Hidden Markov Models

Statistical Models for Sequencial Data

Diego Milone
d.milone⊛ieee.org

Tópicos Selectos en Aprendizaje Maquinal
Doctorado en Ingeniería, FICH-UNL

December 3, 2010

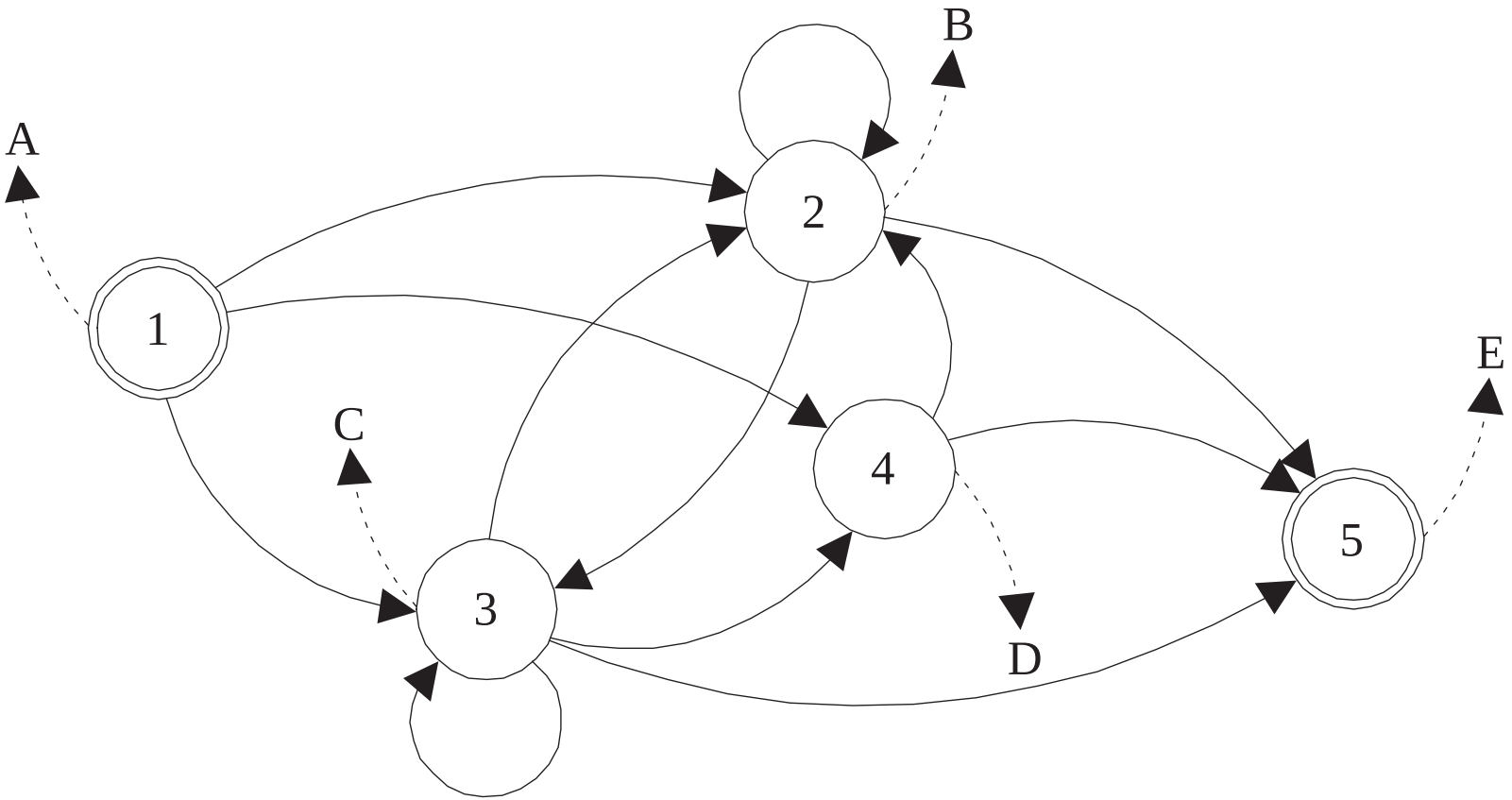
Outline

sinc(i)

Outline

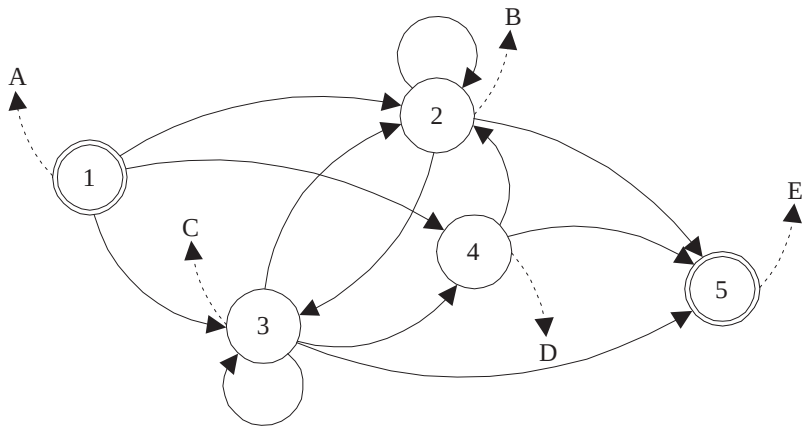
sinc(i)

Deterministic Finite State Automata



Deterministic Finite State Automata

How does this model work?

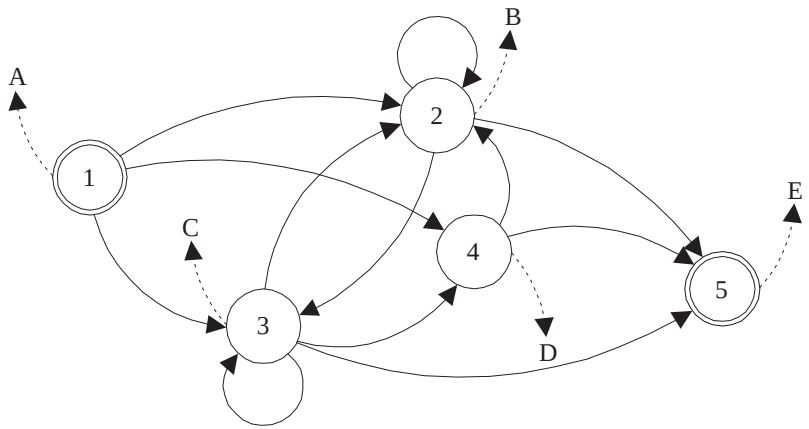


Generative Model: states and transition rules?

Assuming 0 as initial condition in the accumulator:
Which is the output of the model?

Deterministic Finite State Automata

How does this model work?

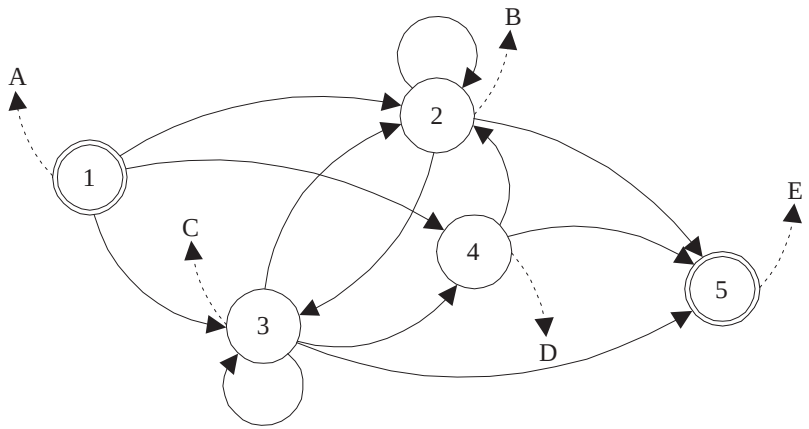


Generative Model: states and transition rules?

Assuming 0 as initial condition in the accumulator:
Which is the output of the model?

Deterministic Finite State Automata

How does this model work?



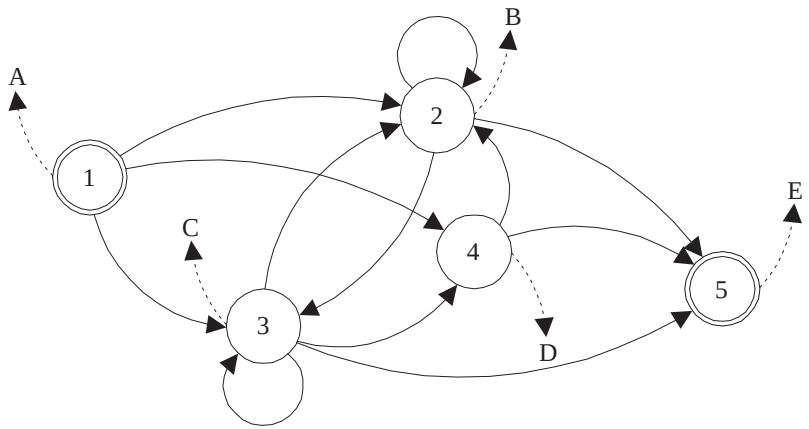
Generative Model: states and transition rules

Suppose:

- no inputs
- each state increase an internal accumulator with half of its number
- self-loops are used at less the state number times
- next state is the closest to current accumulator value

Deterministic Finite State Automata

How does this model work?

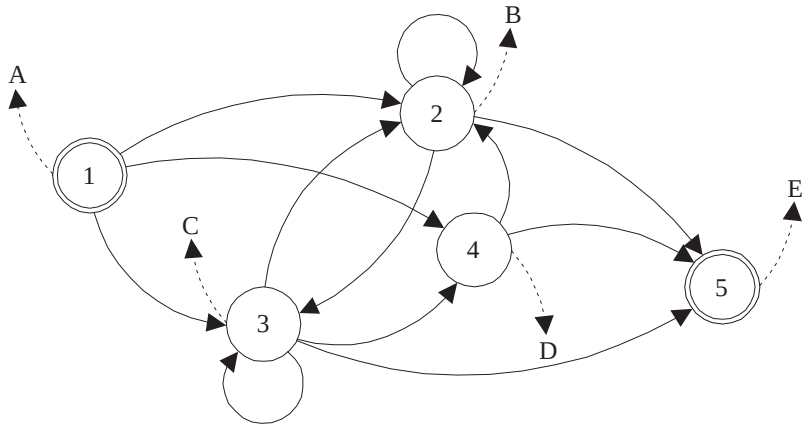


Generative Model: states and transition rules

Assuming 0 as initial condition in the accumulator:
Which is the output of the model?

Deterministic Finite State Automata

How does this model work?



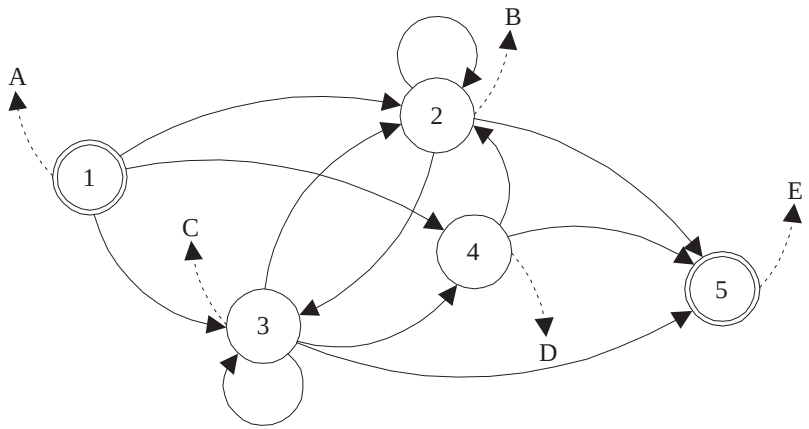
Generative Model: states and transition rules

Assuming 0 as initial condition in the accumulator:
Which is the output of the model?

The sequence of states is: 1, 2, 2, 3, 3, 3, 5

Deterministic Finite State Automata

How does this model work?



Generative Model: states and transition rules

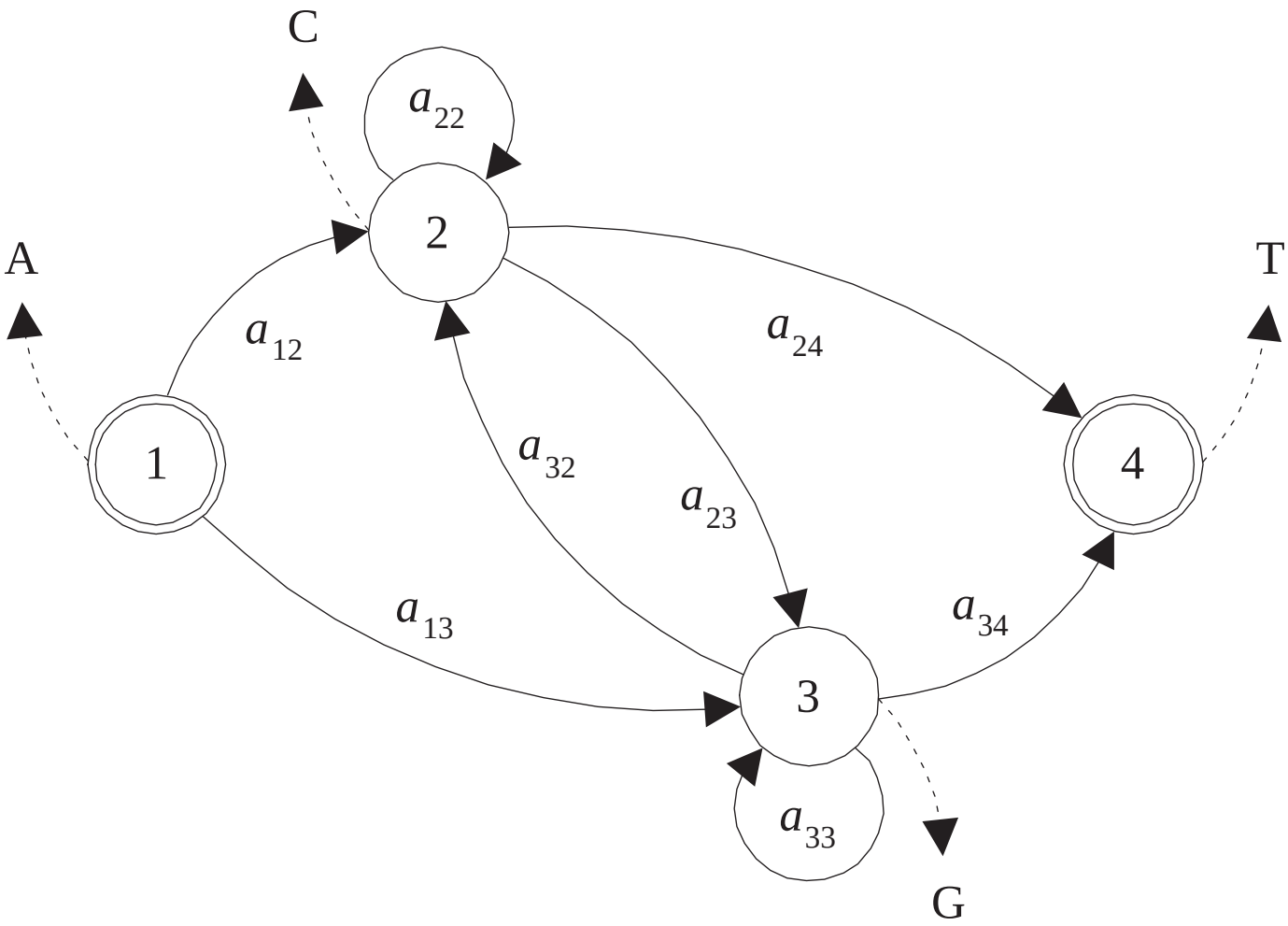
Assuming 0 as initial condition in the accumulator:
Which is the output of the model?

The sequence of states is: 1, 2, 2, 3, 3, 3, 5

thus the outputs are: A, B, B, C, C, C, E

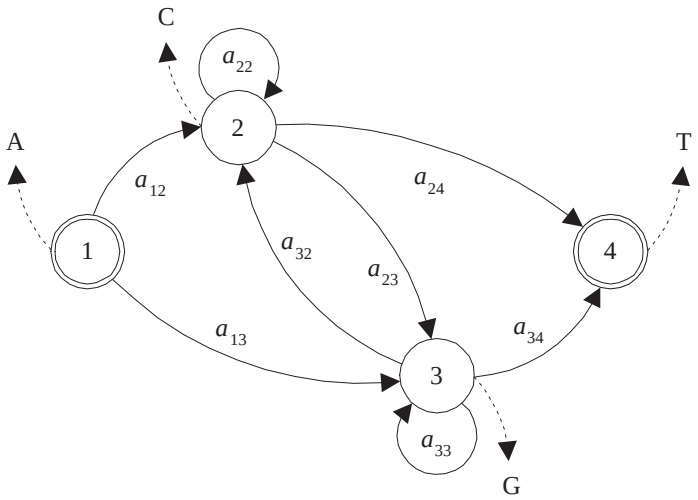
sinc(i)

Probabilistic Finite State Automata



Probabilistic Finite State Automata

How does this model work?



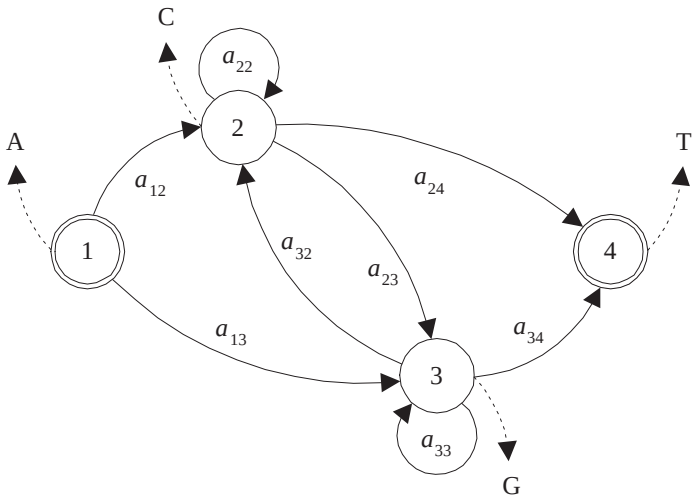
Generative Model: Transition rules?

