Hidden Markov Models

Markov Random Processes

- A random sequence has the Markov property if its distribution is determined solely by its current state. Any random process having this property is called a *Markov random process*.

- For observable state sequences (state is known from data), this leads to a *Markov chain* model.

- For non-observable states, this leads to a *Hidden Markov Model* (HMM).

The casino models

Game:

You bet $1

You roll (always with a fair die)

Casino player rolls (maybe with fair die, maybe with loaded die)

Highest number wins $1

Honest casino: it has one dice:

- Fair die: P(1) = P(2) = P(3) = P(5) =P(6) = 1/6

Crooked casino: it has one dice:

- Loaded die: P(1) = P(2) = … = P(5) = 1/10 P(6) = 1/2

Dishonest casino: it has two dice:

- Fair die: P(1) = P(2) = P(3) = P(5) =P(6) = 1/6
- Loaded die: P(1) = P(2) = … = P(5) = 1/10 P(6) = 1/2

Casino player approximately switches back-&-forth between fair and loaded die once every 20 turns

The casino models

algorithmics genetics group .

Honest casino

2 1 4 3 4 1 6 5 2 5 1 5 1 4 4 6 3 2 4 5 4 3 3 4 6 4 3 5 3 1 5
5 6 1 4 4 4 4 2 4 3 1 3 2 6 5 2 5 4 5 5 3 3 2 2 5 6 4 3 2 2 1 4 6 5 3 1 6 5 2 6 4 5 1 1 4 2 6 4 6 4 6 1 2 1 5 4 5 4 5 1 1 1 1
6 1 2 1 6 1 6 6 3 6 2 5 1 4 6 5 4 2 4 1 4 4 2 1 6 4 4 4 3 5 5 2 1 2 5 2 5 6 1 4 5 2 5 1 6 3 1 1 6 5 6 1 5 4 5 6 2 1 6 1 5 4 6
6 2 5 4 4 5 2 2 2 3 3 3 2 6 1 4 6 4 5 2 5 4 4 2 3 6 6 3 4 6 1 4 2 4 3 3 5 1 6 1 5 1 2 5 5 3 1 6 1 2 2 4 6 3 6 1 3 3 4 4 5 1 6
1 3 3 2 4 4 5 2 1 6 3 4 5 5 1 6 4 3 6 6 5 3 1 2 3 5 5 4 4 3 1 4 4 5 4 1 2 6 1 3 3 1 3 4 1 3 3 6 6 4 2 1 1 5 1 4 3 4 3 4 3 5 1
6 2 4 2 1 4 4 1 1 6 1 6 6 5 1 2 6

Crooked casino:

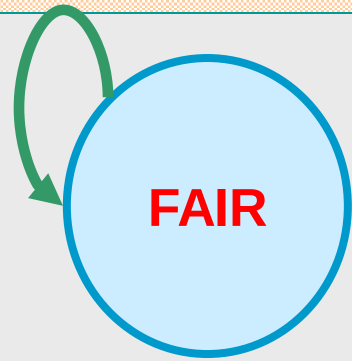
5 6 3 3 4 3 4 6 3 6 6 6 1 2 6 6 6 5 6 6 1 2 2 2 6 2 6 6 6 6 5 2 6
6 3 6 5 6 6 6 5 3 6 1 5 4 1 4 1 3 5 6 6 5 3 1 5 6 2 6 6 5 6 6 6 1 6 3 1 2 1 1 5 6 1 2 2 6 6 6 6 6 5 4 6 6 6 5 6 6 2 6 6 2 6 6 6 1
6 6 6 2 5 6 6 4 5 3 2 4 2 1 5 6 5 3 2 6 1 6 3 6 5 6 6 4 2 5 4 6 6 3 3 6 1 3 6 6 5 6 6 6 4 5 4 6 6 2 2 4 1 5 6 6 1 3 4 6 6 4 4 3 6
2 1 2 2 6 6 4 6 6 4 6 6 4 1 4 6 6 1 1 6 4 2 6 3 6 6 6 6 3 6 6 2 6 6 5 5 6 5 5 6 5 2 6 5 6 6 6 2 1 4 3 4 5 6 6 3 6 3 3 4 6 6 6 3 6
5 5 2 6 6 6 2 3 5 1 6 6 6 6 5 5 4 6 6 5 4 3 2 6 1 4 6 6 6 4 3 6 6 1 6 6 5 6 5 2 6 3 6 1 6 4 2 6 6 2 4 6 6 5 5 6 1 6 6 6 3 1 6 3 6
2 2 6 6 6 6 5

Dishonest casino:

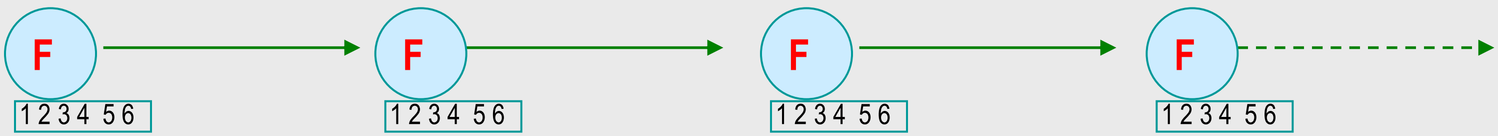
6 6 6 1 5 5 5 4 1 6 6 4 6 4 3 1 5 6 6 3 1 3 3 6 2 6 4 3 6 5 2 3 2
1 6 6 6 2 5 5 5 5 3 6 6 4 6 4 6 1 4 5 2 5 6 2 2 2 3 5 3 5 4 6 4 3 5 6 6 6 6 1 6 6 3 1 2 6 1 1 5 5 2 1 1 5 3 4 2 1 3 3 6 4 2 6
6 3 6 1 1 6 6 6 3 5 5 3 4 5 3 5 4 1 3 3 6 1 2 6 4 3 4 5 3 6 5 6 4 6 1 4 2 6 2 6 1 4 5 6 6 1 3 6 1 6 2 4 1 6 6 6 1 6 6 2 4 4 3
6 2 6 1 6 3 5 5 2 6 5 5 2 6 5 5 3 6 2 6 6 6 4 4 4 3 6 6 6 3 5 6 6 6 5 2 1 6 6 6 4 3 1 6 2 3 1 3 2 1 2 6 6 6 6 6 6 6 6 6 5 6 6 4
6 6 1 6 6 4 3 1 6 3 6 4 4 1 5 2 6 1 3 6 6 6 2 6 6 5 3 6 2 6 3 3 1 2 3 6 3 6 5 6 2 2 5 3 6 5 4 4 5 1 1 2 4 2 1 5 2 6 1 5 6 3 4
6 5 1 5 3 4 4 4 6 3 4 6 2 2 5

The casino models (only one die)

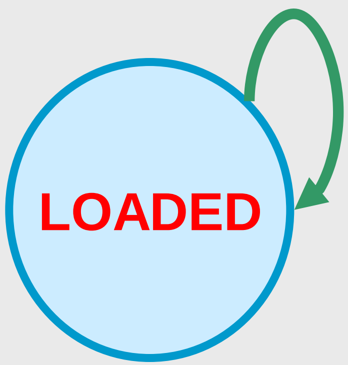
Honest casino



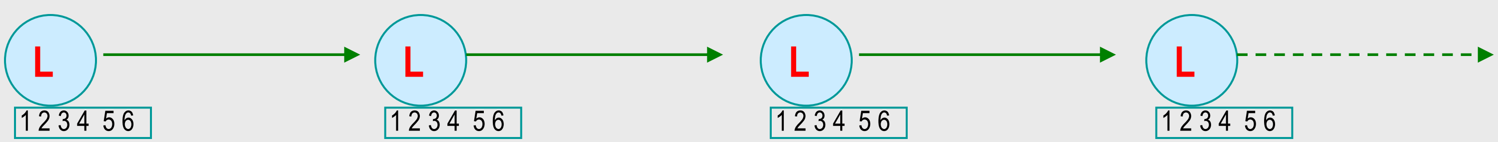
P(1) = 1/6
P(2) = 1/6
P(3) = 1/6
P(4) = 1/6
P(5) = 1/6
P(6) = 1/6



Crooked casino:



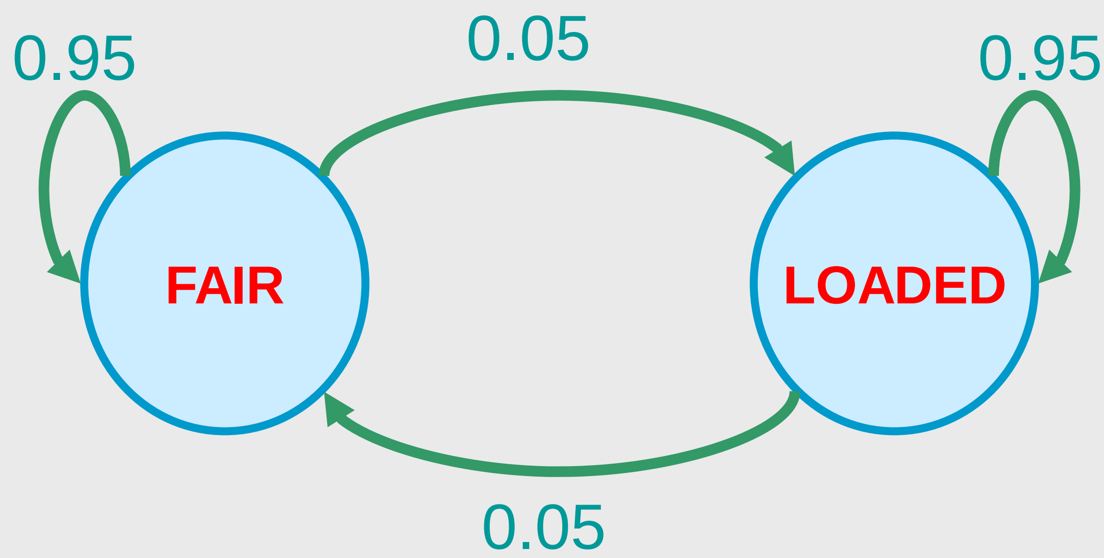
P(1) = 1/10
P(2) = 1/10
P(3) = 1/10
P(4) = 1/10
P(5) = 1/10
P(6) = 1/2



