Outline

- Graphical models (continued from last time)
  - Belief propagation
  - Undirected graphs and factor graphs
- Markov models
- Hidden Markov models

CHAPTER 14:

GRAPHİCAL MODELS – PART 2

Last time

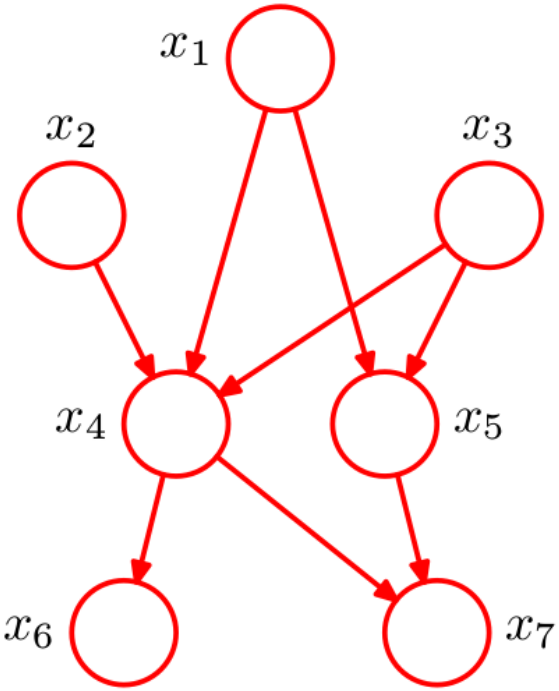
- Joint distributions over random variables
- general expression:

$$p(X_1, X_2, \ldots, X_d) = p(X_1)\, p(X_2|X_1)\, p(X_3|X_1, X_2)\, \cdots\, p(X_d|X_1, X_2, \ldots, X_{d-1})$$

… complex calculations

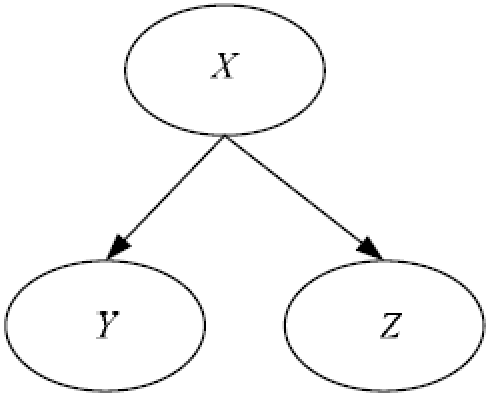
Last time

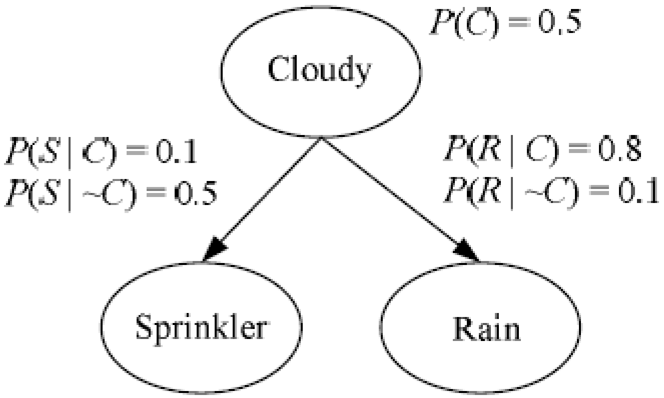
- Graphical models:
  - depict joint distributions
  - conditional independence assumptions
  - simplify distributions



Last time

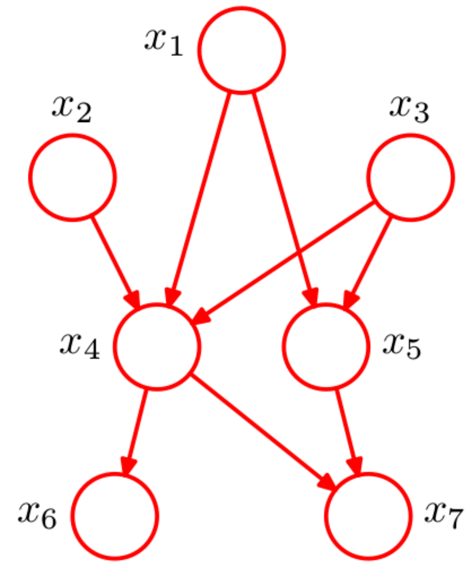
- Probability calculations:
  - Various examples using Bayes' rule





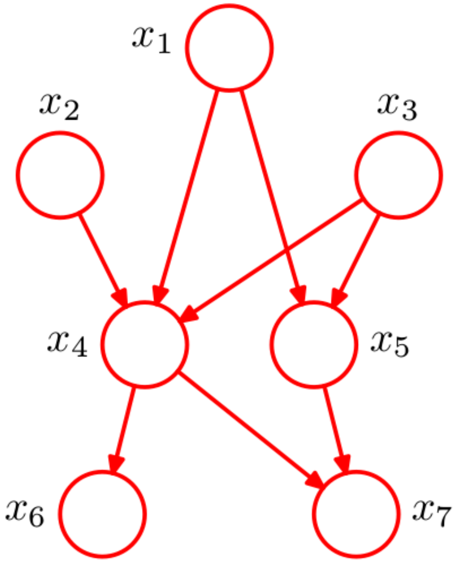
Belief Propagation

- We want a *general algorithm* for finding P(*some variable* | *some evidence*) in a graphical model
- No designated input and output nodes required



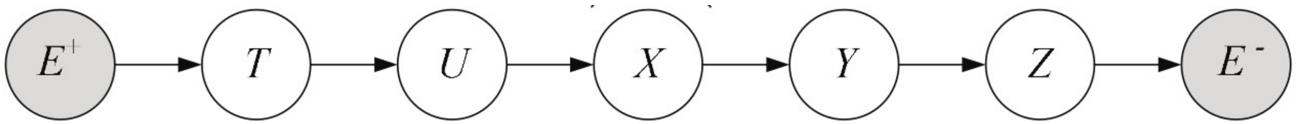
Belief Propagation

- We want a *general algorithm* for finding P(*some variable* | *some evidence*) in a graphical model
- No designated input and output nodes required
- Make use of the structure of the graph
- Different subgraphs imply different calculations



Belief Propagation

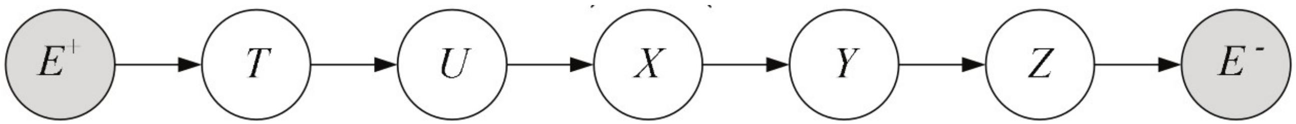
- Simplest case: Chain



- We want to find P(X|E)
  - X: node of interest
  - E: evidence nodes

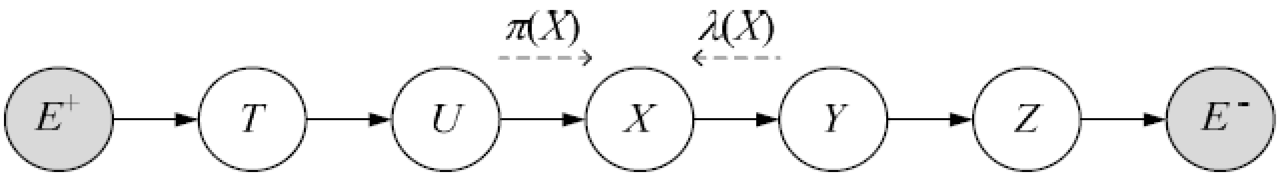
Belief Propagation

- Simplest case: Chain



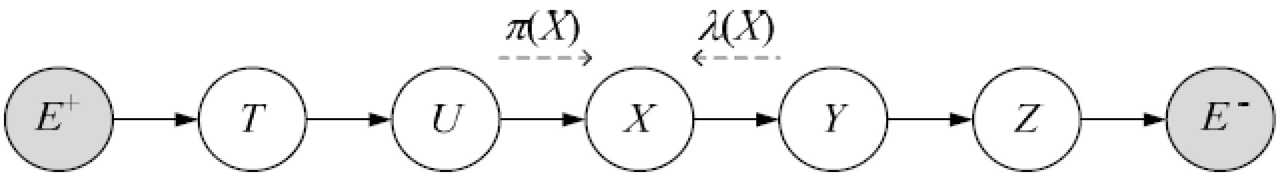
- Every (non-leaf or -root) node splits chain into ancestors and descendants
- Every evidence node blocks any evidence further up/down

Belief Propagation



- Evidence can be thought of as a message being sent up/down the chain
- Nodes:
  - Receive msg from parent/child
  - Processes the msg
  - Sends it onwards to child/parent