How many output sequences can this model give?
Do they all have the same probability?

Given the output sequence: A, C, C, G, C, T.
Which is the probability that it has been generated by this model?

Probabilistic Finite State Automata

How does this model work?



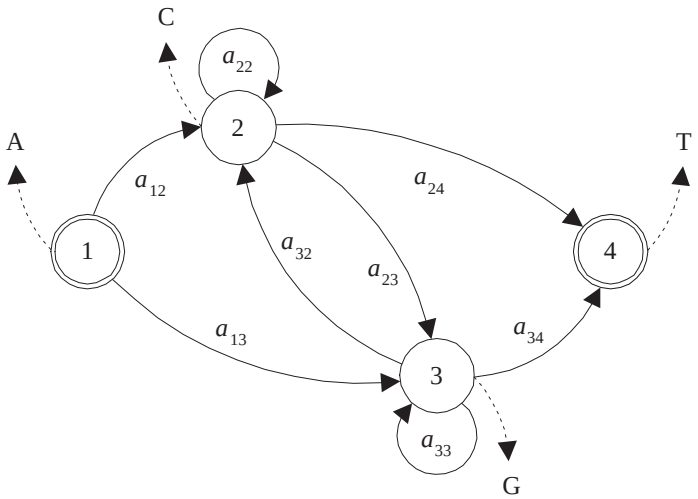
Generative Model: Transition rules?

How many output sequences can this model give?
Do they all have the same probability?

Given the output sequence: A, C, C, G, C, T.
Which is the probability that it has been generated by this model?

Probabilistic Finite State Automata

How does this model work?



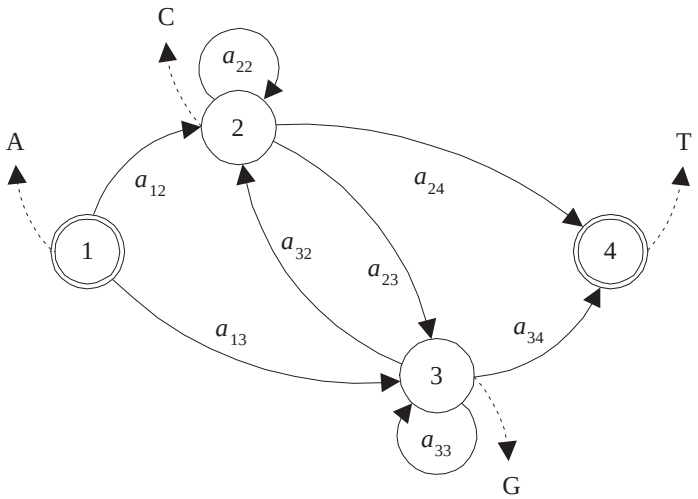
Generative Model: Transition rules?

How many output sequences can this model give?
Do they all have the same probability?

Given the output sequence: A, C, C, G, C, T.
Which is the probability that it has been generated by this model?

Probabilistic Finite State Automata

How does this model work?



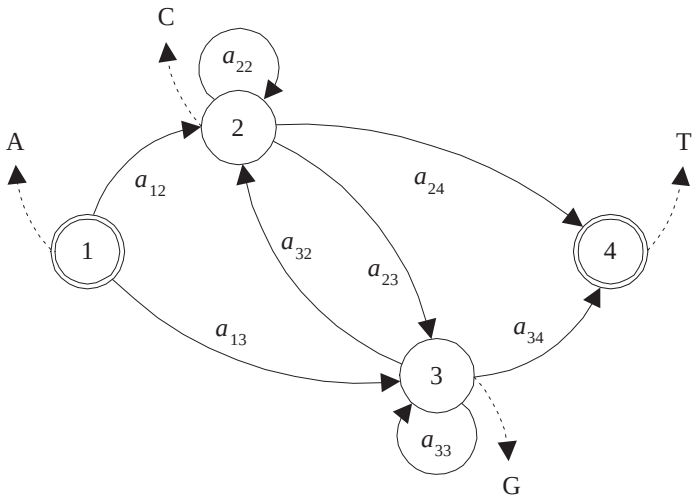
Generative Model: Transition rules?

How many output sequences can this model give?
Do they all have the same probability?

Given the output sequence: A, C, C, G, C, T.
Which is the probability that it has been generated by this model?

Probabilistic Finite State Automata

For example, we can define:



$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 1/4 & 1/4 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

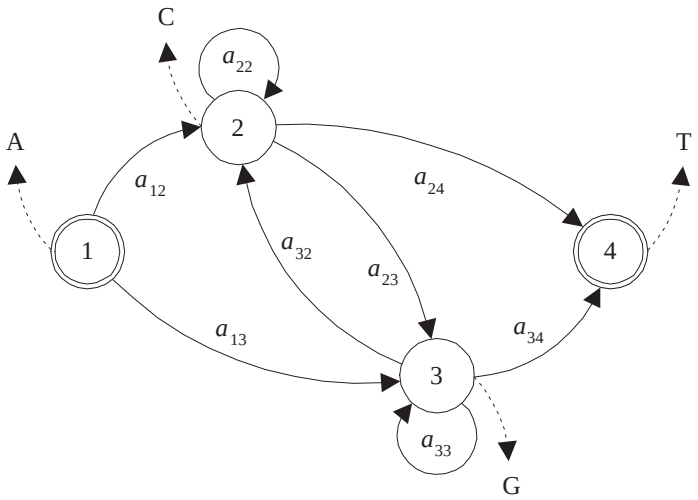
Given the output sequence: A, C, C, G, C, T,
the transitions are: $1 \to 2,\ 2 \to 2,\ 2 \to 3,\ 3 \to 2,\ 2 \to 4,$

Then, the probability

$$p_{122324} = a_{12}a_{22}a_{23}a_{32}a_{24} = \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{128}$$

Probabilistic Finite State Automata

For example, we can define:



$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 1/4 & 1/4 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

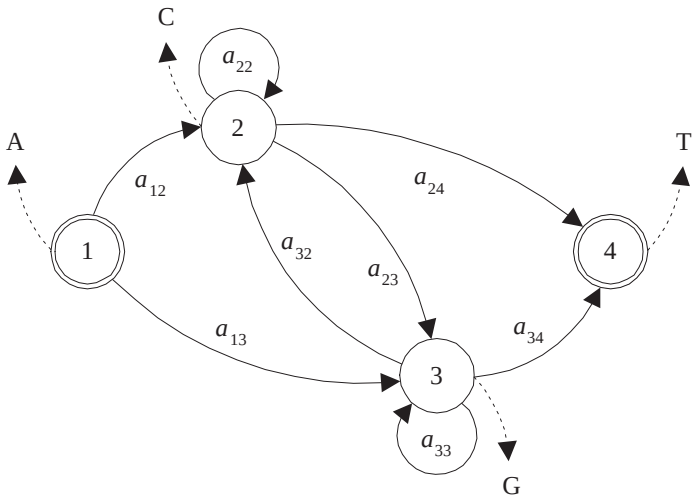
Given the output sequence: A, C, C, G, C, T,
the transitions are: $1 \to 2,\ 2 \to 2,\ 2 \to 3,\ 3 \to 2,\ 2 \to 4,$

Then, the probability

$$p_{122324} = a_{12}a_{22}a_{23}a_{32}a_{24} = \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{128}$$

Probabilistic Finite State Automata

For example, we can define:



$$A = \{a_{ij}\} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 1/4 & 1/4 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

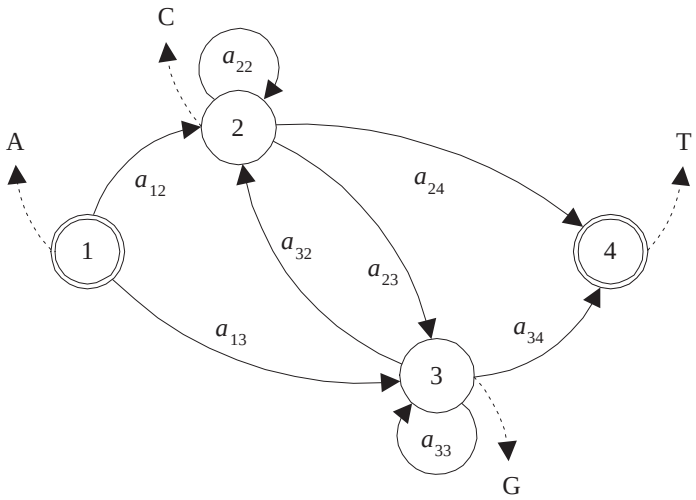
Given the output sequence: A, C, C, G, C, T,
the transitions are: $1 \to 2,\ 2 \to 2,\ 2 \to 3,\ 3 \to 2,\ 2 \to 4,$

Then, the probability

$$p_{122324} = a_{12}a_{22}a_{23}a_{32}a_{24} = \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{128}$$

Observable vs Hidden States

The observable model

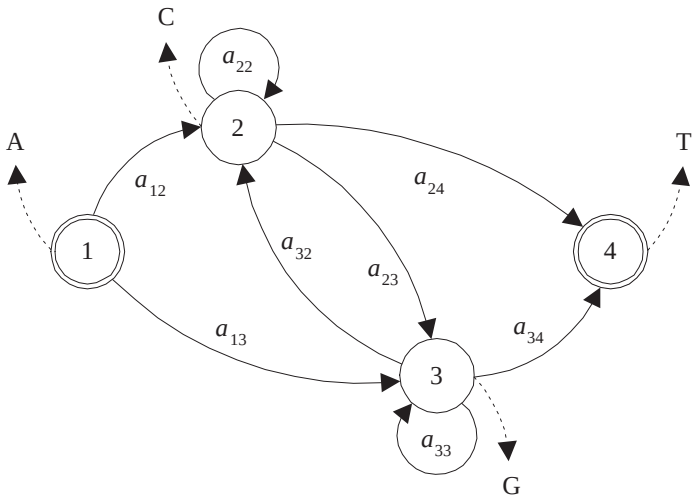


In the *observable* model, the state is completely determined at each instance of time.

For example, in this elemental model always the first symbol will be A and the last one will be T.

Observable vs Hidden States

The observable model

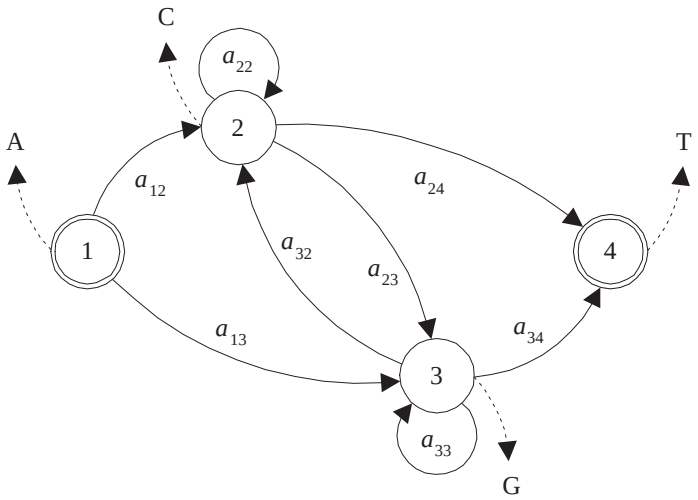


In the *observable* model, the state is completely determined at each instance of time.

For example, in this elemental model always the first symbol will be A and the last one will be T.

Observable vs Hidden States

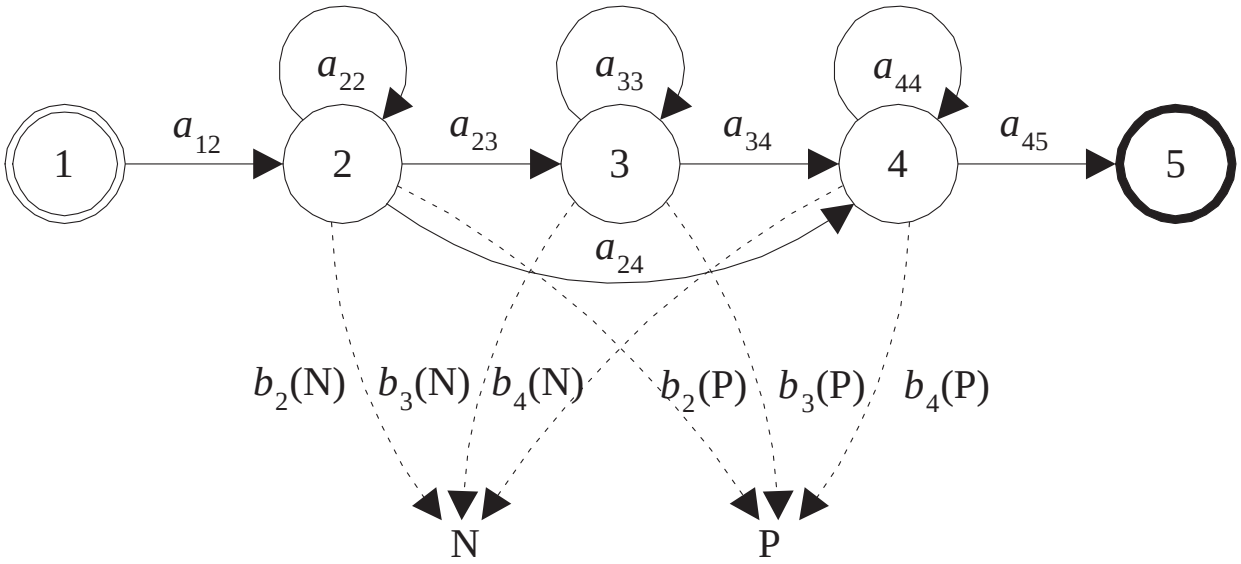
The observable model



In the *observable* model, the state is completely determined at each instance of time.

For example, in this elemental model always the first symbol will be A and the last one will be T.

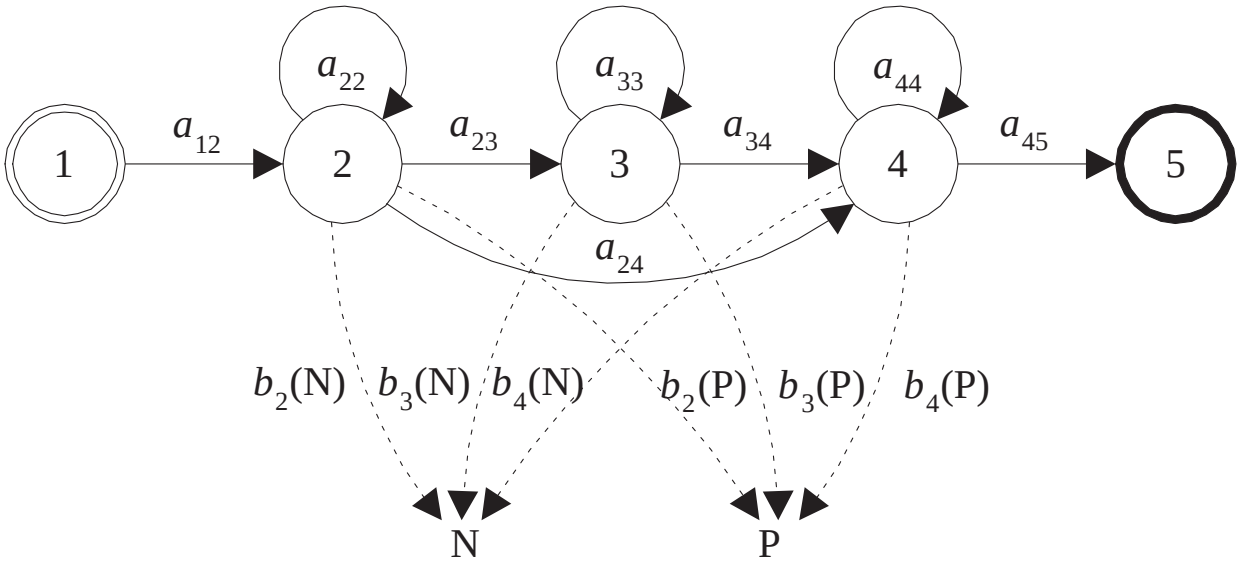
Hidden Markov Models



In a *hidden* model, states must be inferred from observations. Thus, states are *hidden* or *latent* variables of the model.

In other words, each state can emit all the output symbols, and the observation is a probabilistic function of state.

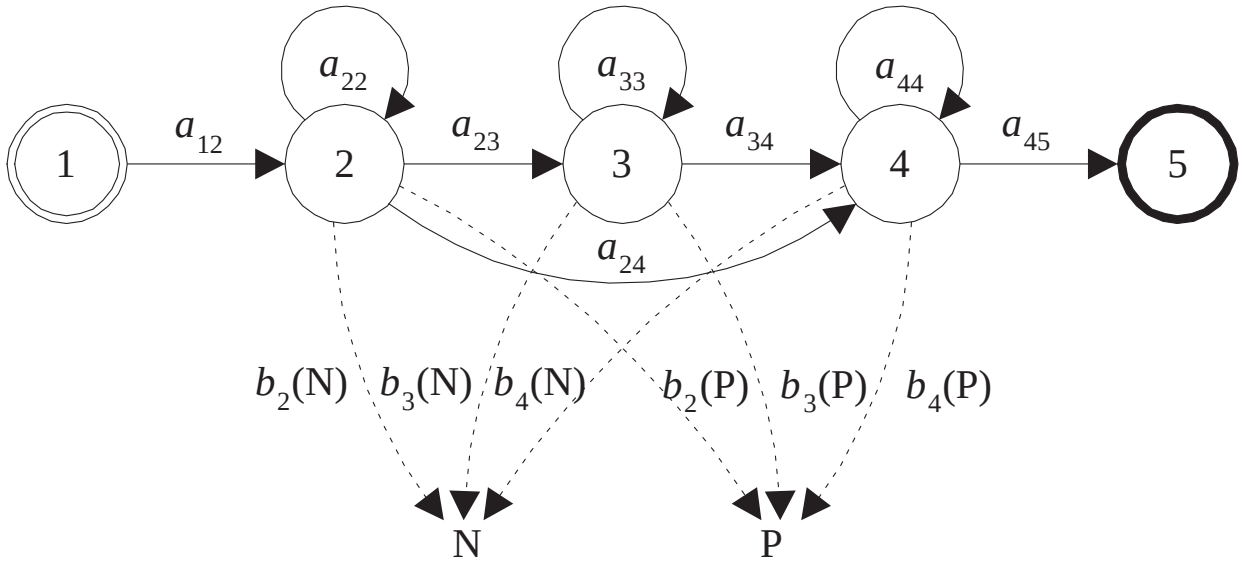
Hidden Markov Models



In a *hidden* model, states must be inferred from observations. Thus, states are *hidden* or *latent* variables of the model.