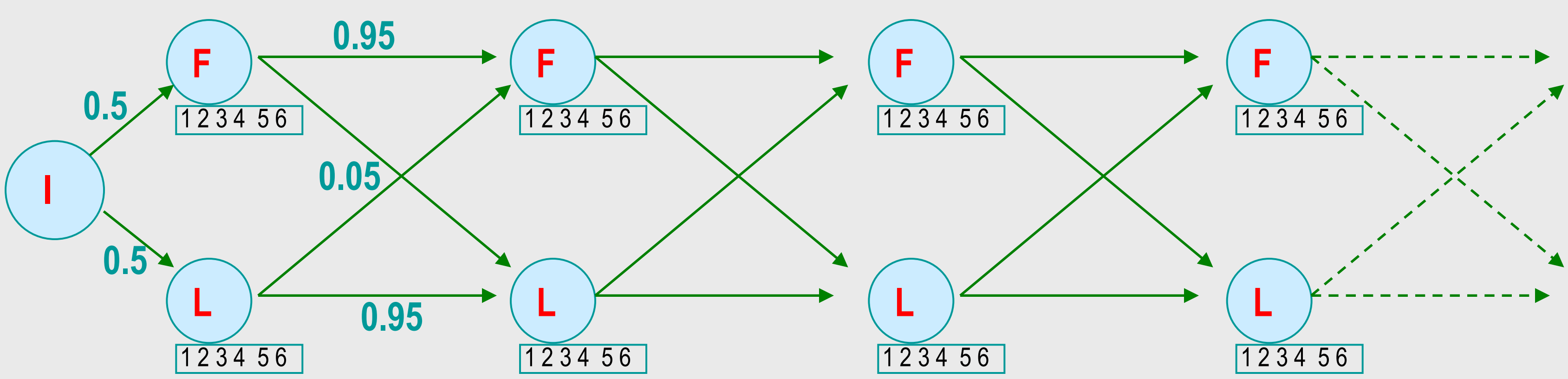
The casino models

Dishonest casino:

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6



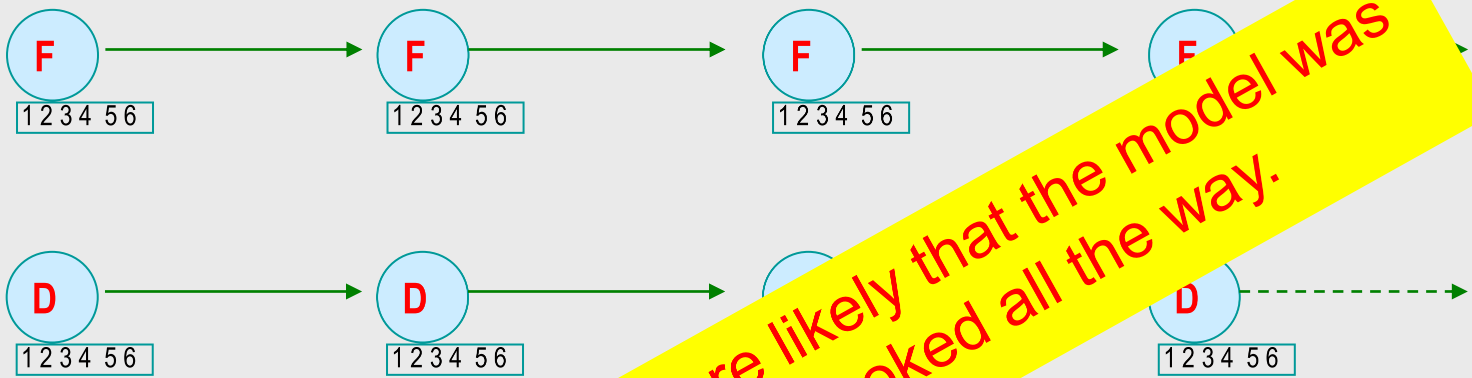
P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



The honest and crooked casino models

P(1) = 1/6
P(2) = 1/6
P(3) = 1/6
P(4) = 1/6
P(5) = 1/6
P(6) = 1/6

P(1) = 1/10
P(2) = 1/10
P(3) = 1/10
P(4) = 1/10
P(5) = 1/10
P(6) = 1/2



Let the sequence of roll… 1, 2, 1, 5, 6, 2, 1, 6, 2, 4

Therefore, it is after all 6.8 times more likely that the model was honest all the way, than that it was crooked all the way.

- Then, what is the lik… π = F, F, F, F, F, F, F, F, F, F?

$P(x|H)=$ … $(4) = (1/6)^{10}=1.7\times10^{-8}$

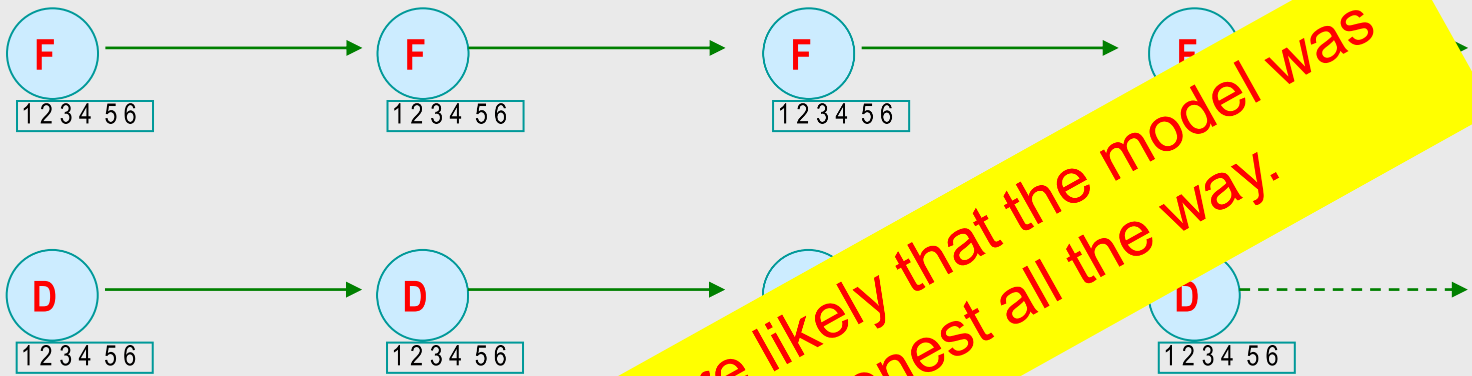
- A… …od of π = L, L, L, L, L, L, L, L, L, L?

…$(x|C)= P(1) \ldots P(4) = (1/10)^8 \times (1/2)^2 = 2.5 \times 10^{-9}$

The honest and crooked casino models

P(1) = 1/6
P(2) = 1/6
P(3) = 1/6
P(4) = 1/6
P(5) = 1/6
P(6) = 1/6

P(1) = 1/10
P(2) = 1/10
P(3) = 1/10
P(4) = 1/10
P(5) = 1/10
P(6) = 1/2



Let the sequence of rol… 1, 6, 6, 5, 6, 2, 6, 6, 3, 6

- Then, what is the lik… = F, F, F, F, F, F, F, F, F, F?

P(x|H)= … (4) = (1/6)^10=1.7x10^{-8}

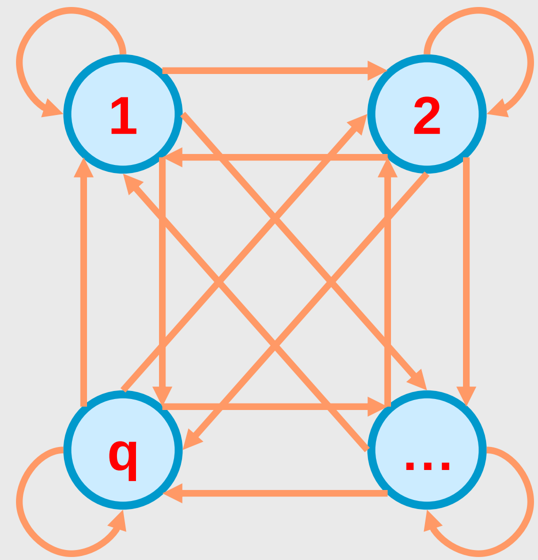
- A… …od of π = L, L, L, L, L, L, L, L, L, L?

…(x|C)= P(1) … P(4) = $(1/10)^4 \times (1/2)^6 = 1.6 \times 10^{-6}$

Therefore, it is after all 98 times more likely that the model was crooked all the way, than that it was honest all the way.

Representation of a HMM

algorithmics genetics group.



Definition: A hidden Markov model (HMM)

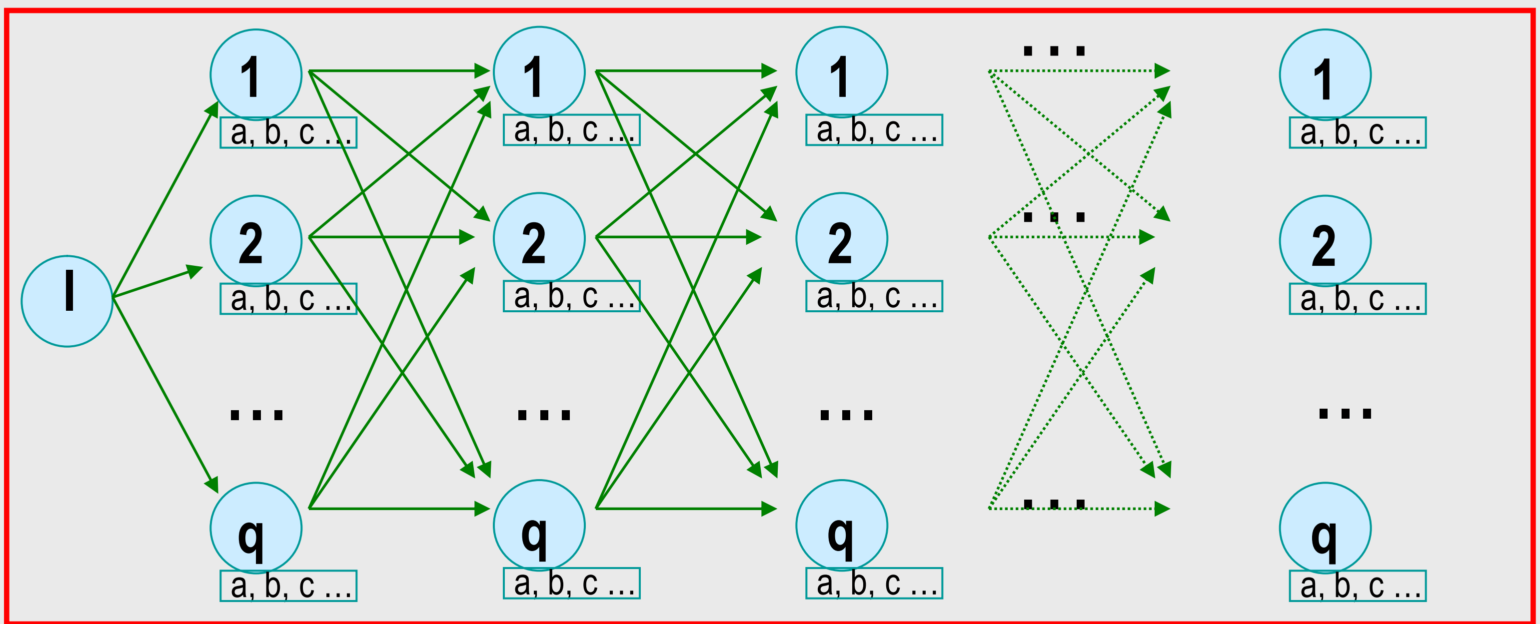
- **Alphabet** $\Sigma = \{a,b,c,\ldots\} = \{ b_1, b_2, \ldots, b_M \}$
- **Set of states** $Q = \{ 1, \ldots, q \}$
- **Transition probabilities** between any two states:

  p_{ij} = transition prob from state i to state j

  $p_{i1} + \ldots + p_{iq} = 1$, for all states i = 1…q
- **Start probabilities** p_{0i} such that $p_{01} + \ldots + p_{0q} = 1$
- **Emission probabilities** within each state

  $e_i(b) = P(x = b \mid q = i)$

  $e_i(b_1) + \ldots + e_i(b_M) = 1$, for all states i = 1…q



General questions

Evaluation problem: how likely is this sequence, given our model of how the casino works?

