# Chapter 2: Discrete-time Markov chains

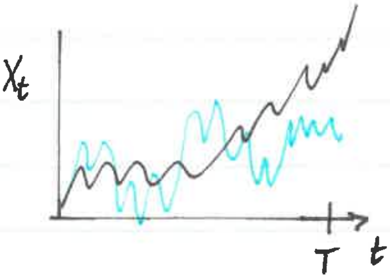
GS Chap.6

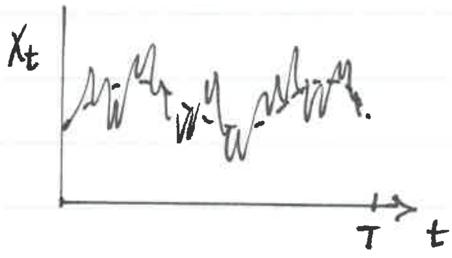
## 2.1. Introduction

Stochastic process   στοχαστικός

Continuous space

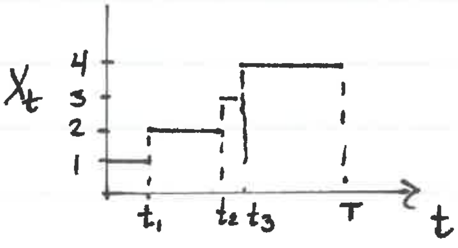
Continuous time

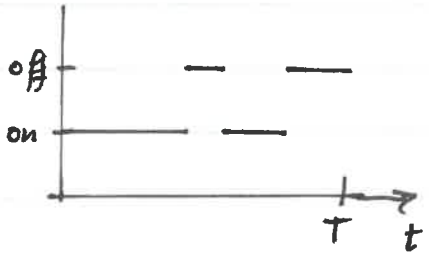




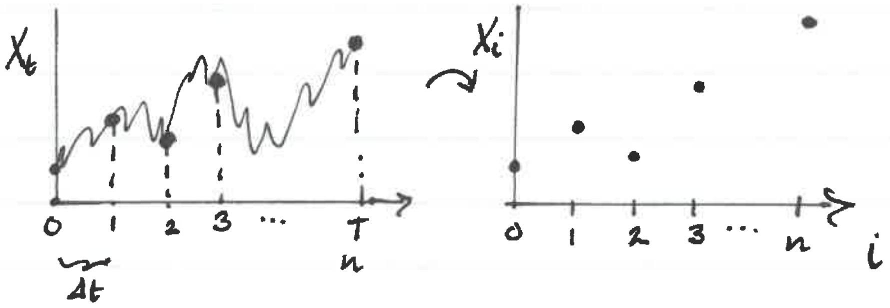
Discrete space

Continuous time





Discrete time



- Trajectory: $\{X_t\}_{t=0}^{T} \longrightarrow \{X_i\}_{i=0}^{n}$ discrete time Sequence of RVs
- State space: $X_i \in \mathcal{X}$   Assume discrete / countable
- Modelling:
  - Full model: Full correlation model

$$p(X_0, X_1, \ldots, X_n) = p(X_0)\, p(X_1|X_0)\, p(X_2|X_0 X_1)\, p(X_3|X_0 X_1 X_2) \cdots$$

cf Chap. 1

  - Memoryless:

$$p(X_0, X_1, \ldots, X_n) = p(X_0)\, p(X_1)\, p(X_2) \cdots$$

independent states

  - Markov:

$$p(X_0, X_1, \ldots, X_n) = p(X_0)\, p(X_1|X_0)\, \underbrace{p(X_2|X_1)}_{X_{i+1} \text{ depends on } X_i} \cdots p(X_n|X_{n-1})$$

W4, L1

## 2.2. Markov chains

- **Def.** : $\{X_i\}_{i=0}^{n}$ is a Markov chain if

$$P(X_n = x_n \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} | x_{n-1}) = \underbrace{P(X_n = x_n \mid X_{n-1} = x_{n-1}}_{\text{transition probability}}$$

for $\underbrace{\text{all } n \geq 1}_{\text{all sequences}}$ and all $\underbrace{x_0, x_1, \dots, x_n}_{\text{all values}}$.

