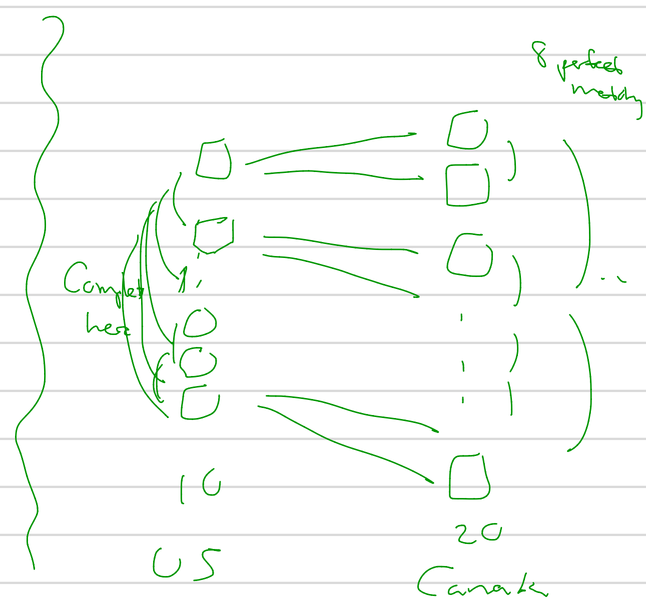
$$\begin{bmatrix} 1 & 0 \end{bmatrix} \quad " \quad = \begin{bmatrix} .9 & , & .1 \end{bmatrix}$$

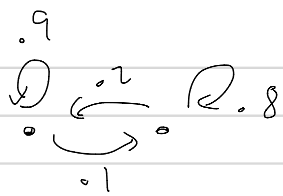
Directed Graph

TL:



$$\begin{bmatrix} 9 & 0 & 1 \\ 0 & 9 & 1 \\ 1 & 1 & 8 \end{bmatrix}$$





$$P = \begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}$$

$$(x_1 \quad x_2) \xrightarrow{C_P} [.9x_1 + .2x_2 \quad .1x_1 + .8x_2]$$

$$\langle [x_1 \quad x_2], [y_1 \quad y_2] \rangle = \alpha x_1 y_1 + \beta x_2 y_2 \quad \text{for some } \alpha, \beta \in \mathbb{R}_{>0}$$

$$\langle [x_1 \; x_2], C_P [y_1 \; y_2] \rangle = \langle [x_1 \quad x_2], [.9y_1 + .2y_2 \quad .1y_1 + .8y_2] \rangle$$

$x_1 y_2 (\alpha \cdot 0.2) \leftarrow$ $= \alpha x_1 (.9y_1 + .2y_2) + \beta$ etc.

$$\langle C_P [x_1 \; x_2], [y_1 \quad y_2] \rangle = \langle [.9x_1 + .2x_2 \quad .1x_1 + .8x_2], (y_1 \quad y_2) \rangle$$

$x_1 y_2 (\beta \cdot 0.1) \leftarrow$ $= \alpha(.9x_1 + .2x_2) y_1 + \beta(.1x_1 + .8x_2) y_2$

$\langle \vec{x}^T, \vec{y}^T P \rangle = \langle \vec{x}^T P, \vec{y}^T \rangle$ then $\beta = 2\alpha$

$\langle \vec{x}, P\vec{y} \rangle = \langle P\vec{x}, \vec{y} \rangle$ then $\beta = \frac{1}{2}\alpha$

$\langle \, , \, \rangle_{\pi^{-1}}$ $\langle \, , \, \rangle_{\pi}$

So: if π = stationary dist $= [\frac{2}{3}, \frac{1}{3}]$

$$\langle \vec{x}, \vec{y} \rangle_\pi = x(1)y(1)\pi(1) + x(2)y(2)\pi(2) =$$

If for $\nu : \{1, \dots, n\} \to \mathbb{R}_{>0}$

$$\langle \vec{x}, \vec{y} \rangle_\nu = x(1)y(1)\nu(1) + \dots + x(n)\,y(n)\,\nu(n)$$

"weighted dot product" $= x^T \begin{bmatrix} \nu(1) & & 0 \\ & \nu(2) & \\ 0 & & \ddots \\ & & & \nu(n) \end{bmatrix} \vec{y}$

Then for Markov mat, P, irreducible, station dist π

$$\langle \vec{x}, P\vec{y} \rangle_\pi = \langle P\vec{x}, \vec{y} \rangle_\pi$$

and

$$\langle \vec{x}^T P, \vec{y}^T \rangle_{\pi^{-1}} = \langle \vec{x}^T, \vec{y}^T P \rangle_{\pi^{-1}}$$

Soon: SVD of P:

"left" eigenvalues, eigenvectors

"right" " , "

same / different:

$$P = \begin{bmatrix} .8 & .1 \\ .2 & .8 \end{bmatrix}$$

e.g. $\vec{\pi}^T P = \vec{\pi}^T$, $P \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