GIVEN a HMM M and a sequence x, FIND Prob[x | M]

Decoding problem: what portion of the sequence was generated with the fair die, and what portion with the loaded die?

GIVEN a HMM M, and a sequence x,

FIND the sequence π of states that maximizes P[x, π | M]

Learning problem: how “loaded” is the loaded die? How “fair” is the fair die? How often does the casino player change from fair to loaded, and back? Are there only two dies?

GIVEN a HMM M, with unspecified transition/emission probs. θ , and a sequence x,

FIND parameters θ that maximize P[x | θ]

Evaluation problem: Forward Algorithm

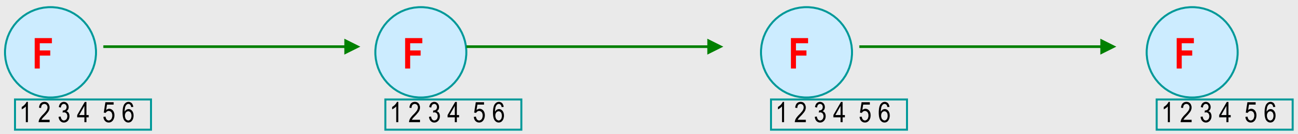
We want to calculate

$P(x \mid M)$ = probability of a sequence x, given the HMM M

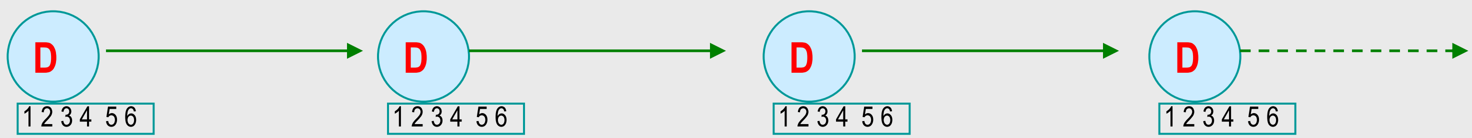
= Sum over all possible ways of generating x:

Given x= 1, 4, 2, 3, 6, 6, 3…, how many ways generate x?

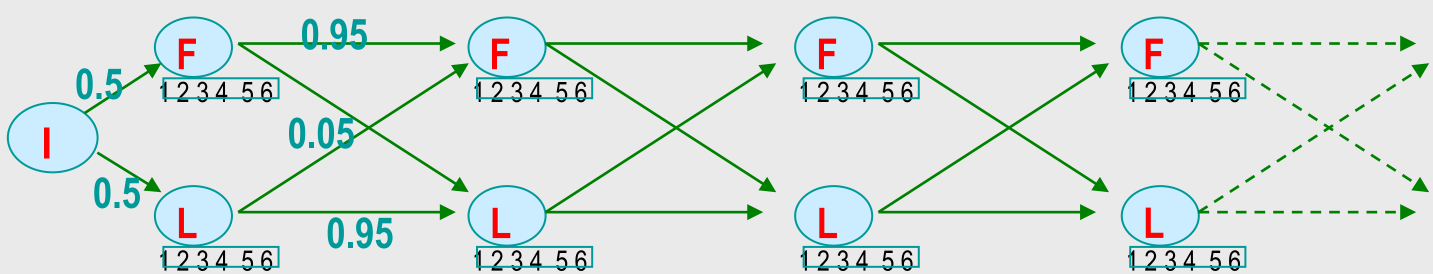
- Honest casino: only one way



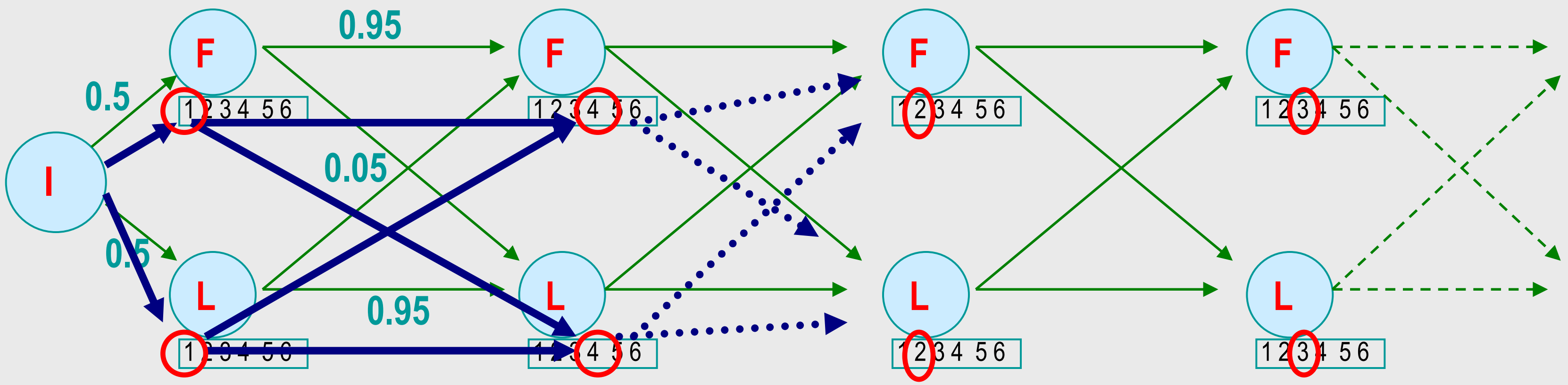
- Crooked casino: only one way



- Dishonest casino: ?



Evaluation problem: Forward Algorithm



We want to calculate

P(x | D) = probability of x, given the HMM D

= Sum over all possible ways of generating x:

Given x= 1, 4, 2, 3, 6, 6, 3…, how many ways generate x? $2^{|x|}$

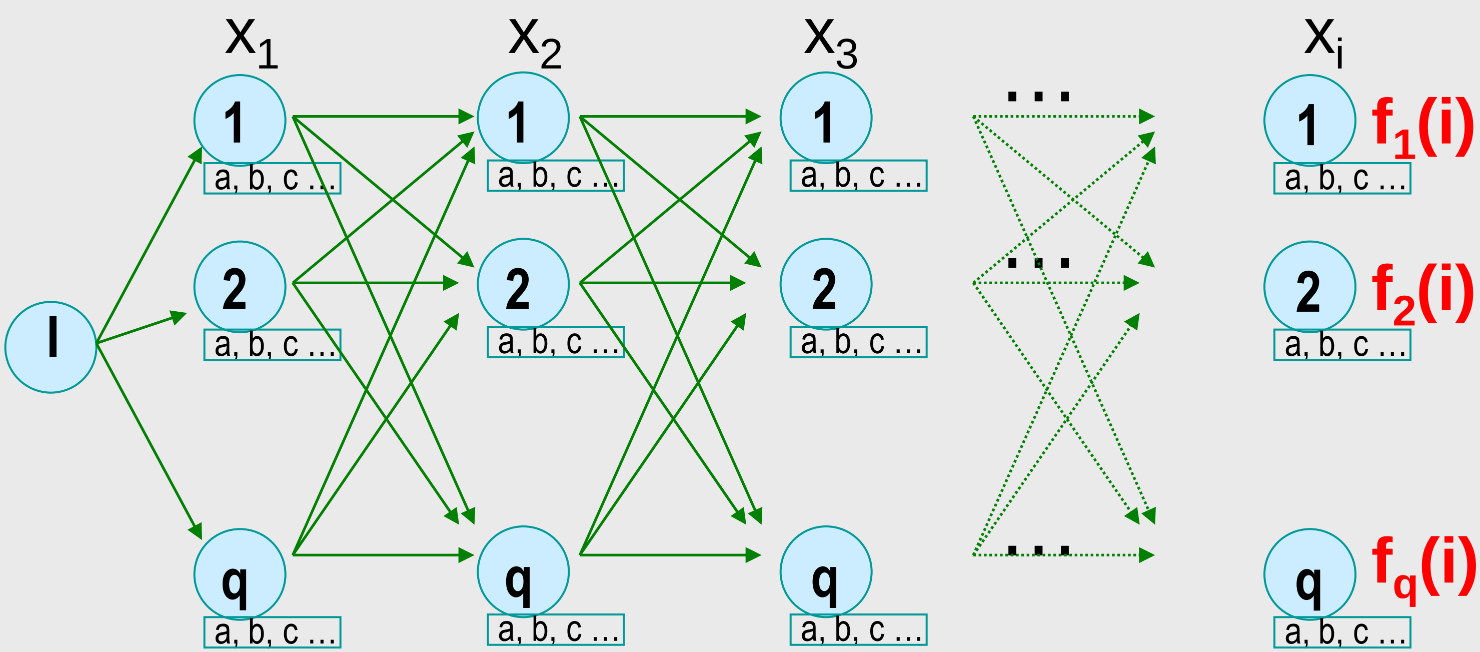
Naïve computation is very expensive: given |x| characters and N states, there are $N^{|x|}$ possible state sequences. Even small HMMs, |x|=10 and N=10, contain 10 billion different paths!

