CS145: Probability & Computing

Lecture 17: Markov Chains, Multi-step Transition Distributions

Brown University Computer Science

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

CS145: Lecture 17 Outline

- Discrete Time Markov Chains
- Examples of Markov Chains
- Multi-step State Transitions

Andrey Markov

Discrete Time, Finite Stochastic Processes

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \quad \cdots$$

- A *discrete time stochastic process* associates a random variable X_t with a sequence of "time" locations: $t = 0, 1, 2, \ldots$
- We will assume that X_t is a discrete random variable with a *finite number of possible outcomes*. (Can be generalized.)
- Sometimes we call X_t the *state* of the process at time t.
- These random variables have a joint distribution:

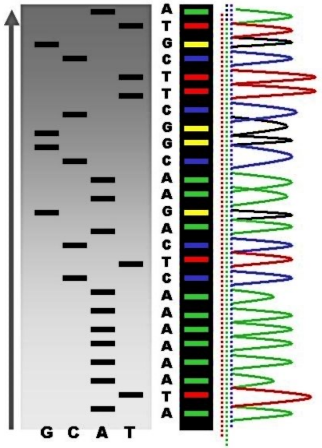
$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$$

In most real applications, time points are *not* independent.

What is Discrete Time?

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \quad \cdots$$

- A *regular sampling of real times* in the world:
  One variable every second, hour, day, year, or …
- The *steps taken by some computational process*:
  One variable for every "iteration" of some algorithm.
- Any other data with "sequential" structure.
  In computational biology, we may model
  genetic sequences (DNA, proteins, …)



Wikipedia

Probability Distributions and Sequences

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \quad \cdots$$

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$$

- Suppose $X_0=1$ is fixed, and for times $t = 1,\ldots,n$:

$$X_t \in \{1, 2, \ldots, m\}$$

- We assign probability to each possible *discrete sequence*:

There are m^n possible sequences of length n.

- Expensive to enumerate these probabilities in a table!

- Often not needed.

Finite Markov Chains

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \quad \cdots$$

- **Markov assumption:** The probability of the next state depends only on the current state, and not the sequence of steps taken to reach the current state:

$$P(X_{t+1} = j \mid X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i_t)$$

- We define a Markov chain via a *state transition matrix*:

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$$X_t \in \{1, \ldots, m\} \qquad \sum_{j=1}^{m} p_{ij} = 1$$

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Joint Distribution of Markov Sequences

- **Joint Distribution:** The Markov assumption implies that

$$P(X_0, X_1, \ldots, X_n) = P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_0) \qquad \textit{(any process)}$$

$$= P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1}) \qquad \textit{(Markov process)}$$

- **Initial state:** From some (possibly degenerate) distribution

$$P(X_0)$$

- **State transition matrix:**

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$$X_t \in \{1, \ldots, m\} \qquad \sum_{j=1}^{m} p_{ij} = 1$$

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

State Transition Diagrams

State transition diagram:

- A *directed graph* with one node for each of *m* possible states
- Draw an edge from node *i* to node *j* if $p_{ij} > 0$
- A sample from a Markov process is then a *record of nodes visited in a random walk in this graph*

State transition matrix:

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$$X_t \in \{1, \ldots, m\} \qquad \sum_{j=1}^{m} p_{ij} = 1$$

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

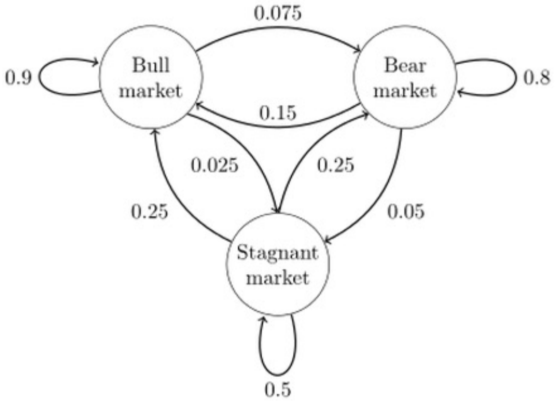
State Transition Diagrams

State transition diagram:

- A *directed graph* with one node for each of *m* possible states
- Draw an edge from node *i* to node *j* if $p_{ij} > 0$
- A sample from a Markov process is then a *record of nodes visited in a random walk in this graph*

Example: *Bull and Bear Markets (Wikipedia)*

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$



Markov Property and Independence

$$P(X_0, X_1, \ldots, X_n) = P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1})$$

- For most choices of the state transition matrix, the states at different times are *not* independent. *This is useful!*

$$p_{X_s X_t}(x_s, x_t) \neq p_{X_s}(x_s) p_{X_t}(x_t)$$

- But conditioned on the value of the present state, *the past and future of a Markov process are independent*:

$$Y_t = \{X_0, X_1, \ldots, X_{t-1}\} \qquad Z_t = \{X_{t+1}, X_{t+2}, \ldots, X_n\}$$

$$p_{Y_t Z_t}(y_t, z_t \mid X_t = x_t) = p_{Y_t}(y_t \mid X_t = x_t) p_{Z_t}(z_t \mid X_t = x_t)$$

CS145: Lecture 17 Outline

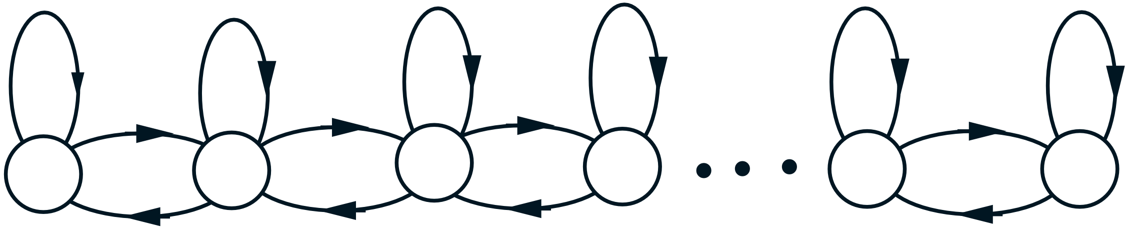
- Discrete Time Markov Chains
- Examples of Markov Chains
- Multi-step State Transitions

Machine Repair

- Let X_i be the number of days (including today) some machine has been broken, or $X_i=0$ if the machine is currently working.
- A machine that is working today will be broken tomorrow with probability p. Otherwise, with probability $1-p$, it keeps working.
- On any given day, a broken machine is repaired with probability r. Otherwise, with probability $1-r$, it remains broken.
- After being broken for m days, it is always replaced with a working machine.

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

A Checkout Line (Queue)



Let X_i be the number of customers in a *line (queue)* during (short) time period i, where exactly one event happens:

- With probability s, a customer is served and leaves the queue, unless already there are no customers ($X_i=0$).
- With probability r, a new customer joins the queue, unless the queue is already at its maximum capacity m.
- Otherwise, the number of customers remains the same ($X_{i+1}=X_i$).

This is a discrete Markov process with $m+1$ states.

English Text: Are characters independent?

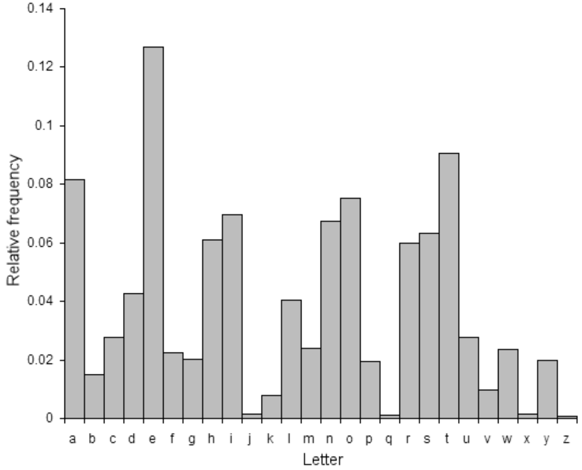
- Assume letters are independent and equally common:

$$P(X_t) = \frac{1}{27} \text{ for } X_t \in \{a, b, c, ..., z, \_\}$$

uzlpcbizdmddk njsdzyyvfgxbgjjgbtsak rqvpgnsbyputvqqdtmgltz ynqotqigexjumqphujcfwn ll jiexpyqzgsdllgcoluphl sefsrvqqytjakmav bfusvirsjl wprwqt