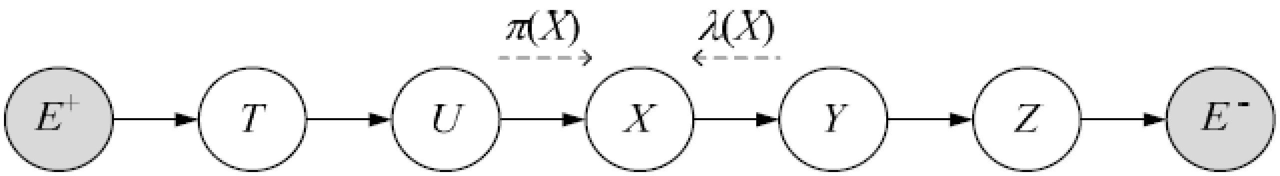
Belief Propagation



- Every node stores:
  - evidence via parent: $\pi(X) \equiv P(X|E^+)$
  - evidence via child: $\lambda(X) \equiv P(E^-|X)$

$$P(X|E) = \alpha\pi(X)\lambda(X)$$

Belief Propagation



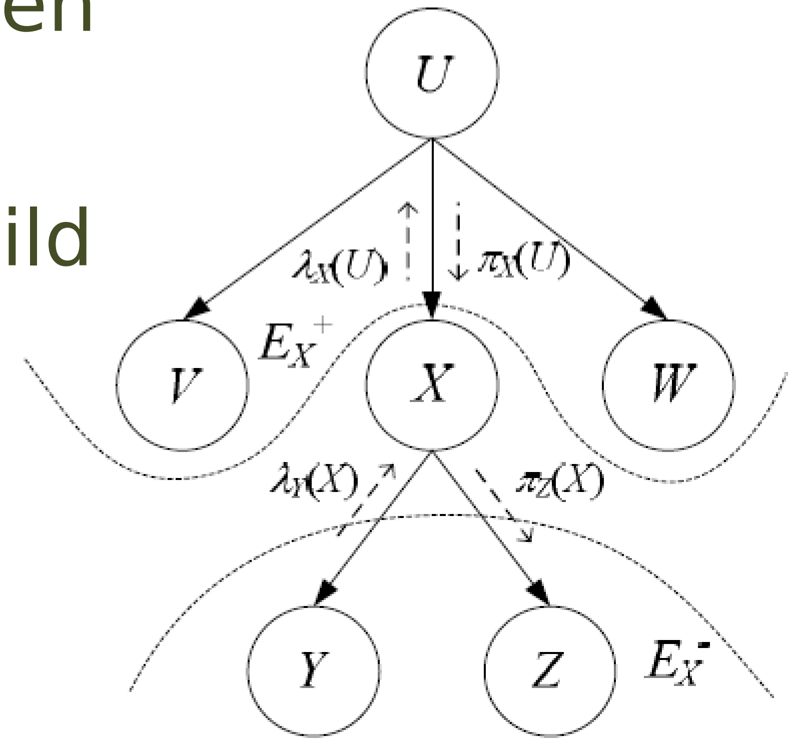
- Evidence is propagated recursively

$$\pi(X) = \sum_{U} P(X|U)\pi(U)$$

$$\lambda(X) = \sum_{Y} P(Y|X)\lambda(Y)$$

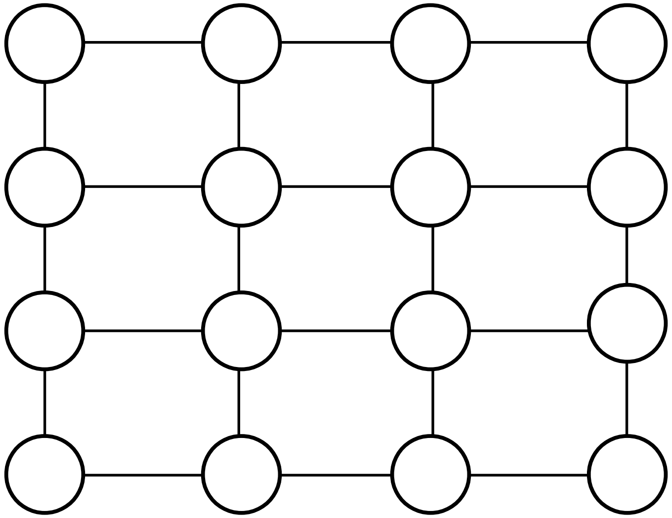
Trees

- One parent, multiple children
- Same logic as for chains
  - different λ from each child
  - Different π to each child



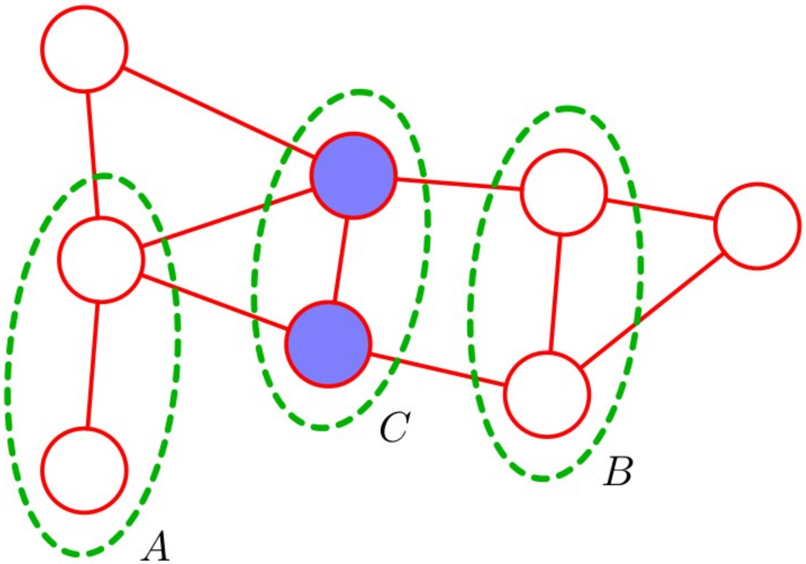
Undirected Graphs: Markov Random Fields

- So far we looked at directed models
- For symmetric interactions, undirected models can be more convenient.
- E.g. a model of pixels in an image



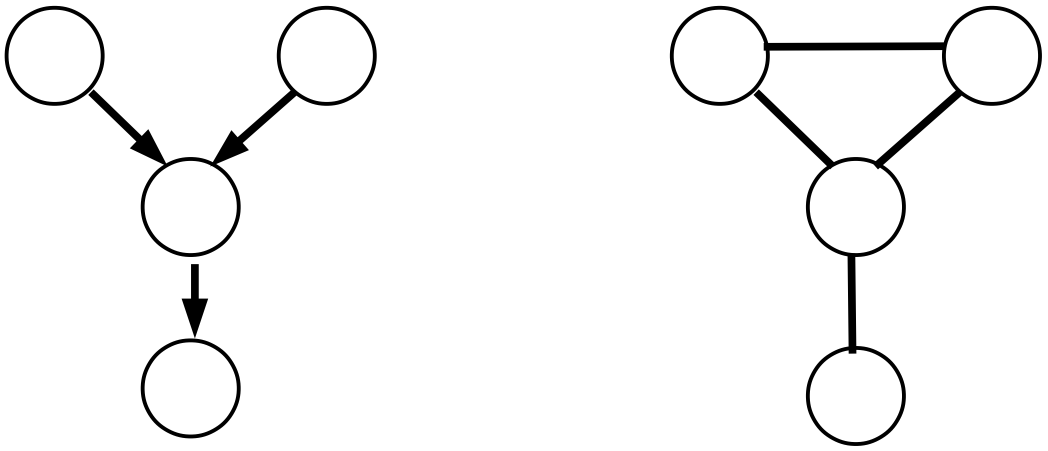
Undirected Graphs: Markov Random Fields

- *A* and *B* are conditionally independent if removing *C* makes them unconnected.



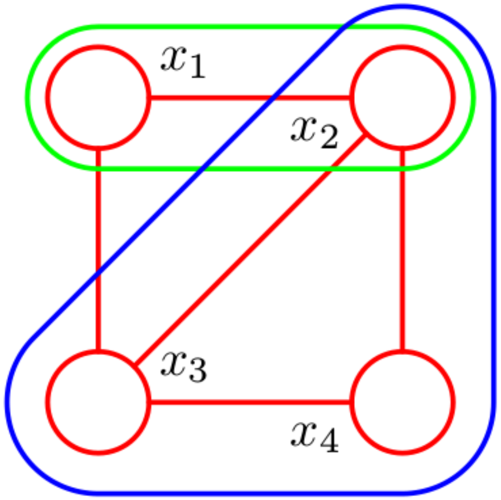
From directed to undirected

- Drop directions of edges
- *Marry* the parents of nodes (*moralization*)