Evaluation problem: Forward Algorithm

P(x) = probability of x, given the HMM D

= Sum over all possible ways of generating x:

$$= \Sigma_\pi P(x, \pi) = \Sigma_\pi P(x \mid \pi) P(\pi)$$

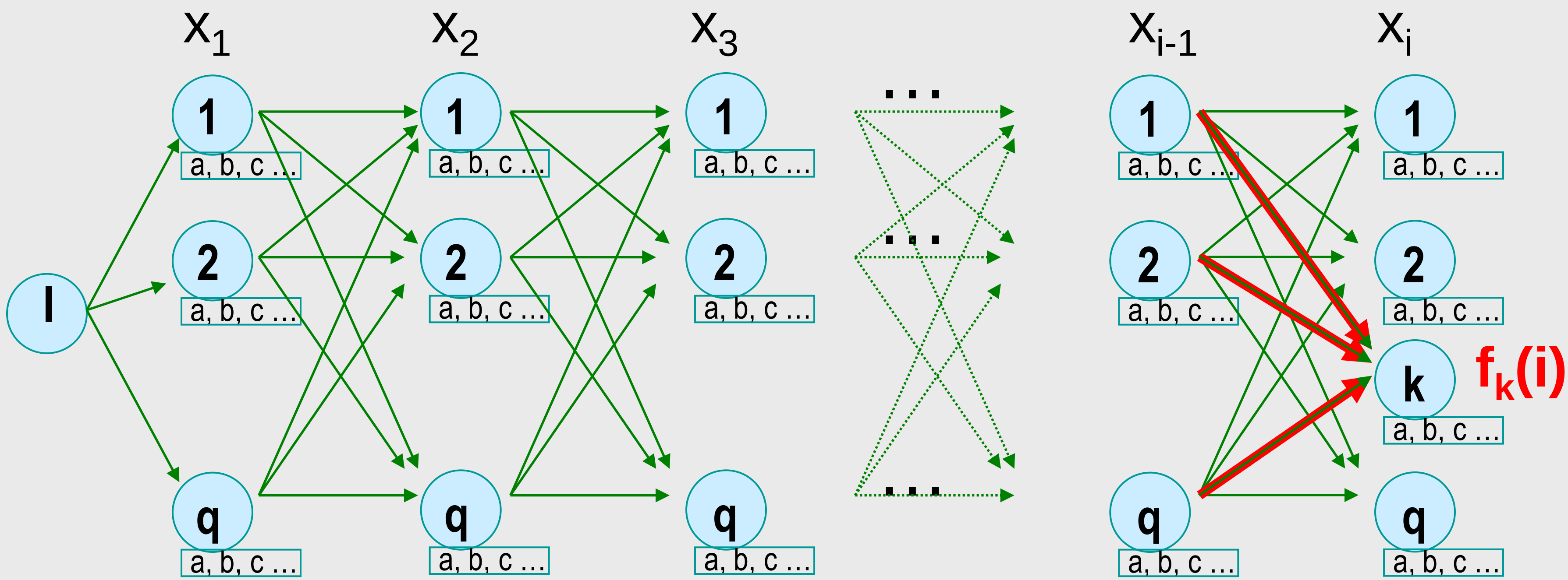


The probability of prefix $x_1x_2\ldots x_i$

Then, define

$$f_k(i) = P(x_1\ldots x_i, \pi_i = k) \quad \text{(the forward probability)}$$

Evaluation problem: Forward Algorithm



The forward probability recurrence:

$f_k(i) = P(x_1 \ldots x_i, \pi_i = k)$

$= \Sigma_{h=1..q} P(x_1 \ldots x_{i-1}, \pi_{i-1} = h) p_{hk}\, e_k(x_i)$

$= e_k(x_i)\, \Sigma_{h=1..q} P(x_1 \ldots x_{i-1}, \pi_{i-1} = h) p_{hk}$

$= e_k(x_i)\, \Sigma_{h=1..q}\ f_h(i-1)\, p_{hk}$

with $f_0(0) = 1$

$f_k(0) = 0$, for all $k > 0$

and cost

space: $O(Nq)$

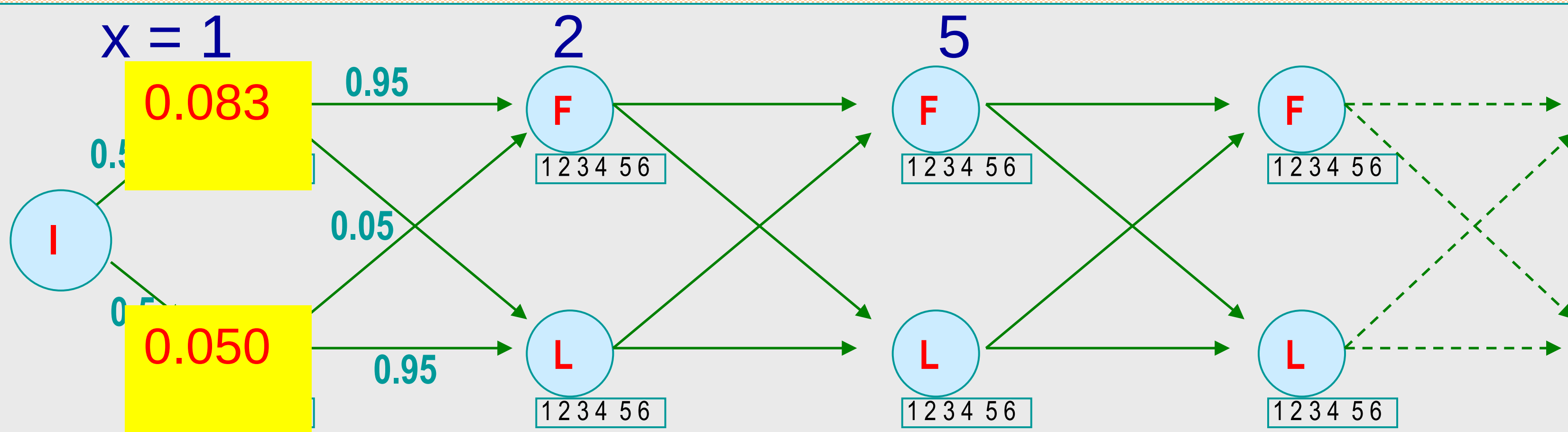
time: $O(Nq^{2)}$

The dishonest casino model

algorithmics genetics group .

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



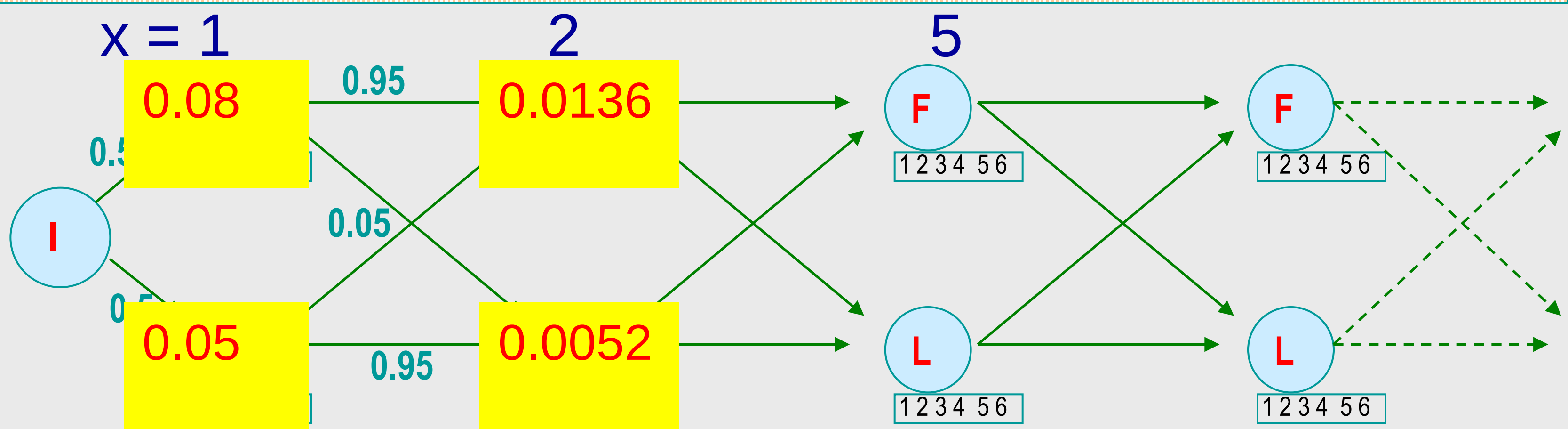
$$f_k(i) = e_k(x_i)\ \Sigma_{h=1..q}\ f_h(i-1)\ p_{hk}$$

$f_k(i)$	0	1	2	5
F	0	1/12 0.08		
L	0	1/20 0.05		

The dishonest casino model

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



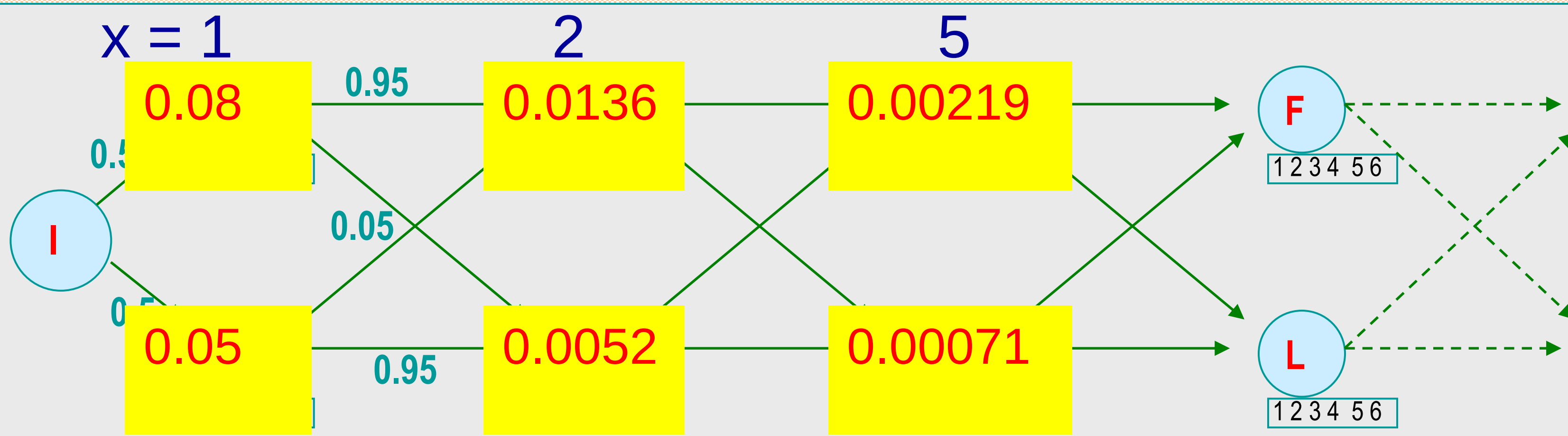
$$f_k(i) = e_k(x_i)\ \Sigma_{h=1..q}\ f_h(i-1)\ p_{hk}$$