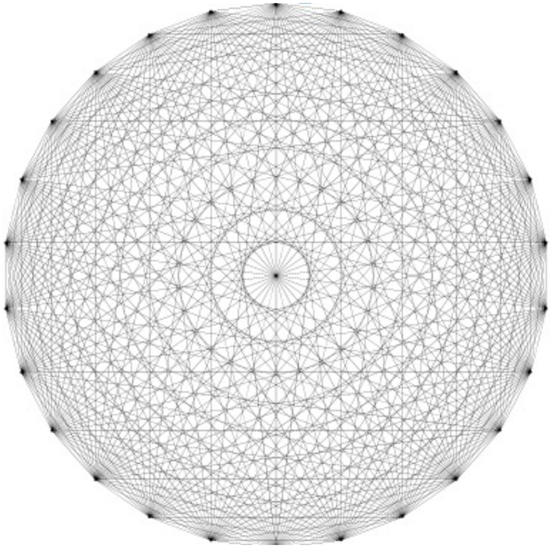
- Assume letters are independent and follow frequencies of real text:

saade ve mw hc n entt da k eethetocusosselalwo gx fgrsnoh,tvettaf aetnlbilo fc lhd okleutsndyeoshtbogo eet ib nheaoopefni ngent



English Text: Markov Models

- A first-order Markov model encodes probability of each letter, given previous letter
- A second-order (bigram) Markov model encodes probability of each letter, given previous two letters (state is letter pairs)
- A third order (trigram) Markov model encodes probability of each letter, given previous three letters (state is letter triples)



English Text: Markov Models

Examples from Programming Pearls, J. Bentley, Sec. 15.3

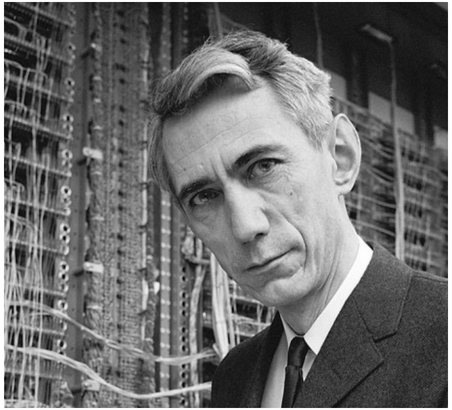
Order-1: t I amy, vin. id wht omanly heay atuss n macon aresethe hired boutwhe t, tl, ad torurest t plur I wit hengamind tarer-plarody thishand.

Order-2: Ther I the heingoind of-pleat, blur it dwere wing waske hat trooss. Yout lar on wassing, an sit." "Yould," "I that vide was nots ther.

Order-3: I has them the saw the secorrow. And wintails on my my ent, thinks, fore voyager lanated the been elsed helder was of him a very free bottlemarkable,

Order-4: His heard." "Exactly he very glad trouble, and by Hopkins! That it on of the who difficentralia. He rushed likely?" "Blood night that.

Claude Shannon's Markov chain simulator (1948): *To construct [an order 1 model] for example, one opens a book at random and selects a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter recorded, etc. It would be interesting if further approximations could be constructed, but the labor involved becomes enormous at the next stage.*



CS145: Lecture 17 Outline

- Discrete Time Markov Chains
- Examples of Markov Chains
- Multi-step State Transitions

Finite Markov Chains

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \quad \cdots$$

Finite State: $X_t \in \{1, \ldots, m\}$

Markov Property: *Given the current state, the past & future are independent.*

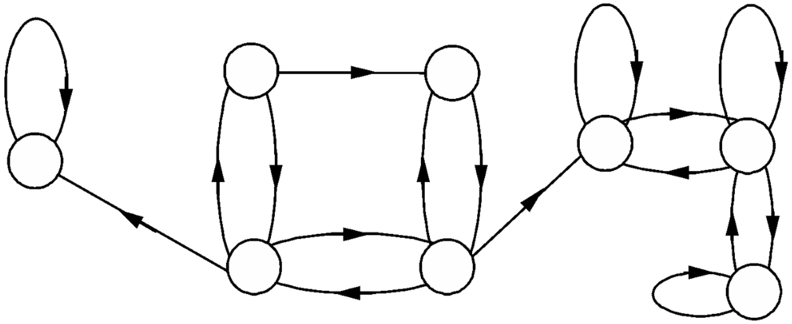
$$P(X_0, X_1, \ldots, X_n) = P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1})$$