Undirected Graphs: Markov Random Fields

- Instead of parent/child:
  - Cliques
  - Maximal cliques



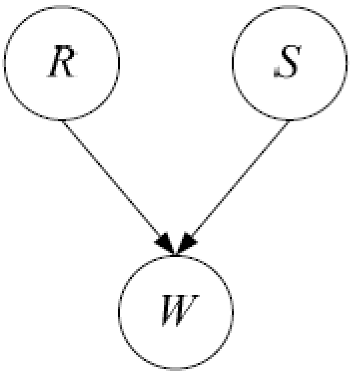
Undirected Graphs: Markov Random Fields

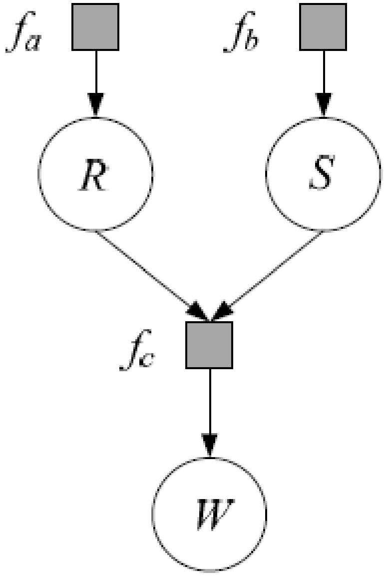
- Instead of conditional probabilities:
- Potential function $\psi_c(X_c)$
  - Function of variables *X* in the clique *C*
- The joint prob. is defined in terms of the max clique potentials

$$p(X) = \frac{1}{Z} \prod_C \psi_C(X_C)$$

Factor Graphs

- Define new factor nodes and write the joint in terms of them



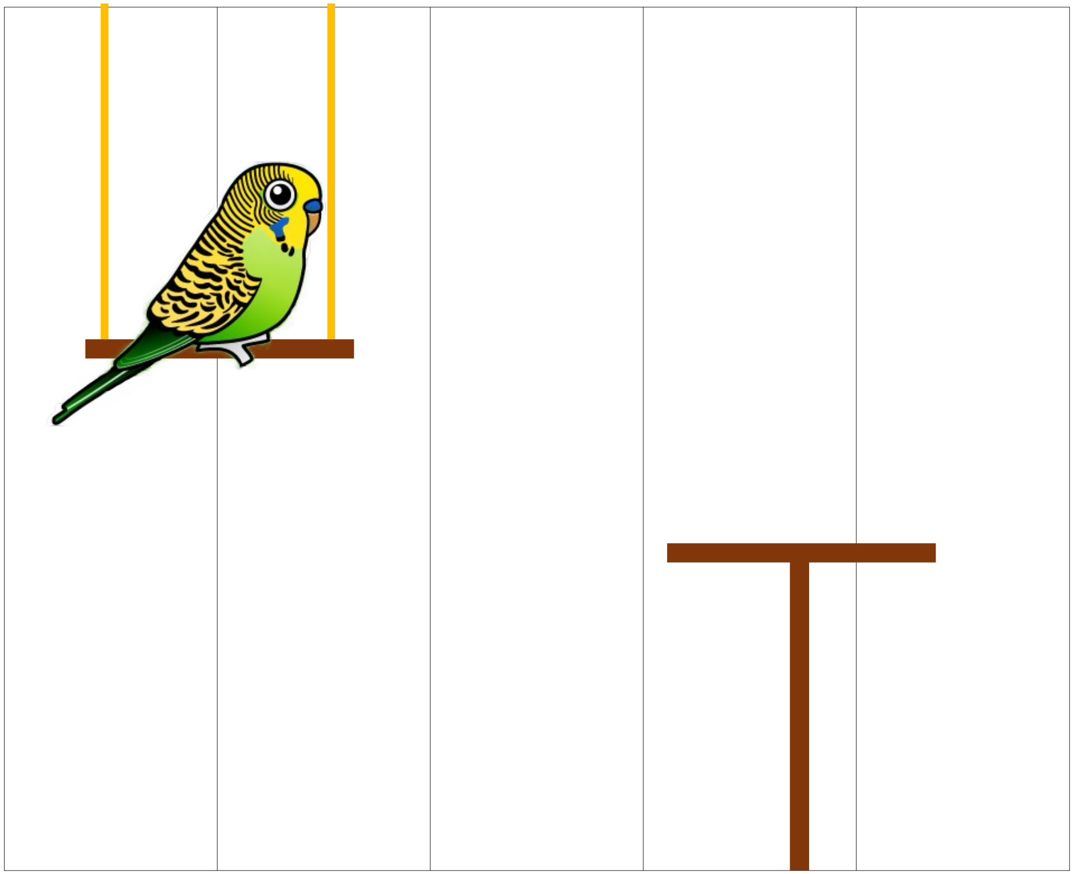


$$p(X) = \frac{1}{Z} \prod_{s} f_s(X_s)$$

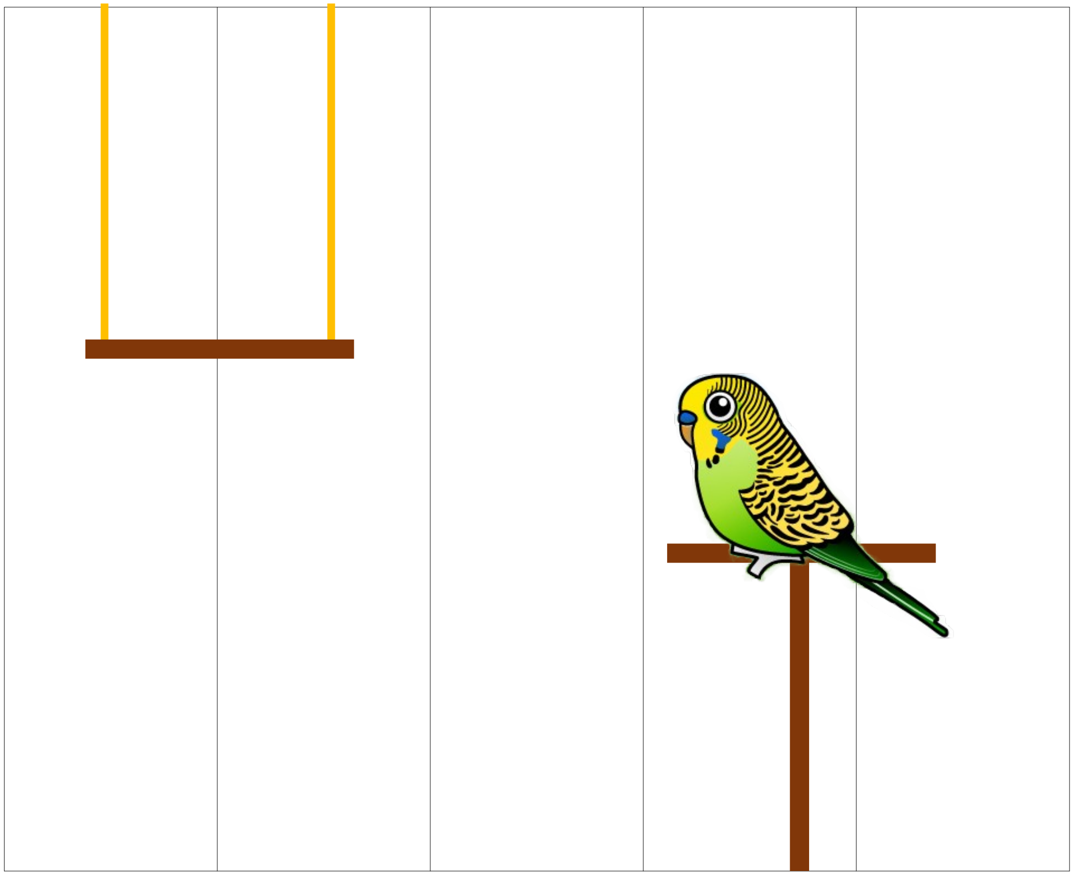
CHAPTER 15:

MARKOV MODELS

Billy the Budgie

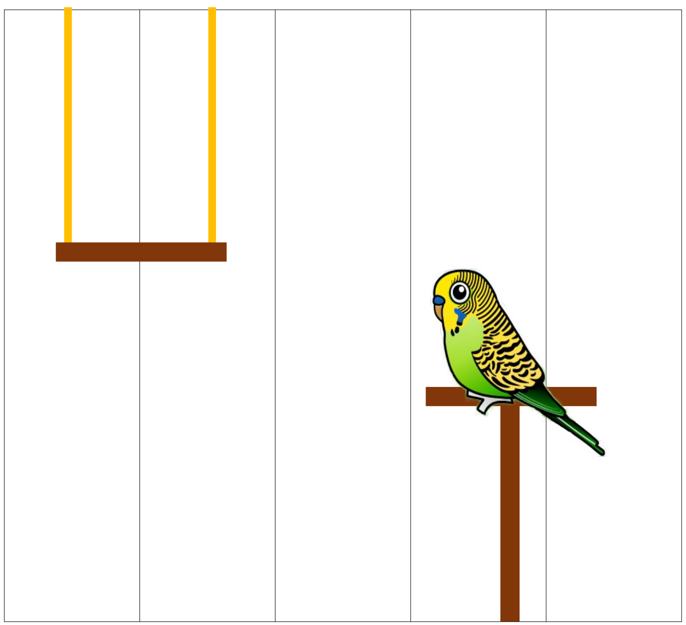


Billy the Budgie



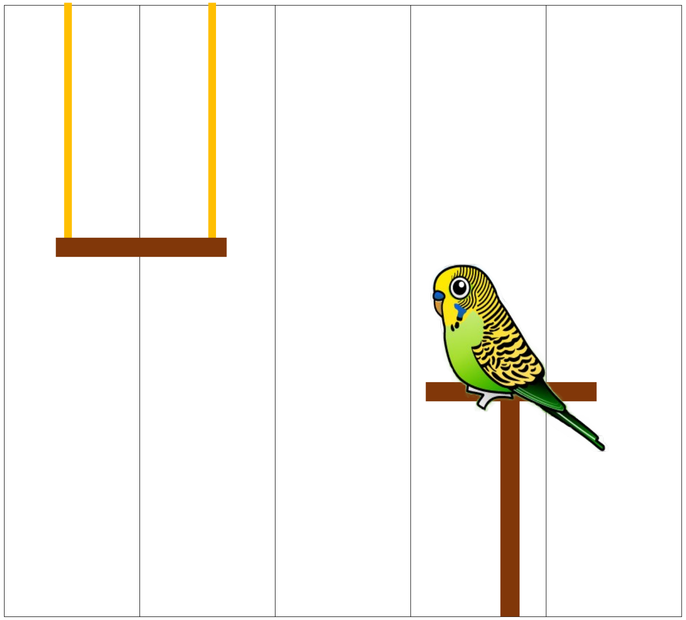
Billy the Budgie

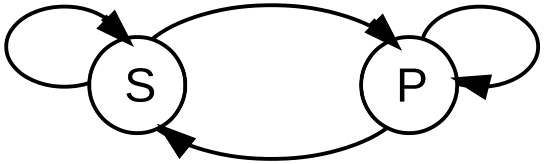
- Observations over time: PPSSPPPSSPPPSPPPP.....



