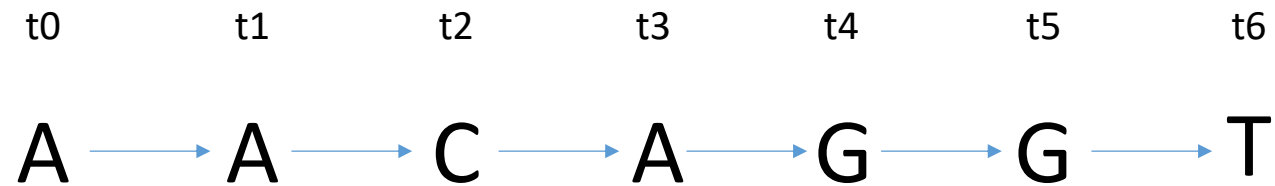


Session 2

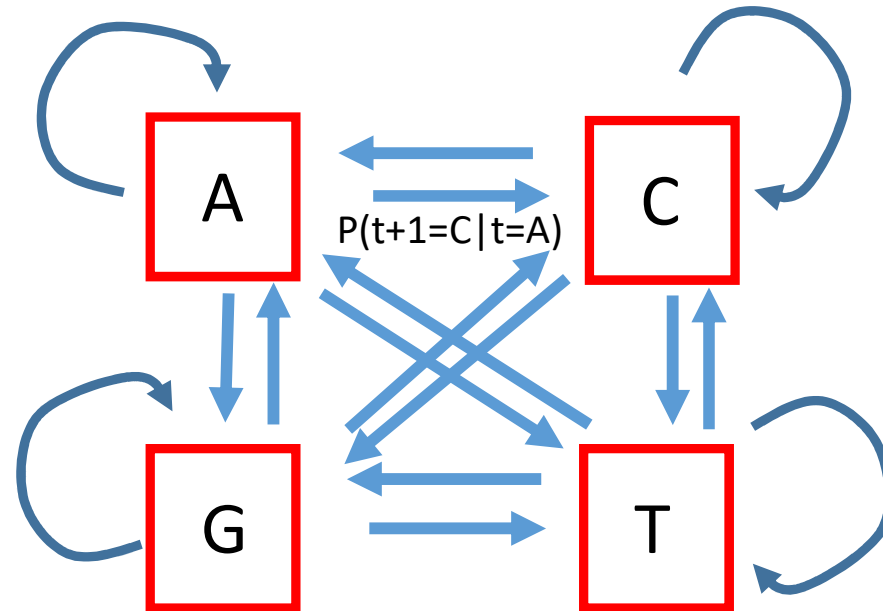
Hidden Markov Models applied to Clustering of Biological Sequences

Markov model

Markov chain



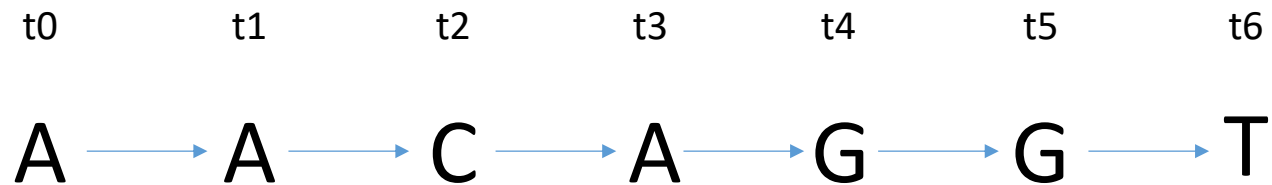
Four possible
categories



Probability
transition matrix

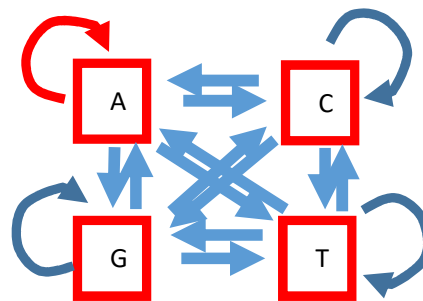
How to identify a pattern of characters?

