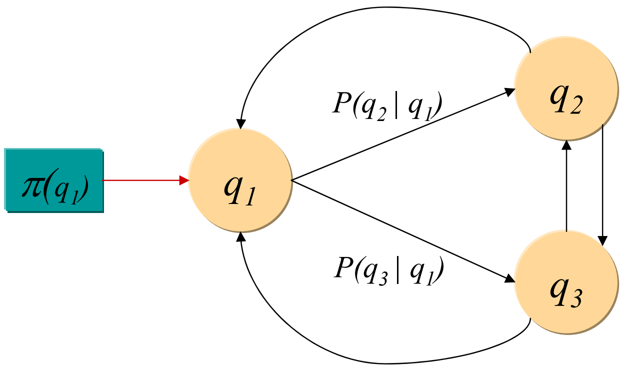
Hidden Markov Models

Markov model

The Markov chain is the tuple

$$M = (Q, P, \pi)$$

where:
Q is the set of states
P is the probability matrix of state transition
π is the vector of initial probabilities to start states



A Markov chain is traversed from state to state, producing a sequence of states.

Sequence Probability

The probability of a sequence (of states) $x = x_1, x_2, \ldots, x_n$ is

$$P(x) = P(x_n, x_{n-1}, ..., x_2, x_1)$$

Using $P(x, y) = P(x|y)P(y)$ this can be re-written as follows

$$P(x) = P(x_n, x_{n-1}, x_{n-2}..., x_2, x_1) =$$
$$P(x_n|x_{n-1}, x_{n-2}..., x_2, x_1)P(x_{n-1}, x_{n-2}..., x_2, x_1) =$$
$$P(x_n|x_{n-1}, x_{n-2}..., x_2, x_1)P(x_{n-1}|x_{n-2}..., x_2, x_1)P(x_{n-2}..., x_2, x_1) =$$
$$P(x_n|x_{n-1}, x_{n-2}..., x_2, x_1)P(x_{n-1}|x_{n-2}..., x_2, x_1)P(x_{n-2}|x_{n-3}..., x_2, x_1)...P(x_2|x_1)P(x_1)$$

but this can be simplified for a Markov chain model, because
the probability of any next state depends ONLY on the previous state

$$P(x_n|x_{n-1}, x_{n-2}..., x_2, x_1) = P(x_n|x_{n-1})$$

Markov Process

Hence,

$$P(x) =$$
$$P(x_n|x_{n-1}, x_{n-2}..., x_2, x_1)P(x_{n-1}|x_{n-2}..., x_2, x_1)P(x_{n-2}|x_{n-3}..., x_2, x_1)...P(x_2|x_1)P(x_1)$$

can be re-written as follows:

$$P(x) = P(x_n|x_{n-1})P(x_{n-1}|x_{n-2})P(x_{n-2}|x_{n-3})...P(x_2|x_1)P(x_1)$$

and thus re-written in terms of the transition probabilities $a_{s,t}$ $(= P(t|s))$

$$P(x) = a_{x_{n-1},x_n} a_{x_{n-2},x_{n-1}} a_{x_{n-3},x_{n-2}} a_{x_{n-4},x_{n-3}} ... a_{x_1,x_2} P(x_1)$$

Sequence Probability

So the probability of any given sequence x

x_1	x_2	x_3	...	...	...	x_{n-2}	x_{n-1}	x_n

can be found by:

- multiplying the probabilities of the individual transitions, and
- the probability of starting in the state x_1

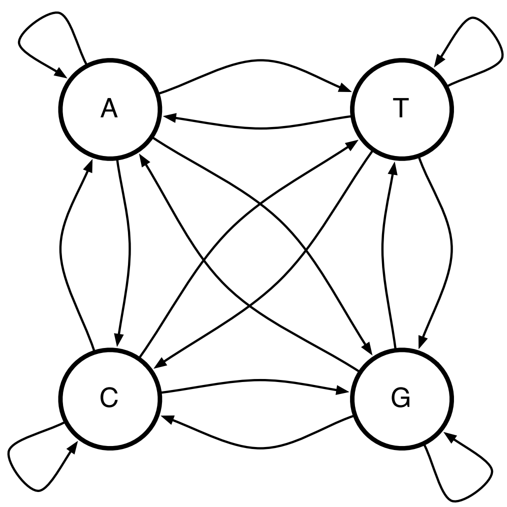
$$P(x) = a_{x_{n-1},x_n} \cdot a_{x_{n-2},x_{n-1}} \cdot a_{x_{n-3},x_{n-2}} \cdot a_{x_{n-4},x_{n-3}} \dots a_{x_1,x_2} \cdot P(x_1)$$

