CHI
CES Engineering

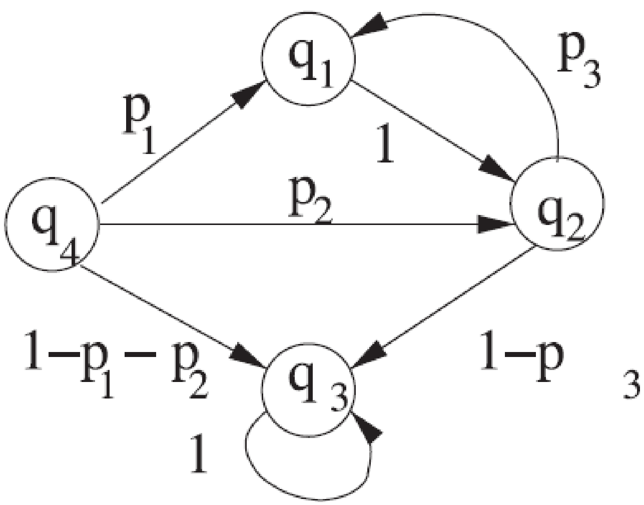
SIMULATION SYSTEMS

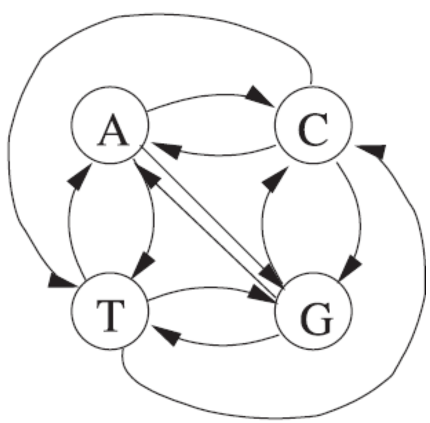
MARKOV MODELS

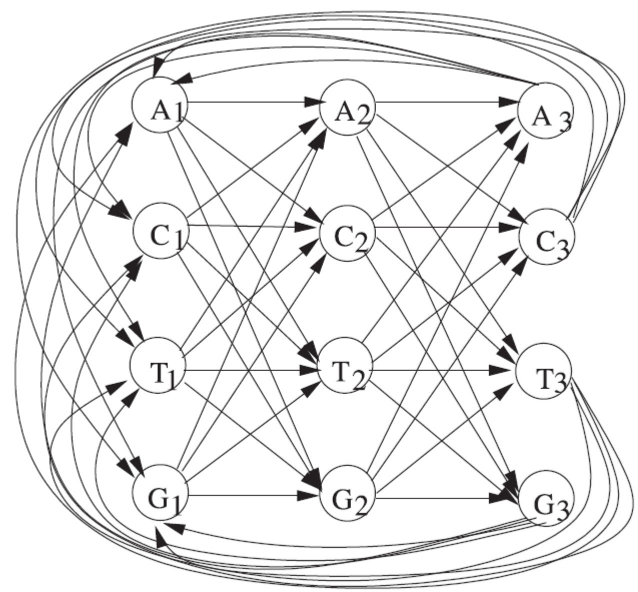
Prof. Yasser Mostafa Kadah

Definition

- A Markov model is generally represented as a graph containing a set of states represented as nodes and a set of transitions with probabilities represented by weighted edges.

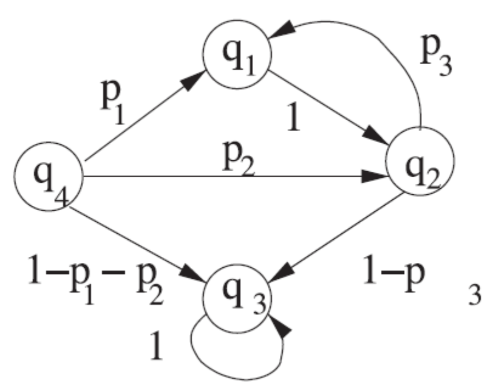






Simulation of Markov Models

- We simulate a Markov model by starting at some state and moving to successive neighboring states by choosing randomly among neighbors according to their labeled probabilities.
  - For example, if we start in state q_4, then we would have probability p_1 of moving to q_1, p_2 of moving to q_2, and $(1-p_1-p_2)$ of moving to q_3. If we move to q_2, then we have probability p_3 of moving to q_1 and $(1-p_3)$ of moving to q_3, and so on. The result is a walk through the state set (e.g., q1; q2; q1; q2; q3; q3; . . .).
  - Resulting sequence of states is called a "**Markov chain**"