Billy the Budgie

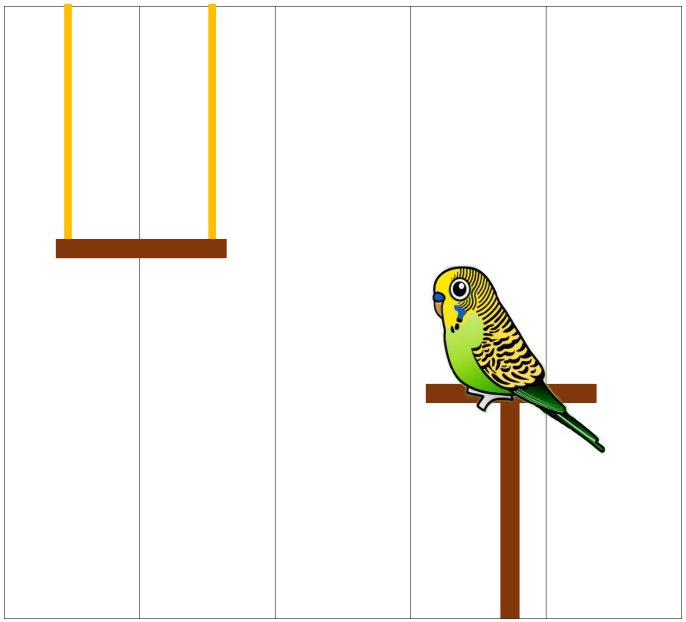
- Observations over time:
  PPSSPPPSSPPPSPPPP.....

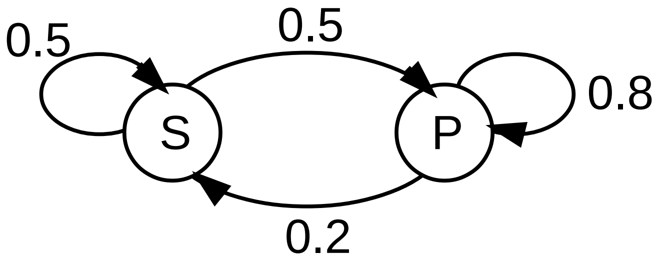




Billy the Budgie

- Observations over time: PPSSPPPSSPPPSPPPP.....





To

From		Swing	Perch
	Swing	0.5	0.5
	Perch	0.2	0.8

Markov Process

- Regularly spaced discrete sequences
  - Time (seconds, minutes, years…)
  - DNA or protein sequences
  - Letters/speech sounds
- Observation at $t+1$ is <u>dependent</u> on previous observation(s)

Markov Process

- General case:

$$P(q_{t+1} = S_j | q_t = S_i, q_{t-1} = S_k, \cdots)$$

- First order Markov Chain:

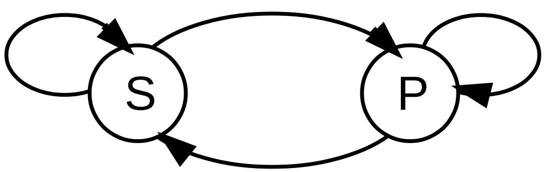
$$P(q_{t+1} = S_j | q_t = S_i, q_{t-1} = S_k, \cdots) = P(q_{t+1} = S_j | q_t = S_i)$$

- The state at *t+1* depends only on the state at *t*

Transition probabilities

- Independent of time
- Transition matrix **A**
- Rows sum to 1

$$\mathbf{A} = [a_{ij}]$$



To

From		Swing	Perch
	Swing	0.5	0.5
	Perch	0.2	0.8

Initial probabilities

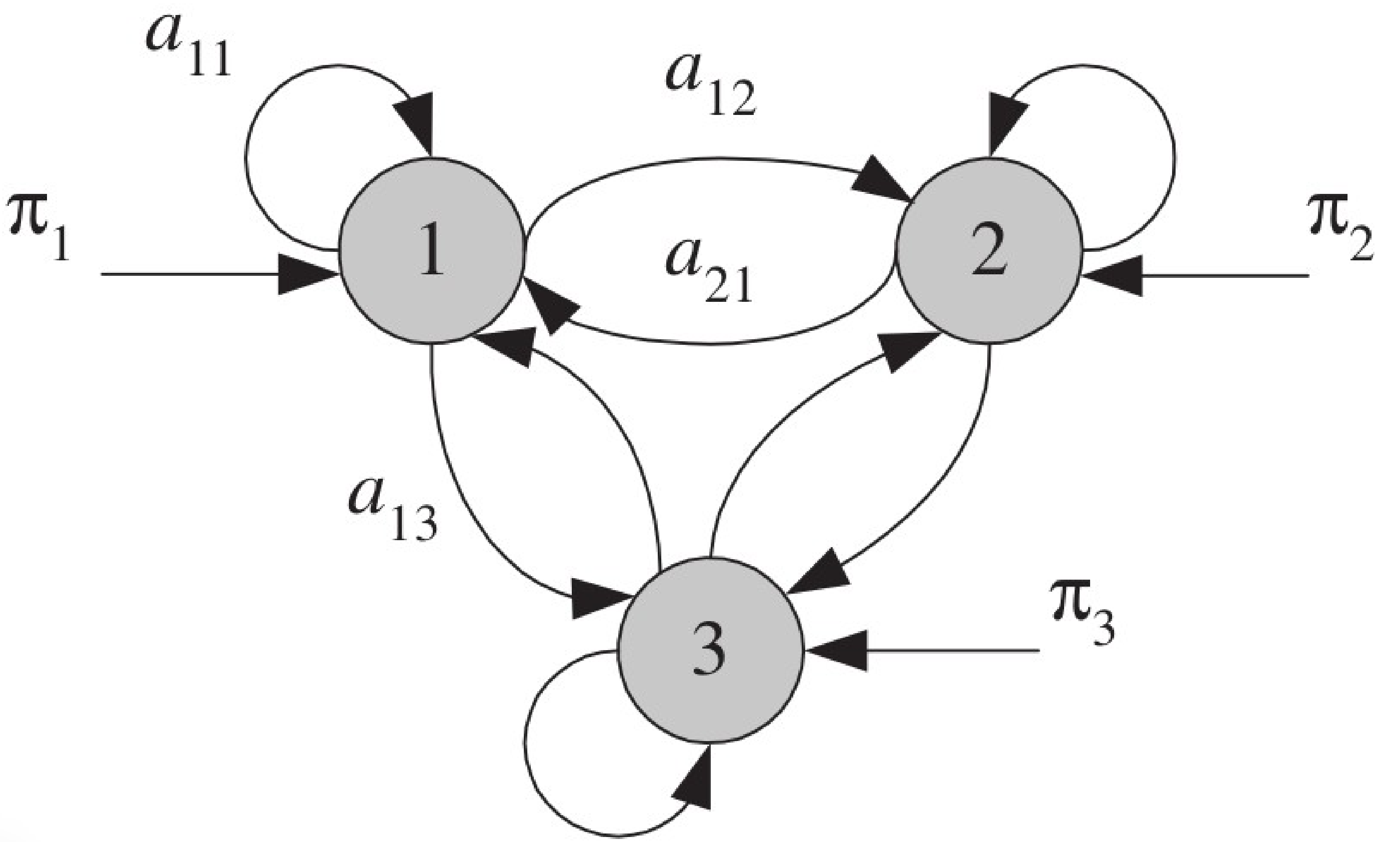
- We need a way to model the beginning of the sequence of observations
- Vector of initial probabilities

$$\mathbf{\Pi} = [\pi_i]$$

$$\pi_i \equiv P(q_1 = S_i)$$

$$\sum_{i=1}^{N} \pi_i = 1$$

Stochastic automaton



Note: The circles here are not different variables, they are *states* of the same variable

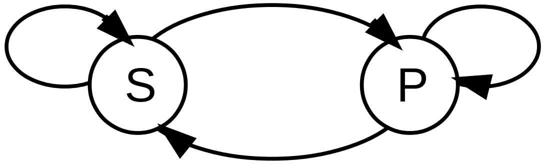
Observed sequences

- Given:
  - Sequence of states: $Q = \{q_1, q_2 \ldots q_{T-1}, q_T\}$
  - Observed sequence $O = Q$
- Probability of O is the product of transition probabilities:

$$P(O = Q|\mathbf{A}, \mathbf{\Pi}) = P(q_1) \prod_{t=2}^{T} P(q_t|q_{t-1}) = \pi_{q_1} a_{q_1 q_2} \cdots a_{q_{T-1} q_T}$$