$$P(x) = (\prod_{i=2}^{n} a_{x_{i-1},x_i})P(x_1)$$

Markov Chains: a simple Markov model for DNA

$$P(x = T|x = A) = a_{A,T}$$

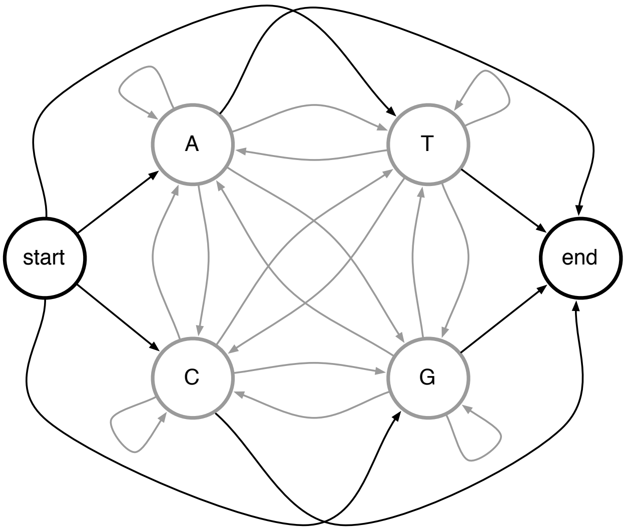
- 4 states A,C,G,T
- Transition between states occur with particular probabilities
- Each arrow has a probability parameter associated with it



$$a_{s,t} = P(x_i = t|x_{i-1} = s)$$

probability of making a transition from state s to state t

Include BEGIN and END state

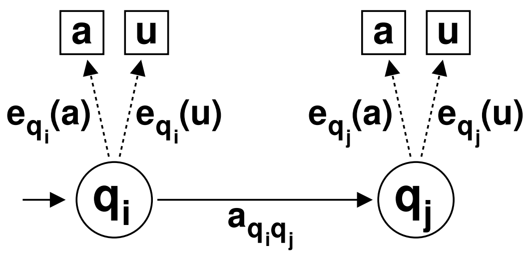


Two new states can be added to the Markov model.
These are treated as silent states, since they do not add to the sequence.

Hidden Markov Model (HMM)

Hidden Markov Models (HMMs) resemble Markov Models in having a finite number of **states** connected by **transitions**.

But the major difference between the two is that the states of the Hidden Markov Models are not associated with one symbol but with more than one symbol. Each state q can emit a symbol x with a probability given by the distribution of emission probabilities $e_q(x)$.



Hidden Markov Model (HMM)