In other words, each state can emit all the output symbols, and the observation is a probabilistic function of state.

Hidden Markov Models

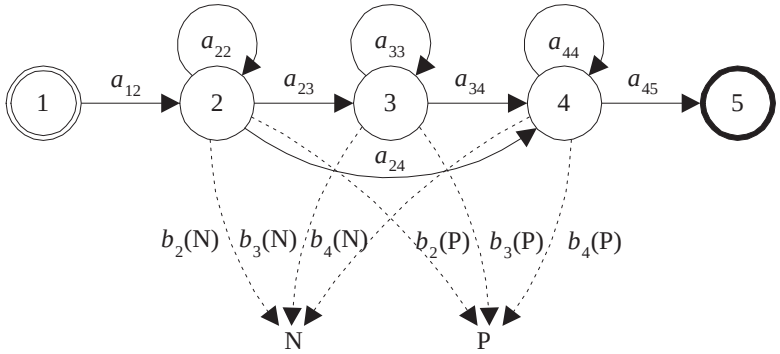


In a *hidden* model, states must be inferred from observations. Thus, states are *hidden* or *latent* variables of the model.

In other words, each state can emit all the output symbols, and the observation is a probabilistic function of state.

Hidden Markov Models

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

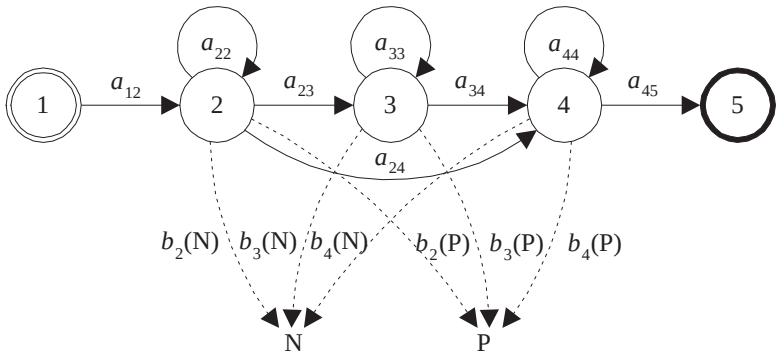
Given the output sequence: N, N, P, N.
Which is the probability that it has been generated by this model?

But, we don't know the sequence of states... it is hidden...
Two solutions:

1. Use the law of alternatives, as a special case of the law of total probability.
2. Use just the most probable sequence of hidden states.

Hidden Markov Models

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

Given the output sequence: N, N, P, N.

Which is the probability that it has been generated by this model?

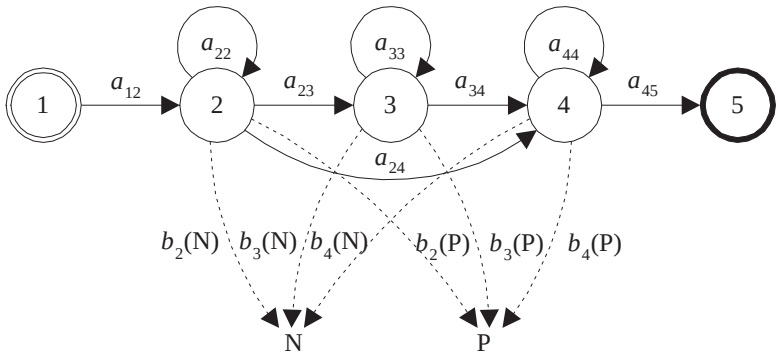
But, we don't know the sequence of states... it is hidden...

Two solutions:

1. Use the law of alternatives, as a special case of the law of total probability.
2. Use just the most probable sequence of hidden states.

Hidden Markov Models

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

Given the output sequence: N, N, P, N.
Which is the probability that it has been generated by this model?

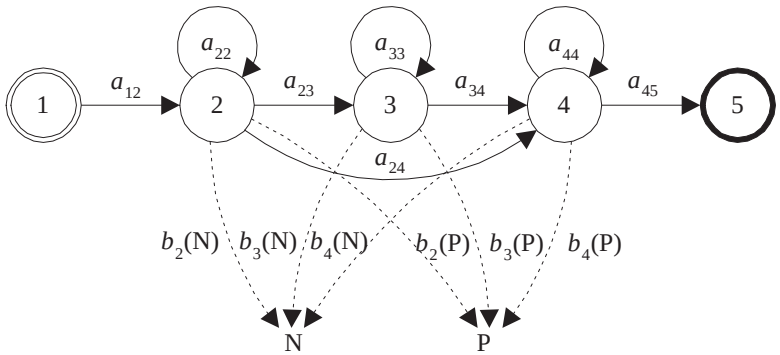
But, we don't know the sequence of states... it is hidden...

Two solutions:

1. Use the law of alternatives, as a special case of the law of total probability.
2. Use just the most probable sequence of hidden states.

Hidden Markov Models

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

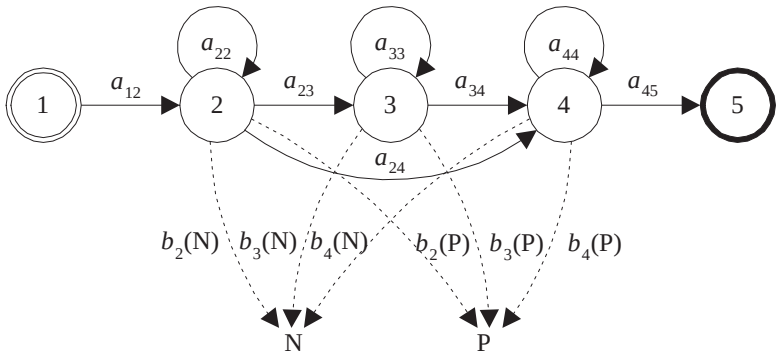
Given the output sequence: N, N, P, N.
Which is the probability that it has been generated by this model?

But, we don't know the sequence of states... it is hidden...
Two solutions:

1. Use the law of alternatives, as a special case of the law of total probability.
2. Use just the most probable sequence of hidden states.

Hidden Markov Models

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

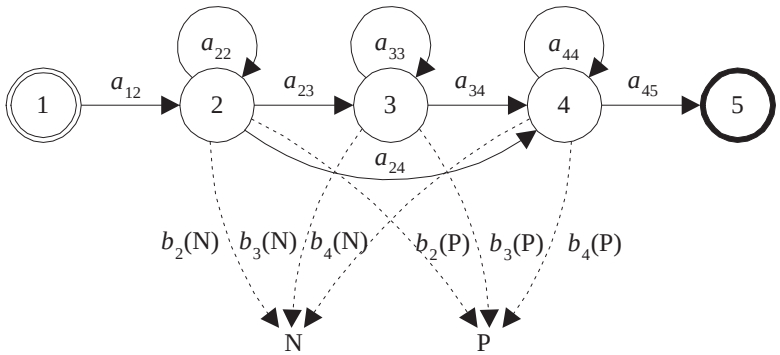
Given the output sequence: N, N, P, N.
Which is the probability that it has been generated by this model?

But, we don't know the sequence of states... it is hidden...
Two solutions:

1. Use the law of alternatives, as a special case of the law of total probability.
2. Use just the most probable sequence of hidden states.

Hidden Markov Models: all the alternative paths

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

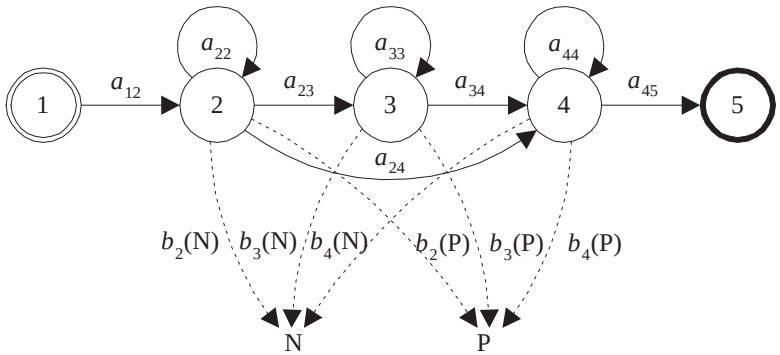
$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

Given the output sequence: N, N, P, N.

State Sequence	Transition Probability	Emission Probability	Sequence Probability
1, 2, 2, 2, 4, 5	$1 \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{2}{3}\frac{2}{3} = \frac{4}{81}$	$\frac{1}{1296}$

Hidden Markov Models: all the alternative paths

The hidden model



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 1/3 & 1/5 & 2/3 & 0 \\ 0 & 2/3 & 4/5 & 1/3 & 0 \end{bmatrix}$$

Given the output sequence: N, N, P, N.

State Sequence	Transition Probability	Emission Probability	Sequence Probability
1, 2, 2, 2, 4, 5	$1\ \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{2}{3}\frac{2}{3} = \frac{4}{81}$	$\frac{1}{1296}$
1, 2, 2, 3, 4, 5	$1\ \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{4}{5}\frac{2}{3} = \frac{8}{135}$	$\frac{1}{1080}$

Hidden Markov Models: all the alternative paths

Given the output sequence $\mathbf{x} =$ N, N, P, N.

$\mathbf{q} = 1, q_1, q_2, q_3, q_4, 5$	$\prod_t a_{q_{t-1}q_t}$	$\prod_t b_{q_t}(x_t)$	$\prod_t b_{q_t}(x_t)\, a_{q_{t-1}q_t}$
1, 2, 2, 2, 4, 5	$1\ \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{2}{3}\frac{2}{3} = \frac{4}{81}$	$\frac{1}{1296}$
1, 2, 2, 3, 4, 5	$1\ \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{4}{5}\frac{2}{3} = \frac{8}{135}$	$\frac{1}{1080}$
1, 2, 2, 4, 4, 5	$1\ \frac{1}{4}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{32}$	$\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{2}{3} = \frac{2}{81}$	$\frac{1}{1296}$
1, 2, 3, 3, 4, 5	$1\ \frac{1}{4}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{32}$	$\frac{1}{3}\frac{1}{5}\frac{4}{5}\frac{2}{3} = \frac{8}{225}$	$\frac{1}{900}$
1, 2, 3, 4, 4, 5	$1\ \frac{1}{4}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{32}$	$\frac{1}{3}\frac{1}{5}\frac{1}{3}\frac{2}{3} = \frac{2}{135}$	$\frac{1}{2160}$
1, 2, 4, 4, 4, 5	$1\ \frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{16}$	$\frac{1}{3}\frac{2}{3}\frac{1}{3}\frac{2}{3} = \frac{4}{81}$	$\frac{1}{324}$

Total Probability $\sum = \frac{77}{10800} \approx 0.007$

$$\Pr(\mathbf{x}|A,B) = \sum_{\forall \mathbf{q}} \prod_t b_{q_t}(x_t)\, a_{q_{t-1}q_t}$$

Hidden Markov Models: all the alternative paths

Given the output sequence $\mathbf{x} =$ N, N, P, N.

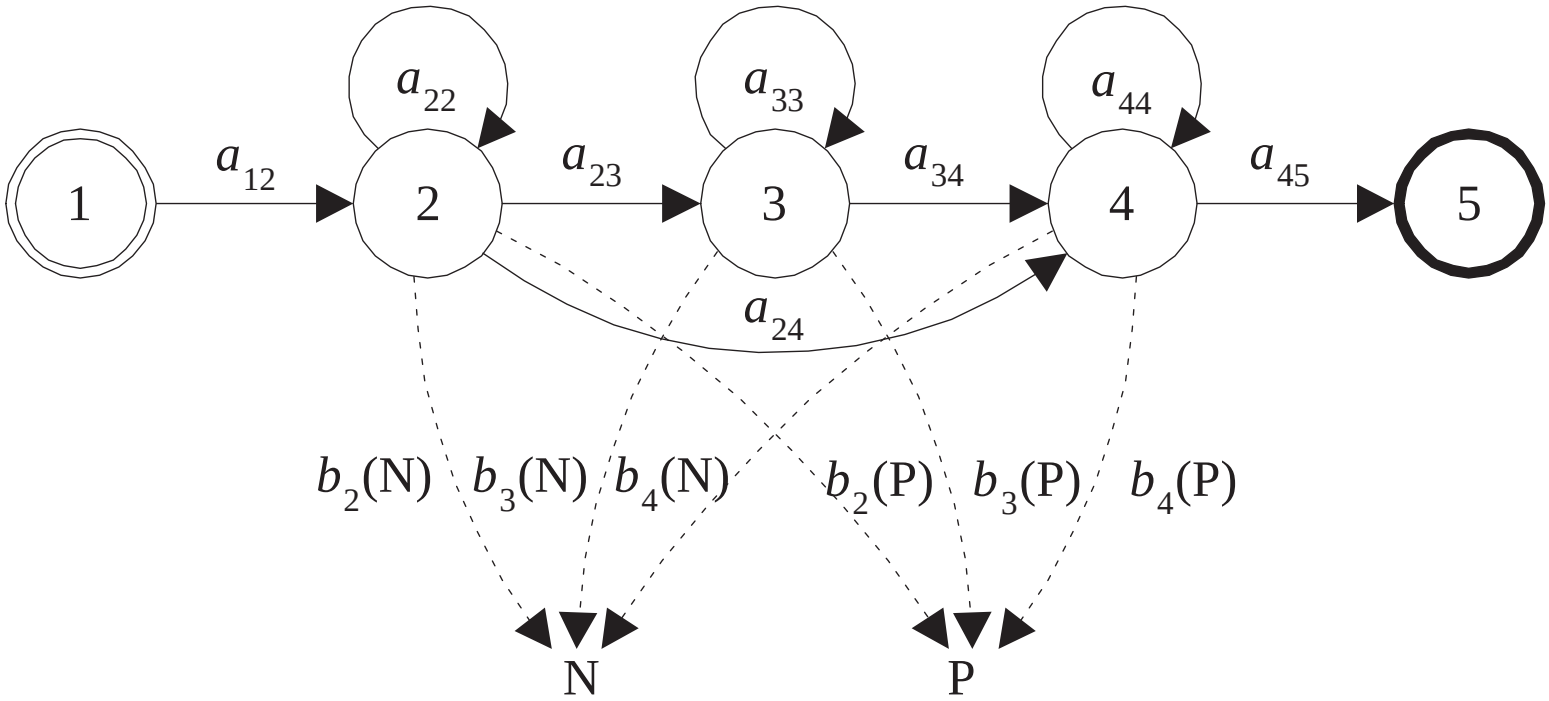
$\mathbf{q} = 1, q_1, q_2, q_3, q_4, 5$	$\prod_t a_{q_{t-1}q_t}$	$\prod_t b_{q_t}(x_t)$	$\prod_t b_{q_t}(x_t)\, a_{q_{t-1}q_t}$
1, 2, 2, 2, 4, 5	$1\ \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{2}{3}\frac{2}{3} = \frac{4}{81}$	$\frac{1}{1296}$
1, 2, 2, 3, 4, 5	$1\ \frac{1}{4}\frac{1}{4}\frac{1}{2}\frac{1}{2} = \frac{1}{64}$	$\frac{1}{3}\frac{1}{3}\frac{4}{5}\frac{2}{3} = \frac{8}{135}$	$\frac{1}{1080}$
1, 2, 2, 4, 4, 5	$1\ \frac{1}{4}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{32}$	$\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{2}{3} = \frac{2}{81}$	$\frac{1}{1296}$
1, 2, 3, 3, 4, 5	$1\ \frac{1}{4}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{32}$	$\frac{1}{3}\frac{1}{5}\frac{4}{5}\frac{2}{3} = \frac{8}{225}$	$\frac{1}{900}$
1, 2, 3, 4, 4, 5	$1\ \frac{1}{4}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{32}$	$\frac{1}{3}\frac{1}{5}\frac{1}{3}\frac{2}{3} = \frac{2}{135}$	$\frac{1}{2160}$
1, 2, 4, 4, 4, 5	$1\ \frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{16}$	$\frac{1}{3}\frac{2}{3}\frac{1}{3}\frac{2}{3} = \frac{4}{81}$	$\frac{1}{324}$