$f_k(i)$	0	1	2	5
F	0	1/12 0.083	0.0136	
L	0	1/20 0.05	0.0052	

The dishonest casino model

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



$$f_k(i) = e_k(x_i) \Sigma_{h=1..q} \; f_h(i-1) \, p_{hk}$$

$f_k(i)$	0	1	2	5
F	0	1/12 0.08	0.0136	0.002197
L	0	1/20 0.05	0.0052	0.000712

Then P(125) = 0.003909

The dishonest casino model

4 6 6 1 6 1 5 4 3 5 6 3 2 1 2 2 3 5 6 5 1 4 6 1 1 6 1 3 5 3 3 5 6 2 3 5 5 2 2 2 2 3 4 5 3 2 5 5 5 4 6 5 3 4 2 6 6 1 2 5 6 4 2
3 3 2 1 1 6 1 2 5 4 4 4 4 4 2 4 6 4 3 2 2 2 3 4 5 6 1 5 1 5 1 6 3 2 3 3 4 2 1 6 1 1 3 5 2 5 6 3 3 2 6 4 3 3 5 3 2 6 3 2 1 6 6
3 6 1 4 3 4 3 1 1 3 1 4 3 3 5 5 4 … 5 1 2 6 3 5 6 4 1 6 2 6 5 5 4 5 5 2 2 2 2 5 4 3 4
1 6 3 3 4 6 3 1 4 5 6 4 2 6 1 6 2 … 2 6 3 2 6 3 2 2 6 3 4 5 4 2 2 6 5 1 3 6 4 1 1 2 1
1 5 3 1 3 3 5 2 3 1 1 6 3 3 6 3 2 … 4 5 6 5 2 6 1 3 2 2 3 5 3 5 6 2 4 1 3 3 1 4 1 5 6
1 5 2 4 1 4 1 1 5 1 3 3 3 1 6 2 3 … 6 3 2 6 5 2 4 4 2 4 4 4 5 6 4 3 6 5 5 6 3 5 3 3 1
6 4 3 6 5 1 6 1 3 2 1 4 4 1 4 2 5 6 6 4 2 6 5 4 4 4 3 4 6 2 5 6 1 6 5 5 1 1 3 2 4 5 5 2 6 2 6 3 1 1 5 6 4 6 5 1 6 3 1 3 1 6 6
1 5 6 1 4 6 4 4 6 3 2 6 5 3 1 1 4 2 3 3 6 3 5 1 3 6 1 2 6 3 2 1 3 2 5 4 5 1 6 2 3 6 1 2 6 1 2 5 4 2 4 6 6 1 1 2 3 1 2

Honest casino

Prob (S | Honest casino model) = exp (-896)

Prob (S | Dishonest casino model)= exp (-916)

6 6 6 2 4 5 2 1 5 3 5 6 6 3 1 5 2 3 6 3 6 4 1 3 6 6 5 5 2 3 1 2 5 2 4 3 3 6 6 2 6 1 6 6 6 2 6 4 4 6 2 3 1 1 2 1 3 5 1 2 1 6 2 1
6 3 6 6 2 6 2 6 6 6 1 6 6 6 3 3 6 4 6 6 6 4 5 5 4 4 5 5 4 3 5 1 6 2 4 6 1 6 6 4 6 6 6 2 5 6 4 6 4 1 6 5 4 5 3 2 1 1 6 5 4 3 6 3
2 6 1 2 3 3 6 3 6 4 3 1 1 1 5 5 … 5 2 4 5 1 5 2 3 1 6 2 1 5 1 1 6 6 1 4 4 3 1 6 5 6
6 6 1 1 1 6 6 1 4 5 5 3 6 1 2 … 3 6 6 6 1 4 6 2 5 6 5 6 6 6 6 6 6 1 1 1 5 4 5 6
4 1 6 2 3 1 6 6 4 2 6 5 6 6 6 … 6 6 6 4 6 1 1 1 4 6 3 1 1 2 6 4 4 6 6 6 2 6 1 6
1 1 5 6 6 2 5 6 3 5 6 6 3 1 4 … 5 4 3 4 2 2 5 6 6 3 4 6 6 1 5 1 6 3 2 5 1 6 4 6
6 4 1 6 6 3 4 5 1 6 5 6 6 2 4 4 3 3 5 3 4 5 1 2 5 2 2 6 6 2 6 6 5 6 1 5 1 5 4 1 6 4 6 1 6 6 6 2 5 4 3 4 6 4 2 6 6 3 4 3 4 3 1 5
5 4 6 4 3 2 6 6 4 5 5 5 4 6 5 2 2 4 6 5 3 6 2 2 2 6 1 5 6 2 3 6 5 6 6 6 4 6 5 3 6 6 6 3 4 2 2 2 5 6 6 4

Dishonest casino

Prob (S | Honest casino model) = exp (-896)

Prob (S | Dishonest casino model)= exp (-847)

General questions

Evaluation problem: how likely is this sequence, given our model of how the casino works?

GIVEN a HMM M and a sequence x, FIND Prob[x | M]

Decoding problem: what portion of the sequence was generated with the fair die, and what portion with the loaded die?

GIVEN a HMM M, and a sequence x,

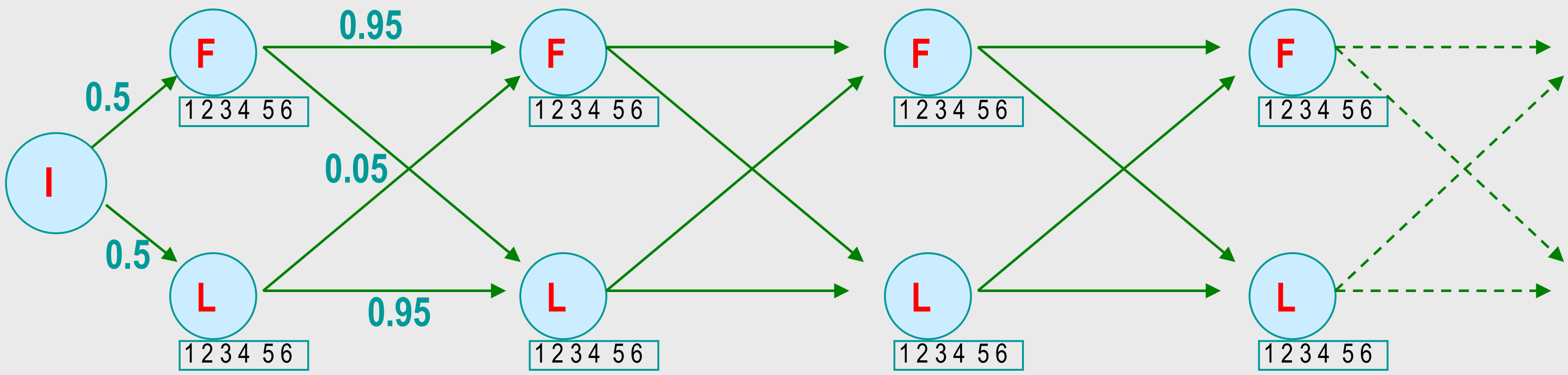
FIND the sequence π of states that maximizes P[x, π | M]

Learning problem: how "loaded" is the loaded die? How "fair" is the fair die? How often does the casino player change from fair to loaded, and back?

GIVEN a HMM M, with unspecified transition/emission probs. θ , and a sequence x,

FIND parameters θ that maximize P[x | θ]

Decoding problem



We want to calculate path π^* such that

$\pi^* = \text{argmax}_\pi P(x, \pi \mid M)$

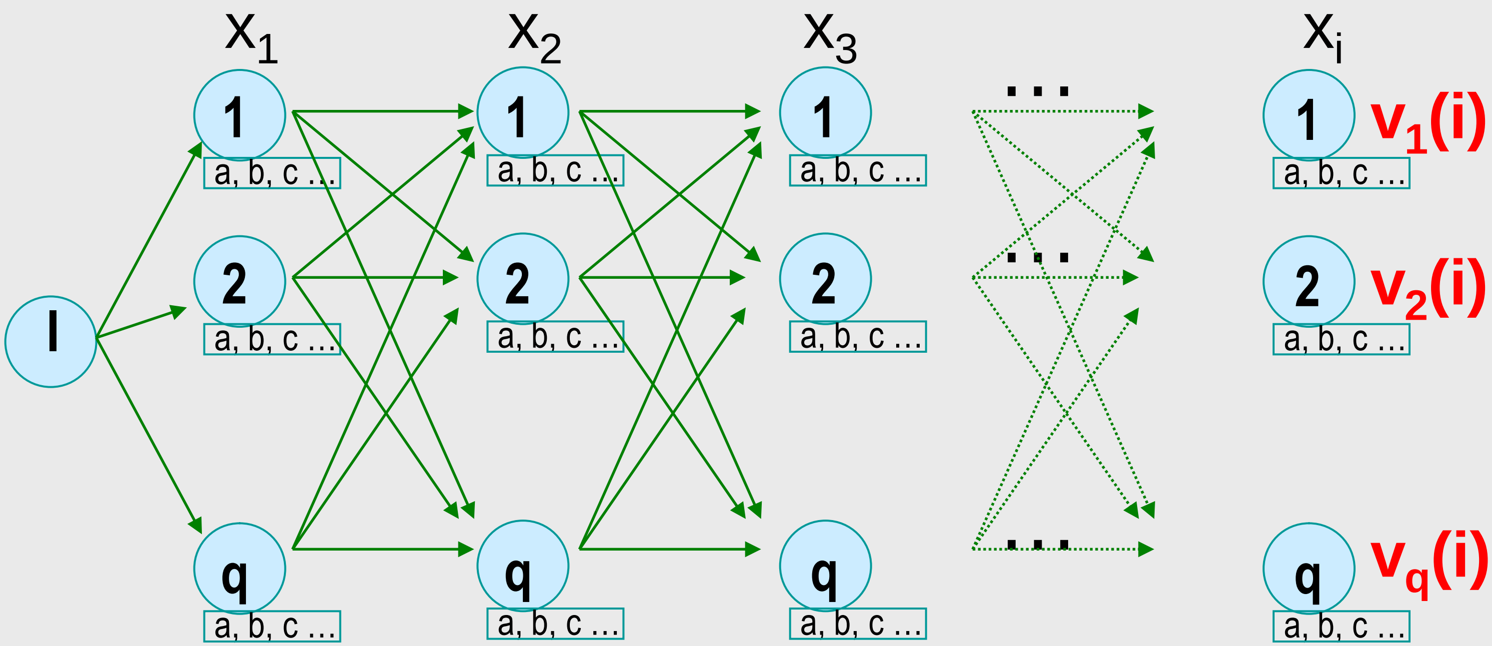
= the sequence π of states that maximizes $P(x, \pi \mid M)$

Naïve computation is very expensive: given $|x|$ characters and N states, there are $N^{|x|}$ possible state sequences.