- Assumptions/restrictions :
  - $X_i \in \mathcal{X}$ discrete/countable
  - $P(X_n = b \mid X_{n-1} = a)$ does not depend on $\underset{\text{time}}{\underline{n}}$

$$\Rightarrow \quad P(X_n = b \mid X_{n-1} = a) = P(X_1 = b \mid X_0 = a) \quad \forall n$$

Time-invariant, time independent, homogeneous.

- Transition matrix :

$$P(X_n = j \mid X_{n-1} = i) = P(i \to j) = \Pi_{ij} \text{ or } P_{ij}$$

  - $|\mathcal{X}| \times |\mathcal{X}|$ matrix
  - $1 \geq \Pi_{ij} \geq 0 \quad \forall\, i, j$
  - $\sum_j \Pi_{ij} = 1 \quad \forall\, i$

$\left.\right\}$ Stochastic matrix

$$\Pi = (\Pi_{\underset{\text{row}}{i}\overset{\text{column}}{j}}) = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & \cdots \\ \Pi_{21} & \Pi_{22} & \Pi_{23} & \cdots \\ \vdots & & & \end{pmatrix} \quad \text{row sum} = 1$$

- Note : Different convention used in physics (column notation)

$$P(i \to j) = \tilde{\Pi}_{ji}$$

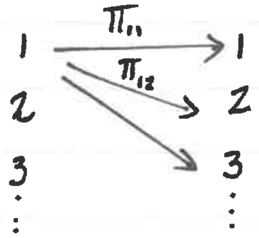
$$\textstyle\sum_j \tilde{\Pi}_{ji} = 1 \quad \text{column sum} = 1$$

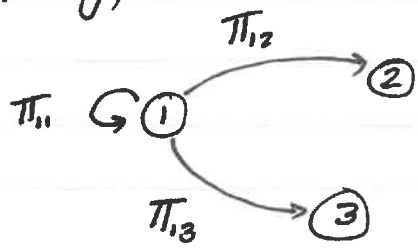
$$\Rightarrow \quad \tilde{\Pi} = \Pi^T$$

W4, h1

- Graphical representation:

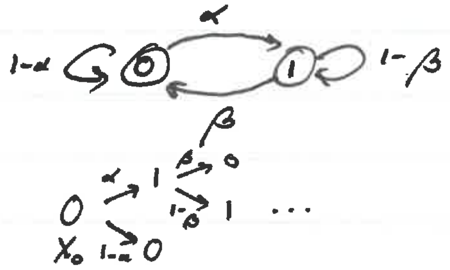
$$\Pi_{ij} = P(i \to j)$$





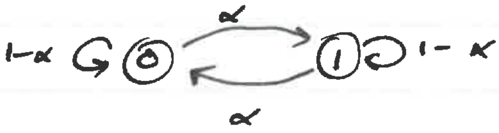
$\sum_j \Pi_{ij} = 1$
all arrows out sum to 1

- Example: 2-state Markov chain



$$\Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \begin{matrix} \to \sum \cdots = 1 \\ \to \sum \cdots = 1 \end{matrix}$$

- Example: Symmetric case $\beta = \alpha$



$$\Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

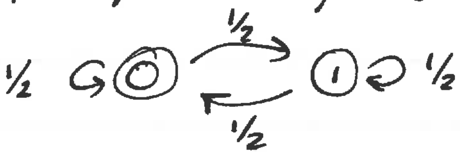
- $\alpha \approx 0$ : 0000 ··· 1111 ···· 0100 ·· persistence

low jump prob ⇒ long sequences of 0's and 1's

- $\alpha \approx 1$ : 0101 001011 ··· anti-persistence

high jump prob ⇒ alternating sequences

- Example: $\alpha = \beta = 1/2$



$$\Pi = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

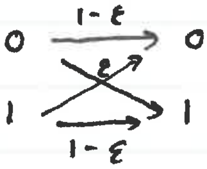
Coin tossing! No correlation.