State Transition Matrix:

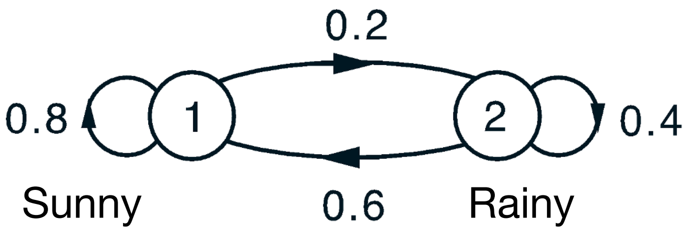
$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

State Transition Diagram:



Example: Sunny or Rainy?



	S	R
S	0.8	0.2
R	0.6	0.4

$$P$$

Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

$$r_{ij}(n) = \mathbf{P}(X_n = j \,|\, X_0 = i) \qquad \text{where } r_{ij}(1) = p_{ij}$$

State transition matrix:

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$$X_t \in \{1, \ldots, m\} \qquad \sum_{j=1}^{m} p_{ij} = 1$$

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

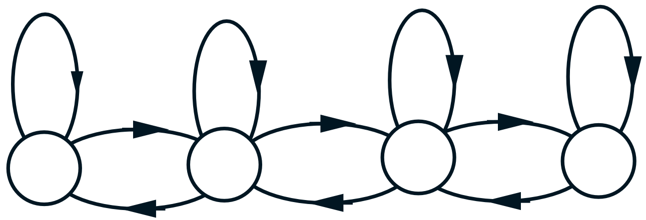
$$r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i) \qquad \text{where } r_{ij}(1) = p_{ij}$$

- First consider the special case where *n=2*:

$$P(X_2 = j \mid X_0 = i) = \sum_{k=1}^{m} P(X_2 = j, X_1 = k \mid X_0 = i)$$

$$P(X_2 = j \mid X_0 = i) = \sum_{k=1}^{m} P(X_2 = j \mid X_1 = k) P(X_1 = k \mid X_0 = i)$$

$$r_{ij}(2) = \sum_{k=1}^{m} p_{ik} p_{kj} = \sum_{k=1}^{m} r_{ik}(1) p_{kj}$$



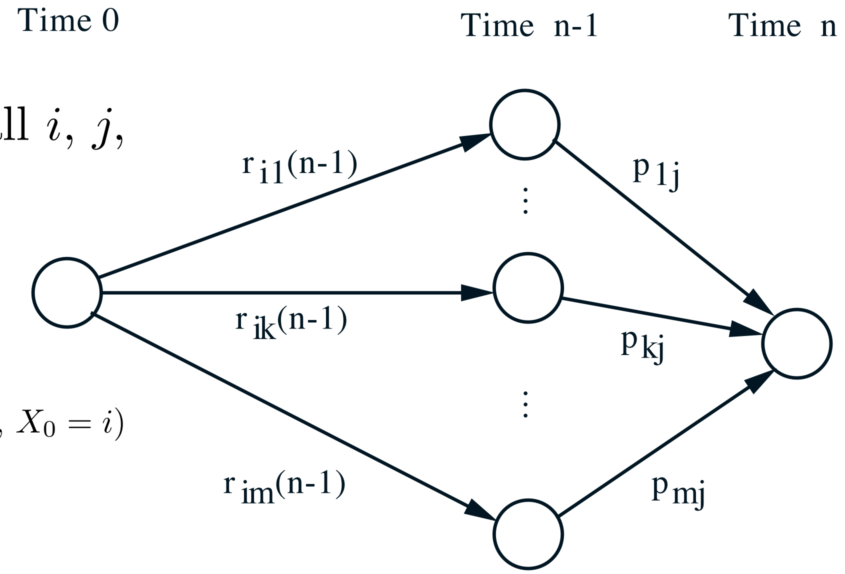
Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

$$r_{ij}(n) = \mathbf{P}(X_n = j \,|\, X_0 = i) \qquad \text{where } r_{ij}(1) = p_{ij}$$