Total Probability $\sum = \frac{77}{10800} \approx 0.007$

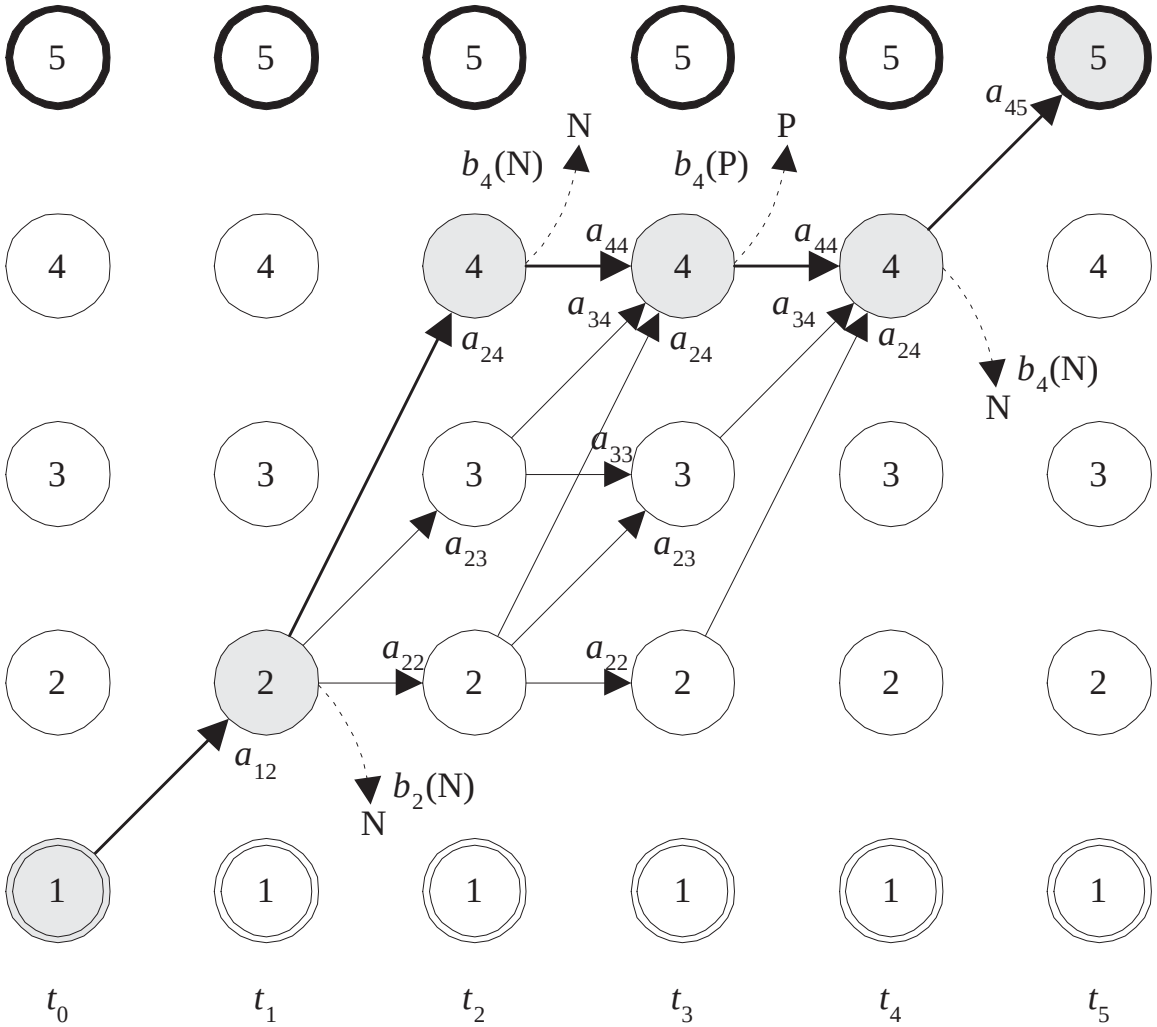
$$\Pr(\mathbf{x}\,|A, B) = \sum_{\forall \mathbf{q}} \prod_t b_{q_t}(x_t)\, a_{q_{t-1}q_t}$$

sinc(i)

HMM through Trellis: the most probable state sequence



HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

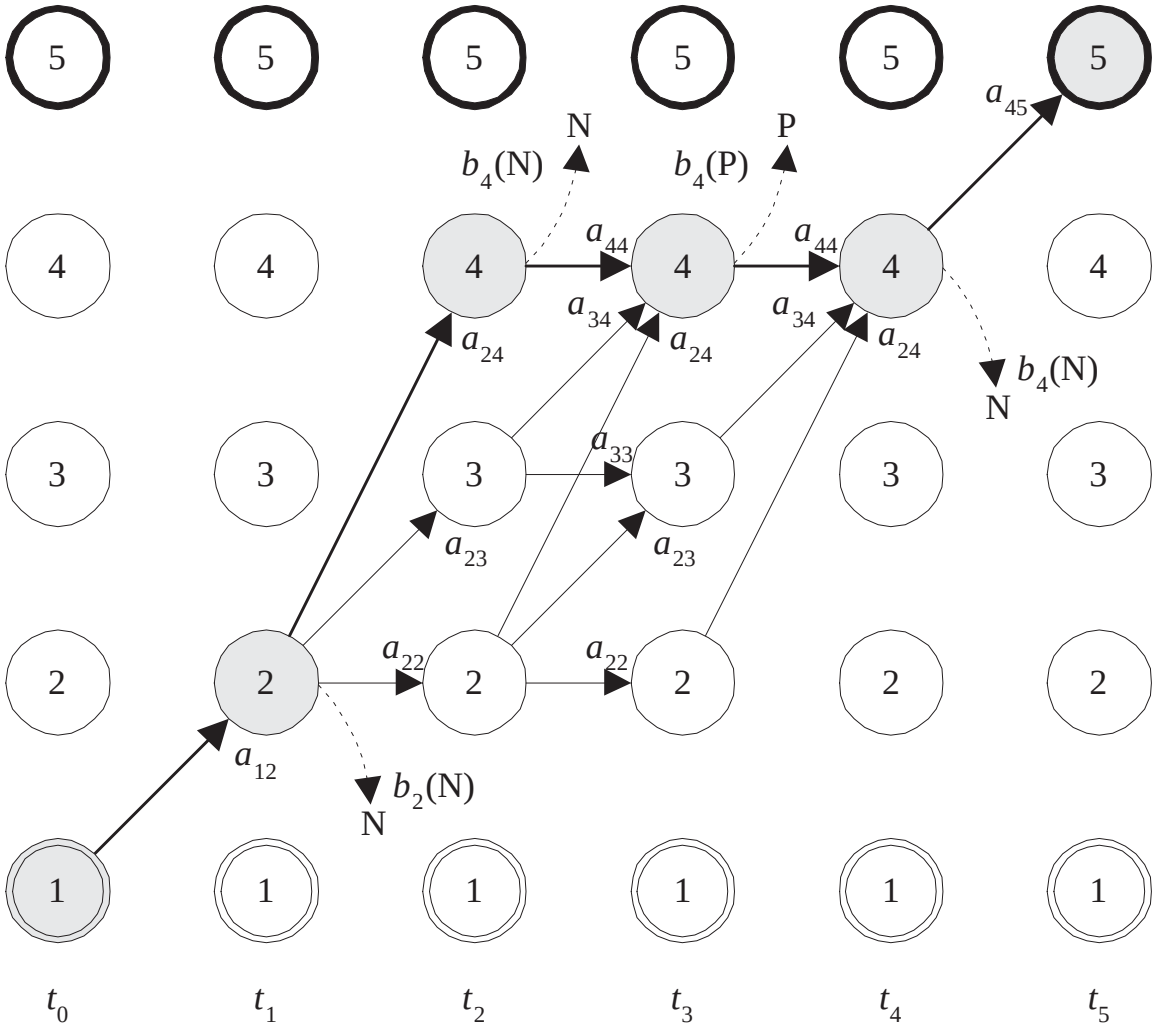
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

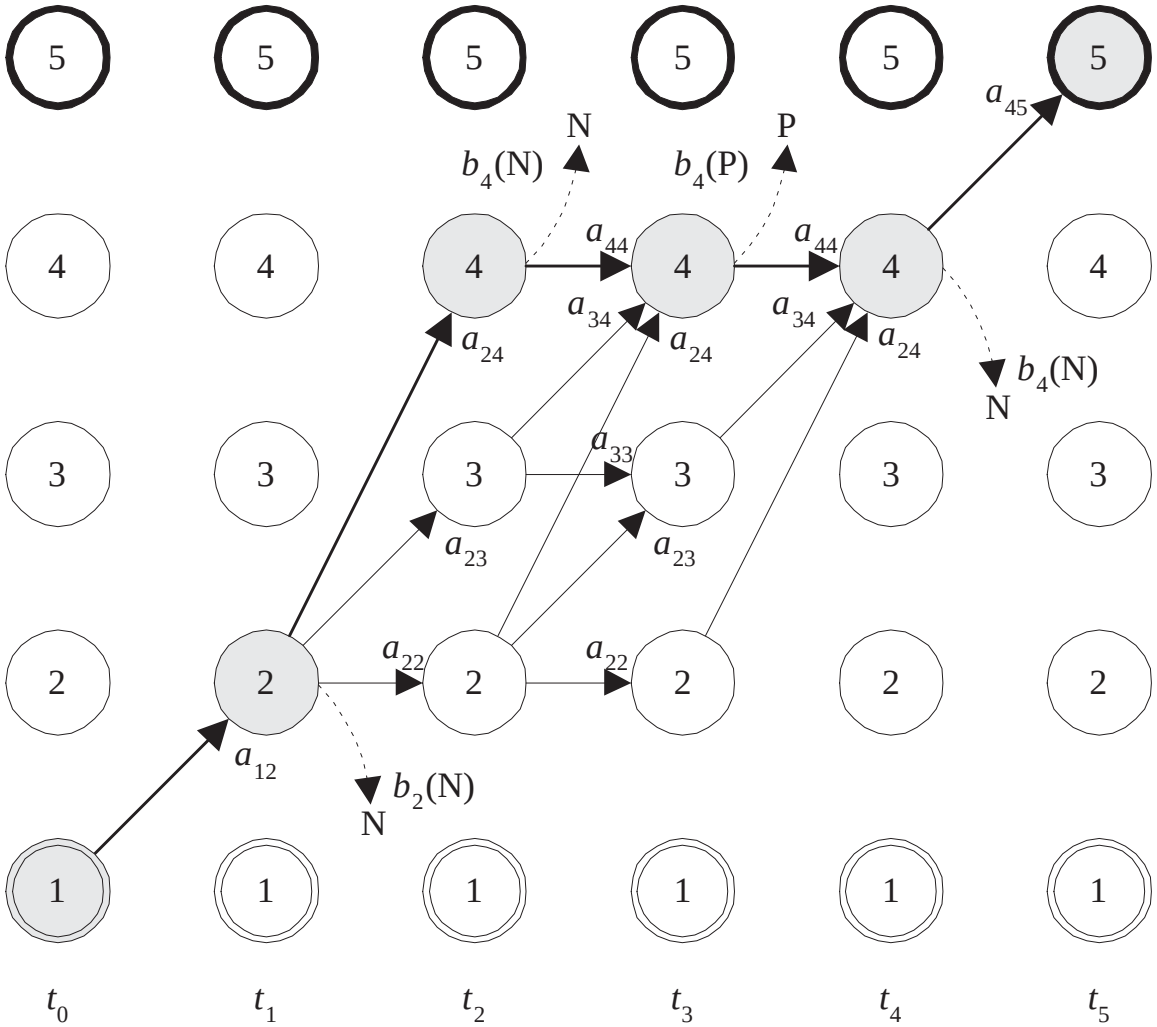
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

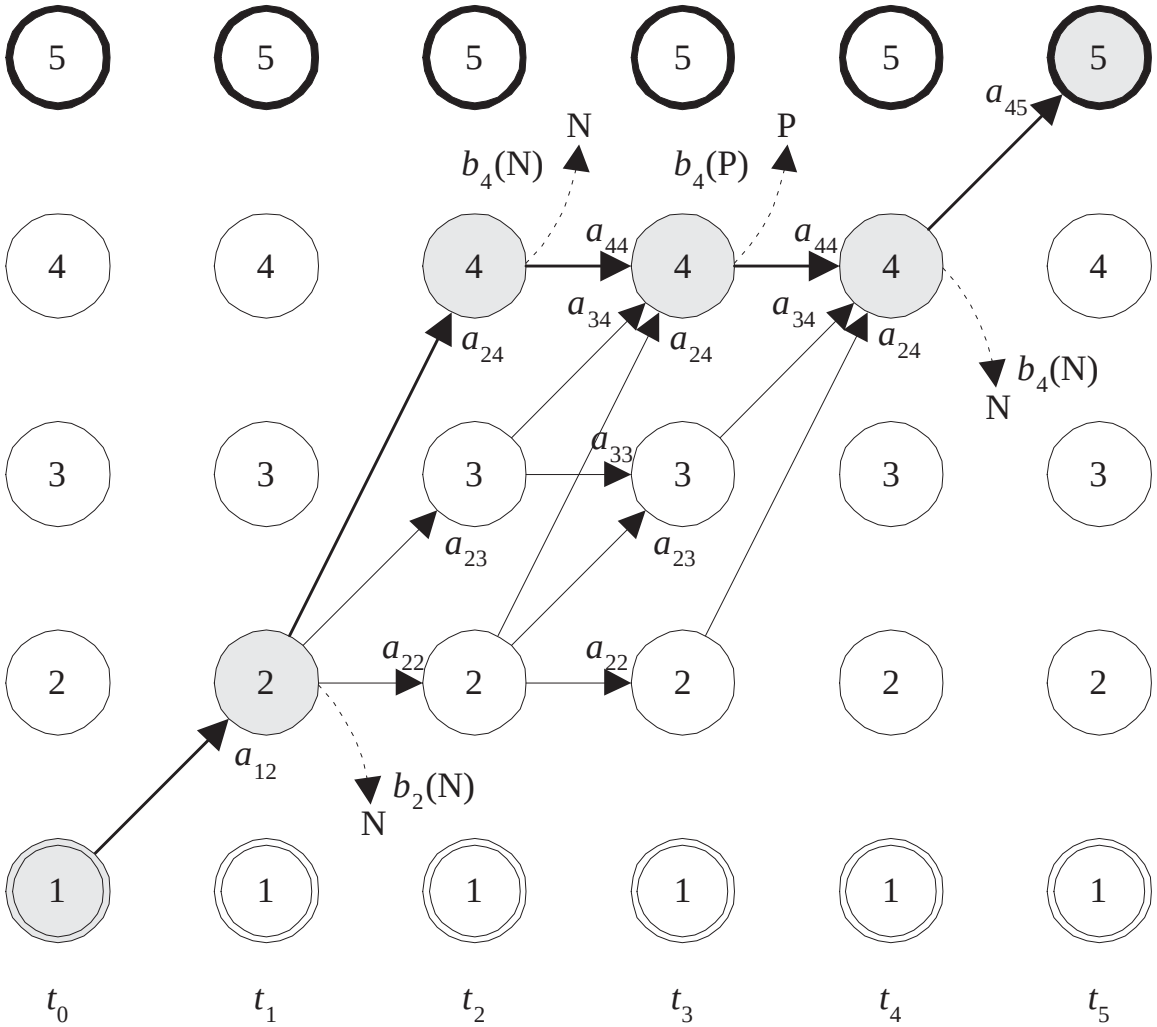
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

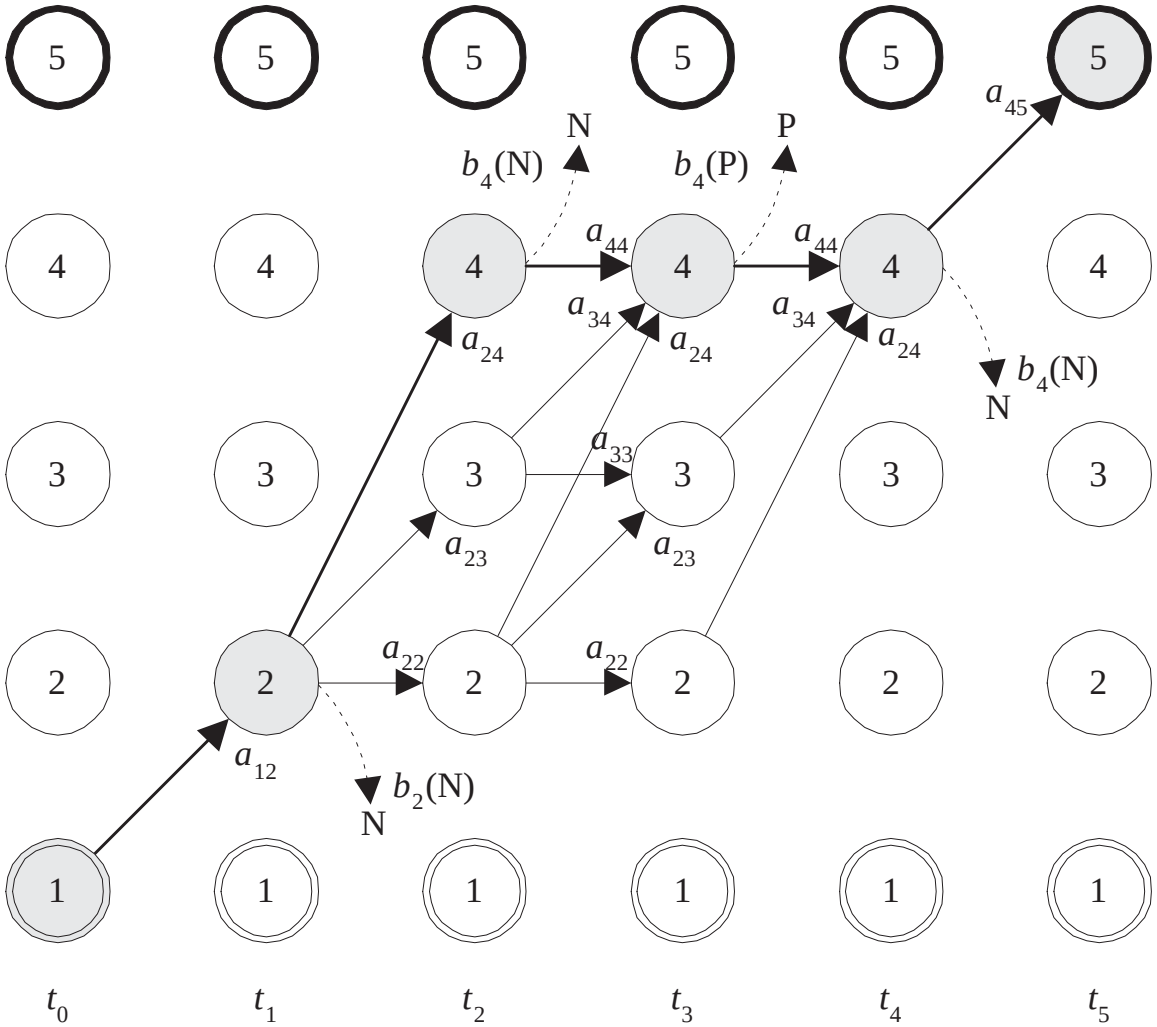
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