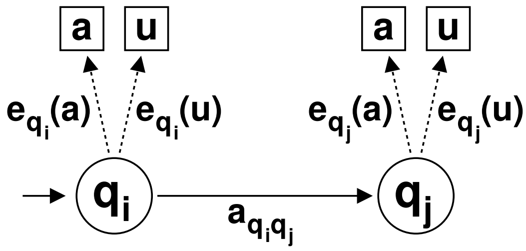
The HMM is a tuple

$$\mathcal{M} = (\Sigma, Q, A, E)$$

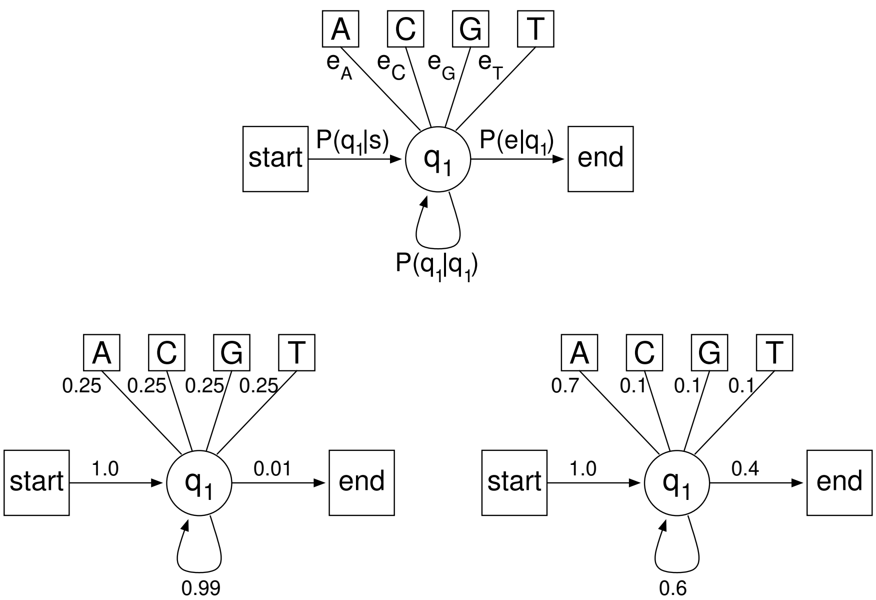
There are a finite number of states $\boldsymbol{Q}$ in the model

At a given time j, each new state q_j is entered from a previous state q_i, based upon a transition probability $a_{q_i,q_j} = P(q_j|q_i)$ from the probability distribution $\boldsymbol{A}$, which only depends on the previous state q_i (the Markovian property)

After each transition a symbol y_j from $\boldsymbol{\Sigma}$ is produced based on the current state q_j, with emission probability $e_{q_j}(y_j) = P(y_j|q_j)$

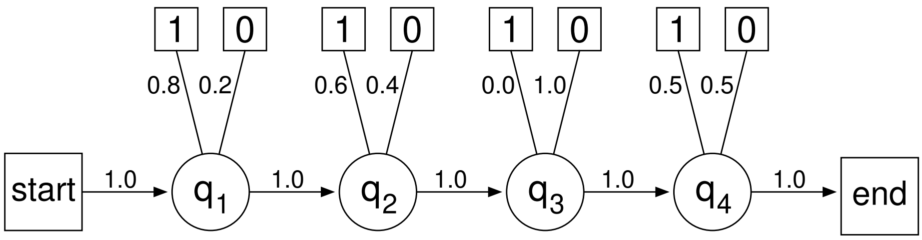


Example 1 (HMM)



- AAAATGGTGAACCTGTCGTTCCG
- GAAA

Example 2 (HMM)



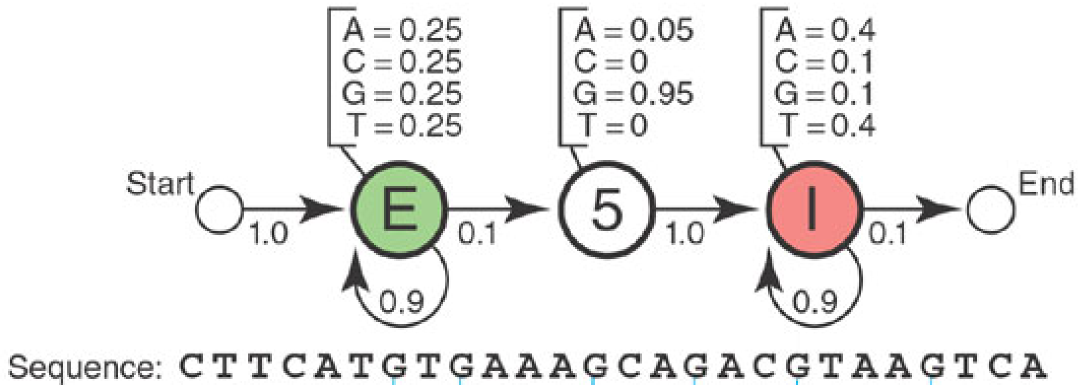
Can this HMM produce the following emitted sequences?