Evaluation problem: Viterbi algorithm

$\pi^* = \text{argmax}_\pi P(x, \pi \mid M)$

$= $ the sequence π of states that maximizes $P(x, \pi \mid M)$

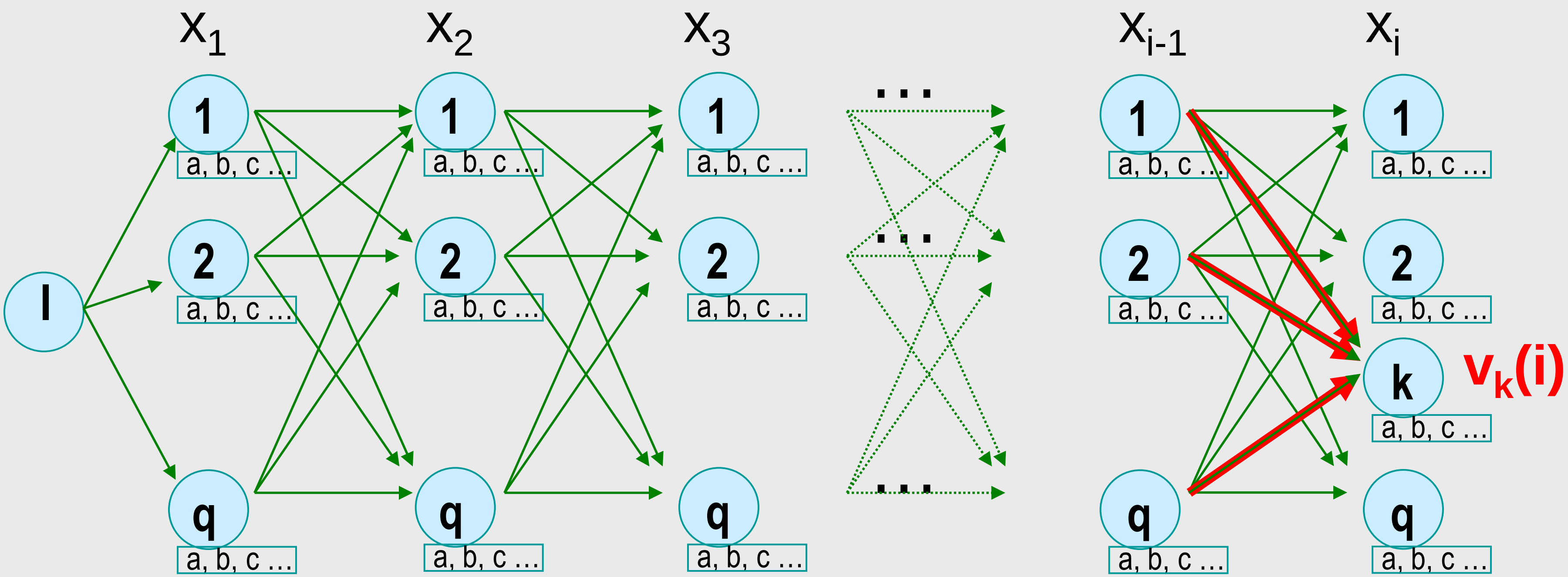


The sequence of states that maximizes $x_1x_2\ldots x_i$

Then, define

$$v_k(i) = \text{argmax}_\pi P(x_1\ldots x_i, \pi_i = k)$$

Evaluation problem: Viterbi algorithm



The forward probability recurrence:

$v_k(i) = \text{argmax}_{\pi}\ P(x_1 \ldots x_i,\ \pi_i = k)$

$= \max_h [\text{argmax}_{\pi}\ P(x_1 \ldots x_{i-1},\ \pi_{i-1} = h) p_{hk}\, e_k(x_i)]$

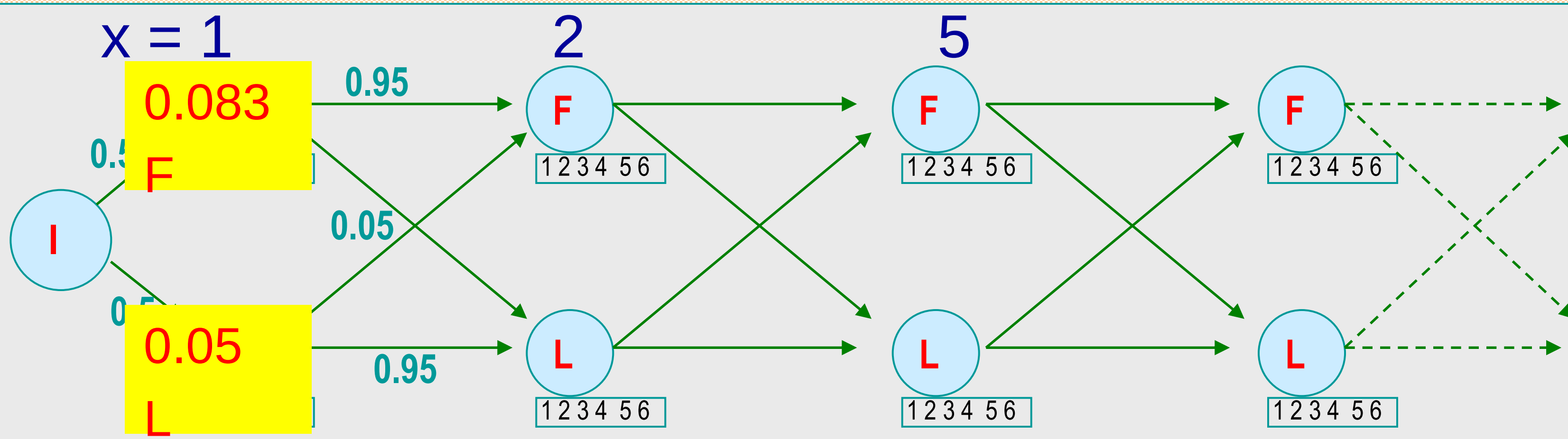
$= e_k(x_i) \max_h [p_{hk}\, \text{argmax}_{\pi}\ P(x_1 \ldots x_{i-1},\ \pi_{i-1} = h)]$

$= e_k(x_i) \max_h [p_{hk}\, v_h(i-1)]$

The dishonest casino model

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



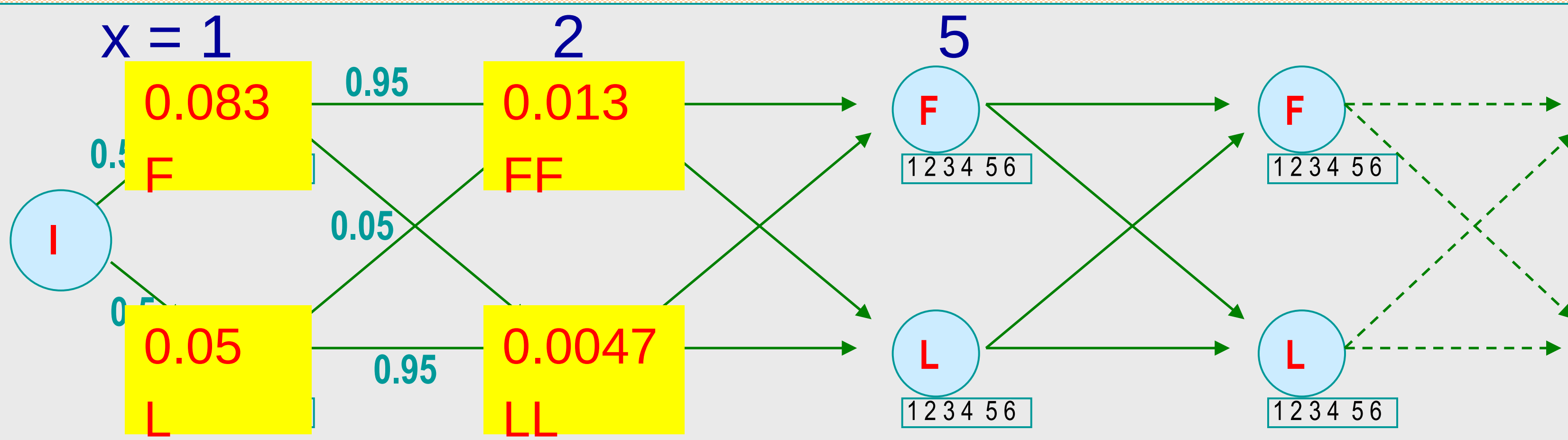
$$f_k(i) = e_k(x_i) \max_{h=1..q} v_h(i-1)\, p_{hk}$$