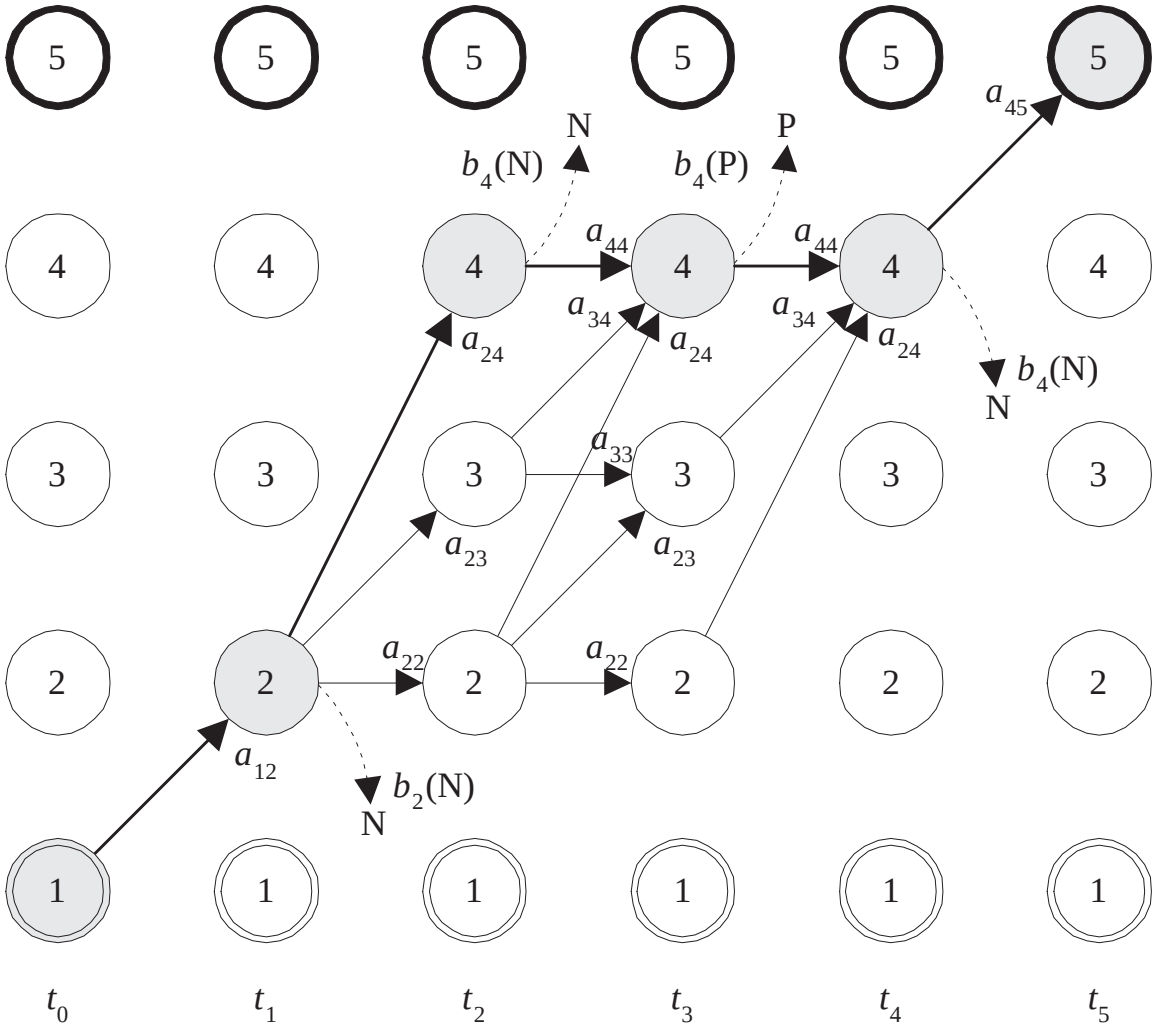
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$$p_{12} = \frac{1}{3}$$

$$p_{122} = \frac{1}{36}$$
$$p_{123} = \frac{1}{60}$$
$$p_{124} = \frac{1}{9}$$

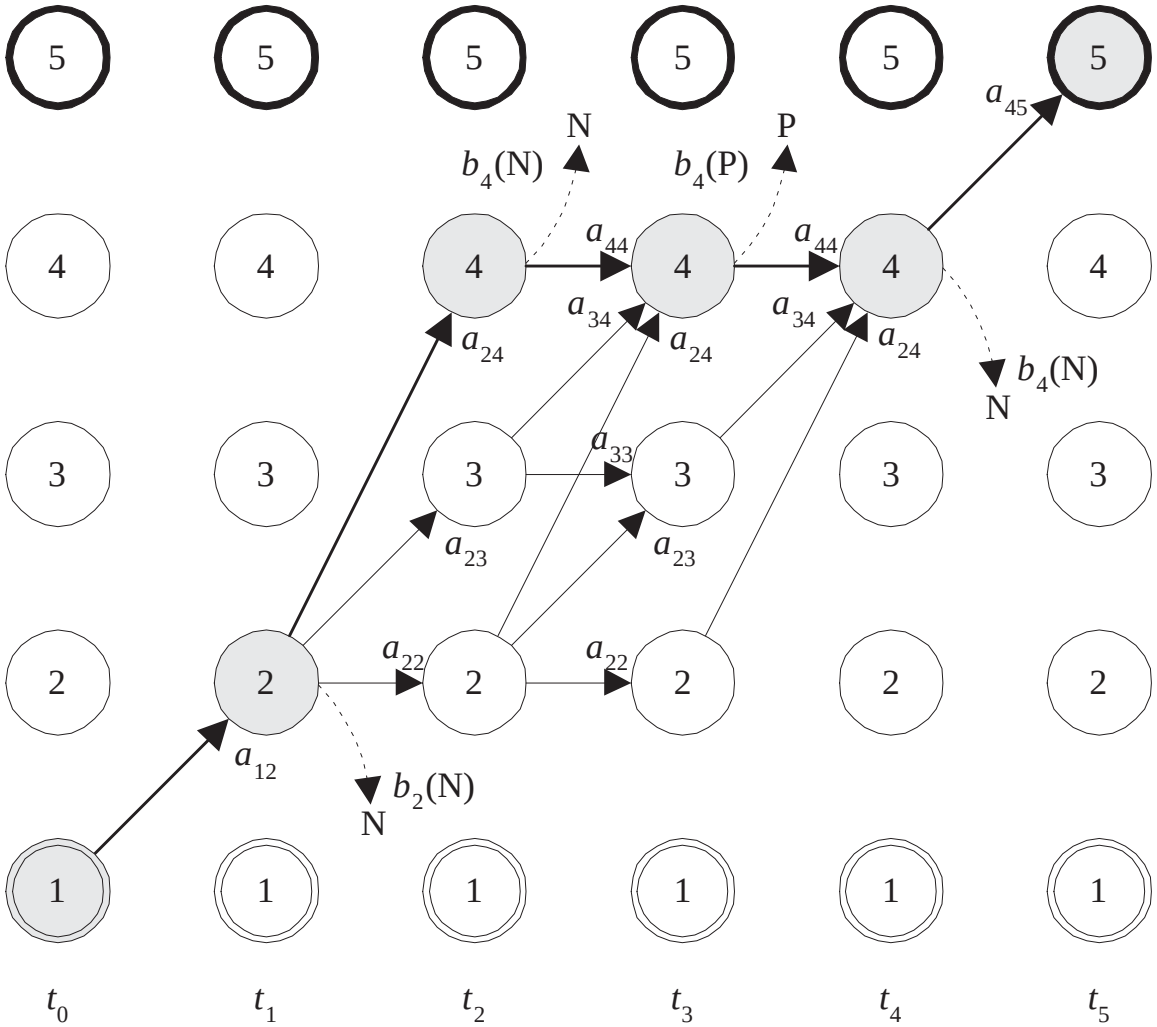
$$p_{1222} = \frac{1}{216}$$
$$p_{1223} = \frac{1}{180}$$
$$p_{1224} = \frac{1}{216}$$
$$p_{1233} = \frac{1}{600}$$
$$p_{1234} = \frac{1}{360}$$
$$p_{1244} = \frac{1}{54}$$

$$p_{12224} = \frac{1}{1296}$$
$$p_{12?34} =$$
$$p_{12?44} =$$

$$p_{12??45} =$$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

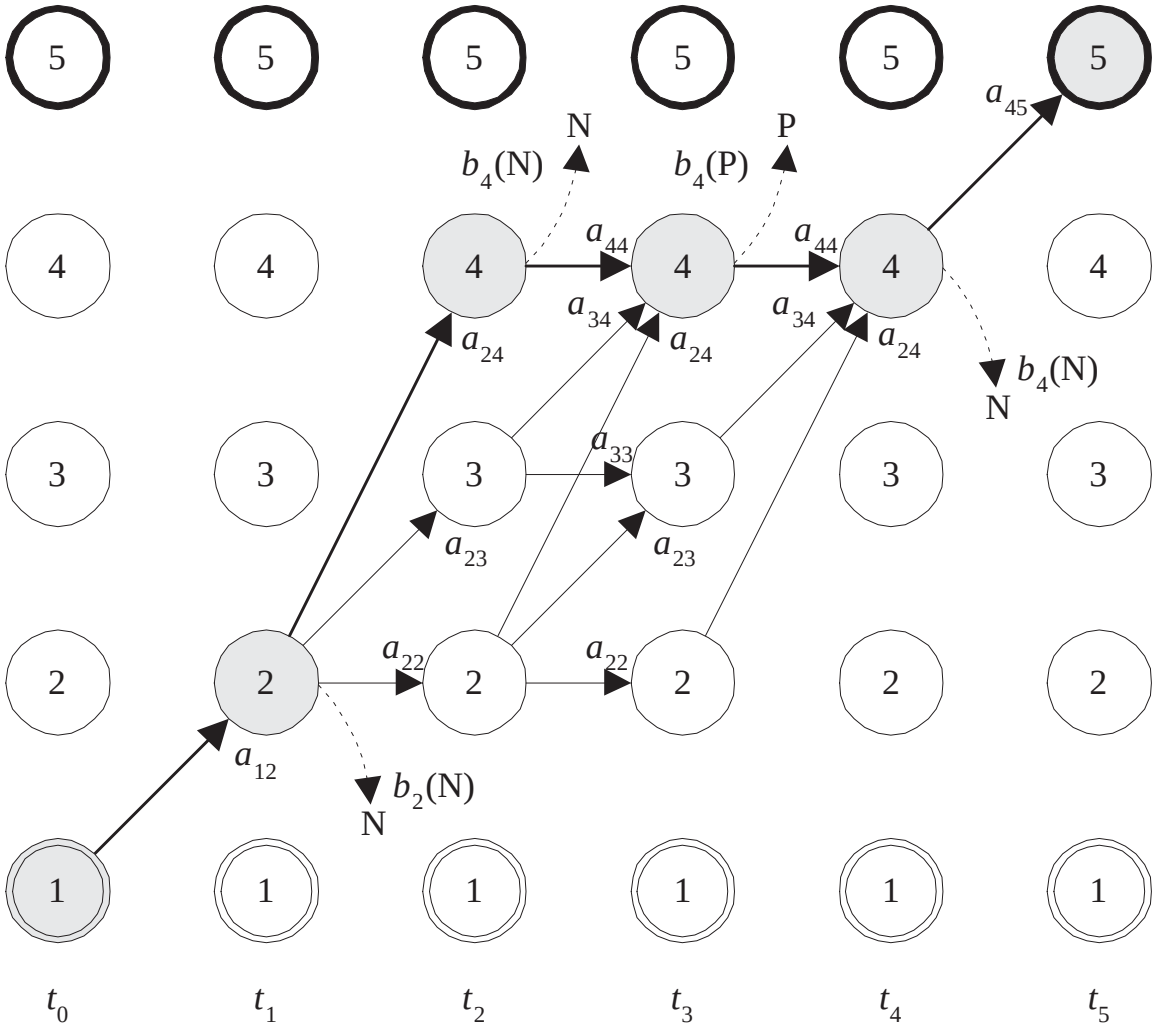
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

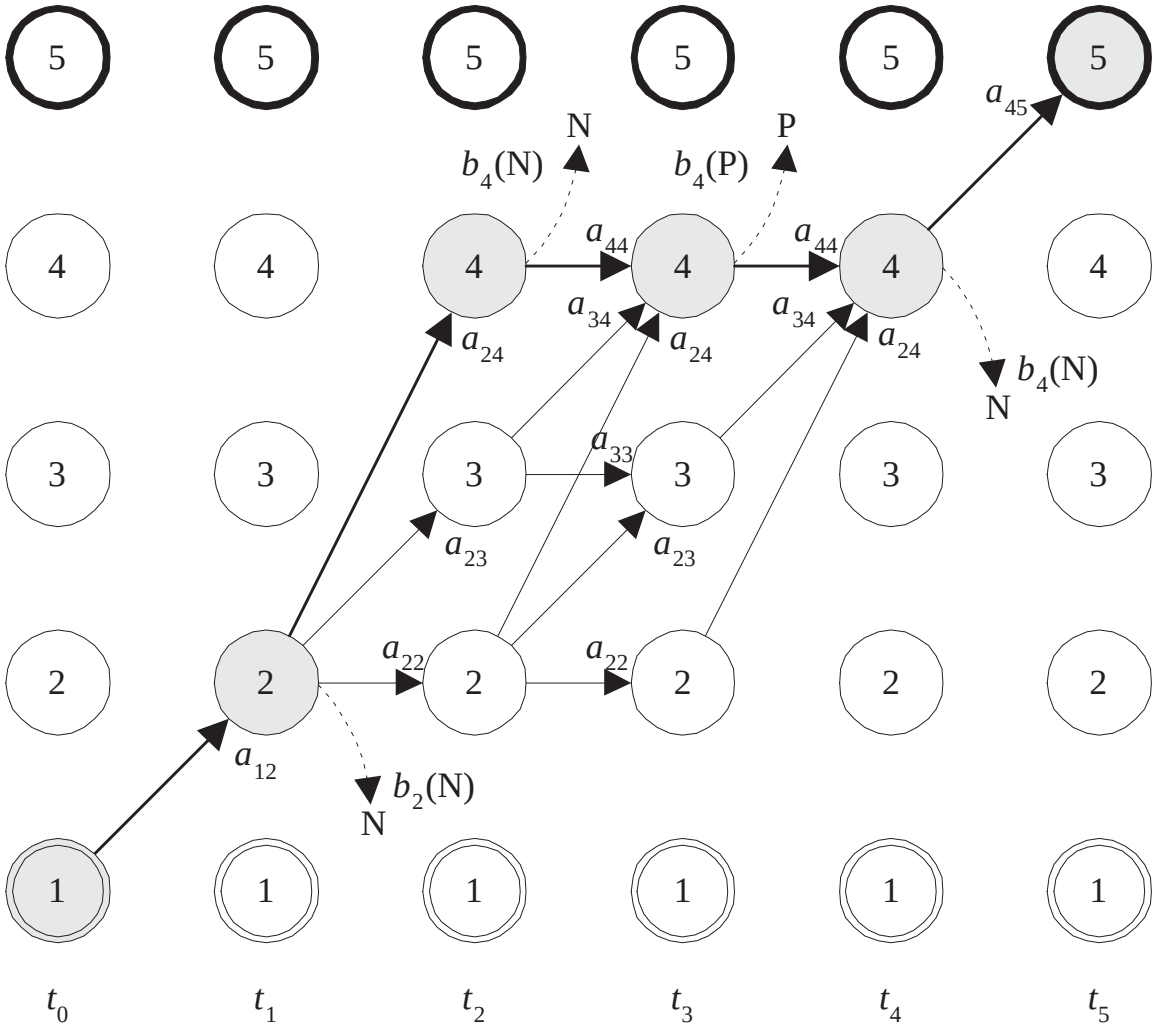
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} =$
$\ldots \max\{p_{1223}, p_{1233}\}$
$p_{12?44} =$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

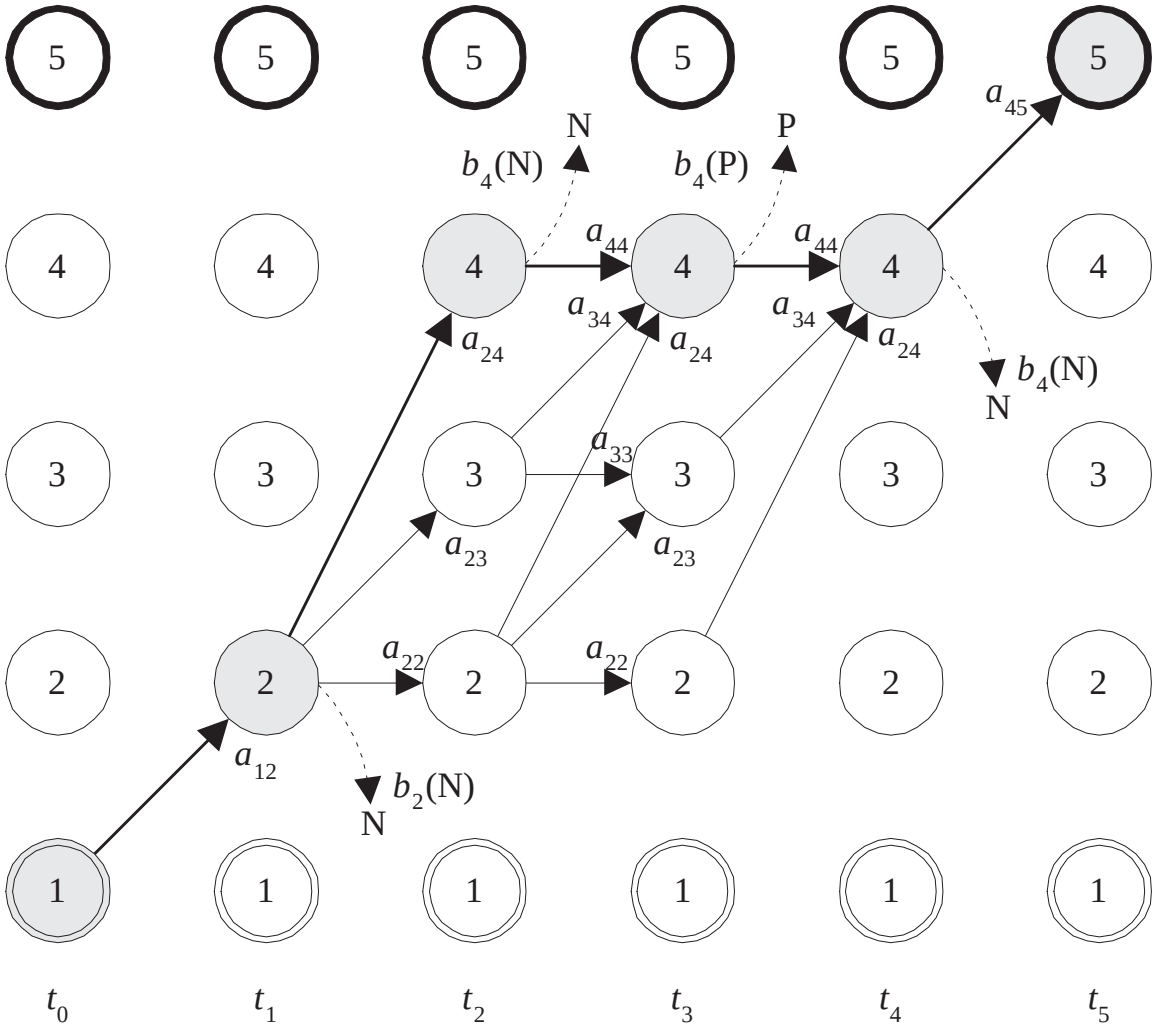
$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} = p_{12234} = \frac{1}{2160}$
$p_{12?44} =$
$\ldots \max\{p_{1224}, p_{1234}, p_{1244}$