Markov chain

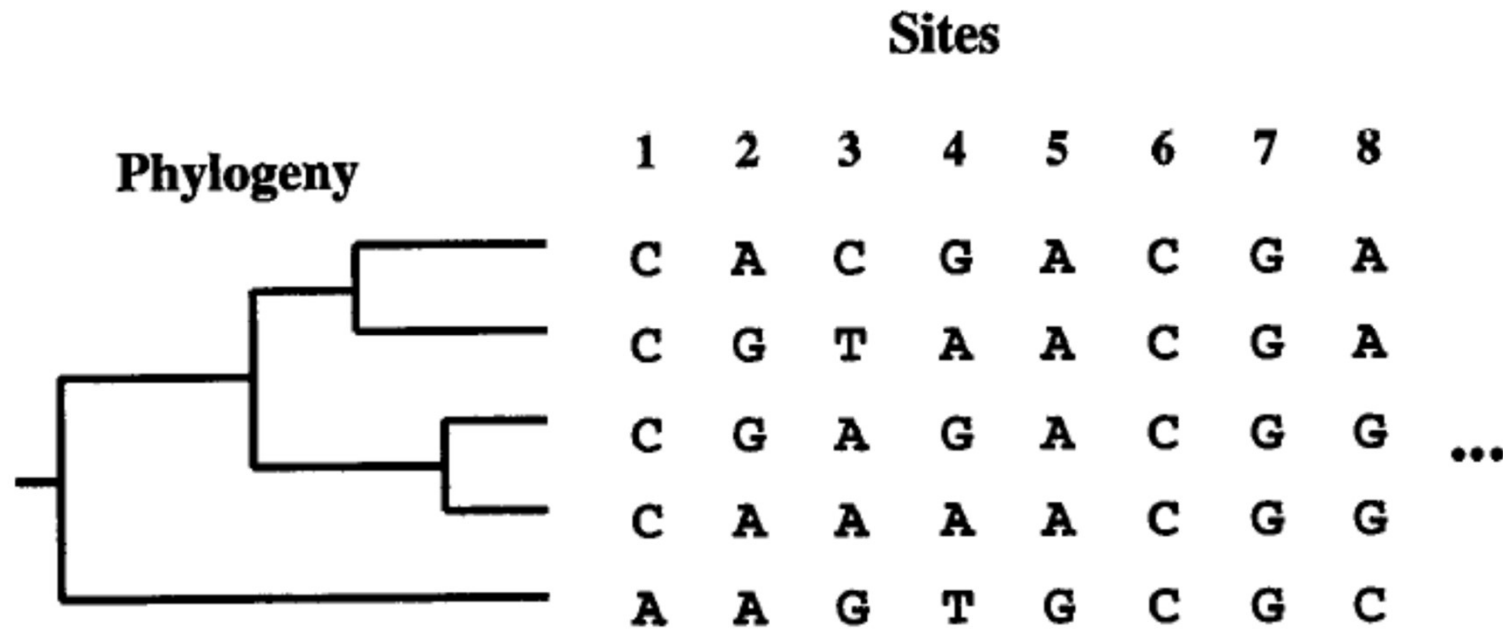


$$P(AACAGGT) = P(A)P(A|A)P(C|A)P(A|C)P(G|A)P(G|G)P(T|G)$$

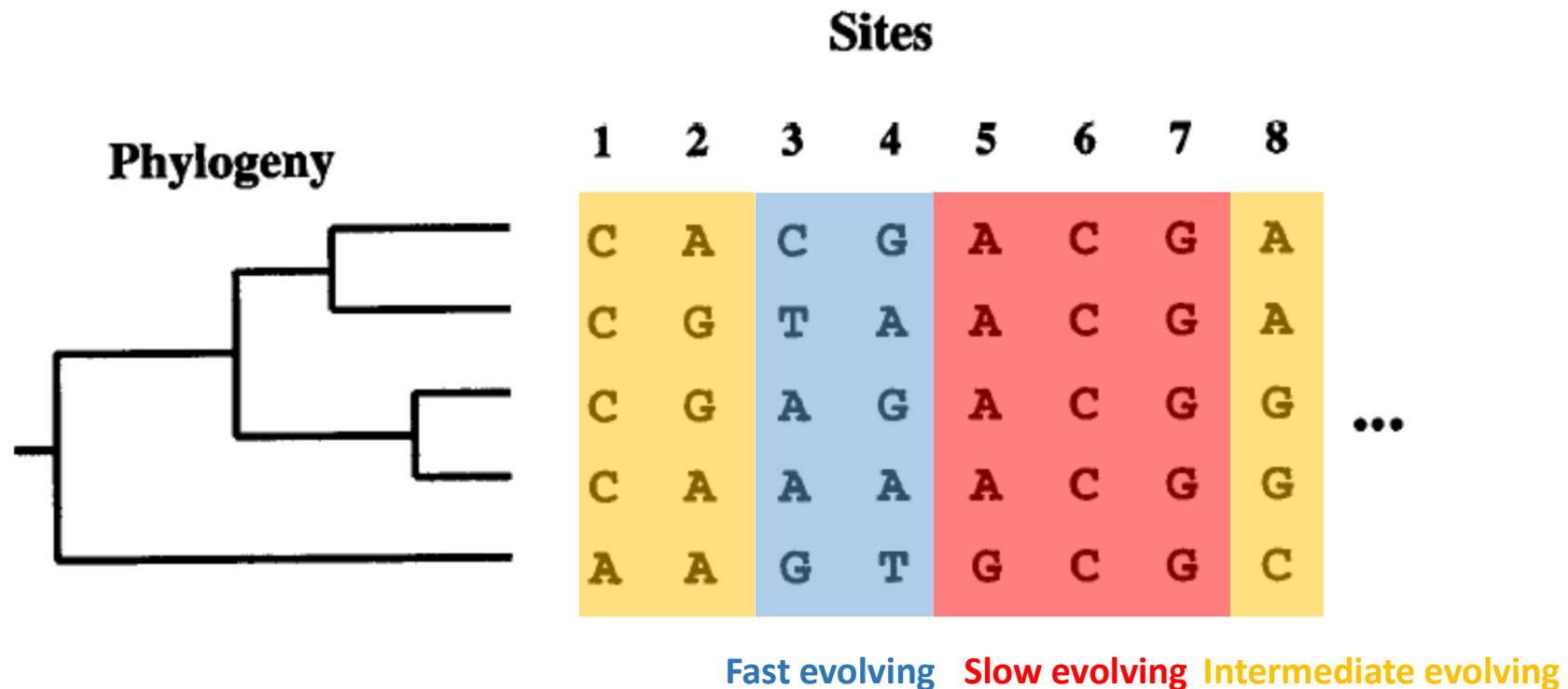
Prior