$f_k(i)$	0	1	2	5
F	0	1/12 0.08		
L	0	1/20 0.05		

The dishonest casino model

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



$$f_k(i) = e_k(x_i)\ \max_{h=1..q}\ v_h(i-1)\ p_{hk}$$

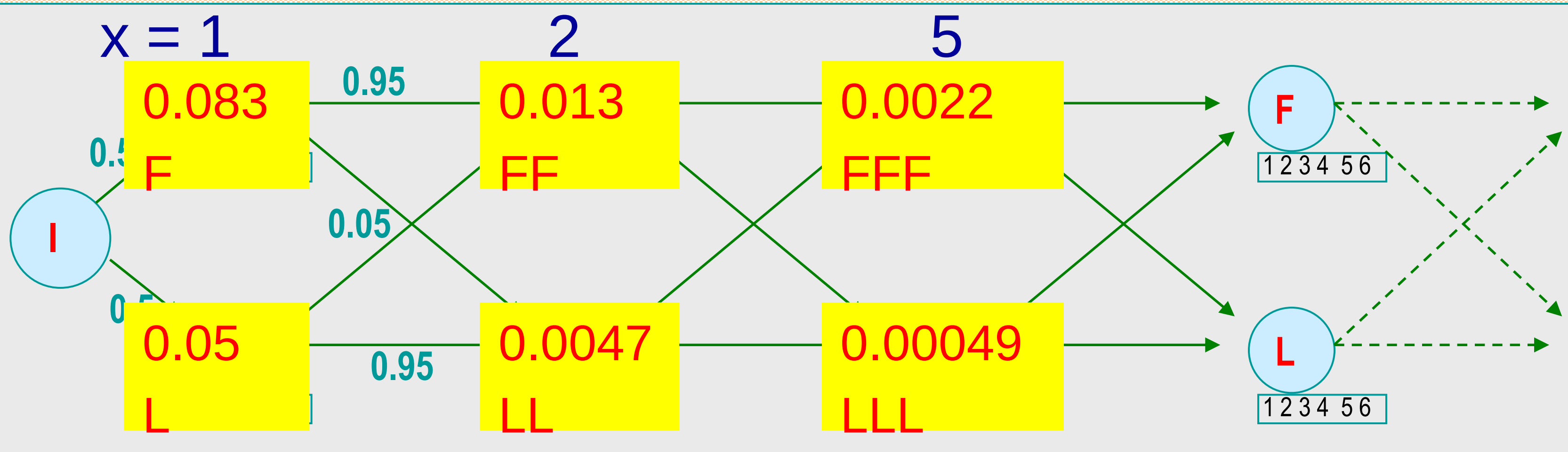
$f_k(i)$	0	1	2	5
F	0	1/12	max(0.013, 0.0004)	
		0.08	0.013	
L	0	1/20	max(0.00041,0.0047)	
		0.05	0.0047	

The dishonest casino model

algorithmics genetics group .

P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2



$$f_k(i) = e_k(x_i)\ \max_{h=1..q}\ v_h(i-1)\ p_{hk}$$

$f_k(i)$	0	1	2	5
F	0	1/12	max(0.013, 0.0004)	max(0.0022,0.000043)
		0.08	0.0126	0.0022
L	0	1/20	max(0.00041,0.0047)	max(0.000068,00049)
		0.05	0.0047	0.00049

Then, the most probable path is FFF !

The dishonest casino model

algorithmics genetics group

Dishonest casino sequence of values:

6 6 6 2 4 5 2 1 5 3 5 6 6 3 1 5 2 3 6 3 6 4 1 3 6 6 5 5 2 3 1 2 5 2 4 3 3 6 6 2 6 1 6 6 6 2 6 4 4 6 2 3 1 1 2 1 3 5 1 2 1 6 2 1
6 3 6 6 2 6 2 6 6 6 1 6 6 6 3 3 6 4 6 6 6 4 5 5 4 4 5 5 4 3 5 1 6 2 4 6 1 6 6 4 6 6 6 2 5 6 4 6 4 1 6 5 4 5 3 2 1 1 6 5 4 3 6 3
2 6 1 2 3 3 6 3 6 4 3 1 1 1 5 5 3 2 1 1 2 4 3 2 1 2 4 6 6 3 6 4 6 1 4 6 6 6 6 5 2 4 5 1 5 2 3 1 6 2 1 5 1 1 6 6 1 4 4 3 1 6 5 6
6 6 1 1 1 6 6 1 4 5 5 3 6 1 2 6 1 2 6 1 4 6 6 6 6 3 6 4 5 1 4 6 5 6 5 5 6 6 3 6 3 6 6 6 1 4 6 2 5 6 5 6 6 6 6 6 6 1 1 1 5 4 5 6
4 1 6 2 3 1 6 6 4 2 6 5 6 6 6 5 4 5 3 3 3 4 2 4 1 6 6 1 4 6 6 6 6 1 1 5 5 4 6 6 6 6 6 4 6 1 1 1 4 6 3 1 1 2 6 4 4 6 6 6 2 6 1 6
1 1 5 6 6 2 5 6 3 5 6 6 3 1 4 5 6 6 1 6 4 5 1 4 1 3 3 6 6 6 6 3 3 2 6 2 2 1 4 5 5 4 3 4 2 2 5 6 6 3 4 6 6 1 5 1 6 3 2 5 1 6 4 6
6 4 1 6 6 3 4 5 1 6 5 6 6 2 4 4 3 3 5 3 4 5 1 2 5 2 2 6 6 2 6 6 5 6 1 5 1 5 4 1 6 4 6 1 6 6 6 2 5 4 3 4 6 4 2 6 6 3 4 3 4 3 1 5
5 4 6 4 3 2 6 6 4 5 5 5 4 6 5 2 2 4 6 5 3 6 2 2 2 6 1 5 6 2 3 6 5 6 6 6 4 6 5 3 6 6 6 3 4 2 2 2 5 6 6 4

Dishonest casino sequence of states:

2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

General questions

Evaluation problem: how likely is this sequence, given our model of how the casino works?

GIVEN a HMM M and a sequence x, FIND Prob[x | M]

Decoding problem: what portion of the sequence was generated with the fair die, and what portion with the loaded die?

GIVEN a HMM M, and a sequence x,

FIND the sequence π of states that maximizes P[x, π | M]

Learning problem: how "loaded" is the loaded die? How "fair" is the fair die? How often does the casino player change from fair to loaded, and back?

GIVEN a HMM M, with unspecified transition/emission probs. θ , and a sequence x,

FIND parameters θ that maximize P[x | θ]

Learning problem

How "loaded" is the loaded die? How "fair" is the fair die? How often does the casino player change from fair to loaded, and back?

GIVEN a HMM M, with unspecified transition/emission probs. θ , and a sequence x,

FIND parameters θ that maximize $P[\, x \mid \theta \,]$

We need a training data set. It could be:

- A sequence of pairs $(x,\pi) = (x_1,\pi_1), (x_2,\pi_2), \ldots ,(x_n,\pi_n)$ where we know the set of values and the states.

- A sequence of singles $x = x_1,x_2, \ldots ,x_n$ where we only know the set of values.

Learning problem: given $(x,\pi)^{i=1..n}$

From the training set we can define:

- H_{ki} as the number of times the transition from state k to state i appears in the training set.
- $J_l(x)$ as the number of times the value x is emitted by state l.

For instance, given the training set: Fair die, Loaded die

1 2 5 2 3 6 4 5 1 2 6 4 3 6 5 6 4 2 6 3 2 3 1 6 4 5

3 2 4 2 4 6 5 4 1 6 2 3 6 3 2 6 6 3 2 6 3 1 2 4 1 5 4 6 3

2 3 1 4 6 3 5 1 3 2 4 6 4 3 6 6 6 2 0 6 5 4 1 2 3 2 1 4 6 5 4

$H_{FF} = 51$ $H_{FL} = 4$ $H_{LF} = 4$ $H_{LL} = 26$

$J_F(1) = 10$ $J_F(2) = 11$ $J_F(3) = 9$

$J_F(4) = 12$ $J_F(5) = 8$ $J_F(6) = 6$

$J_L(1) = 0$ $J_L(2) = 5$ $J_L(3) = 6$

$J_L(4) = 3$ $J_L(5) = 1$ $J_L(6) = 14$

Learning problem: given $(x,\pi)^{i=1..n}$

From the training set we have computed:

- H_{ki} as the number of times the transition from state k to state i appears in the training set.
- $J_l(r)$ as the number of times the value r is emitted by state l.

And we estimate the parameters of the HMM as

- $p_{kl} = H_{ki} / (H_{k1} + \dots + H_{kq})$.

$H_{FF} = 51$	$H_{FL} = 4$	$H_{LF} = 4$	$H_{LL} = 26$
$p_{FF} = 51/85=0.6$	$p_{FL} = 0.045$	$p_{LF} = 0.045$	$p_{LL} = 0.31$

- $e_l(r) = J_l(r) /(J_1(r) +\dots+ J_q(r))$

$J_F(1) = 10$	$e_F(1)= 10/56$	$J_F (2) = 11$	$e_F(2)= 2/56$	$J_F(3)= 9$	$e_F(3)= 3/56$
$J_F (4) = 12$	$e_F(4)= 0.21$	$J_F (5) = 8$	$e_F(5)= 0.054$	$J_F (6) = 6$	$e_F(6)=0.1$
$J_L(1) = 0$	$e_L(6)= 0/29$	$J_L (2) = 5$	$e_L(6)= 5/29$	$J_L(3)=6$	$e_L(6)= 6/29$
$J_L (4) = 3$	$e_L(6)=0.1$	$J_L (5) = 1$	$e_L(6)=0.034$	$J_L (6) = 14$	$e_L(6)=0.48$

Learning problem: given $x^{i=1..n}$

To choose the parameters of HMM that maximize

$P(x^1) \times P(x^2) \times \ldots \times P(x^n)$

that implies

The use of standard (iterative) optimization algorithms:

- Determine initial parameters values
- Iterate until $P(x^1) \times P(x^2) \times \ldots \times P(x^n)$ becomes smaller that some predeterminated threshold

but

the algorithm may converge to a point close to a local maximum, not to a global maximum.

Learning problem: algorithm

From the training $x^{i=1..n}$ we estimate M^0:

- p_{ki} as the probability of transitions.
- $e_l(r)$ as the probability of emissions.