1 1 1 1 1	no
0 0 0 0	yes
1 0 0 1	yes
1 1 1 1	no
1 0	no

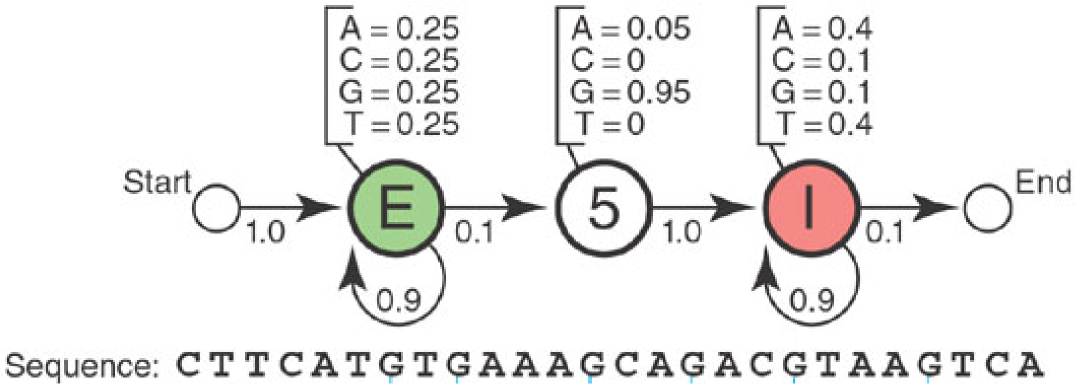
HMMs - splice site finding example (Eddy, 2004)

Given:

- Exon frequencies (A: 0.25, C: 0.25, G: 0.25, T: 0.25)
- Intron frequencies (A: 0.4, C: 0.1, G: 0.1, T: 0.4)
- Splice site donor (G: 0.95, A: 0.05)
- A sequence with exon, (unknown) splice site, intron
- then we can draw a simple HMM:

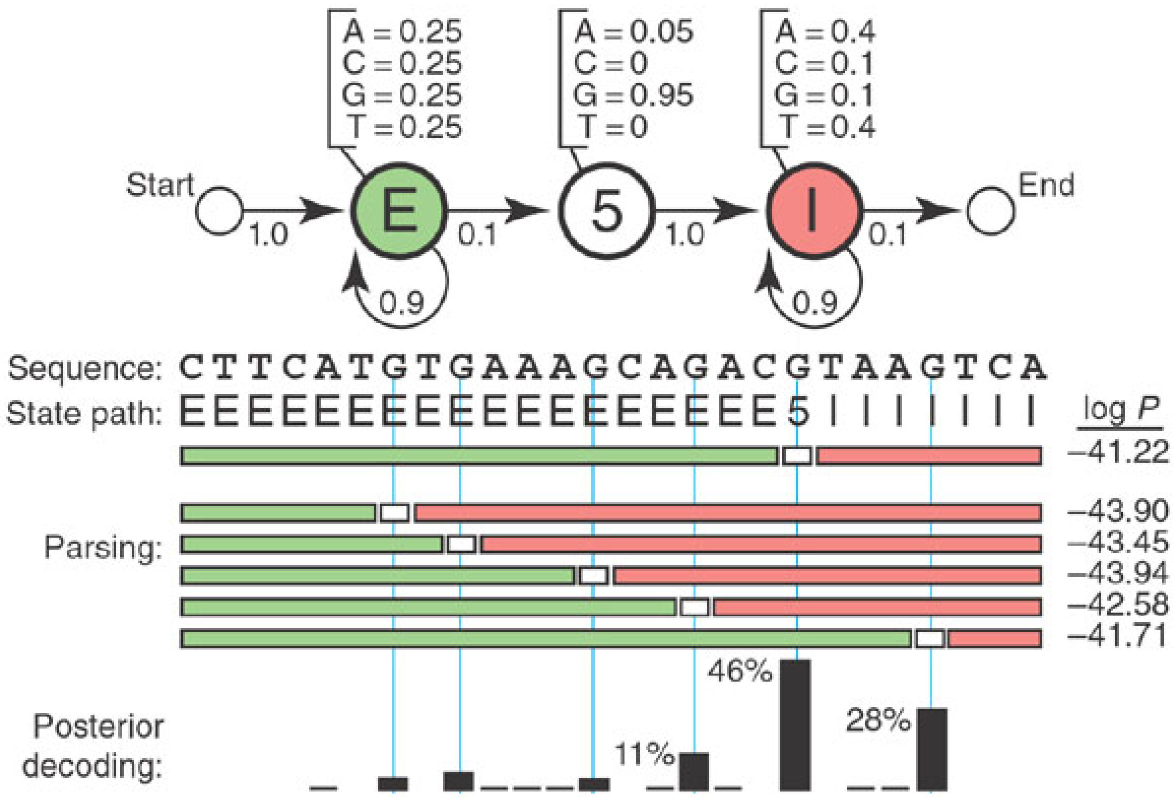


HMMs - splice site finding example (Eddy, 2004)