- Remark: Independent Markov chain if $\Pi_{ij}$ doesn't depend on $i$: $P(i \to j) = P(j)$ All rows are the same

$$P(X_0, X_1, X_2, \ldots) = P(X_0)\, P(X_1 | X_0)\, P(X_2 | X_1) \cdots$$
$$= P(X_0)\ P(X_1)\ \ P(X_2)\ \cdots$$ iid sequence

W4, h1

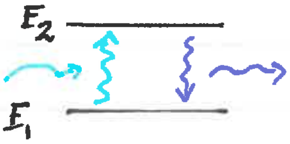
- Example: Binary symmetric channel (information theory)



Bit wrongly copied/transmitted with prob ε
↪ error probability

$$\Pi = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$$ Symmetric Markov chain

- Example: Laser model (physics)

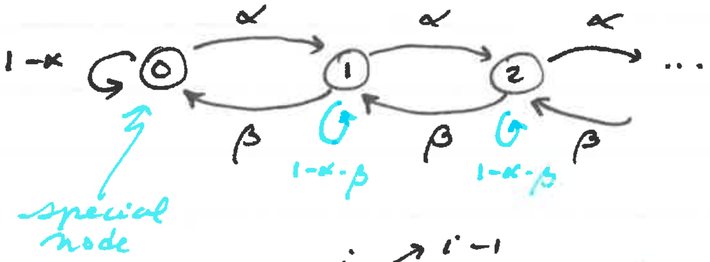


$E_2 > E_1$

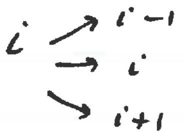
$$P(E_1 \to E_2) = \frac{e^{-\beta E_2}}{e^{-\beta E_1}} = e^{-\beta(E_2-E_1)} = e^{-\beta \Delta E}$$

$$P(E_2 \to E_1) = 1 - e^{-\beta \Delta E} \qquad \Delta E > 0$$

- Example: Population model
  - $X_i \in \{0, 1, 2, \dots\}$
  - $P(i \to i+1) = \alpha$ birth
  - $P(i \to i-1) = \beta$ death



See CW2



population stays same

- Also models: Queues, random walks, ...

## 2.3. Probability propagation

$$P(X_0, X_1) = P(X_0)\, P(X_1 | X_0)$$

$$\Rightarrow \quad P(X_1 = j) = \sum_i P(X_0 = i)\, P(X_1 = j | X_0 = i)$$

$$\Rightarrow \quad P(X_{n+1} = j) = \sum_i P(X_n = i)\, P(X_{n+1} = j | X_n = i)$$

$$= \sum_i P(X_n = i)\, \Pi_{ij}$$

Chapman-Kolmogorov equation

- Matrix notation:

$$\bar{p}_n \qquad (\bar{p}_n)_i = \bar{p}_n(i) = P(X_n = i) \qquad \sum_i \bar{p}_n(i) = 1$$

$$\Pi \qquad (\Pi)_{ij} = P(X_{n+1} = j | X_n = i) \qquad \sum_j \Pi_{ij} = 1$$

$$\bar{p}_{n+1} = \bar{p}_n \Pi$$

Chapman-Kolmogorov

- Propagation:

$$\bar{p}_1 = \bar{p}_0 \Pi$$

$$\bar{p}_2 = \bar{p}_1 \Pi = \bar{p}_0 \Pi \Pi = \bar{p}_0 \Pi^2$$

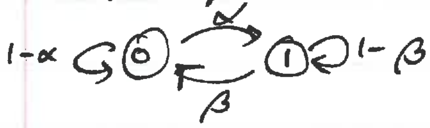
$$\vdots$$

$$\bar{p}_n = \bar{p}_{n-1} \Pi = \bar{p}_0 \underbrace{\Pi^n}_{n\text{-step transition probability}}$$

$$P(X_n = j) = \sum_i P(X_0 = i)\, (\Pi^n)_{ij} = \sum_i P(X_0 = i)\, \underbrace{P(X_n = j | X_0 = i)}_{n\text{-step}}$$

$$\Rightarrow \quad P(X_n = j | X_0 = i) = (\Pi^n)_{ij}$$

- Example: 2-state Markov chain



$$\Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$\begin{pmatrix} p_{n+1}(0) & p_{n+1}(1) \end{pmatrix} = \begin{pmatrix} p_n(0) & p_n(1) \end{pmatrix} \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$\Rightarrow \quad p_{n+1}(0) = (1-\alpha)\, p_n(0) + \beta\, p_n(1)$$

$$p_{n+1}(1) = \alpha\, p_n(0) + (1-\beta)\, p_n(1)$$

W4, h2

- Note: Physics notation

$$\begin{pmatrix} | \\ p_{n+1} \\ | \end{pmatrix} = \begin{pmatrix} \Pi \end{pmatrix}\begin{pmatrix} | \\ p_n \\ | \end{pmatrix} \qquad (-p_{n+1}-) = (-p_n-)\begin{pmatrix} \Pi \end{pmatrix}$$