$p_{12??45} =$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

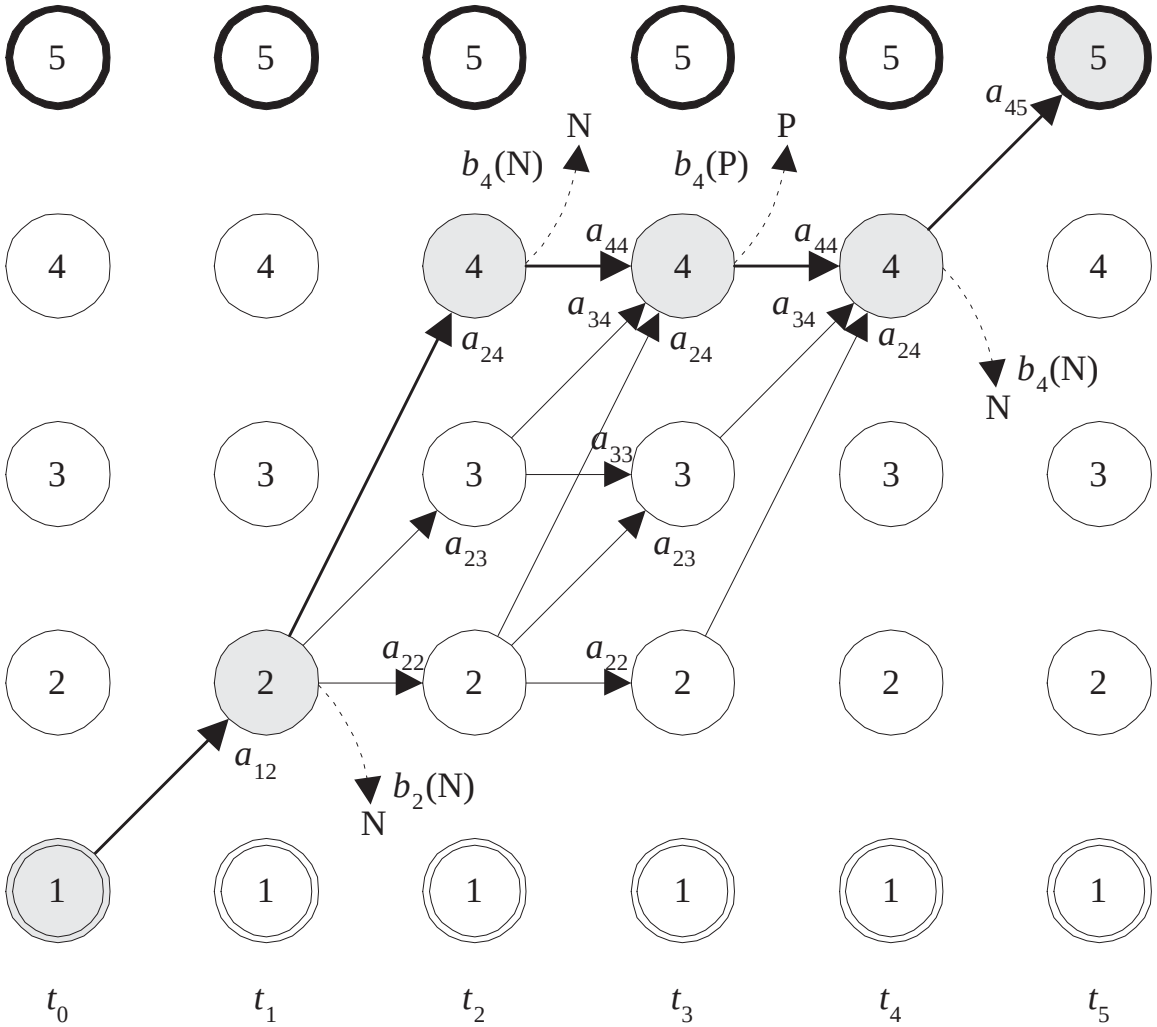
$p_{12224} = \frac{1}{1296}$
$p_{12?34} = p_{12234} = \frac{1}{2160}$
$p_{12?44} = p_{12444} = \frac{1}{162}$

$p_{12??45} =$

$\ldots \max \ldots$

Viterbi

HMM through Trellis: the most probable state sequence



$p_{12} = \frac{1}{3}$

$p_{122} = \frac{1}{36}$
$p_{123} = \frac{1}{60}$
$p_{124} = \frac{1}{9}$

$p_{1222} = \frac{1}{216}$
$p_{1223} = \frac{1}{180}$
$p_{1224} = \frac{1}{216}$
$p_{1233} = \frac{1}{600}$
$p_{1234} = \frac{1}{360}$
$p_{1244} = \frac{1}{54}$

$p_{12224} = \frac{1}{1296}$
$p_{12?34} = p_{12234} = \frac{1}{2160}$
$p_{12?44} = p_{12444} = \frac{1}{162}$

$p_{12??45} = p_{124445} = \frac{1}{324}$

Viterbi

sinc(*i*)

Hidden Markov Models: sum vs max

$$\Pr(\mathbf{x}\,|A,B) = \sum_{\forall \mathbf{q}} \prod_{t} b_{q_t}(x_t)\, a_{q_{t-1}q_t} = \frac{77}{10800} \approx 0.007$$

$$\Pr(\mathbf{x}\,|A,B) \approx \max_{\forall \mathbf{q}} \prod_{t} b_{q_t}(x_t)\, a_{q_{t-1}q_t} = \frac{1}{324} \approx 0.003$$

Simple Ideas for Parameter Estimation

Given a sequence $\mathbf{x}$ of emitted (observed) data...
How can we estimate the transition probabilities A in the *observable* model?

1. Count all the transitions from state i to state j: $\gamma(i, j)$
2. Count all the times that the model was in state i: $\gamma(i)$
3. Then estimate

$$\tilde{a}_{ij} = \frac{\gamma(i, j)}{\gamma(i)}$$

Simple Ideas for Parameter Estimation

Given a sequence $\mathbf{x}$ of emitted (observed) data...
How can we estimate the transition probabilities A in the *observable* model?

1. Count all the transitions from state i to state j: $\gamma(i, j)$
2. Count all the times that the model was in state i: $\gamma(i)$
3. Then estimate

$$\tilde{a}_{ij} = \frac{\gamma(i, j)}{\gamma(i)}$$

Simple Ideas for Parameter Estimation

Given a sequence $\mathbf{x}$ of emitted (observed) data...
How can we estimate the transition probabilities A in the *observable* model?

1. Count all the transitions from state i to state j: $\gamma(i, j)$
2. Count all the times that the model was in state i: $\gamma(i)$
3. Then estimate

$$\tilde{a}_{ij} = \frac{\gamma(i, j)}{\gamma(i)}$$

Simple Ideas for Parameter Estimation

Given a sequence $\mathbf{x}$ of emitted (observed) data...
How can we estimate the transition probabilities A in the *observable* model?

1. Count all the transitions from state i to state j: $\gamma(i, j)$
2. Count all the times that the model was in state i: $\gamma(i)$
3. Then estimate

$$\tilde{a}_{ij} = \frac{\gamma(i, j)}{\gamma(i)}$$

Simple Ideas for Parameter Estimation

And how can we estimate the transition probabilities A in the *hidden* model?

1. Find the most probable sequence of hidden states (Viterbi)
2. Make the counts $\gamma(i,j)$ and $\gamma(i)$ as before
3. Estimate $\tilde{a}_{ij} = \frac{\gamma(i,j)}{\gamma(i)}$

And the emission probabilities B?

1. Find the most probable sequence of hidden states (Viterbi)
2. Count all the times that the model was in state j but emitting the output k: $\delta_j(k)$
3. Count all the times that the model was in state j: $\gamma(j)$
4. Estimate $\tilde{b}_j(k) = \frac{\delta_j(k)}{\gamma(j)}$

Simple Ideas for Parameter Estimation

And how can we estimate the transition probabilities A in the *hidden* model?

1. Find the most probable sequence of hidden states (Viterbi)
2. Make the counts $\gamma(i,j)$ and $\gamma(i)$ as before
3. Estimate $\tilde{a}_{ij} = \frac{\gamma(i,j)}{\gamma(i)}$

And the emission probabilities B?

1. Find the most probable sequence of hidden states (Viterbi)
2. Count all the times that the model was in state j but emitting the output k: $\delta_j(k)$
3. Count all the times that the model was in state j: $\gamma(j)$
4. Estimate $\tilde{b}_j(k) = \frac{\delta_j(k)}{\gamma(j)}$

Simple Ideas for Parameter Estimation

And how can we estimate the transition probabilities A in the *hidden* model?

1. Find the most probable sequence of hidden states (Viterbi)
2. Make the counts $\gamma(i,j)$ and $\gamma(i)$ as before
3. Estimate $\tilde{a}_{ij} = \frac{\gamma(i,j)}{\gamma(i)}$

And the emission probabilities B?

1. Find the most probable sequence of hidden states (Viterbi)
2. Count all the times that the model was in state j but emitting the output k: $\delta_j(k)$
3. Count all the times that the model was in state j: $\gamma(j)$
4. Estimate $\tilde{b}_j(k) = \frac{\delta_j(k)}{\gamma(j)}$

Simple Ideas for Parameter Estimation

And how can we estimate the transition probabilities A in the *hidden* model?

1. Find the most probable sequence of hidden states (Viterbi)
2. Make the counts $\gamma(i,j)$ and $\gamma(i)$ as before
3. Estimate $\tilde{a}_{ij} = \frac{\gamma(i,j)}{\gamma(i)}$

And the emission probabilities B?

1. Find the most probable sequence of hidden states (Viterbi)
2. Count all the times that the model was in state j but emitting the output k: $\delta_j(k)$
3. Count all the times that the model was in state j: $\gamma(j)$
4. Estimate $\tilde{b}_j(k) = \frac{\delta_j(k)}{\gamma(j)}$

sinc(i)

Outline

sinc(i)

Definition

To model a sequence $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$, a Hidden Markov Model can be defined with the structure

$$\Theta = \langle \mathcal{Q}, \mathbf{A}, \boldsymbol{\pi}, \mathcal{B} \rangle,$$

where:

i) $\mathcal{Q} = \{Q \in [1 \ldots N_Q]\}$ is the set of states.

ii) $\mathbf{A} = [a_{ij} = \Pr(Q_t = j \,|Q_{t-1} = i)], \forall i, j \in \mathcal{Q}$, is the matrix of transition probabilities, where $Q_t \in \mathcal{Q}$ is the model state at time $t \in [1 \ldots T]$, $a_{ij} \geq 0 \ \forall i, j$ and $\sum_j a_{ij} \doteq 1 \ \forall i$.

iii) $\boldsymbol{\pi} = [\pi_j = \Pr(Q_1 = j)]$ is the initial state probability vector. In the case of left-to-right HMM this vector is $\boldsymbol{\pi} = \boldsymbol{\delta}_1$.

iv) $\mathcal{B} = \{b_k(\mathbf{x}_t) = \Pr(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = k)\}, \forall k \in \mathcal{Q}$, is the set of observation (or emission) probability distributions.

sinc(i)

Definition

To model a sequence $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$, a Hidden Markov Model can be defined with the structure

$$\Theta = \langle \mathcal{Q}, \mathbf{A}, \boldsymbol{\pi}, \mathcal{B} \rangle,$$

where:

i) $\mathcal{Q} = \{Q \in [1 \ldots N_Q]\}$ is the set of states.

ii) $\mathbf{A} = [a_{ij} = \Pr(Q_t = j \,|Q_{t-1} = i)], \forall i, j \in \mathcal{Q}$, is the matrix of transition probabilities, where $Q_t \in \mathcal{Q}$ is the model state at time $t \in [1 \ldots T]$, $a_{ij} \geq 0 \ \forall i, j$ and $\sum_j a_{ij} \doteq 1 \ \forall i$.

iii) $\boldsymbol{\pi} = [\pi_j = \Pr(Q_1 = j)]$ is the initial state probability vector. In the case of left-to-right HMM this vector is $\boldsymbol{\pi} = \boldsymbol{\delta}_1$.

iv) $\mathcal{B} = \{b_k(\mathbf{x}_t) = \Pr(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = k)\}, \forall k \in \mathcal{Q}$, is the set of observation (or emission) probability distributions.

sinc(i)

Definition

To model a sequence $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$, a Hidden Markov Model can be defined with the structure

$$\Theta = \langle \mathcal{Q}, \mathbf{A}, \boldsymbol{\pi}, \mathcal{B} \rangle,$$

where:

i) $\mathcal{Q} = \{Q \in [1 \ldots N_Q]\}$ is the set of states.

ii) $\mathbf{A} = [a_{ij} = \Pr(Q_t = j \,|Q_{t-1} = i)], \forall i, j \in \mathcal{Q}$, is the matrix of transition probabilities, where $Q_t \in \mathcal{Q}$ is the model state at time $t \in [1 \ldots T]$, $a_{ij} \geq 0 \ \forall i, j$ and $\sum_j a_{ij} \doteq 1 \ \forall i$.

iii) $\boldsymbol{\pi} = [\pi_j = \Pr(Q_1 = j)]$ is the initial state probability vector. In the case of left-to-right HMM this vector is $\boldsymbol{\pi} = \boldsymbol{\delta}_1$.

iv) $\mathcal{B} = \{b_k(\mathbf{x}_t) = \Pr(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = k)\}, \forall k \in \mathcal{Q}$, is the set of observation (or emission) probability distributions.

sinc(i)

Definition

To model a sequence $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$, a Hidden Markov Model can be defined with the structure

$$\Theta = \langle \mathcal{Q}, \mathbf{A}, \boldsymbol{\pi}, \mathcal{B} \rangle,$$

where:

i) $\mathcal{Q} = \{Q \in [1 \ldots N_Q]\}$ is the set of states.

ii) $\mathbf{A} = [a_{ij} = \Pr(Q_t = j \,|Q_{t-1} = i)], \forall i, j \in \mathcal{Q}$, is the matrix of transition probabilities, where $Q_t \in \mathcal{Q}$ is the model state at time $t \in [1 \ldots T]$, $a_{ij} \geq 0 \; \forall i, j$ and $\sum_j a_{ij} \doteq 1 \; \forall i$.

iii) $\boldsymbol{\pi} = [\pi_j = \Pr(Q_1 = j)]$ is the initial state probability vector. In the case of left-to-right HMM this vector is $\boldsymbol{\pi} = \boldsymbol{\delta}_1$.

iv) $\mathcal{B} = \{b_k(\mathbf{x}_t) = \Pr(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = k)\}, \forall k \in \mathcal{Q}$, is the set of observation (or emission) probability distributions.

sinc(i)

CHMM with Gaussian Mixture Models

For continuous HMM using Gaussian Mixture Models we define

$$b_j(\mathbf{x}_t) = \sum_{k=1}^{N_c} c_{jk} b_{jk}(\mathbf{x}_t) \quad \forall j \in \mathcal{Q};\ N_c < \infty$$

where

i) $b_{jk}(\mathbf{x}_t)$ are normal distributions

$$\mathcal{N}(\mathbf{x}_t, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk}) = \frac{1}{(2\pi)^{N_x} |\mathbf{U}_{jk}|^{\frac{1}{2}}} e^{-\frac{1}{2}\left[(\mathbf{x}_t - \boldsymbol{\mu}_{jk})^{\mathrm{T}} \mathbf{U}_{jk}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_{jk})\right]},$$

ii) $c_{jk} \in \mathbb{R}^{+0}$ are the mixture weights, satisfying $\sum_{k=1}^{N_c} c_{jk} \doteq 1 \ \ \forall j \in \mathcal{Q}$,

iii) $\boldsymbol{\mu}_{jk} \in \mathbb{R}^{N_x}$ are the mean vectors and,

iv) $\mathbf{U}_{jk} \in \mathbb{R}^{N_x \times N_x}$ are the covariance matrices, with $\int_{-\infty}^{+\infty} b_j(\mathbf{x}_t)\, d\mathbf{x}_t \doteq 1 \ \ \forall j \in \mathcal{Q}$

CHMM with Gaussian Mixture Models

For continuous HMM using Gaussian Mixture Models we define

$$b_j(\mathbf{x}_t) = \sum_{k=1}^{N_c} c_{jk} b_{jk}(\mathbf{x}_t) \quad \forall j \in \mathcal{Q};\ N_c < \infty$$

where

i) $b_{jk}(\mathbf{x}_t)$ are normal distributions

$$\mathcal{N}(\mathbf{x}_t, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk}) = \frac{1}{(2\pi)^{N_x} |\mathbf{U}_{jk}|^{\frac{1}{2}}} e^{-\frac{1}{2}\left[(\mathbf{x}_t - \boldsymbol{\mu}_{jk})^{\mathrm{T}} \mathbf{U}_{jk}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_{jk})\right]},$$

ii) $c_{jk} \in \mathbb{R}^{+0}$ are the mixture weights, satisfying $\sum_{k=1}^{N_c} c_{jk} \doteq 1 \quad \forall j \in \mathcal{Q}$,

iii) $\boldsymbol{\mu}_{jk} \in \mathbb{R}^{N_x}$ are the mean vectors and,

iv) $\mathbf{U}_{jk} \in \mathbb{R}^{N_x \times N_x}$ are the covariance matrices, with $\int_{-\infty}^{+\infty} b_j(\mathbf{x}_t)\, \mathrm{d}\mathbf{x}_t \doteq 1 \quad \forall j \in \mathcal{Q}$

sinc(i)

Main Hypothesis for HMM

The *first-order Markov* hypothesis states that the history has no influence on the chain's future evolution if the present is specified.