- Imagine the HMM produces sequences, instead of taking input.
- These sequences will occur with certain probabilities (but possibly produced by several unknown (= *hidden*) state paths).
- The HMM is a full probabilistic model.
- We can use the model to retrieve the state path through the HMM and also the probability of a sequence.

HMMs - example (Eddy, 2004)



Typical problems associated with HMMs

Given a sequence $\boldsymbol{y} = (y_1 y_2 y_3 \ldots y_n)$ and a model $\mathcal{M}$

???

Given the sequence $\boldsymbol{y}$, what was the optimal state sequence?

???

What is the probability of the observed sequence $\boldsymbol{y}$?

???

What is the most probable state for a particular observation $\boldsymbol{y_i}$?

Example: HMM for the Fair Bet Casino

$\mathcal{M} = (\boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{A}, \boldsymbol{E})$

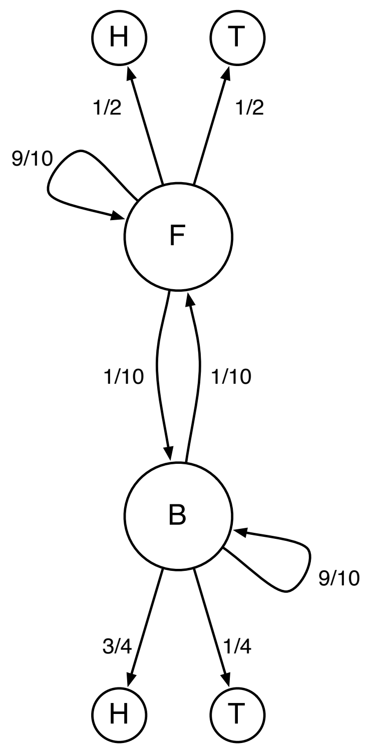
- $\mathbf{Q} = \{F, B\}$ – F for Fair and B for Biased coin
- $\mathbf{\Sigma} = \{H, T\}$ – **H**eads and **T**ails
- transition probabilities:

$$\mathbf{A} = \left\{ \begin{aligned} a_{0F} &= P(F|0) = 0.5[= \pi_F], \\ a_{0B} &= P(B|0) = 0.5[= \pi_B], \\ a_{FF} &= P(F|F) = 0.9, \\ a_{BF} &= P(F|B) = 0.1, \\ a_{FB} &= P(B|F) = 0.1, \\ a_{BB} &= P(B|B) = 0.9 \end{aligned} \right\}$$

  Note, that a_{0q} reflect the initial probabilities of starting in states q (coming from the start state 0).

- emission probabilities:

$$\mathbf{E} = \left\{ \begin{aligned} e_F(T) &= P(T|F) = 1/2, \\ e_F(H) &= P(H|F) = 1/2, \\ e_B(T) &= P(T|B) = 1/4, \\ e_B(H) &= P(H|B) = 3/4 \end{aligned} \right\}$$



Example: HMM for the Fair Bet Casino (II)