Physics – column      Maths – row

## 2.4 Classification of states

- Persistent / recurrent / transient
- Periodic / aperiodic
- Reducible / irreducible

} Reading GS Sec. 6.2, 6.3

## 2.5 Ergodic Markov chains

- Stationary distribution : $p^* = p^* \Pi$
  - Eigenvector of eigenvalue 1
  - $p_0 = p^* \Rightarrow p_1 = p_0 \Pi = p^* \Pi = p^*$

    $p_2 = p_1 \Pi = \cdots = p^*$

    $\Rightarrow p_n = p^* \ \forall \ n$

    start at $p^*$ / stay at $p^*$
  - $\Pi$ can have many stationary distribution (degeneracy)

- Limiting distribution : $p_\infty = \lim\limits_{n\to\infty} p_n = \lim\limits_{n\to\infty} p_0 \Pi^n$
  - Might not exist
  - Can depend on choice of $p_0$
  - Interesting case: $p_\infty$ exists and independent of $p_0$

    $\Rightarrow \quad p_n \to p_\infty = p^*$   Ergodic

- Proposition: If $\{X_i\}_{i=0}^{n}$ is aperiodic and irreducible, then

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} p_0 \Pi^n = p^*$$

for all initial distribution $p_0$.

W4, h2
W5, h1

Example :

- $0 < \alpha, \beta < 1$ : $\bar{p}^* = \left( \frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right)$ unique (ergodic)

  $\bar{p}_0 \to \bar{p}^*$

- $\alpha = \beta = 0$ :

$\Pi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

  - $\bar{p}_0 \Pi = \bar{p}_0 \Rightarrow$ any $\bar{p}_0$ is a stationary distribution
  - Reducible Markov chain

- $\alpha = \beta = 1$ :

$\bar{p}_0 = (a, b)$

$\to \bar{p}_1 = (b, a)$

$\to \bar{p}_2 = (a, b)$

  - $\bar{p}^* = (\frac{1}{2}, \frac{1}{2})$ stationary
  - Periodic Markov $\Rightarrow$ not ergodic
  - Only $\bar{p}^* = \bar{p}_0$ stationary
  - Other $\bar{p}_0$'s $\not\to \bar{p}^*$

- Interpretation :
  - $p_0 \to p_n \to p^* \quad \forall \, p_0$
  - $p^*(i) =$ long time occupation in state $i$

    $=$ fraction of time spent in $i$

    $= \dfrac{\text{no. times } i \text{ is visited}}{n}$ as $n \to \infty$

- Ergodic theorem : If $X_0, X_1, \ldots, X_n$ is an ergodic Markov chain, then

$$\underbrace{\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(X_i)}_{\text{time average}} = \underbrace{E_{p^*}[g(X)]}_{\text{state average w/r } p^*} \quad \text{in probability}$$

  - Generalization of Law Large Numbers to Markov chains

    cf Week 2 notes
  - Empirical (time) distribution :

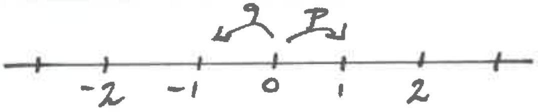
$$\frac{1}{n} \sum_{i=0}^{n} \delta_{X_i, j} \xrightarrow{n\to\infty} p^*(j) \quad \text{in prob.}$$