Markov Model Components

- A state set $Q=\{q_1; q_2; \ldots ; q_n\}$
- A starting distribution $Pr\{q(0)=q_i\}= p_i$
  - Represented by a vector $\vec{p}$
- A set of transition probabilities:

$$Pr\{q(n+1)=q_j \mid q(n)=q_i\}= p_{ij}$$

  - Represented by a matrix **P**

This is the definition of the First Order Markov Model: probability of entering each possible next state dependent only on the current state

Higher Order Markov Models

- k^{th} Order Markov Model:

$$\Pr\{q(n)= q_{i,n} \mid q(n-1)= q_{i,(n-1)}, q(n-2)= q_{i,(n-2)}, \dots, q(n-k)=q_{i,(n-k)}\}= p_{i,j}$$

  - Probability of next state depends on previous k states
- Note: Any k^{th}-order Markov model can be transformed into a first order Markov model by defining a new state set $Q'=Q^k$ (i.e., each state in Q' is a set of k states in Q), with current state in Q' being the last k states visited in Q.
  - Then a Markov chain in the k^{th}-order model Q—$q_1; q_2; q_3; q_4; \dots$—becomes the chain $\{q_1; q_2; \dots; q_k\}; \{q_2; q_3; \dots; q_k\}; \{q_3; q_4; \dots; q_k\}; \dots$ in Q'
  - Ignore higher-order Markov models when talking about theory

Time Evolution of Markov Models

- Although the behavior of Markov models is random, it is also in some ways predictable
- Suppose we have a two-state model: Q={q1; q2}, with initial probabilities p_1 and p_2 and transition probabilities p_{11}, p_{12}, p_{21}, and p_{22}
  - Step 0: $$\begin{bmatrix} Pr\{q(0)=q_1\} \\ Pr\{q(0)=q_2\} \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$
  - After 1 step: $$\begin{bmatrix} Pr\{q(1)=q_1\} \\ Pr\{q(1)=q_2\} \end{bmatrix} = \begin{bmatrix} p_1 p_{11} + p_2 p_{21} \\ p_1 p_{12} + p_2 p_{22} \end{bmatrix}$$

$$\begin{bmatrix} Pr\{q(2)=q_1\} \\ Pr\{q(2)=q_2\} \end{bmatrix} = \begin{bmatrix} (p_1 p_{11} + p_2 p_{21}) p_{11} + (p_1 p_{12} + p_2 p_{22}) p_{21} \\ (p_1 p_{11} + p_2 p_{21}) p_{12} + (p_1 p_{12} + p_2 p_{22}) p_{22} \end{bmatrix}$$

Time Evolution of Markov Models

- Matrix Notation:

$$\begin{bmatrix} Pr\{q(i+1)=q_1\} \\ Pr\{q(i+1)=q_2\} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \begin{bmatrix} Pr\{q(i)=q_1\} \\ Pr\{q(i)=q_2\} \end{bmatrix}$$

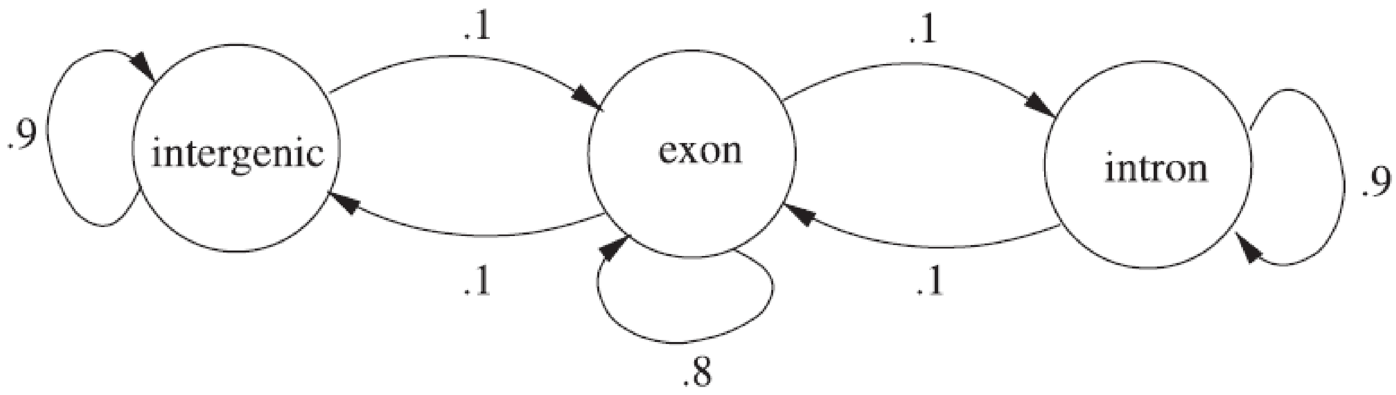
- Distribution after n steps:

$$\begin{bmatrix} Pr\{q(n)=q_1\} \\ Pr\{q(n)=q_2\} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \times \cdots \times \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$

Time Evolution: Example

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

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$$



Time 0 $\begin{bmatrix} .2 \\ .3 \\ .5 \end{bmatrix}$

Time 1

$$P = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .9 & 0 \\ .1 & 0 & .9 \end{bmatrix} \begin{bmatrix} .2 \\ .3 \\ .5 \end{bmatrix} = \begin{bmatrix} .24 \\ .29 \\ .47 \end{bmatrix}$$

Time 2

$$P = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .9 & 0 \\ .1 & 0 & .9 \end{bmatrix} \begin{bmatrix} .24 \\ .29 \\ .47 \end{bmatrix} = \begin{bmatrix} .268 \\ .285 \\ .447 \end{bmatrix}$$

Time n

$$P = \begin{bmatrix} .8 & .1 & .1 \\ .1 & .9 & 0 \\ .1 & 0 & .9 \end{bmatrix}^n \begin{bmatrix} .2 \\ .3 \\ .5 \end{bmatrix}$$

Chapman–Kolmogorov Equations

- Generalization of how the distribution of states of a Markov model evolves over time
- Suppose we have a Markov model with |Q| states where we define $p_{ij}(n)$ to be the probability of going from state i to state j in exactly n steps

$$p_{ij}(n+m) = \sum_{k=1}^{|Q|} p_{ik}(n)\, p_{kj}(m)$$

for all $n \geq 0$, $m \geq 0$, and any states i and j.

That is, the probability of getting to state j from state i in (n+m) steps is the sum over all possible intermediate states k of the probability of getting from i to k in n steps, then from k to j in the remaining m steps

Stationary Distributions

- Look at the evolution of Markov model over really long time scale for previous example:

$$\begin{bmatrix} .2 \\ .3 \\ .5 \end{bmatrix} \rightarrow \begin{bmatrix} .24 \\ .29 \\ .47 \end{bmatrix} \rightarrow \begin{bmatrix} .268 \\ .285 \\ .447 \end{bmatrix} \rightarrow \cdots \rightarrow \begin{bmatrix} .33333 \\ .33333 \\ .33333 \end{bmatrix} \rightarrow \begin{bmatrix} .33333 \\ .33333 \\ .33333 \end{bmatrix}$$

  - Convergence on a single probability distribution that will not change on further multiplication
  - Always converge to the same final distribution vector, regardless of our starting point (initial distribution)

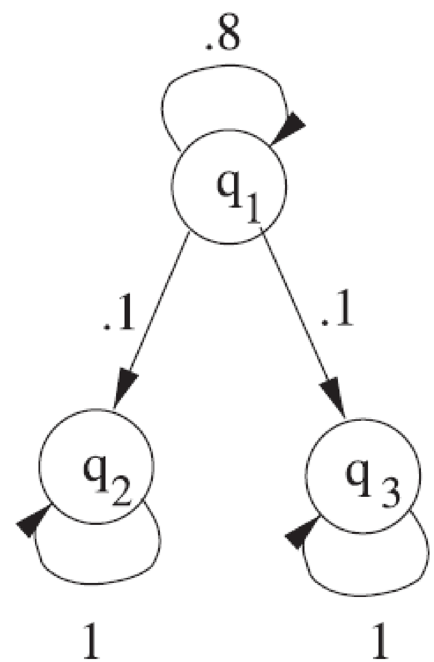
This vector on which the state distribution converges after a large number of steps is called the stationary distribution