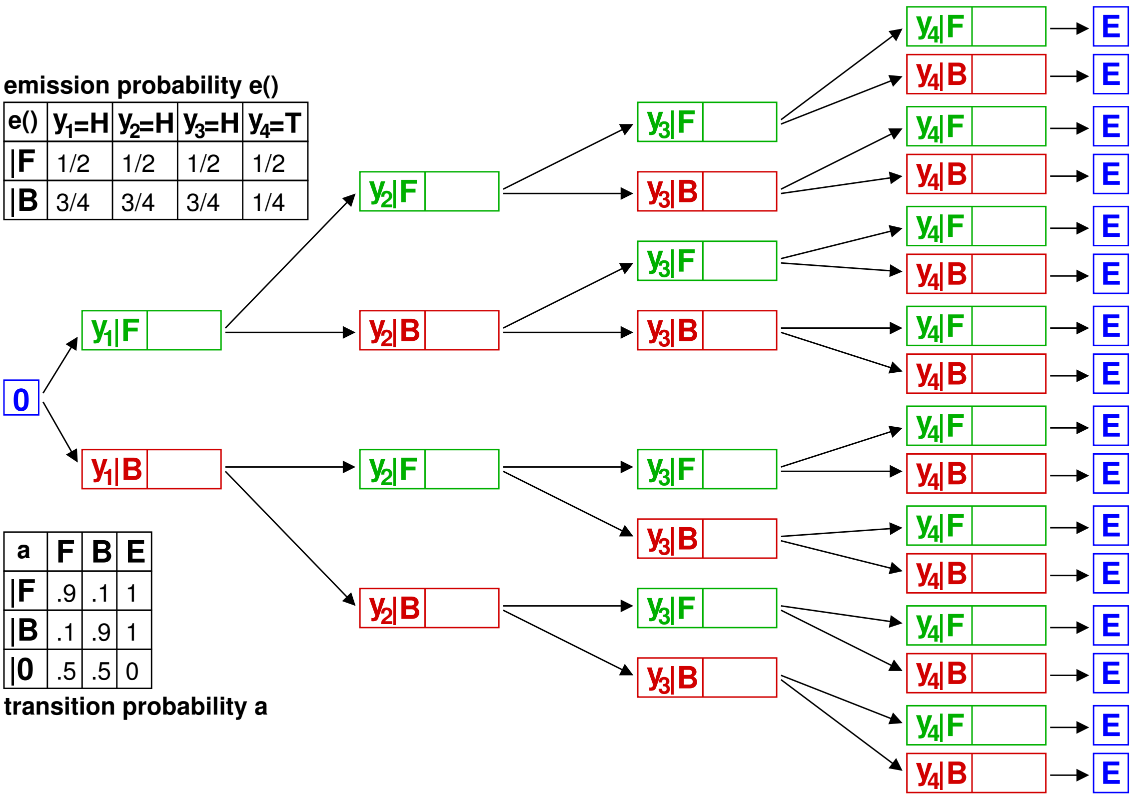
Given an emitted sequence $y = (HHHT)$ we can easily construct all possible state paths through that HMM M.

- For simplicity we add a start state **0** using π as transition probabilities
- and an end state **E**, where the HMM ends after having emitted the last sign in **y**.

All possible paths for y=HHHT in a tree-shaped graph

emission probability e()

e()	y_1=H	y_2=H	y_3=H	y_4=T
|F	1/2	1/2	1/2	1/2
|B	3/4	3/4	3/4	1/4



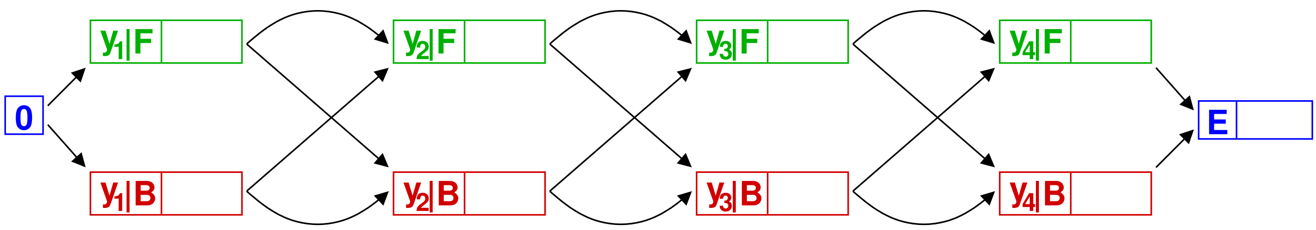
a	F	B	E
|F	.9	.1	1
|B	.1	.9	1
|0	.5	.5	0

transition probability a

Due to the Markov property identical states (same color) for the same positions can be collapsed.

All possible paths for y=HHHT in a reduced state graph

- Due to the Markov property, that transition to a new state of the (hidden) Markov model only depends on the current state,
- the state graph can be very much reduced
- by collapsing identical states (same color) for the same positions
- thus, producing a linear HMM where certain states associated with certain positions.