W5, h1

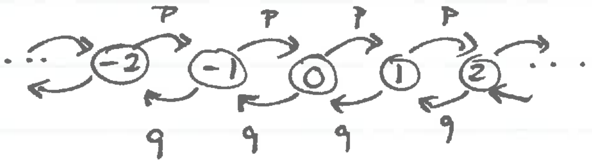
# 2.6 Applications

## 2.6.1. Random walk

GS P216



$p+q=1$



$X_0 \to X_1 \to \cdots \to X_n$

$X_i \in \mathbb{Z}$

$$\bar{p}_n = \begin{pmatrix} \vdots \\ p_n(-1) \\ p_n(0) \\ p_n(1) \\ \vdots \end{pmatrix} \qquad \Pi = \begin{pmatrix} \ddots & \ddots & \ddots & & & \\ & q & 0 & p & & 0 \\ & & q & 0 & p & \\ & 0 & & q & 0 & p \\ & & & \ddots & \ddots & \ddots \end{pmatrix} \leftarrow 0$$

- Chapman-Kolmogorov equations:

$$p_{n+1}(i) = p \quad p_n(i-1) + q \quad p_n(i+1) \qquad i \in \mathbb{Z}$$

position at time $n+1$

- Sum representation: $X_{n+1} = X_n + \xi_n$

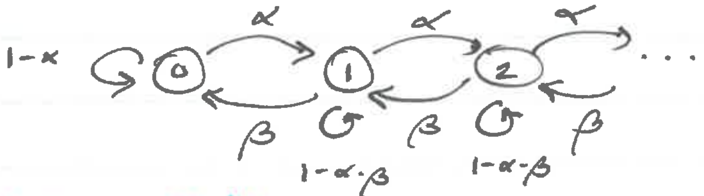
$\xi_n \in \{-1, 1\}$

$P(\xi_n = 1) = p$

$P(\xi_n = -1) = q = 1-p$

To be studied in Chap 4

## 2.6.2 Population / queuing model

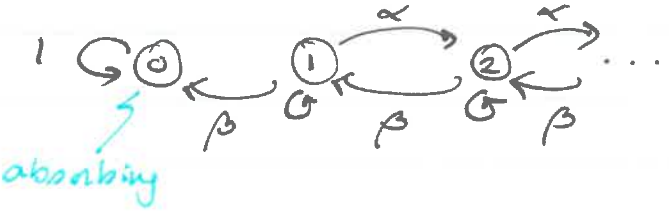


$P(i \to i+1) = \alpha$ births/arrivals

$P(i \to i-1) = \beta$ deaths/service

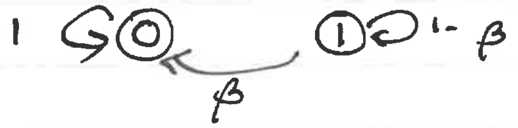
Studied in CW2.

- With absorbing state at 0:



- If $X_n = 0$ for some $n$, then $X_m = 0 \quad \forall m > n$

## 2.6.3 Absorbing Markov chain



$\beta > 0$

$\Pi = \begin{pmatrix} 1 & 0 \\ \beta & 1-\beta \end{pmatrix}$   $p^* = (1\ 0)$   Check: $(1\ 0)\begin{pmatrix} 1 & 0 \\ \beta & 1-\beta \end{pmatrix} = (1\ 0)$

unique

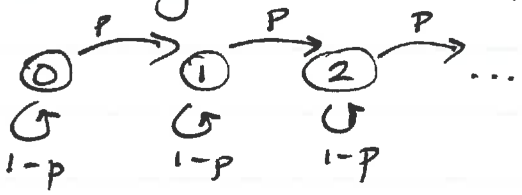
$1\ 1\ \underline{0}\ 0\ \cdots$   $\Rightarrow$ Possible sequences/trajectories:

absorb

$\underbrace{1\,1\cdots 1}_{n_1}\ \underbrace{0\cdots\cdots 0}_{n_0}$   $n_1 + n_0 = n$

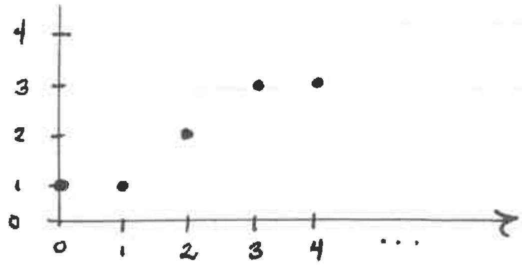
## 2.6.4 Counting or Bernoulli process

GS p218



$\Pi = \begin{pmatrix} 1-p & p & 0 & \cdots \\ 0 & 1-p & p & \cdots \\ \vdots & & \ddots & \ddots \end{pmatrix}$