Mini-exercise

- Calculate the probability of the sequence: PPSSPPP



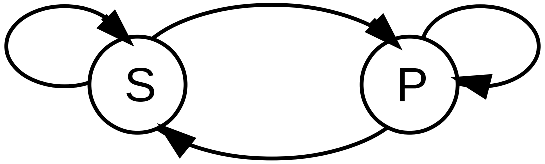
Π
0.5
0.5

To / From

A	Swing	Perch
Swing	0.5	0.5
Perch	0.2	0.8

Mini-exercise

- Calculate the probability of the sequence: PPSSPPP
- $\pi_P \cdot a_{PP} \cdot a_{PS} \cdot a_{SS} \cdot a_{SP} \cdot a_{PP} \cdot a_{PP}$
- $0.5 \cdot 0.8 \cdot 0.2 \cdot 0.5 \cdot 0.5 \cdot 0.8 \cdot 0.8$
- $= 0.0128$



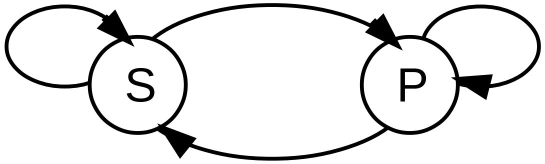
Π
0.5
0.5

To / From

A	Swing	Perch
Swing	0.5	0.5
Perch	0.2	0.8

Mini-exercise II

- Given the sequence: SSSPPSSSSPSSSSPP
- Estimate the transition probabilities



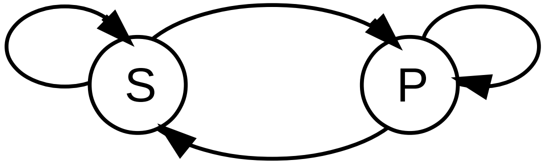
To

From

A	Swing	Perch
Swing		
Perch		

Mini-exercise II

- Sequence: SSSPPSSSPSSSPP
- S* = 9, SS = 6/9, SP = 3/9
- P* = 4, SS = 2/4, SP = 2/4



To (columns), From (rows)

Counts	Swing	Perch
Swing	6	3
Perch	2	2

To (columns), From (rows)

A	Swing	Perch
Swing	2/3	1/3
Perch	1l2	1/2

Estimating Probabilities

- Initial probabilities

$$\hat{\pi}_i = \frac{\#\{\text{sequences starting with } S_i\}}{\#\{\text{sequences}\}} = \frac{\sum_k 1(q_1^k = S_i)}{K}$$

- Transition probabilities

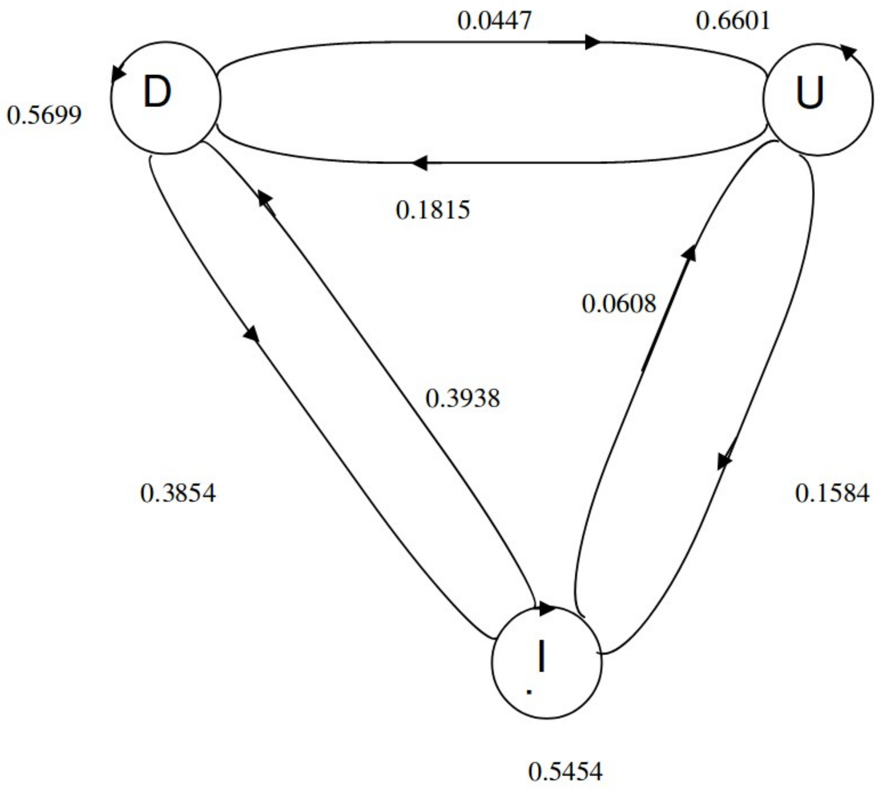
$$\hat{a}_{ij} = \frac{\#\{\text{transitions from } S_i \text{ to } S_j\}}{\#\{\text{transitions from } S_i\}} = \frac{\sum_k \sum_{t=1}^{T-1} 1(q_t^k = S_i \text{ and } q_{t+1}^k = S_j)}{\sum_k \sum_{t=1}^{T-1} 1(q_t^k = S_i)}$$