Finding the optimal state path: Viterbi algorithm

Using this representation and the emitted sequence $Y = y_1, y_2 \ldots y_L$ we can use the Viterbi Algorithm to find the optimal state path π^*.

Viterbi Algorithm

1. initialize start state ($k = 0$), disable other states before position i=1:

$$v_{k,0} = \begin{cases} 1.0 & \text{if } k = 0 \text{ (start state)} \\ 0.0 & \text{otherwise} \end{cases}$$

2. compute for any state k and position $i = 1 \ldots L$:

$$v_{k,i} = \max_{q \in Q} \left(v_{q,i-1} \cdot a_{q,k} \right) \cdot e_k(y_i)$$

3. result in end state E (i.e. probability of the maximal state path π^*):

$$P(y, \pi^*) = v_{E,L+1} = \max_{q \in Q} \left(v_{q,L} \right)$$

(end of forward part)

4. to get the maximal state path π^*: backtrace (backward part)

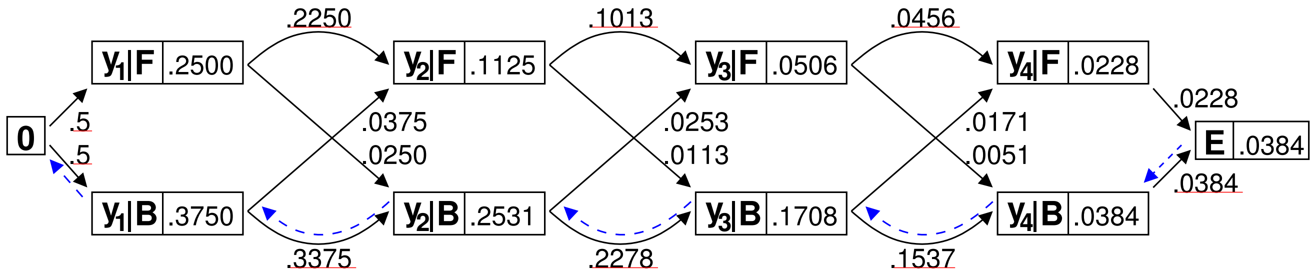
Viterbi algorithm - finding the optimal state path

transition probability a

a	F	B	E
|F	.9	.1	1
|B	.1	.9	1
|0	.5	.5	0

emission probability e()

e()	y_1=H	y_2=H	y_3=H	y_4=T
|F	1/2	1/2	1/2	1/2
|B	3/4	3/4	3/4	1/4



Typical problems associated with HMMs

Given an observation sequence $\boldsymbol{y} = (y_1 y_2 y_3 \ldots y_n)$ and a model $\mathcal{M}$

Viterbi algorithm

Given the observation sequence $\boldsymbol{y}$, what was the optimal state sequence?

???

What is the probability of the observed sequence $\boldsymbol{y}$?

???

What is the most probable state for a particular observation $\boldsymbol{y_i}$?

Finding the overall probablility: Forward algorithm

Using the graph representation and the emitted sequence $Y = y_1, y_2 \dots y_L$ we can use the so-called Forward Algorithm to compute the overall probability of all state paths emitting Y.

Forward Algorithm

1. initialize start state ($k = 0$), disable other states before position i=1:

$$f_{k,0} = \begin{cases} 1.0 & \text{if } k = 0 \text{ (start state)} \\ 0.0 & \text{otherwise} \end{cases}$$

2. compute for any state k and position $i = 1 \dots L$:

$$f_{k,i} = \sum_{q \in Q} (f_{q,i-1} \cdot a_{q,k}) \cdot e_k(y_i)$$

3. result in end state E (i.e. probability of the HMM generating Y):

$$P(y) = f_{E,L+1} = \sum_{q \in Q} (f_{q,L})$$

Note, $f_{k,i}$ are called forward-probabilities.