- Computed recursively via the *Chapman-Kolmogorov equation*:

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj}, \qquad \text{for } n > 1, \quad \text{and all } i,\ j,$$



$$\mathbf{P}(X_n = j \,|\, X_0 = i) = \sum_{k=1}^{m} \mathbf{P}(X_{n-1} = k \,|\, X_0 = i) \cdot \mathbf{P}(X_n = j \,|\, X_{n-1} = k,\, X_0 = i)$$

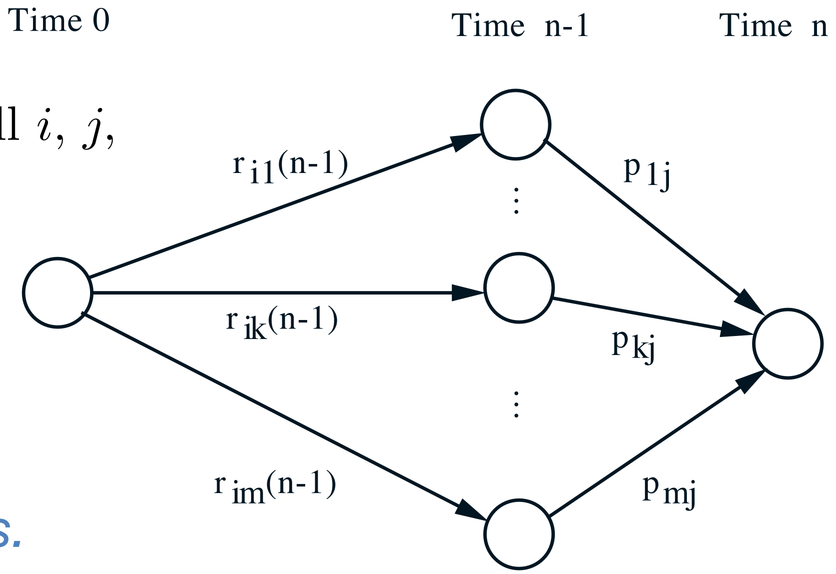
Multi-step State Transitions

- Given the current state, we would like to predict what state we will be in at multiple steps into the future:

$$r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i) \qquad \text{where } r_{ij}(1) = p_{ij}$$

- Computed recursively via the *Chapman-Kolmogorov equation*:

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj}, \qquad \text{for } n > 1, \quad \text{and all } i,\ j,$$



- With random initial state:

$$\mathbf{P}(X_n = j) = \sum_{i=1}^{m} \mathbf{P}(X_0 = i) r_{ij}(n)$$

Marginal distribution of state after n steps.

Reminder: Matrix Multiplication

$$y = Ax = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i .$$

$$y^T = x^T A = x^T \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x^T a_1 & x^T a_2 & \cdots & x^T a_n \end{bmatrix}$$

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

Zico Kolter, 2012

State Transitions & Matrix Multiplication

$$\pi_{ti} = P(X_t = i) \qquad p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$$\pi_{1j} = \sum_{i=1}^{m} p_{ij}\pi_{0i}$$

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$\pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T$$

Textbook convention:

$$\pi_1^T = \pi_0^T P$$

Each row of P sums to one.

Alternative convention:

$$\pi_1 = P^T \pi_0$$

Each column of P^T sums to one.

Multi-Step State Transitions

$$\pi_{ti} = P(X_t = i) \qquad p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$$\pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T \qquad P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

State Distribution after *n* time steps:

$$\pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n$$

$$\pi_n = P^T \pi_{n-1} = P^T P^T \pi_{n-2}^T = (P^n)^T \pi_0$$

P^n multiplies the square matrix P by itself n times.

This is not equivalent to raising the entries of P to the power n.