Feb 4: Markov matrix P, irreducible

Rem: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ periodic, period 2 Adj

$\lambda = 1$

$\lambda = -1 + 2(.000001)$

$\left\{ \begin{bmatrix} .000001 & 1-.000001 \\ 1-.000001 & .000001 \end{bmatrix} \right.$ is "not aperiodic" almost periodic

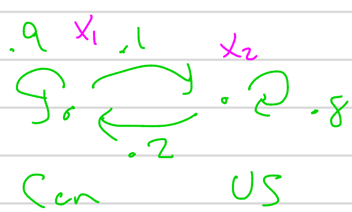
$\begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}$ much faster "mixing" $\left. \right\} \begin{matrix} \lambda = 1 \\ \lambda = .7 \end{matrix}$

$$Tr(A) = \sum_{i=1}^{n} \lambda_i$$

Claim: Define: P Markov matrix, π stationary:

$$\text{Entropy}_{\pi}(\vec{x}) = -\sum_{i=1}^{n} x_i \log \pi_i \quad \left(\begin{matrix} n = \dim \\ \text{matrix } P \end{matrix}\right)$$

e.g. $\begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}$, $\pi = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$



$\text{Entropy}_{\pi}[x_1, x_2] = -x_1 \log \frac{2}{3} - x_2 \log \frac{1}{3}$

$([x_1 \; x_2]$ stochastic$)$

$[1 \; 0] \to [.9 \; .1] \to \ldots \to \ldots \lim \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$

"Proof" 2 lines: (of a well known theory)

If $p, q : \underset{\mathcal{X}}{\text{Set}} \to [0,1]$, $\sum_{x \in \mathcal{X}} p(x) = \sum_{x \in \mathcal{X}} q(x) = 1$

Define

$$D(p \| q) = \sum_{x \in \mathcal{X}} p_x \log \frac{p_x}{q_x}$$

e.g. P $n \times n$ Markov mat: $\mathcal{X} : \{1, 2, \ldots, n\}$

$$D(p \| q) = \sum_{i=1}^{n} p(i) \log \frac{p(i)}{q(i)}$$

e.g.

$$\begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}, \quad D([x_1 \ x_2] \,|\, [x_1' \ x_2']) = -x_1 \log \frac{x_1}{x_1'} - x_2 \log \frac{x_2}{x_2'}$$

$$D([x_1 \ x_2] \,|\, [\tfrac{2}{3} \ \tfrac{1}{3}]) = -x_1 \log \frac{x_1}{2/3} - x_2 \log \frac{x_2}{1/3}$$

e.g.

$$D([\tfrac{2}{3} \ \tfrac{1}{3}] \,|\, [\tfrac{2}{3} \ \tfrac{1}{3}]) = -\tfrac{2}{3}\, 0 - \tfrac{1}{3} \cdot 0$$

$$= 0$$

We want $[x_1, x_2]$ stochastic: $D([x_1 \ x_2] \,|\, [\tfrac{2}{3} \ \tfrac{1}{3}]) \geq 0$

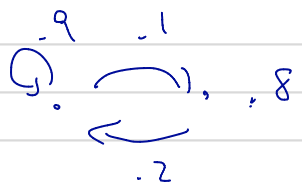
equality iff $[x_1 \ x_2] = [2/3 \ 1/3]$

Rem: $D(p \| q) \neq D(q \| p)$ in general

Rem: $D(p \| q) = \sum p_i \log \frac{p_i}{q_i} \overset{?}{\geq} 0$

$$= E_{[p_1, \dots, p_n]} \left(\sum \log \frac{p_i}{q_i} \right)$$

$=$



$x_1 = .6 \leftarrow C$

$x_2 = .4 \leftarrow U$

$CC \leftarrow (.6)(.9)$

$CU \leftarrow (.6)(.1)$

$UC \leftarrow (.4)(.9)$

$UU \leftarrow (.4)(.1)$

$\swarrow \text{proj}_1$ $\qquad$ $\searrow \text{proj}_2$

1st letter $\qquad$ 2nd letter

C, U $\qquad$ C, U

Feb 6: Remarks on Info Thy:

Gung hay fat choy! Mandarin ? Shing Yeen Kuai

Refs: Cover & Thomas p. 28, 81, 82. [Seach textbook for "26.3"]

Basic idea: (State set is fixed) (Just as in Markov chain)

Subtle ↓↓↓

$$D(\mu_{n+1} \| \mu'_{n+1}) = D(\mu_n \| \mu'_n)$$

$D(P\|P)=0 \rightsquigarrow + D(\underset{\text{"}P\text{"}}{\mu_{n+1}|\mu_n} \| \underset{\text{"}P\text{"}}{\mu'_{n+1}|\mu'_n})$

$D(\cdot\|\cdot) \geq 0$ any other $\rightsquigarrow - D(\underbrace{\mu_n|\mu_{n+1} \| \mu'_{n+1}|\mu'_n}_{\text{something like this}})$

"relative-entropy" or "Kullback-Leibler" distance. See also Wikipedia

$$\boxed{\begin{aligned} H(X,Y) &= H(X) + H(Y;X) \\ &= H(Y) + H(X;Y) \end{aligned}}$$