Stationary Distributions

- Will this property of convergence on a unique stationary distribution regardless of starting point work for any example?
  - Answer is NO. It is possible that final vector is not unique!
- Example:

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

Start		Final
$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$	$\Rightarrow$	$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$
$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\Rightarrow$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$



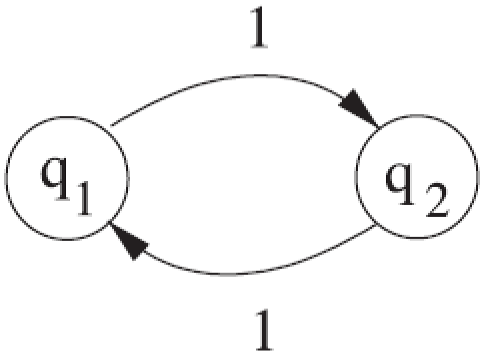
Stationary Distributions

- A Markov model is not even guaranteed to converge on any vector
- Example:

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

Initial Probability

$\begin{bmatrix} p \\ 1-p \end{bmatrix}$



$$\begin{bmatrix} p \\ 1-p \end{bmatrix} \rightarrow \begin{bmatrix} 1-p \\ p \end{bmatrix} \rightarrow \begin{bmatrix} p \\ 1-p \end{bmatrix} \rightarrow \begin{bmatrix} 1-p \\ p \end{bmatrix} \rightarrow \begin{bmatrix} p \\ 1-p \end{bmatrix} \rightarrow \cdots$$

Ergodicity

- Ergodicity means that for any two states q_i and q_j there is some sequence of transitions with nonzero probability that go from q_i to q_j
  - Ergodic Markov chain is also sometimes called irreducible
- Example1 Example 2 Example 3

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

Ergodic ⇩ Unique Stationary Distribution

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

Not Ergodic ⇩ No Unique Stationary Distribution

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

Not Ergodic ⇩ No Stationary Distribution

Eigenvalues and Stationary Distribution

- Markov model will converge to a unique stationary distribution if its transition matrix has exactly one eigenvector with eigenvalue $\lambda_1=1$ and has $|\lambda_i| < 1$ for every other eigenvector
  - Similar to Power Method of computing maximum eigenvalue and its corresponding eigenvector
  - Converges to this eigenvector after all eigenvalues die out after k-iterations: $\lambda_i^k = 1$ (i=1) or 0 (otherwise)

Eigenvalues and Stationary Distribution

- If a Markov model is not ergodic, then its state set can be partitioned into discrete graph components unreachable from one another
  - Each such component will have its own eigenvector with eigenvalue 1
  - Depending on which component we start in, we may converge on any of them
- Example: Nonergodic Markov model 2

$$\begin{bmatrix} .8 & 0 & 0 \\ .1 & 1 & 0 \\ .1 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Are both eigenvector with eigenvalue=1

Test of Markov Model Convergence

- A Markov model is guaranteed to converge on a stationary distribution if there exists some integer N > 0 such that,

$$\min_{i,j} \; p_{ij}(N) = \delta, \quad \delta > 0.$$

  - That is, there is some number of steps N such that no matter where we start, we have some bounded nonzero probability of getting to any given ending position in exactly N steps.

Mixing Time

- Informally, the mixing time is the time needed for the Markov model to get close to its stationary distribution
- if we want to run the model long enough for the transients to die away by some factor r, then we need to run for a number of rounds k such that,
  - Assume λ_1=1 and $|\lambda_i| < 1$, $i \neq 1$

$$|\lambda_2|^k = r$$

$$k \log \lambda_2 = \log r$$

$$k = \frac{\log r}{\log \lambda_2}.$$

Assignments

- For each of the following models:
  - Determine whether the following Markov models have stationary distributions
  - Estimate stationary distributions (if available)
  - Compare Stationary distributions to eigenvector corresponding to maximum eigenvalue (if available)
  - Estimate mixing time for transients to die out by a factor of 1/100

$$P_1 = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0 & 0.6 & 0.2 \\ 0 & 0.2 & 0.7 \end{bmatrix} \qquad P_3 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}$$

Assume initial state of q_1 for all models

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.8 & 0.3 \\ 0 & 0.2 & 0.7 \end{bmatrix} \qquad P_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$