- Birth/death process w/o deaths: $P(i \to i+1) = p$
  $P(i \to i-1) = 0$



$X_0 = 1$
$X_1 = 1$ no jump
$X_2 = 2$ 1 jump
$X_3 = 3$ 1 jump
$X_4 = 3$ no jump

- Sum representation:

$$X_{n+1} = X_n + \xi_n$$

$\xi_n \in \{0,1\}$
$P(\xi_n = 1) = p$
$P(\xi_n = 0) = 1-p$
Bernoulli RV

$\Rightarrow X_n = \#$ jumps from $X_0$
$= \#\ \xi_n = 1$'s " $X_0$

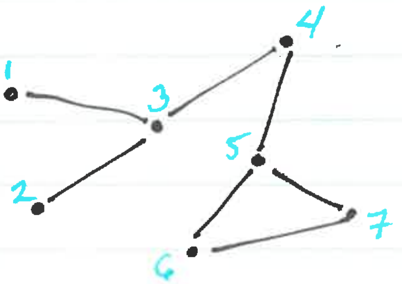
$$\Rightarrow P(X_n = j \mid X_0 = i) = \binom{n}{j-i} p^{j-i} (1-p)^{n-j+i}$$

n steps
j-i 1's
n-(j-i) 0's

Binomial distribution

WS, 1/12

## 2.6.5 Random walk on graphs



vertices/nodes

- Graph: $G = (V, E)$

edges

- Undirected
- Connected

- Adjacency matrix: $A_{ij} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{if } i \not\sim j \end{cases}$ connected / non-connected
  - $|V| \times |V|$ matrix
  - Symmetric: $A = A^T$ since $i \sim j \Leftrightarrow j \sim i$

- Node degree: $k_i = \#$ links to node $i$

$$= \sum_j A_{ij}$$

- Degree list: $\bar{k} = (k_1, k_2, \dots, k_{|V|})$
- Number of edges: $M = \frac{1}{2} \sum_i k_i = \frac{1}{2} \sum_{ij} A_{ij}$

- Uniform random walk (URW):
  1. Start at some node $X_0 = i$
  2. Choose node connected to $i$ (random, uniform)

     $k_i$ of them $\Rightarrow P(i \to j) = \frac{1}{k_i}$ for $j \sim i$
  3. Repeat

- Trajectory/path: $X_0 \to X_1 \to X_2 \to \dots \to X_n$
- Transition matrix:

$$\Pi_{ij} = P(i \to j) = \frac{A_{ij}}{k_i}$$

  - $0 \leq \Pi_{ij} \leq 1 \quad \forall\, i,j$
  - $\sum_j \Pi_{ij} = \frac{1}{k_i} \sum_j A_{ij} = \frac{k_i}{k_i} = 1 \quad \forall\, i$

Stochastic matrix

  - $\Pi$ ergodic if $G$ connected

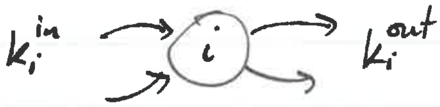
W5, h2

- Stationary distribution: $p_i^* = \frac{k_i}{2M}$ See CW2
  - Normalization: $\sum_i p_i^* = \frac{1}{2M} \sum_i k_i = \frac{2M}{2M} = 1 \quad \forall i$
  - Ergodic theorem:

$$\frac{1}{n} \sum_{i=0}^{n-1} \delta_{X_i, j} = \frac{\#\text{ visit node } j}{n} \xrightarrow{n \to \infty} p_j^*$$

  - Can use URW to estimate $k_i$ and $\bar{k}$ without actually counting # links!