For a discrete-time, first order, Markov chain, the probabilistic dependence is truncated to

$$\Pr\left(Q_t = j \,|Q_{t-1} = i, Q_{t-2} = k, \ldots\right) = \Pr\left(Q_t = j \,|Q_{t-1} = i\right)$$

The *output independence* hypothesis states that neither chain evolution nor past observations influence the present observation if the last chain transition is specified. Given the sequences of states $\mathbf{q}$

$$\Pr\left(\mathbf{X} \,|\mathbf{q}\right) = \prod_t \Pr\left(\mathbf{X}_t = \mathbf{x}_t \,|\mathbf{q}\right) = \prod_t \Pr\left(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = q_t\right)$$

Main Hypothesis for HMM

The *first-order Markov* hypothesis states that the history has no influence on the chain's future evolution if the present is specified.
For a discrete-time, first order, Markov chain, the probabilistic dependence is truncated to

$$\Pr\left(Q_t = j \,|Q_{t-1} = i, Q_{t-2} = k, \ldots\right) = \Pr\left(Q_t = j \,|Q_{t-1} = i\right)$$

The *output independence* hypothesis states that neither chain evolution nor past observations influence the present observation if the last chain transition is specified. Given the sequences of states $\mathbf{q}$

$$\Pr\left(\mathbf{X} \,|\mathbf{q}\right) = \prod_t \Pr\left(\mathbf{X}_t = \mathbf{x}_t \,|\mathbf{q}\right) = \prod_t \Pr\left(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = q_t\right)$$

Main Hypothesis for HMM

The *first-order Markov* hypothesis states that the history has no influence on the chain's future evolution if the present is specified.
For a discrete-time, first order, Markov chain, the probabilistic dependence is truncated to

$$\Pr\left(Q_t = j \,|Q_{t-1} = i, Q_{t-2} = k, \ldots\right) = \Pr\left(Q_t = j \,|Q_{t-1} = i\right)$$

The *output independence* hypothesis states that neither chain evolution nor past observations influence the present observation if the last chain transition is specified. Given the sequences of states $\mathbf{q}$

$$\Pr\left(\mathbf{X} \,|\mathbf{q}\right) = \prod_t \Pr\left(\mathbf{X}_t = \mathbf{x}_t \,|\mathbf{q}\right) = \prod_t \Pr\left(\mathbf{X}_t = \mathbf{x}_t \,|Q_t = q_t\right)$$

sinc(i)

HMM Likelihood

The HMM likelihood can be defined using the probability of the observed data given the model

$$
\begin{aligned}
\mathcal{L}_\Theta(\mathbf{X}) = \sum_{\forall \mathbf{q}} \mathcal{L}_\Theta(\mathbf{X}, \mathbf{q}) \quad &\triangleq \quad \sum_{\forall \mathbf{q}} \Pr\left(\mathbf{X}, \mathbf{q} \,|\Theta\right) \\
&= \quad \sum_{\forall \mathbf{q}} \left\{\Pr\left(\mathbf{X} \,|\mathbf{q}, \Theta\right) \Pr\left(\mathbf{q} \,|\Theta\right)\right\} \\
&= \quad \sum_{\forall \mathbf{q}} \left\{\prod_t \Pr\left(\mathbf{x}_t \,|q_t, \Theta\right) \prod_{t=2} \Pr\left(q_t \,|q_{t-1}, \Theta\right)\right\} \\
&= \quad \sum_{\forall \mathbf{q}} \left\{\prod_t b_{q_t}\left(\mathbf{x}_t\right) \prod_{t=2}^{T} a_{q_{t-1} q_t}\right\} \\
&= \quad \sum_{\forall \mathbf{q}} \prod_t a_{q_{t-1} q_t} b_{q_t}\left(\mathbf{x}_t\right)
\end{aligned}
$$

where $a_{01} = \pi_1 = 1$.

HMM: Auxiliary Function

Define the auxiliary function using the log-likelihood

$$\mathcal{O}(\Theta, \tilde{\Theta}) \triangleq \frac{1}{\Pr(\mathbf{X}|\Theta)} \sum_{\forall \mathbf{q}} \Pr(\mathbf{X}, \mathbf{q}|\Theta) \log \Pr(\mathbf{X}, \mathbf{q}|\tilde{\Theta})$$

and replacing $b_j(\mathbf{x}_t) = \sum_k c_{jk} b_{jk}(\mathbf{x}_t)$ in $\Pr(\mathbf{X}, \mathbf{q}|\Theta) = \prod_t a_{q_{t-1}q_t} b_{q_t}(\mathbf{x}_t)$ we can expand

$$\Pr(\mathbf{X}, \mathbf{q}|\Theta) = \sum_{k_1} \sum_{k_2} \cdots \sum_{k_T} \left\{ \prod_t b_{q_t k_t}(\mathbf{x}_t) a_{q_{t-1}q_t} \right\} c_{q_1 k_1} c_{q_2 k_2} \cdots c_{q_T k_T}$$

HMM: Auxiliary Function

Define the auxiliary function using the log-likelihood

$$\mathcal{O}(\Theta, \tilde{\Theta}) \triangleq \frac{1}{\Pr(\mathbf{X}|\Theta)} \sum_{\forall \mathbf{q}} \Pr(\mathbf{X}, \mathbf{q}|\Theta) \log \Pr(\mathbf{X}, \mathbf{q}|\tilde{\Theta})$$

and replacing $b_j(\mathbf{x}_t) = \sum_k c_{jk} b_{jk}(\mathbf{x}_t)$ in $\Pr(\mathbf{X}, \mathbf{q}|\Theta) = \prod_t a_{q_{t-1}q_t} b_{q_t}(\mathbf{x}_t)$ we can expand

$$\Pr(\mathbf{X}, \mathbf{q}|\Theta) = \sum_{k_1} \sum_{k_2} \cdots \sum_{k_T} \left\{ \prod_t b_{q_t k_t}(\mathbf{x}_t) a_{q_{t-1}q_t} \right\} c_{q_1 k_1} c_{q_2 k_2} \cdots c_{q_T k_T}$$

HMM: Auxiliary Function

... and thus

$$\Pr\left(\mathbf{X}\,|\Theta\right) = \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{c}} \prod_{t} b_{q_t k_t}\left(\mathbf{x}_t\right) a_{q_{t-1} q_t} c_{q_t k_t} = \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{c}} \Pr\left(\mathbf{X}, \mathbf{q}, \mathbf{c}\,|\Theta\right)$$

where $\mathbf{c}$ are all the sequences like $c_{q_1 k_1}, c_{q_2 k_2}, \ldots, c_{q_T k_T}$.

Also we can write

$$\log \Pr\left(\mathbf{X}, \mathbf{q}, \mathbf{c}\,\middle|\tilde{\Theta}\right) = \sum_{t} \log \tilde{b}_{q_t k_t}\left(\mathbf{x}_t\right) + \sum_{t} \log \tilde{a}_{q_{t-1} q_t} + \sum_{t} \log \tilde{c}_{q_t k_t}$$

and the auxiliary function reads

HMM: Auxiliary Function

... and thus

$$\Pr\left(\mathbf{X}\,|\Theta\right) = \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{c}} \prod_{t} b_{q_t k_t}\left(\mathbf{x}_t\right) a_{q_{t-1} q_t} c_{q_t k_t} = \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{c}} \Pr\left(\mathbf{X}, \mathbf{q}, \mathbf{c}\,|\Theta\right)$$

where $\mathbf{c}$ are all the sequences like $c_{q_1 k_1}, c_{q_2 k_2}, \ldots, c_{q_T k_T}$.
Also we can write

$$\log \Pr\left(\mathbf{X}, \mathbf{q}, \mathbf{c}\,\middle|\tilde{\Theta}\right) = \sum_{t} \log \tilde{b}_{q_t k_t}\left(\mathbf{x}_t\right) + \sum_{t} \log \tilde{a}_{q_{t-1} q_t} + \sum_{t} \log \tilde{c}_{q_t k_t}$$

and the auxiliary function reads

HMM: Auxiliary Function

$$
\begin{aligned}
\mathcal{O}(\Theta, \tilde{\Theta}) &= \frac{1}{p(\mathbf{X}|\Theta)} \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{c}} \Pr\left(\mathbf{X}, \mathbf{q}, \mathbf{c} \,|\Theta\right) \\
&\quad \times \left\{ \sum_t \log \tilde{b}_{q_t k_t}\left(\mathbf{x}_t\right) + \sum_t \log \tilde{a}_{q_{t-1} q_t} + \sum_t \log \tilde{c}_{q_t k_t} \right\} \\
&= \mathcal{O}_b(\Theta, \tilde{b}_{jk}) + \mathcal{O}_a(\Theta, \tilde{a}_{ij}) + \mathcal{O}_c(\Theta, \tilde{c}_{jk})
\end{aligned}
$$

where

HMM: Auxiliary Function

$$\mathcal{O}_b(\Theta, \tilde{b}_{jk}) = \sum_j \sum_{\forall \mathbf{c}} \sum_t \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right) \log \tilde{b}_{jk}\left(\mathbf{x}_t\right)$$

$$\mathcal{O}_a(\Theta, \tilde{a}_{ij}) = \sum_i \sum_j \sum_t \sum_{\forall \mathbf{c}} \Pr\left(q_{t-1} = i, q_t = j, \mathbf{c} \,|\mathbf{X}, \Theta\right) \log \tilde{a}_{ij}$$

$$\mathcal{O}_c(\Theta, \tilde{c}_{jk}) = \sum_j \sum_{\forall \mathbf{c}} \sum_t \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right) \log \tilde{c}_{jk}$$

Outline

sinc(i)

HMM Training: Transition Probabilities

We are looking for the extrema but subject to constraints

$$\sum_j \tilde{a}_{ij} \doteq 1 \quad \forall i \in \mathcal{Q}$$

Using Lagrange multipliers

$$\nabla_{\tilde{a}} \left(\mathcal{O}_a(\Theta, \tilde{a}_{ij}) - \sum_i \ell_i \left(\sum_j \tilde{a}_{ij} - 1 \right) \right) = 0$$

replacing and deriving

$$\sum_i \sum_j \sum_t \sum_{\forall \mathbf{c}} \left\{ \Pr\left(q_{t-1} = i, q_t = j, \mathbf{c} \,|\mathbf{X}, \Theta\right) \frac{1}{\tilde{a}_{ij}} \right\} - \ell_i = 0$$

(we can maximize all the terms in i separately)

HMM Training: Transition Probabilities

We are looking for the extrema but subject to constraints

$$\sum_j \tilde{a}_{ij} \doteq 1 \quad \forall i \in \mathcal{Q}$$

Using Lagrange multipliers

$$\nabla_{\tilde{a}} \left(\mathcal{O}_a(\Theta, \tilde{a}_{ij}) - \sum_i \ell_i \left(\sum_j \tilde{a}_{ij} - 1 \right) \right) = 0$$

replacing and deriving

$$\sum_i \sum_j \sum_t \sum_{\forall \mathbf{c}} \left\{ \Pr\left(q_{t-1} = i, q_t = j, \mathbf{c} \,|\mathbf{X}, \Theta\right) \frac{1}{\tilde{a}_{ij}} \right\} - \ell_i = 0$$

(we can maximize all the terms in i separately)

sinc(i)

HMM Training: Transition Probabilities

Multiplying both terms by $\tilde{a}_{ij}$

$$\sum_{j}\sum_{t}\sum_{\forall \mathbf{c}} \Pr\left(q_{t-1}=i, q_t=j, \mathbf{c}\,|\mathbf{X},\Theta\right) = \sum_{j} \tilde{a}_{ij}\ell_i$$

and then

$$\begin{aligned}\ell_i &= \sum_{j}\sum_{t}\sum_{\forall \mathbf{c}} \Pr\left(q_{t-1}=i, q_t=j, \mathbf{c}\,|\mathbf{X},\Theta\right) \\ &= \sum_{t}\sum_{\forall \mathbf{c}} \Pr\left(q_{t-1}=i, \mathbf{c}\,|\mathbf{X},\Theta\right)\end{aligned}$$

HMM Training: Transition Probabilities

Replacing we have

$$\begin{aligned}\tilde{a}_{ij} &= \frac{\sum_t \sum_{\forall \mathbf{c}} \Pr\left(q_{t-1}=i, q_t=j, \mathbf{c} \,|\mathbf{X}, \Theta\right)}{\sum_t \sum_{\forall \mathbf{c}} \Pr\left(q_{t-1}=i, \mathbf{c} \,|\mathbf{X}, \Theta\right)} \\ &= \frac{\sum_t \frac{\Pr\left(\mathbf{X}, q_{t-1}=i, q_t=j \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)}}{\sum_t \frac{\Pr\left(\mathbf{X}, q_{t-1}=i, |\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)}}\end{aligned}$$

HMM Training: Observation Probabilities

Now, we are subject to constraints

$$\sum_k \tilde{c}_{jk} = 1 \quad \forall j,$$

thus, we start from

$$\nabla_{\tilde{c}} \left(\mathcal{O}_c(\Theta, \tilde{c}_{kj}) - \sum_j \ell_j \left(\sum_k \tilde{c}_{jk} - 1 \right) \right) = 0,$$

HMM Training: Observation Distributions

Again, we derive the auxiliary function, we obtain the Lagrange multipliers and the reestimation formula results

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t = j, k_t = k \,|\Theta)}{\Pr(\mathbf{X}\,|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t = j \,|\Theta)}{\Pr(\mathbf{X}\,|\Theta)}}$$

HMM Training: Observation Distributions

To obtain $\boldsymbol{\mu}_{jk}$ and $\mathbf{U}_{jk}$ we use $\nabla_{\tilde{b}}\mathcal{O}_b(\Theta, \tilde{b}_{jk}) = 0$

$$0 = \frac{\partial \mathcal{O}_b(\Theta, \tilde{b}_{jk})}{\partial \tilde{\boldsymbol{\mu}}_{jk}} = \sum_j \sum_{\forall \mathbf{c}} \sum_t \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right) \tilde{\mathbf{U}}_{jk}^{-1}(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})$$

and therefore

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr\left(\mathbf{X}, q_t = j, k_t = k \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)} \mathbf{x}_t}{\sum_t \frac{\Pr\left(\mathbf{X}, q_t = j, k_t = k \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)}}$$

sinc(i)

HMM Training: Observation Distributions

To obtain $\boldsymbol{\mu}_{jk}$ and $\mathbf{U}_{jk}$ we use $\nabla_{\tilde{b}}\mathcal{O}_b(\Theta, \tilde{b}_{jk}) = 0$

$$0 = \frac{\partial \mathcal{O}_b(\Theta, \tilde{b}_{jk})}{\partial \tilde{\boldsymbol{\mu}}_{jk}} = \sum_j \sum_{\forall \mathbf{c}} \sum_t \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right) \tilde{\mathbf{U}}_{jk}^{-1}(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})$$

and therefore

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr\left(\mathbf{X}, q_t = j, k_t = k \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)} \mathbf{x}_t}{\sum_t \frac{\Pr\left(\mathbf{X}, q_t = j, k_t = k \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)}}$$

HMM Training: Observation Distributions

In a similar way, from $\nabla_{\tilde{b}} \mathcal{O}_b(\Theta, \tilde{b}_{jk}) = 0$ we obtain

$$0 = \frac{\partial \mathcal{O}_b(\Theta, \tilde{b}_{jk})}{\partial \tilde{\mathbf{U}}_{jk}^{-1}} =$$

$$= \sum_j \sum_{\forall \mathbf{c}} \sum_t \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right) \frac{1}{2} \tilde{\mathbf{U}}_{jk}^{-1} - (\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})^{\mathrm{T}}$$

and then

$$\tilde{\mathbf{U}}_{jk}^{-1} = \frac{\sum_t \frac{\Pr\left(\mathbf{X}, q_t = j, k_t = k \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)} (\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})^{\mathrm{T}}}{\sum_t \frac{\Pr\left(\mathbf{X}, q_t = j, k_t = k \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)}}$$

sinc(*i*)

Viterbi: approximation

A good approximation is

$$\Pr\left(\mathbf{X}\,|\Theta\right) \approx \max_{\forall \mathbf{q}} \left\{\Pr\left(\mathbf{X}\,|\mathbf{q},\Theta\right)\Pr\left(\mathbf{q}\,|\Theta\right)\right\}$$

We define the *best path probability* as the probability of the most likely state sequence at time t, which has generated the observation $\mathbf{X}^t$ (until time t) and ends in state i

$$\lambda_t(j) \triangleq \max_{\forall \mathbf{q}^{t-1}} \left\{\Pr\left(\mathbf{q}^{t-1}, q_t = j, \mathbf{X}^t\,|\Theta\right)\Pr\left(\mathbf{q}^{t-1}\,|\Theta\right)\right\}; \quad \forall j \in \mathcal{Q}$$

with $\lambda_0(j) = 1\ \forall j \in \mathcal{Q}$.