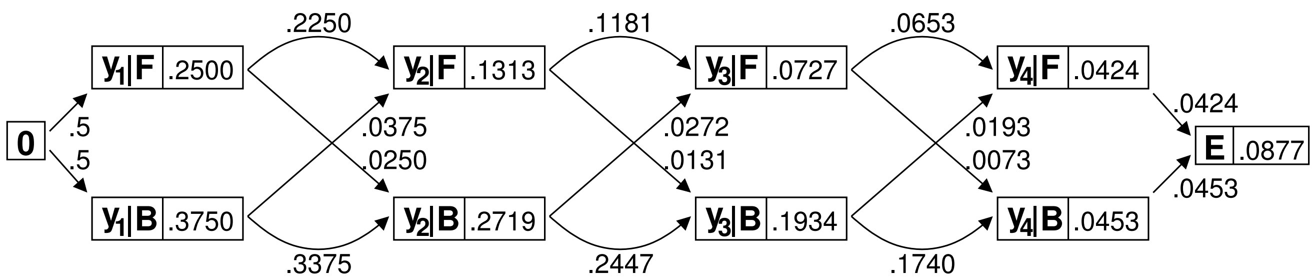
Forward algorithm - finding the overall sequence probability

transition probability a

a	F	B	E
|F	.9	.1	1
|B	.1	.9	1
|0	.5	.5	0

emission probability e()

e()	y_1=H	y_2=H	y_3=H	y_4=T
|F	1/2	1/2	1/2	1/2
|B	3/4	3/4	3/4	1/4



Typical problems associated with HMMs

Given an observation sequence $\boldsymbol{y} = (\boldsymbol{y_1 y_2 y_3 \ldots y_n})$ and a model $\mathcal{M}$

Viterbi algorithm

Given the observation sequence $\boldsymbol{y}$, what was the optimal state sequence? (attempt to recover the hidden part of the model = posterior decoding)

Forward algorithm

What is the probability of the observed sequence $\boldsymbol{y}$?

???

What is the most probable state for a particular observation $\boldsymbol{y_i}$?

Backward algorithm – most probable state for observation

Using the graph representation and the emitted sequence $Y = y_1, y_2 \ldots y_L$ we can use the so-called Backward Algorithm to compute the backward probabilities $b_{l,i}$.

Backward Algorithm

1. initialize last emmitting states ($i = L$):
$$b_{k,L} = \{ \; 1.0 \quad \text{for all } k \in Q$$
2. compute for any state k and position $i = L - 1 \ldots 1$:
$$b_{k,i} = \sum_{q \in Q} (a_{k,q} \cdot e_q(y_{i+1}) \cdot b_{q,i+1})$$
3. result in start state 0, position $i = 0$:
$$P(y) = b_{0,0} = \sum_{q \in Q} (a_{0,q} \cdot e_q(y_1) \cdot b_{q,1})$$

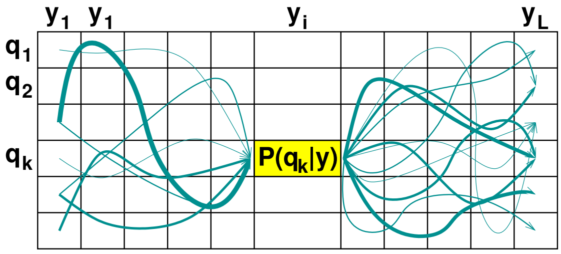
Note, $b_{k,i}$ are called backward-probabilities, and $P(y)$ is the same as in the forward algorithm.

Backward algorithm – most probable state for observation

Finally, the probability that a state q_k produced the output y_i is computed by

$$P(\pi_i = q_k | y) = \frac{P(y, \pi_i = q_k)}{P(y)} = \frac{f_{q_k,i} \cdot b_{q_k,i}}{P(y)}$$

This reflects the sum of probabilities over all paths leading through q_k:



Thus, choosing the path π' of states q_i maximizing the probability at each position i is an alternative to the Viterbi algorithm for posterior decoding.

Typical problems associated with HMMs

Given an observation sequence $\boldsymbol{y} = (\boldsymbol{y_1 y_2 y_3 \ldots y_n})$ and a model $\mathcal{M}$

Viterbi algorithm

Given the observation sequence $\boldsymbol{y}$, what was the optimal state sequence? (attempt to recover the hidden part of the model = posterior decoding)

Forward algorithm

What is the probability of the observed sequence $\boldsymbol{y}$?

Backward algorithm

What is the most probable state for a particular observation $\boldsymbol{y_i}$?