- Application: Google Page Rank
  - World Wide Web WWW: Collection of web pages + in links and out links



  - Directed graph
  - Google ranking: Importance page i $\propto k_i^{in}$
  - Problems:
    - From i, don't know how many pages point to i
    - WWW graph too big to draw or build/store adjacency matrix
  - Solution: Construct random walk on WWW to estimate $k_i^{in}$

$\Rightarrow$ Visit pages at random to rank them.

cf Mathematica demonstration

WS, h2

# 2.7. Stationary distribution

· Ergodic Markov chain: $X_0 \xrightarrow{\Pi} X_1 \xrightarrow{\Pi} X_2 \xrightarrow{\Pi} \cdots \xrightarrow{\Pi} X_n$

$$\bar{p}_n = \bar{p}_0 \Pi \xrightarrow{n\to\infty} \bar{p}^* \quad \forall \bar{p}_0$$

$$\bar{p}^* = \bar{p}^* \Pi$$

· $p^*(i)$ = long term occupation in state $i$

· Methods to compute $\bar{p}^*$:

1- Linear algebra

- Input matrix $\Pi$ $\qquad v\Pi = v$ vs $\hat{\Pi} d = d$
- Output (left) eigenvector of eigenvalue 1
- Normalize properly

Matlab:

> pimat = [1 ½ ... ; ... ; ...]; $\qquad$ Don't call it pi

> [V,D] = eig(pimat.') $\qquad$ .' = transpose, ' = conjugate transpose

or eig(transpose(pimat))

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \end{pmatrix} \qquad V = \begin{pmatrix} | & | & \\ v_1 & v_2 & \cdots \\ | & | & \end{pmatrix}$$

$\lambda = 1$ $\qquad\qquad$ $\bar{p}^*$

Mathematica: Eigenvectors[pimat // Transpose]

or Eigenvectors[Transpose[pimat]]

or Eigensystem[...]

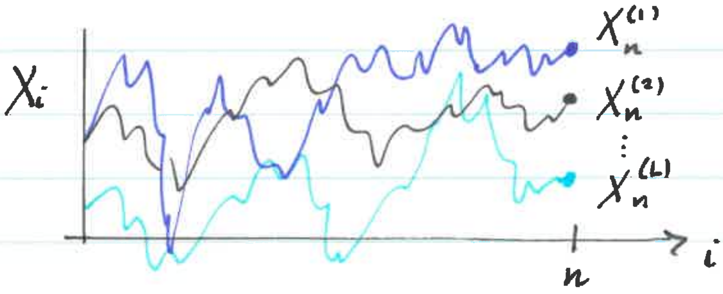
W6, L1

Python :

```
import scipy.linalg as la
   :
pimat = np.array([[a,b,c], [...],[...]])
eigvals, eigvecs = la.eig(np.transpose(pimat))
```

np also good

## 2 - Parallel simulations



- Simulate L copies/realizations/trajectories $\{X_i^{(j)}\}_{i=0}^{n}$ $j=1..L$ of the Markov chain

in parallel or in series

- Keep last state $X_n^{(j)}$ at time n in sample
- Histogram : $\hat{p}_{n,L}(i) \approx p_n(i)$
- Convergence : $\hat{p}_{n,L}(i) \xrightarrow[L\to\infty]{n\to\infty} p^*(i)$
- Parameters :
  - $n$ large
  - $L$ large
  - $X_0$ irrelevant to some extent

See CW2

Pseudo code :

```
n = 100
L = 10^5
xsample = []
for j=1:L
    x = ...                  initial values X0
    for i=1:n
        generate new x       according to Markov chain
    end
    add x to xsample
end
```

W6, h1

2 loops

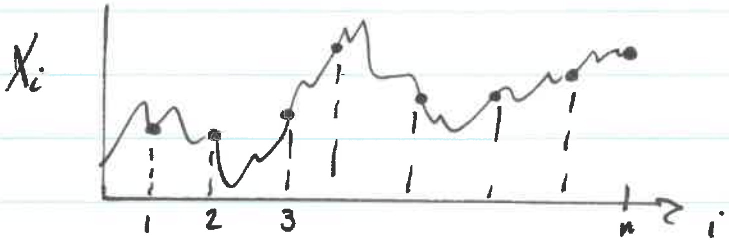
### 3- Ergodic simulation

- Simulate 1 long trajectory : $\{X_i\}_{i=0}^{n}$
- Time-averaged occupation :

$$\hat{p}_n(i) = \frac{1}{n}\sum_{j=0}^{n-1} \delta_{X_j, i}$$

- Convergence/ergodic theorem:

$$\hat{p}_n(i) \xrightarrow{n\to\infty} p^*(i)$$



- Parameters : n large
  $X_0$ irrelevant } see CW2

Pseudocode:

```
n = 1000
xsample=[]      initial state x_0
x= ...
for i=1:n
        generate new x
        add x to xsample
end
```

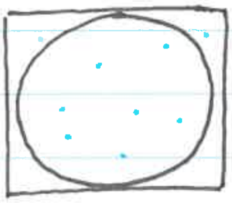
Only 1 loop!