Warning: $\text{Entropy}(X) := \sum_{i \in \text{Im}(X)} -\Pr(X^{-1}(i)) \log(\Pr(X^{-1}(i)))$

$= -\sum_i p_i \log(p_i)$

$X: \Omega \to \mathbb{R}$

$\Omega \to \mathbb{N}$

$\Omega \to \{1, \ldots, n\}$

etc.

is a "discontinuous function" of $X: \Omega \to \mathbb{R}$

e.g. $\Pr(X_n = 0) = \frac{1}{2}$ $\quad \Pr(X_n = \frac{1}{n}) = \frac{1}{2}$ $\quad x=0,\ x=\frac{1}{n}$; $\frac{1}{2}, \frac{1}{2}$

$\text{Entropy}(X_n) = \log_2 2 \text{ bits} = 1 \text{ bit}$

$= \log_e 2 \text{ nats} = \log_e 2 \text{ nats}$

$\text{Entropy}(\lim X_n) = 0.$ Better to think of X as equivalence classes in Ω.

$X_\infty = 0$, 1, X_∞

$\underline{X}$: $\mathrm{Map}_{\underline{X}} : \Omega \to \mathrm{Image}\,\underline{X}$ (Ω-measurable)

or $\underline{X}$: equivalence classes on Ω :

$$\omega_1 \sim_{\underline{X}} \omega_2$$

=

Given $\underline{X}$: $\mathrm{Map}_{\underline{X}} : \Omega \to \mathrm{Image}\,\underline{X}$

and

$\underline{Y}$: $\mathrm{Map}_{\underline{Y}} : \Omega \to \mathrm{Image}\,\underline{Y}$

what is $(\underline{X}, \underline{Y})$? Joint dist of $\underline{X}, \underline{Y}$??

① $\underline{X} \times \underline{Y}$: $\mathrm{Map}_{\underline{X} \times \underline{Y}} : \Omega \to \mathrm{Image}\,\underline{X} \times \mathrm{Image}\,\underline{Y}$

OR

② $\omega_1 \sim_{\underline{X} \times \underline{Y}} \omega_2$ iff $\begin{cases} \omega_1 \sim_{\underline{X}} \omega_2 & \text{AND} \\ \omega_1 \sim_{\underline{Y}} \omega_2 \end{cases}$

Levin, Peres, Wilmer, *Markov Chains and Mixing Times*

page 30, §1.6 REVERSIBILITY AND TIME REVERSALS

P irreducible Markov chain, stationary π:

<u>Def</u>: <u>Time reversal</u> of P is

$$\hat{P}(x,y) := \frac{\pi(y)\,P(y,x)}{\pi(x)}$$

P is reversible:

$$\boxed{\pi(x)\,P(x,y) = \pi(y)\,P(y,x)}$$

"1-step reversibility"

$\Rightarrow$ 2-step "

$\Rightarrow$ 3-step "

$\vdots$

<u>Exercise</u>: If P is irreducible and aperiodic, then for each sequence of states $x_1, \dots, x_k$, we have for all stochastic X_0,

(1) $$\mathrm{Prob}_P(x_1,\dots,x_k) = \lim_{n\to\infty} \underbrace{\mathrm{Prob}(X_n = x_1, X_{n+1} = x_2, \dots, X_{n+k-1} = x_k)}_{\text{steady-state limit}}.$$

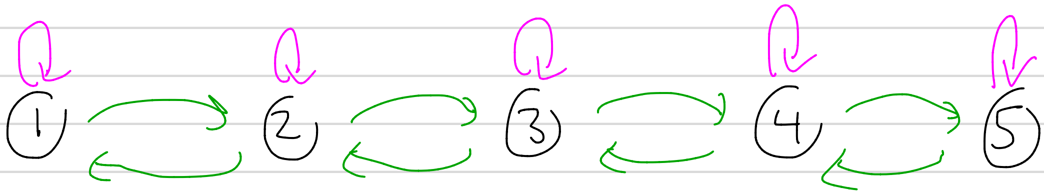
Write

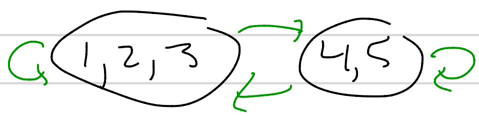
(2) $\mathrm{Prob}_{\hat{P}}(x_1, x_2, \dots, x_k)$ in terms of $\pi(x_1)$, $\pi(x_k)$ and $\mathrm{Prob}_P(x_k, x_{k-1}, \dots, x_1)$.

(3) What if P is periodic (but irreducible)?

Two examples of time reversible Markov chains:

1. Tridiagonal Markov chains:





2. Simulated Annealing / Metropolis–Hastings (?) Algorithm:

Vertices of a graph ?

Search for <u>absolute</u> min of $f: S \to \mathbb{R}$

S = large set, maybe not much structure.

Idea: Start hot, jump around a lot, gradually cool down...

Next time: SVD $A \in \mathbb{R}^{m \times n}$

left eigenpairs

right "