Viterbi: approximation

A good approximation is

$$\Pr\left(\mathbf{X}\,|\Theta\right) \approx \max_{\forall \mathbf{q}} \left\{\Pr\left(\mathbf{X}\,|\mathbf{q},\Theta\right)\Pr\left(\mathbf{q}\,|\Theta\right)\right\}$$

We define the *best path probability* as the probability of the most likely state sequence at time t, which has generated the observation $\mathbf{X}^t$ (until time t) and ends in state i

$$\lambda_t(j) \triangleq \max_{\forall \mathbf{q}^{t-1}} \left\{\Pr\left(\mathbf{q}^{t-1}, q_t = j, \mathbf{X}^t\,|\Theta\right)\Pr\left(\mathbf{q}^{t-1}\,|\Theta\right)\right\}; \quad \forall j \in \mathcal{Q}$$

with $\lambda_0(j) = 1\ \forall j \in \mathcal{Q}$.

Viterbi: recursion

Using an induction procedure we have

$$\begin{aligned}
\lambda_t(j) &= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) \Pr(q_t = j, \mathbf{x}_t | q_{t-1} = i, \Theta)\} \\
&= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) \Pr(q_t = j | q_{t-1} = i, \Theta) \Pr(\mathbf{x}_t | q_t = j, \Theta)\} \\
&= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) \Pr(q_t = j | q_{t-1} = i, \Theta)\} \Pr(\mathbf{x}_t | q_t = j, \Theta) \\
&= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) a_{ij}\} b_j(\mathbf{x}_t)
\end{aligned}$$

and thus

$$\Pr(\mathbf{X} | \Theta) \approx \max_{\forall j \in \mathcal{Q}} \{\lambda_T(j)\}.$$

sinc(i)

Viterbi: recursion

Using an induction procedure we have

$$\begin{aligned}
\lambda_t(j) &= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) \Pr(q_t = j, \mathbf{x}_t | q_{t-1} = i, \Theta)\} \\
&= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) \Pr(q_t = j | q_{t-1} = i, \Theta) \Pr(\mathbf{x}_t | q_t = j, \Theta)\} \\
&= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) \Pr(q_t = j | q_{t-1} = i, \Theta)\} \Pr(\mathbf{x}_t | q_t = j, \Theta) \\
&= \max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) a_{ij}\} b_j(\mathbf{x}_t)
\end{aligned}$$

and thus

$$\Pr(\mathbf{X} | \Theta) \approx \max_{\forall j \in \mathcal{Q}} \{\lambda_T(j)\}.$$

sinc(*i*)

Viterbi: the most probable sequences

To find the best sequence of hidden states we define

$$\xi_t(j) \triangleq \arg\max_{\forall i \in \mathcal{Q}} \{\lambda_{t-1}(i) a_{ij}\}$$

and from

$$\tilde{q}_T = \arg\max_{\forall i \in \mathcal{Q}} \{\lambda_T(i)\}$$

use the inverse recursion

$$\tilde{q}_t = \xi_{t+1}(\tilde{q}_{t+1}); \quad t = T-1, T-2, \ldots, 1$$

Baum-Welch: Forward Probabilities

Define the forward probability as

$$\alpha_t(j) \triangleq \Pr\left(\mathbf{x}_1, \ldots, \mathbf{x}_t, q_t = j \,|\Theta\right)$$

And this can be computed with the recursion

$$\begin{aligned}\alpha_t(j) &= \sum_i \left[\Pr\left(\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}, q_{t-1} = i \,|\Theta\right) \Pr\left(q_t = j \,|q_{t-1} = i\right)\right] \cdot \\ &\quad \cdot \Pr\left(\mathbf{x}_t \,|q_t = j\right) \\ &= b_j(\mathbf{x}_t) \sum_i \alpha_{t-1}(i) a_{ij}\end{aligned}$$

Baum-Welch: Forward Probabilities

Define the forward probability as

$$\alpha_t(j) \triangleq \Pr\left(\mathbf{x}_1, \ldots, \mathbf{x}_t, q_t = j \,|\Theta\right)$$

And this can be computed with the recursion

$$\begin{aligned}
\alpha_t(j) &= \sum_i \left[\Pr\left(\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}, q_{t-1} = i \,|\Theta\right) \Pr\left(q_t = j \,|q_{t-1} = i\right)\right] \cdot \\
&\qquad \cdot \Pr\left(\mathbf{x}_t \,|q_t = j\right) \\
&= b_j(\mathbf{x}_t) \sum_i \alpha_{t-1}(i) a_{ij}
\end{aligned}$$

sinc(i)

Baum-Welch: Forward Probabilities

Then

$$\begin{aligned}\Pr(\mathbf{X}|\Theta) &= \Pr(\mathbf{x}_1,\ldots,\mathbf{x}_T|\Theta) \\ &= \sum_i \Pr(\mathbf{x}_1,\ldots,\mathbf{x}_T, q_T = i|\Theta) \\ &= \sum_i \alpha_T(i)\end{aligned}$$

It is easy to show that the complexity for the forward algorithm is $O(N^2T)$ rather than the exponential one for $\Pr(\mathbf{X}|\Theta) = \sum_{\forall \mathbf{q}} \prod_t a_{q_{t-1}q_t} b_{q_t}(\mathbf{x}_t)$.

This is because we can make full use of partially computed probabilities for the improved efficiency.

sinc(i)

Baum-Welch: Forward Probabilities

Then

$$\begin{aligned}\Pr(\mathbf{X}|\Theta) &= \Pr(\mathbf{x}_1,\ldots,\mathbf{x}_T|\Theta) \\ &= \sum_i \Pr(\mathbf{x}_1,\ldots,\mathbf{x}_T, q_T = i|\Theta) \\ &= \sum_i \alpha_T(i)\end{aligned}$$

It is easy to show that the complexity for the forward algorithm is $O(N^2T)$ rather than the exponential one for $\Pr(\mathbf{X}|\Theta) = \sum_{\forall \mathbf{q}} \prod_t a_{q_{t-1}q_t} b_{q_t}(\mathbf{x}_t)$.

This is because we can make full use of partially computed probabilities for the improved efficiency.

Baum-Welch: Backward Probabilities

Define the backward probability as

$$\beta_t(j) \triangleq \Pr\left(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T, q_t = j \,|\Theta\right)$$

In a similar way as before we obtain the recursion

$$\beta_t(j) = \sum_i a_{ji} b_i(\mathbf{x}_{t+1}) \beta_{t+1}(i)$$

and also

$$\Pr\left(\mathbf{X} \,|\Theta\right) = \sum_i b_i(\mathbf{x}_1) \beta_1(i)$$

Baum-Welch: Backward Probabilities

Define the backward probability as

$$\beta_t(j) \triangleq \Pr\left(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T, q_t = j \,|\Theta\right)$$

In a similar way as before we obtain the recursion

$$\beta_t(j) = \sum_i a_{ji} b_i(\mathbf{x}_{t+1}) \beta_{t+1}(i)$$

and also

$$\Pr\left(\mathbf{X} \,|\Theta\right) = \sum_i b_i(\mathbf{x}_1) \beta_1(i)$$

Baum-Welch: Expected transitions from state i

Define

$$\gamma_t(j) \triangleq \Pr\left(q_t = j \,|\mathbf{X}, \Theta\right)$$

We can write

$$\begin{aligned}
\gamma_t(j) &= \frac{\Pr\left(q_t = j, \mathbf{X} \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)} \\
&= \frac{\Pr\left(\mathbf{x}_1, \ldots, \mathbf{x}_t, q_t = j \,|\Theta\right) \Pr\left(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T, q_t = j \,|\Theta\right)}{\sum_i \Pr\left(q_t = i, \mathbf{X} \,|\Theta\right)} \\
&= \frac{\alpha_t(j)\beta_t(j)}{\sum_i \alpha_t(i)\beta_t(i)}
\end{aligned}$$

Baum-Welch: Expected transitions from state i

Define

$$\gamma_t(j) \triangleq \Pr\left(q_t = j \,|\mathbf{X}, \Theta\right)$$

We can write

$$\begin{aligned}
\gamma_t(j) &= \frac{\Pr\left(q_t = j, \mathbf{X} \,|\Theta\right)}{\Pr\left(\mathbf{X} \,|\Theta\right)} \\
&= \frac{\Pr\left(\mathbf{x}_1, \ldots, \mathbf{x}_t, q_t = j \,|\Theta\right) \Pr\left(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T, q_t = j \,|\Theta\right)}{\sum_i \Pr\left(q_t = i, \mathbf{X} \,|\Theta\right)} \\
&= \frac{\alpha_t(j)\beta_t(j)}{\sum_i \alpha_t(i)\beta_t(i)}
\end{aligned}$$

sinc(i)

Baum-Welch: Expected transitions from state i to state j

Define

$$\gamma_t(i,j) \triangleq \Pr\left(q_{t-1}=i, q_t=j \,|\mathbf{X},\Theta\right)$$

Now we can write

$$\begin{aligned}\gamma_t(i,j) &= \frac{\Pr\left(q_{t-1}=i, q_t=j, \mathbf{X}\,|\Theta\right)}{\Pr\left(\mathbf{X}\,|\Theta\right)} \\ &= \frac{\Pr\left(\mathbf{x}_1,\ldots,\mathbf{x}_t, q_{t-1}=i\,|\Theta\right) a_{ij} b_j(\mathbf{x}_t) \Pr\left(\mathbf{x}_{t+1},\ldots,\mathbf{x}_T, q_t=j\,|\Theta\right)}{\sum_m \sum_n \Pr\left(\mathbf{x}_1,\ldots,\mathbf{x}_{t-1}, q_t=m\,|\Theta\right) a_{mn} b_n(\mathbf{x}_t) \Pr\left(\mathbf{x}_{t+1},\ldots,\mathbf{x}_T, q_t=n\,|\Theta\right)}\end{aligned}$$

Thus

$$\begin{aligned}\gamma_t(i,j) &= \frac{\alpha_{t-1}(i) a_{ij} \sum_k c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_m \sum_n \alpha_{t-1}(m) a_{mn} \sum_k c_{nk} b_{nk}\left(\mathbf{x}_t\right) \beta_t(n)} \\ &= \frac{\alpha_{t-1}(i) a_{ij} \sum_k c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_m \alpha_t(m) \beta_t(m)}\end{aligned}$$

Baum-Welch: Expected transitions from state i to state j

Define

$$\gamma_t(i,j) \triangleq \Pr\left(q_{t-1}=i, q_t=j \,|\mathbf{X},\Theta\right)$$

Now we can write

$$\begin{aligned}\gamma_t(i,j) &= \frac{\Pr\left(q_{t-1}=i, q_t=j, \mathbf{X}\,|\Theta\right)}{\Pr\left(\mathbf{X}\,|\Theta\right)} \\ &= \frac{\Pr\left(\mathbf{x}_1,\ldots,\mathbf{x}_t, q_{t-1}=i\,|\Theta\right) a_{ij} b_j(\mathbf{x}_t) \Pr\left(\mathbf{x}_{t+1},\ldots,\mathbf{x}_T, q_t=j\,|\Theta\right)}{\sum_m \sum_n \Pr\left(\mathbf{x}_1,\ldots,\mathbf{x}_{t-1}, q_t=m\,|\Theta\right) a_{mn} b_n(\mathbf{x}_t) \Pr\left(\mathbf{x}_{t+1},\ldots,\mathbf{x}_T, q_t=n\,|\Theta\right)}\end{aligned}$$

Thus

$$\begin{aligned}\gamma_t(i,j) &= \frac{\alpha_{t-1}(i) a_{ij} \sum_k c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_m \sum_n \alpha_{t-1}(m) a_{mn} \sum_k c_{nk} b_{nk}\left(\mathbf{x}_t\right) \beta_t(n)} \\ &= \frac{\alpha_{t-1}(i) a_{ij} \sum_k c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_m \alpha_t(m) \beta_t(m)}\end{aligned}$$

Baum-Welch: Expected transitions from state i to state j

Define

$$\gamma_t(i,j) \triangleq \Pr\left(q_{t-1}=i, q_t=j \,|\mathbf{X},\Theta\right)$$

Now we can write

$$\begin{aligned}\gamma_t(i,j) &= \frac{\Pr\left(q_{t-1}=i, q_t=j, \mathbf{X}\,|\Theta\right)}{\Pr\left(\mathbf{X}\,|\Theta\right)} \\ &= \frac{\Pr\left(\mathbf{x}_1,\ldots,\mathbf{x}_t, q_{t-1}=i\,|\Theta\right) a_{ij} b_j(\mathbf{x}_t) \Pr\left(\mathbf{x}_{t+1},\ldots,\mathbf{x}_T, q_t=j\,|\Theta\right)}{\sum_m \sum_n \Pr\left(\mathbf{x}_1,\ldots,\mathbf{x}_{t-1}, q_t=m\,|\Theta\right) a_{mn} b_n(\mathbf{x}_t) \Pr\left(\mathbf{x}_{t+1},\ldots,\mathbf{x}_T, q_t=n\,|\Theta\right)}\end{aligned}$$

Thus

$$\begin{aligned}\gamma_t(i,j) &= \frac{\alpha_{t-1}(i) a_{ij} \sum_k c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_m \sum_n \alpha_{t-1}(m) a_{mn} \sum_k c_{nk} b_{nk}\left(\mathbf{x}_t\right) \beta_t(n)} \\ &= \frac{\alpha_{t-1}(i) a_{ij} \sum_k c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_m \alpha_t(m) \beta_t(m)}\end{aligned}$$

B-W: Expected occupations of the Gaussian k in state j

Define

$$\psi_t(j,k) \triangleq \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right)$$

And in a similar way as before, this expected value can be efficiently computed by

$$\psi_t(j,k) = \frac{\sum_i \alpha_{t-1}(i) a_{ij} c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_i \alpha_T(i)}$$

B-W: Expected occupations of the Gaussian k in state j

Define

$$\psi_t(j,k) \triangleq \Pr\left(q_t = j, k_t = k \,|\mathbf{X}, \Theta\right)$$

And in a similar way as before, this expected value can be efficiently computed by

$$\psi_t(j,k) = \frac{\sum_i \alpha_{t-1}(i) a_{ij} c_{jk} b_{jk}\left(\mathbf{x}_t\right) \beta_t(j)}{\sum_i \alpha_T(i)}$$

Baum-Welch: Reestimation formulas

$$\tilde{a}_{ij} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, |\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \gamma_t(i,j)}{\sum_t \gamma_t(i)}$$

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)}{\sum_t \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} \mathbf{x}_t}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)\mathbf{x}_t}{\sum_t \psi_t(j,k)}$$

sinc(i)

Baum-Welch: Reestimation formulas

$$\tilde{a}_{ij} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, |\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \gamma_t(i,j)}{\sum_t \gamma_t(i)}$$

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)}{\sum_t \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} \mathbf{x}_t}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)\mathbf{x}_t}{\sum_t \psi_t(j,k)}$$

Baum-Welch: Reestimation formulas

$$\tilde{a}_{ij} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i,|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \gamma_t(i,j)}{\sum_t \gamma_t(i)}$$

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)}{\sum_t \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} \mathbf{x}_t}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)\mathbf{x}_t}{\sum_t \psi_t(j,k)}$$

sinc(i)

Baum-Welch: Reestimation formulas

$$\tilde{a}_{ij} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, |\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \gamma_t(i,j)}{\sum_t \gamma_t(i)}$$

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)}{\sum_t \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} \mathbf{x}_t}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)\mathbf{x}_t}{\sum_t \psi_t(j,k)}$$

Baum-Welch: Reestimation formulas

$$\tilde{a}_{ij} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i,|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \gamma_t(i,j)}{\sum_t \gamma_t(i)}$$

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)}{\sum_t \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} \mathbf{x}_t}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)\mathbf{x}_t}{\sum_t \psi_t(j,k)}$$

sinc(i)

Baum-Welch: Reestimation formulas

$$\tilde{a}_{ij} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_{t-1}=i,|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \gamma_t(i,j)}{\sum_t \gamma_t(i)}$$

$$\tilde{c}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)}{\sum_t \gamma_t(i)}$$

$$\tilde{\boldsymbol{\mu}}_{jk} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} \mathbf{x}_t}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}} = \frac{\sum_t \psi_t(j,k)\mathbf{x}_t}{\sum_t \psi_t(j,k)}$$

sinc(i)

Baum-Welch: Reestimation formulas

$$\tilde{\mathbf{U}}_{jk}^{-1} = \frac{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)} (\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})^{\mathrm{T}}}{\sum_t \frac{\Pr(\mathbf{X}, q_t=j, k_t=k|\Theta)}{\Pr(\mathbf{X}|\Theta)}}$$

$$= \frac{\sum_t \psi_t(j,k)(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})(\mathbf{x}_t - \tilde{\boldsymbol{\mu}}_{jk})^{\mathrm{T}}}{\sum_t \psi_t(j,k)}$$

Other Topics in the Standard HMM

- Multiple observation sequences
- Scaling factors
- Tied Mixtures
- Log-scale counters
- Decoding composite HMM
- Variants on HMM structures
- Discriminant training
- Implementation issues
- ...
- ...
- ...

Bibliography

- Deller J. R., Proakis, J. G. and Hansen, J. H. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing, New York, 1993.
- Huang, X. D. Ariki, Y. and Jack, M. A. *Hidden Markov Models for Speech Recognition*. Edinburgh University Press, 1990.
- Jelinek F. *Statistical Methods for Speech Recognition*. MIT Press, Cambrige, Massachussets, 1999.
- Rabiner L. R. and Juang B. H. *Fundamentals of Speech Recognition*. Prentice-Hall, 1993.
- Ferguson J. *Hidden Markov Models for Speech*. IDA, Princeton, NJ, 1980.
- Huang X., Acero A., and Hon H.-W. *Spoken Language Processing: A Guide to Theory, Algorithm and System*. Prentice Hall PTR, 2001.
- Apuntes D. Milone (hmmdiet.pdf and hmmsweet.pdf)