# 2.8. Markov chain Monte Carlo (MCMC)

GS Sec 6.14

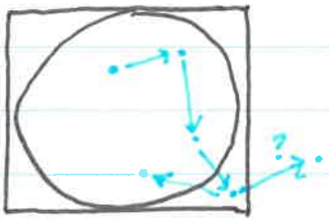
- Example: Estimation of $\pi$

Method 1



- iid sequence of point in square
  $X^{(1)}, X^{(2)}, \dots, X^{(L)}$
- Estimator:
  $$\hat{\pi}_L = \frac{4}{L} \sum_{j=1}^{L} \mathbb{1}(X^{(j)} \text{ in circle})$$
- Normal Monte Carlo

Method 2



- Markov chain in square
  $X_0 \to X_1 \to \cdots \to X_n$
- Estimator:
  $$\hat{\pi}_n = \frac{4}{n} \sum_{i=0}^{n-1} \mathbb{1}(X_i \text{ in circle})$$
- MCMC

- **Goal**: Generate variates (values, realizations) from distribution $f(x)$, $x \in \mathcal{X}$ $\sum_x f(x) = 1$
  target distribution

- Method 1: Transformation of RVs / inversion — see notes
  - Idea: Find "clever" way to generate variates of $f(x)$ from uniform RVs.
  - Problems:
    - Some $f(x)$ difficult to implement
    - Not efficient in high dimensions

- Method 2: MCMC
  - Idea: Generate Markov chain that has $f(x)$ as stationary distribution: $\bar{f} = \bar{f}\Pi$
  - Advantages:
    - Efficient in high dims
    - Many $\Pi$ possible (Gibbs sampler, Metropolis, ...)
    - Requires "less thinking" (black box method)

W6, h2

- Metropolis algorithm:
  - Target distribution: $f(x)$
  - Initial state: $X_0$
  - New state (move or try):

$$X_1 = X_0 + \underbrace{\delta X}_{\text{displacement/jump}} \qquad \delta X \sim q(\delta x) \ \underline{\text{symmetric}}$$

  - Acceptance probability:

$$P(X_0 \to X_1) = \min\left\{1, \frac{f(X_1)}{f(X_0)}\right\}$$

  - $X_1 = X_0$ if move not accepted

- Note: $\Pi_{ij} = P(i \to j) = \min\{1, f(j)/f(i)\}$ is reversible w/r $f$ and has $f$ as its stationary distribution. See CWL

- Pseudo code:

```
nsteps = 10^3
xsample = [ ]
x = ...                        initial value/state
for i = 1: n
        xp = x + random_symmetric_displacemt.        X' = X + δx
                 δx
        r = rand( )
        if   r < min (1, f(xp)/f(x))
                 x = xp                       accept move
        end
        add x to xsample
end
```

Note: No "else" when not accepting move

Example : Estimation of $\pi$ : See CW2

Example: Generate variates of $\mathcal{N}(0,1)$

- $f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$
- $x' = x + \delta x$ , $\delta x \sim \mathcal{U}[-1,1]$

can also use other symmetric dist.

- $P(x \to x') = \min\left(1, \frac{f(x')}{f(x)}\right)$

$$\frac{f(x')}{f(x)} = \frac{1}{\sqrt{2\pi}}\, e^{-x'^2/2}\, \frac{\sqrt{2\pi}}{e^{-x^2/2}} = e^{-\frac{1}{2}(x'^2 - x^2)}$$

See demonstration

- Remarks:
  - Can we use $\delta x \sim \mathcal{U}[0,1]$ ?
  - Which distribution to use for $\delta x$ ?
  - Optimal displacement distribution ?
  - Acceptance rate

- Comparison :

| Method 1: Std MC | Method 2 : MCMC |
|---|---|
| iid variates, sample | MC variates |
| No correlation | Correlated samples |
| Not efficient in high dim | Efficient in high dim |
| "Clever" student needed ! | Easy / general implementation |