Example: Stock price model

- Stock price has three states:
  - Decrease
  - Unchanged
  - Increase

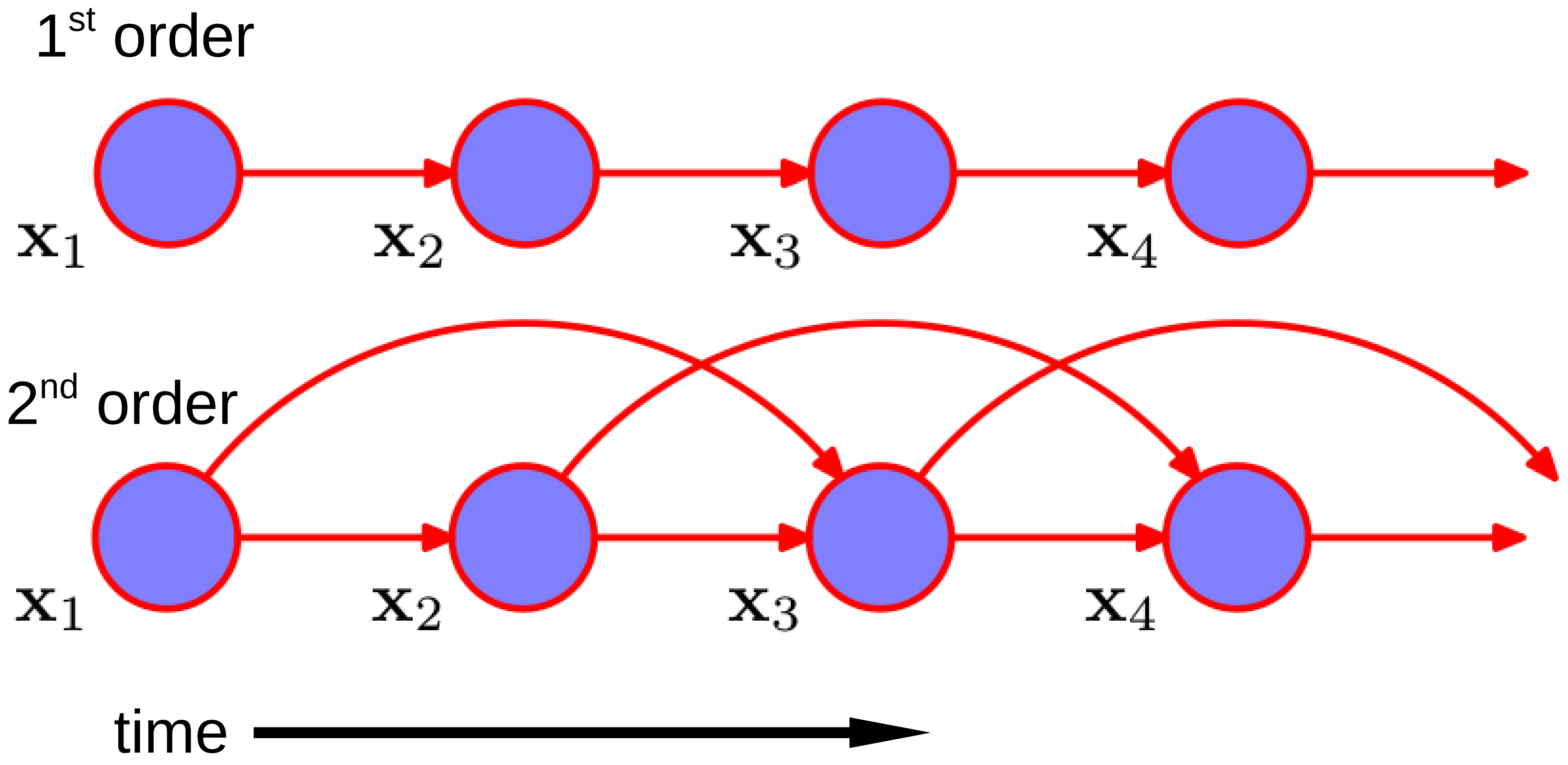
Example: Stock price model

- Stock price has three states:
  - Decrease
  - Unchanged
  - Increase



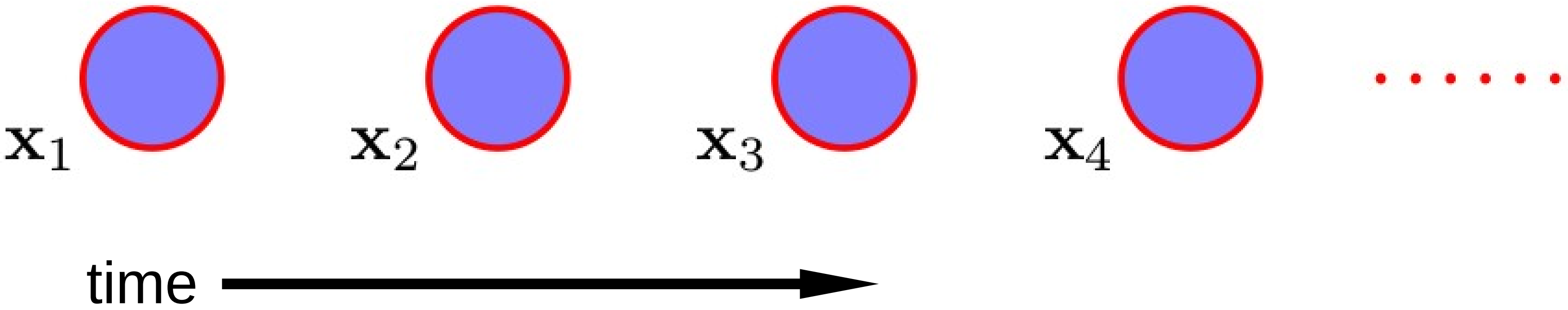
Choji et al 2013

Markov chains as graphical models



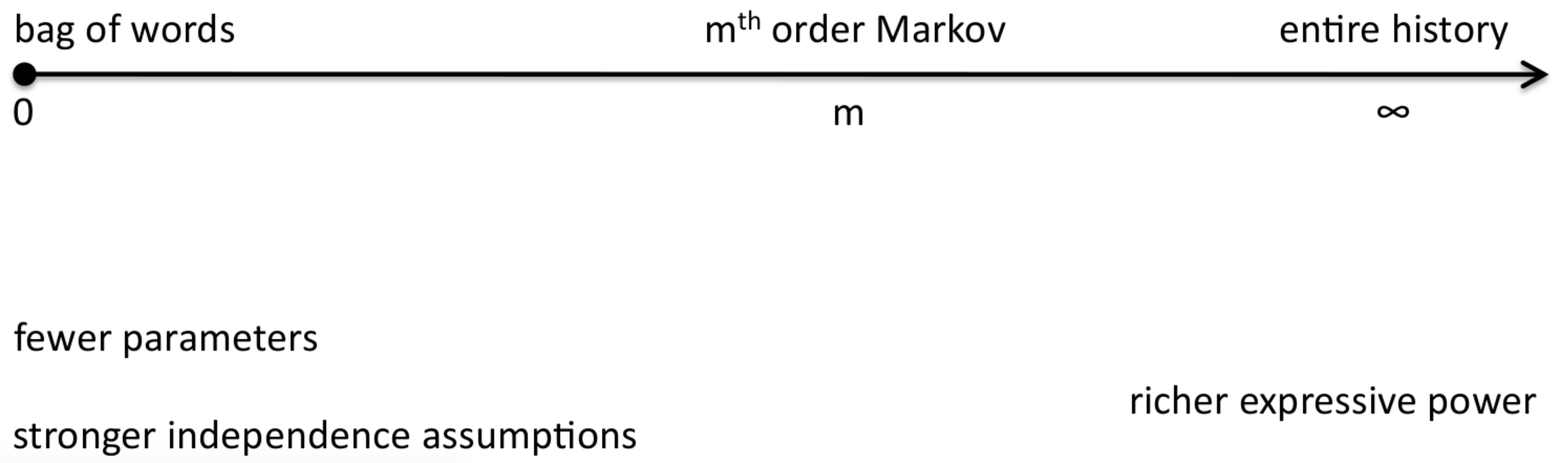
Markov chains as graphical models

- What is this then?



Markov chain: US political blog text

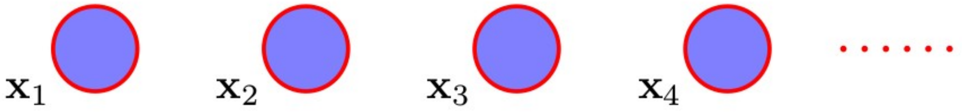
- 2.8M words from blog posts



Markov chain: US political blog text

- 0th order (bag of words, unigram)

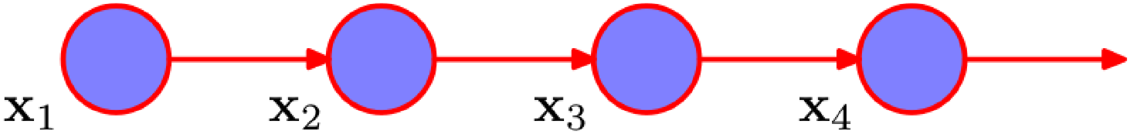
this trying our putting and funny
and among it herring it obama
but certainly foreign my
c on byron again but from i
i so and i chuck yeah the as but but republicans if
this stay oh so or it mccain bush npr this with what
and they right i while because obama



Markov chain: US political blog text

- 1st order (bigram)

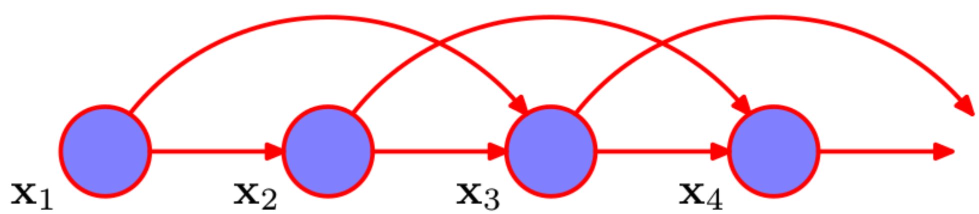
the lack of the senator mccain hadn t keep this story backwards
while showering praise of the kind of gop weakness
it was mistaken for american economist anywhere in the white house press hounded the absence of those he s as a wide variety of this election day after the candidate b richardson was polled ri in hempstead moderated by the convention that he had zero wall street journal argues sounds like you may be the primary
but even close the bill told c e to take the obama on the public schools and romney
fred flinstone s see how a lick skillet road it s little sexist remarks



Markov chain: US political blog text

- 2nd order (trigram)

```
 as i can pin them all none of them want to bet that
any of the might be
 conservatism unleashed into the privacy rule book and
when told about what paul
 fans organized another massive fundraising initiative
yesterday and i don t know what the rams supposedly
want ooh
 but she did but still victory dinner
 alone among republicans there are probably best not
all of the fundamentalist community
 asked for an independent maverick now for
crystallizing in one especially embarrassing
```



Markov chain: US political blog text

- 4th order (5-gram)

```
 he realizes fully how shallow and insincere
conservative behavior has been he realizes that there
is little way to change the situation
 this recent arianna huffington item about mccain
issuing heartfelt denials of things that were actually
true or for that matter about the shi a sunni split and
which side iran was on would get confused about this
any more than someone with any knowledge of us politics
would get confused about whether neo confederates were
likely to be supporting the socialist workers party
 at the end of the world and i m not especially
discouraged now that newsweek shows obama leading by
three now
```

Markov chain: US political blog text

- 99th order (100-gram)

and it would be the work of many hands to catalogue all the ridiculous pronouncements made by this man since his long train of predictions about the middle east has been gaudily disastrously stupefyingly misinformed just the buffoon it seems for the new york times to award with a guest column for if you object to the nyt rewarding failure in quite this way then you re intolerant according to the times editorial page editor andrew rosenthal rosenthal doesn t seem to recognize that his choice of adjectives to describe kristol serious respected are in fact precisely what is at issue for those whom he dismisses as having a fear of opposing views

Hidden Markov Models

Example from Durbin et al

The Occasionally Dishonest Casino

- Game of dice
- Uses a fair die most of the time

Fair

Num	Prob
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

The Occasionally Dishonest Casino

- But will occasionally randomly switch to a loaded die (and randomly back again)

Fair

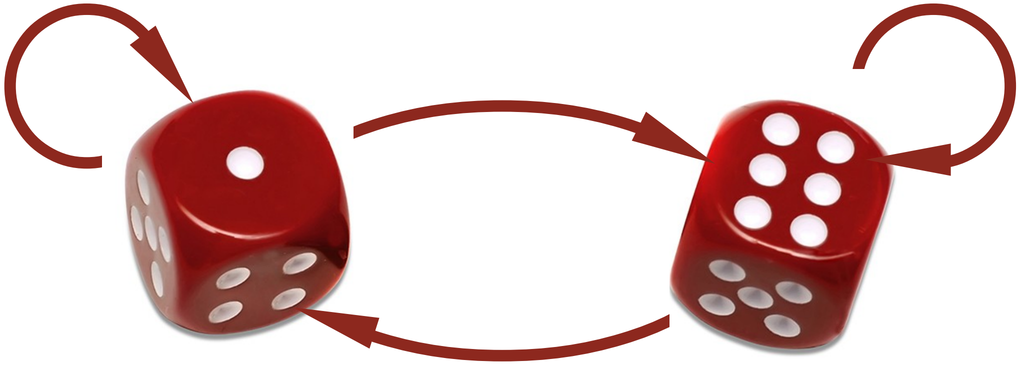
Num	Prob
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Loaded