Do (we have M^s)

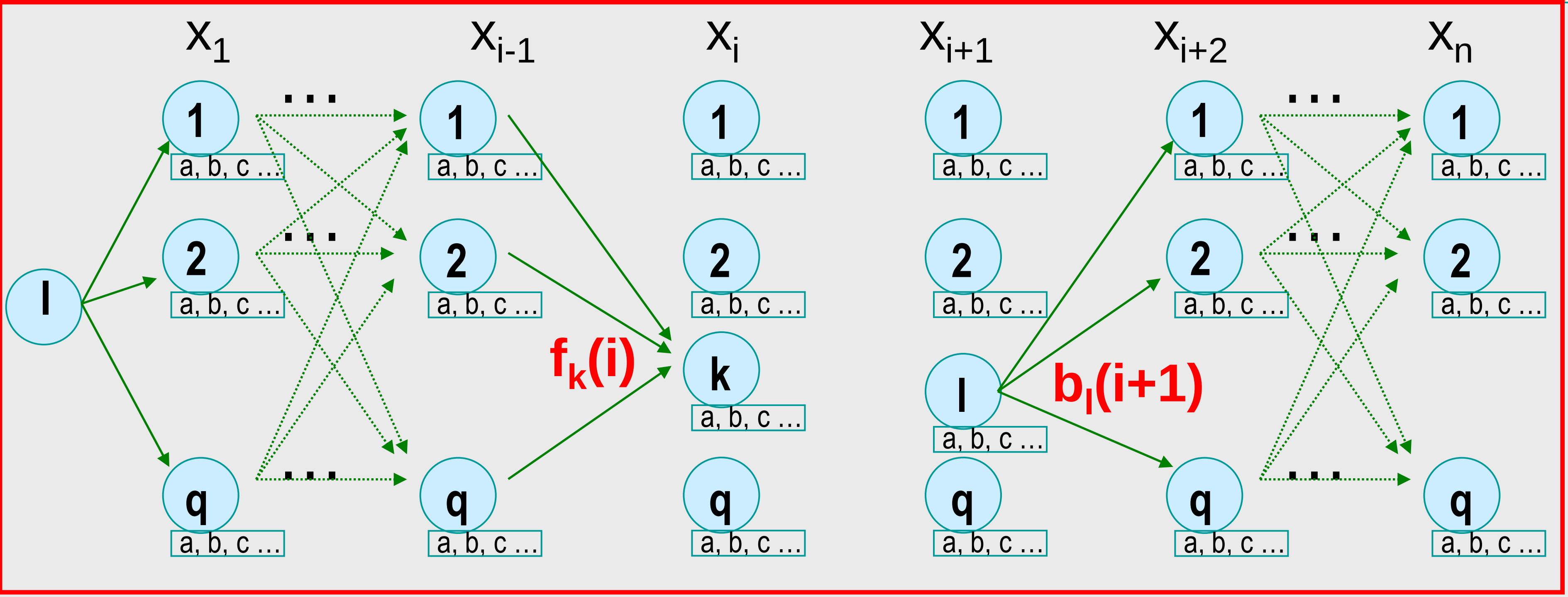
- Compute H_{ki} as the expected number of times the transition from state k to state l is reached.
- Compute $J_l(r)$ as the expected number of times the value r is emitted by state l.
- Compute

  $p_{ki} = H_{ki} / (H_{k1} + \ldots + H_{kq})$ and $e_l(r) = J_l(r) / (J_1(r) + \ldots + J_q(r))$.
- { we have M^{s+1}}

Until some value smaller than the threshold

{ M is close to a local maximum}

Recall forward and backward algorithms



The forward probability recurrence:

$$f_k(i) = P(x_1 \ldots x_i, \pi_i = k) = e_k(x_i) \, \Sigma_{h=1..q} \; f_h(i-1)$$

The backward probability recurrence:

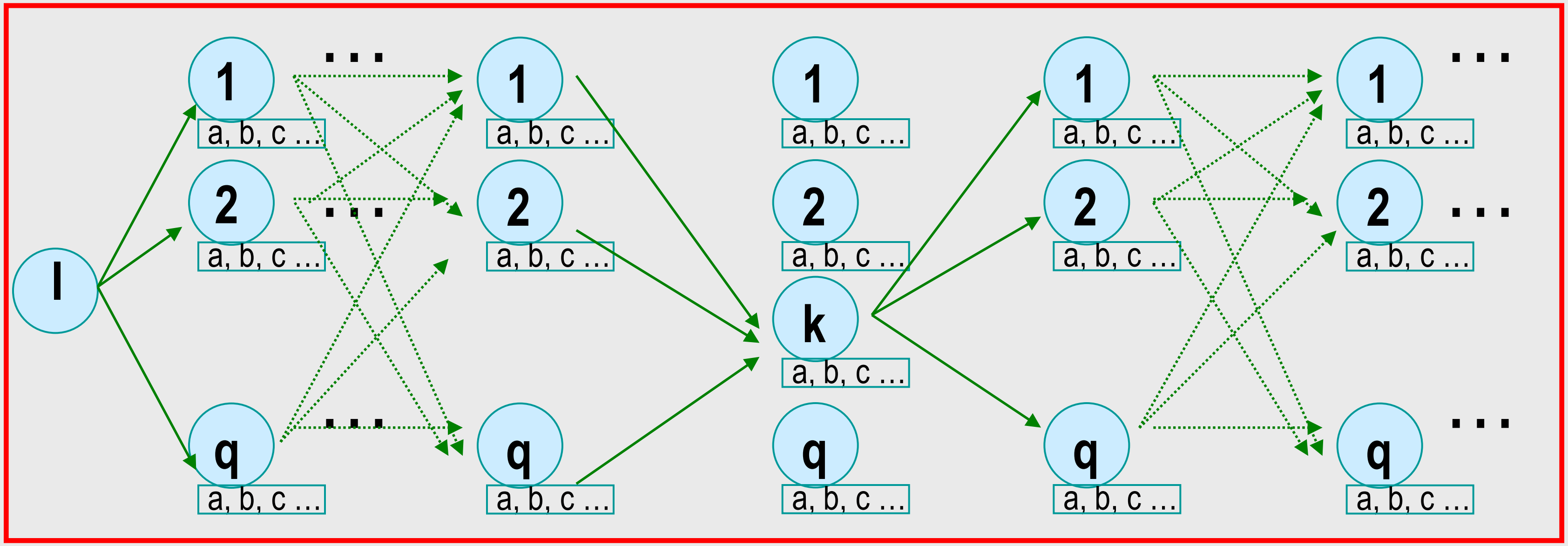
$$b_l(i+1) = P(x_{i+1} \ldots x_n, \pi_{i+1} = l) = \Sigma_{h=1..q} \, p_{lh} \, e_h(x_{i+2}) \, b_h(i+2)$$

Baum-Welch training algorithm

$J_k(r)$ = the expected number of times the value r is emitted by state k.

$$= \sum_{\text{all } x} \sum_{\text{all } i} \text{Prob(state k emits r at step i in sequence x)}$$

$$= \sum_{\text{all } x} \sum_{\text{all } i} \frac{\text{Prob}(x_1 \ldots x_n \mid \text{state k emits } x_i)}{\text{Prob}(x_1 \ldots x_n)} \delta(r = x_i)$$



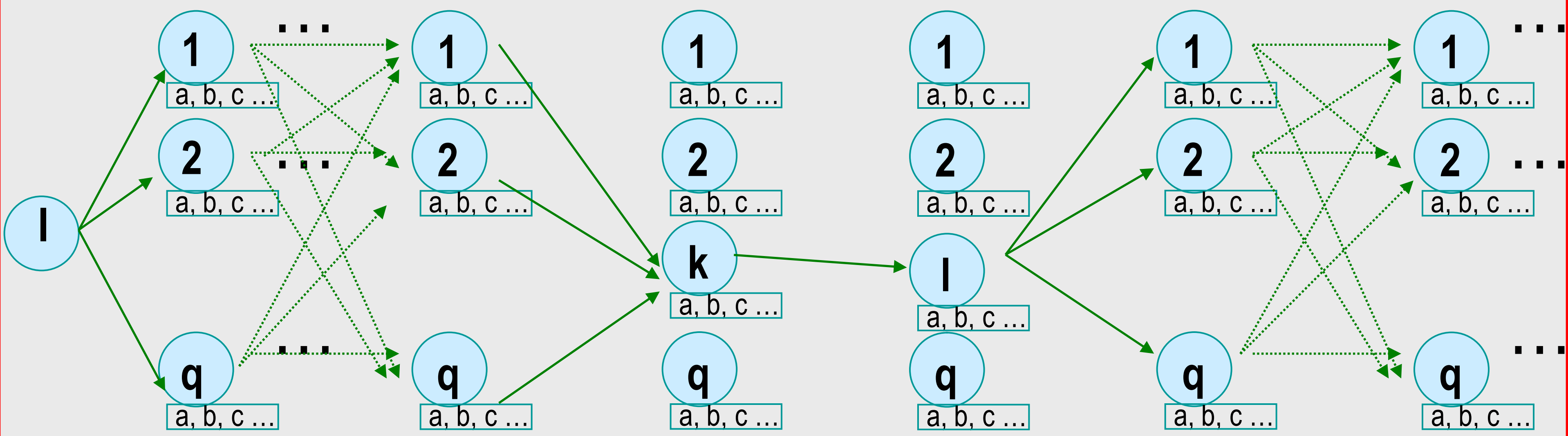
$$= \sum_{\text{all } x} \sum_{\text{all } i} \frac{f_k(i)\, b_k(i)}{\text{Prob}(x_1 \ldots x_n)} \delta(r = x_i)$$

Baum-Welch training algorithm

$H_{kl}(r)$ = as the expected number of times the transition from k to l is reached

$$= \sum_{\text{all } x} \sum_{\text{all } i} \text{Prob(transition from k to l is reached at step i in x)}$$

$$= \sum_{\text{all } x} \sum_{\text{all } i} \frac{\text{Prob}(x_1 \ldots x_n \mid \text{state k reaches state l})}{\text{Prob}(x_1 \ldots x_n)}$$



$$= \sum_{\text{all } x} \sum_{\text{all } i} \frac{f_k(i)\, p_{kl}\, e_l(x_{i+1})\, b_l(i+1)}{\text{Prob}(x_1 \ldots x_n)}$$

Baum-Welch training algorithm

H_{ki} as the expected number of times the transition from state k to state i appears.

$$H_{kl}(r) = \sum_{\textbf{all x}} \sum_{\textbf{all i}} \frac{f_k(i)\, p_{kl}\, e_l(x_{i+1})\, b_l(i+1)}{Prob(x_1 \ldots x_n)}$$

$J_l(r)$ as the expected number of times the value r is emitted by state l.

$$J_l(r) = \sum_{\textbf{all x}} \sum_{\textbf{all i}} \frac{f_k(i)\, b_k(i)}{Prob(x_1 \ldots x_n)} \quad \delta(r = x_i)$$

And we estimate the new parameters of the HMM as

- $p_{kl} = H_{ki} / (H_{k1} + \ldots + H_{kq})$.
- $e_l(r) = J_l(r) / (J_1(r) + \ldots + J_q(r))$

Baum-Welch training algorithm

The algorithm has been applied to the sequences:.

For |S|=500:

- M= 6 N= 2 P0F=0.004434 P0L=0.996566
- PFF=0.198205 PFL=0.802795 PLL=0.505259 PLF= 0.495741
- 0.166657 0.150660 0.054563 0.329760 0.026141 0.277220
- 0.140923 0.095672 0.152771 0.018972 0.209654 0.387008
- pi:
- 0.004434 0.996566

For |S|=50000:

- M= 6 N= 2 0.027532 0.973468
- 0.127193 0.873807
- 0.299763 0.701237
- 0.142699 0.166059 0.097491 0.168416 0.106258 0.324077
- 0.130120 0.123009 0.147337 0.125688 0.143505 0.335341