Num	Prob
1	1/10
2	1/10
3	1/10
4	1/10
5	1/10
6	1/2

The Occasionally Dishonest Casino

- The dice are the states, but we only observe the *rolls*



Fair

Num	Prob
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

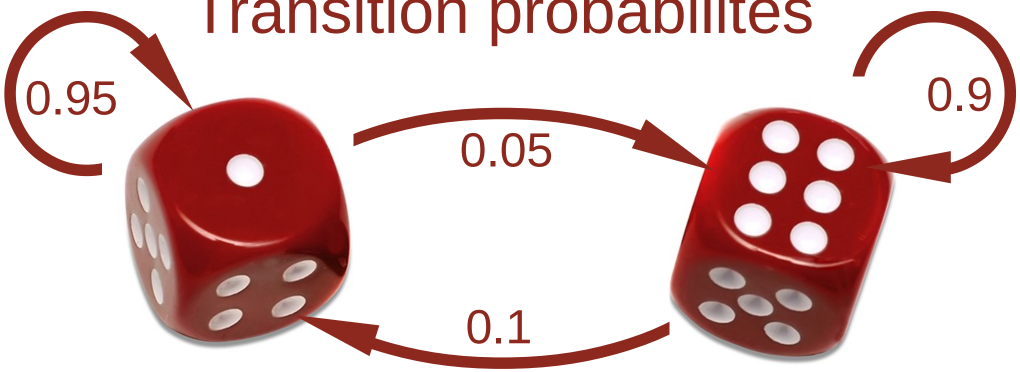
Loaded

Num	Prob
1	1/10
2	1/10
3	1/10
4	1/10
5	1/10
6	1/2

The Occasionally Dishonest Casino

- Hidden Markov Model

Transition probabilites



Fair

Num	Prob
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Emission probabilites

Loaded

Num	Prob
1	1/10
2	1/10
3	1/10
4	1/10
5	1/10
6	1/2

Hidden Markov Models

- An HMM (λ):

1. N: Number of states in the model

   $S = \{S_1, S_2, \ldots, S_N\}$

2. M: Number of distinct observation symbols in the *alphabet*

   $V = \{v_1, v_2, \ldots, v_M\}$

3. State transition probabilities:

   $\mathbf{A} = [a_{ij}]$ where $a_{ij} \equiv P(q_{t+1} = S_j | q_t = S_i)$

4. Observation probabilities:

   $\mathbf{B} = [b_j(m)]$ where $b_j(m) \equiv P(O_t = v_m | q_t = S_j)$

5. Initial state probabilities:

   $\mathbf{\Pi} = [\pi_i]$ where $\pi_i \equiv P(q_1 = S_i)$

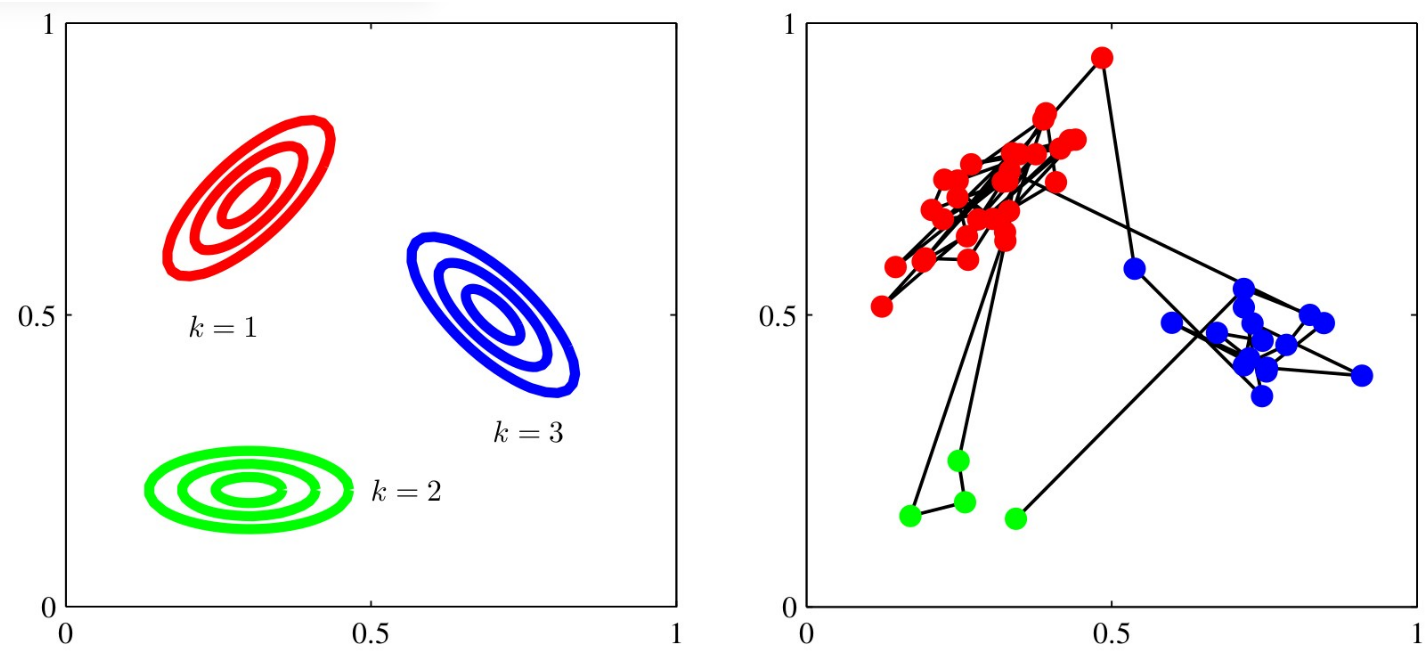
Hidden Markov Models

- Two sources of randomness
  - State (fair or loaded)
  - Emission (roll)
- Many possible state sequences for a sequence of observations, e.g:

FFFLLLLFFFFFFLLLLL
4352662344636461

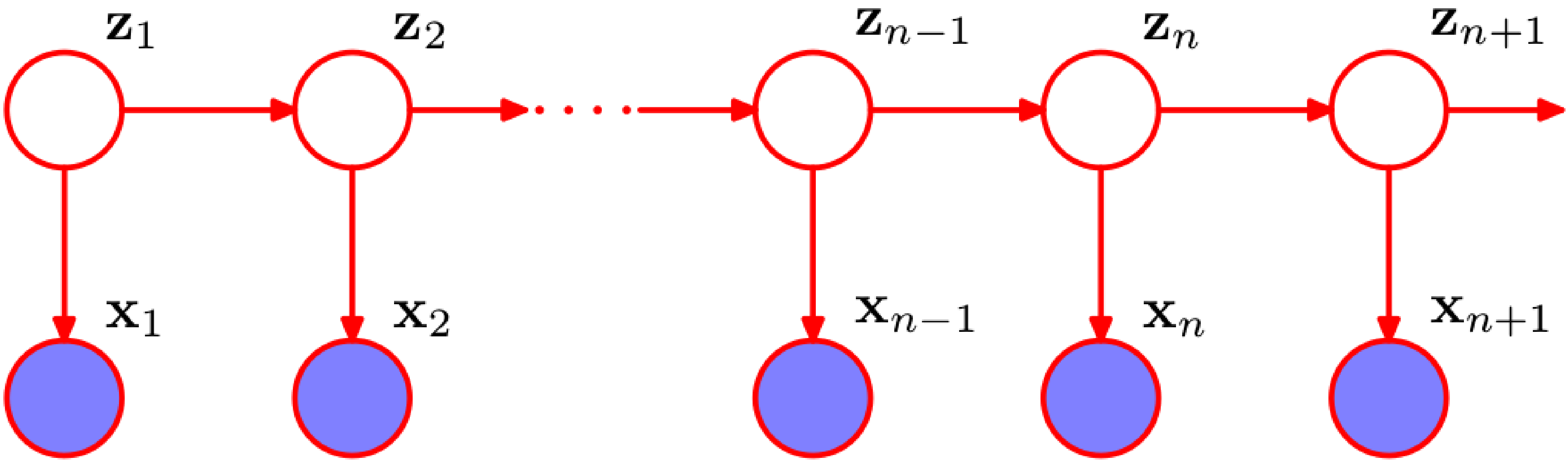
Continuous-valued observations

- HMM sampling from Gaussian distributions

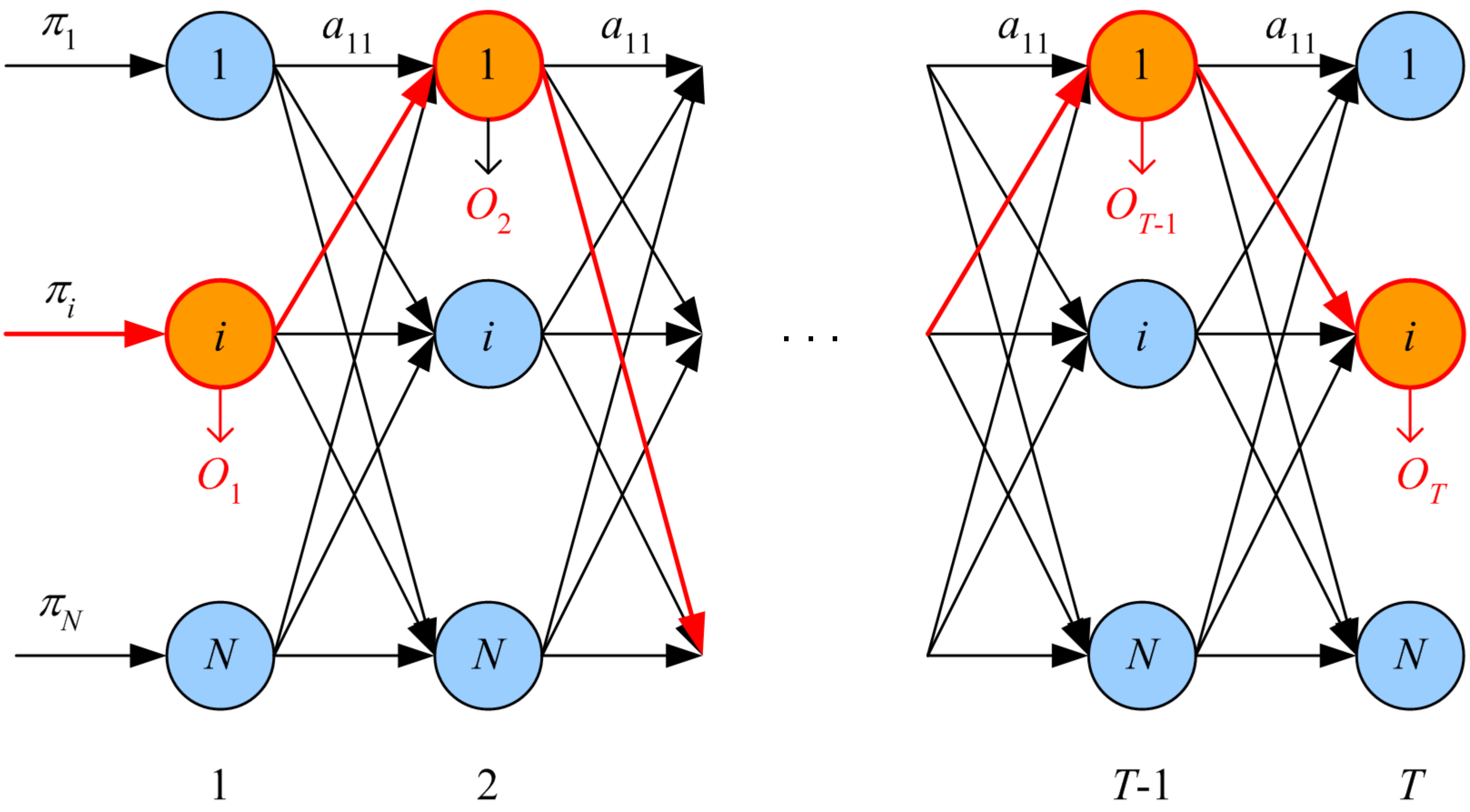


HMMs as graphical models

Simplest case:



HMM unfolded in time



Hidden Markov Models

- Three main questions:
  - What is the probability of a given sequence of observations?
  - What is the most probable *path* given the observations?
  - What are the parameters of the model given some observations?

Hidden Markov Models

- Three main questions:
  - What is the probability of a given sequence of observations? Forward – Backward algorithm
  - What is the most probable *path* given the observations? (Viterbi algorithm)
  - What are the parameters of the model given some observations? (Baum-Welch algorithm)

Probability of a sequence

- Given:
  - Some observed sequence O
  - A known state sequence Q
- Calculate P(O|Q)

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdots b_{q_T}(O_T)$$

But we don’t know Q (the hidden state sequence)!

Probability of a sequence

- We want the joint probability

$$
\begin{aligned}
P(O,Q|\lambda) &= P(q_1)\prod_{t=2}^{T} P(q_t|q_{t-1})\prod_{t=1}^{T} P(O_t|q_t) \\
&= \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T)
\end{aligned}
$$

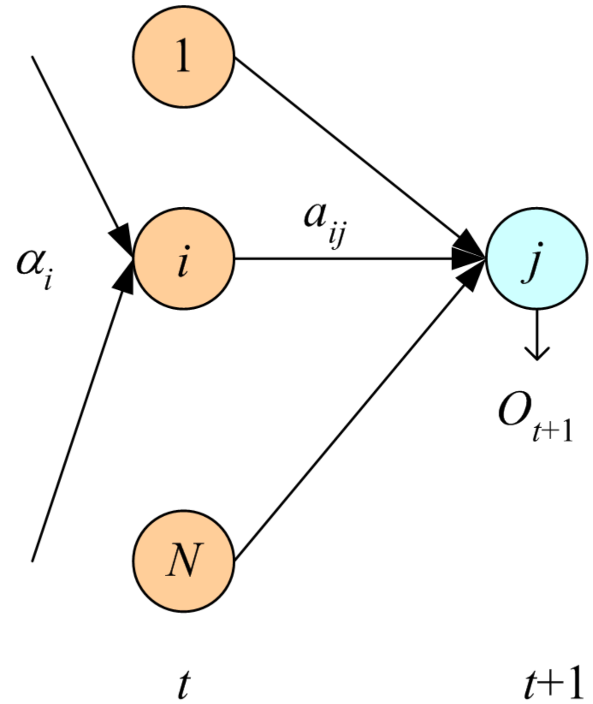
- ... summed over all possible Q

But that's not practical.
Instead use recursive forward-backward procedure

Probability of a sequence

- *Forward* variable α
  - P(observations up to and including j)
- Iterative procedure

$$\alpha_t(i) \equiv P(O_1 \cdots O_t, q_t = S_i | \lambda)$$



Probability of a sequence

- Initialization:

$$
\begin{aligned}
\alpha_1(i) &\equiv P(O_1, q_1 = S_i|\lambda) \\
&= P(O_1|q_1 = S_i, \lambda)P(q_1 = S_i|\lambda) \\
&= \pi_i b_i(O_1)
\end{aligned}
$$

Probability of a sequence

- Initialization:

$$\begin{aligned} \alpha_1(i) &\equiv P(O_1, q_1 = S_i | \lambda) \\ &= P(O_1 | q_1 = S_i, \lambda) P(q_1 = S_i | \lambda) \\ &= \pi_i b_i(O_1) \end{aligned}$$

- Recursion:

$$\begin{aligned} \alpha_{t+1}(j) &\equiv P(O_1 \cdots O_{t+1}, q_{t+1} = S_j | \lambda) \\ &= \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \end{aligned}$$

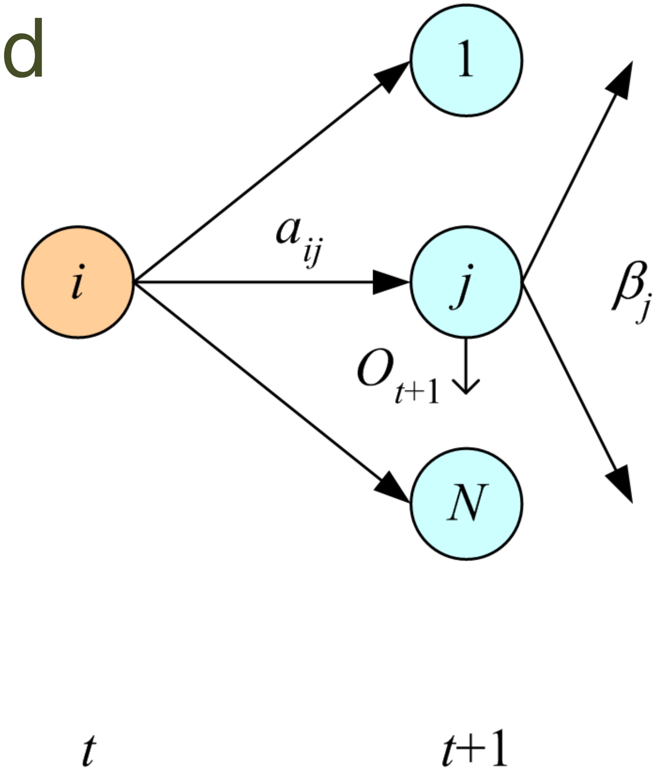
Probability of a sequence

- $\alpha_T(i)$ is the probability of the whole sequence, ending in state *i*
- Sum over all states to get P(O)

$$\begin{aligned} P(O|\lambda) &= \sum_{i=1}^{N} P(O, q_T = S_i|\lambda) \\ &= \sum_{i=1}^{N} \alpha_T(i) \end{aligned}$$

Probability of a sequence

- Similar to (π, λ) in Belief Propagation, we can also calculate the *Backward* variable
- Used in state sequence and parameter estimation



Sources & Resources

- Recommended books:
  - Bishop: Pattern recognition
  - Durbin, Eddy, Krogh Mitchison: Biological sequence analysis
- Explanation and animations of Markov chains:
  - http://setosa.io/ev/markov-chains/