probability



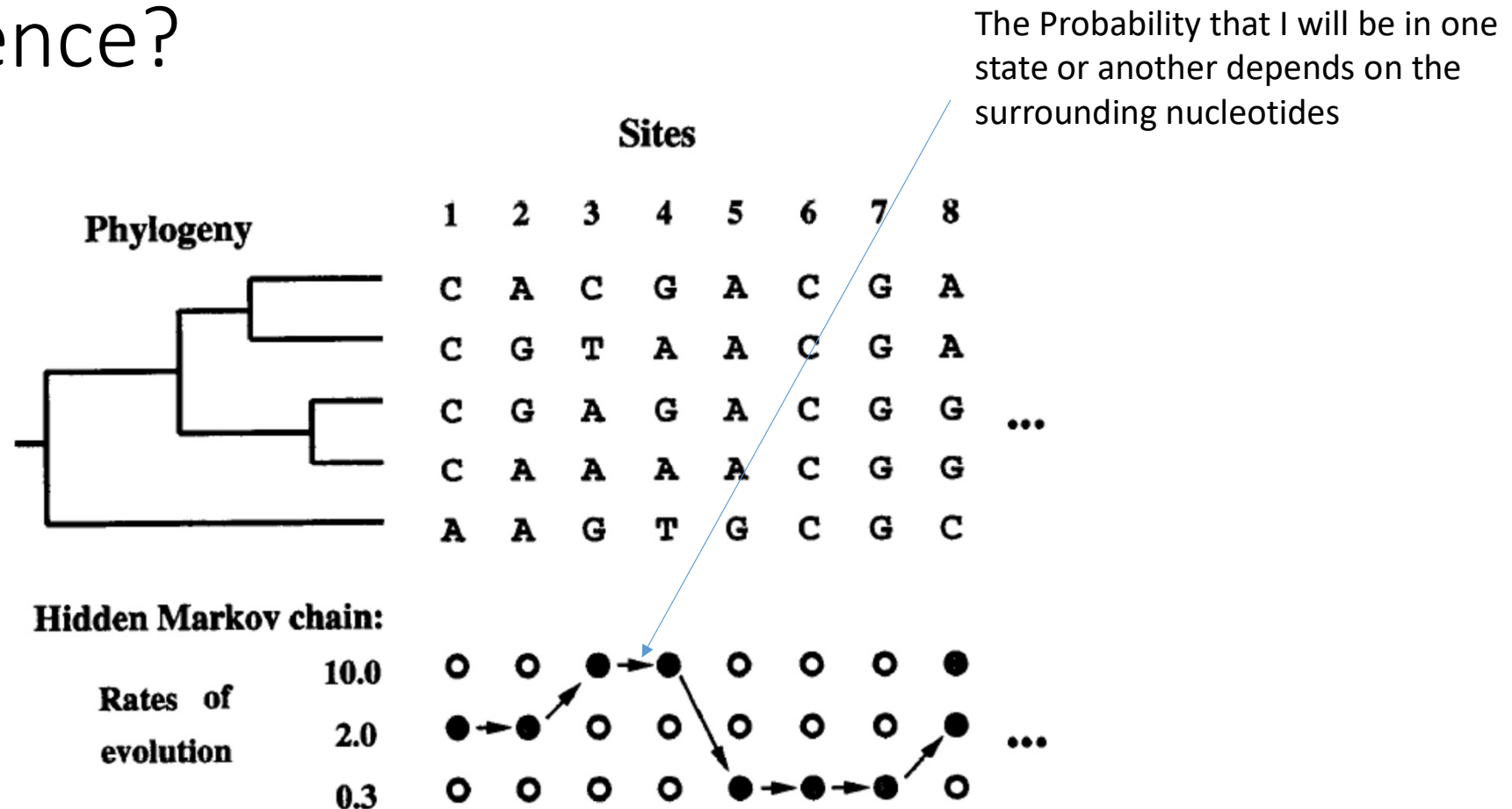
How does the rates of evolution change over a sequence?



How does the rates of evolution change over a sequence?

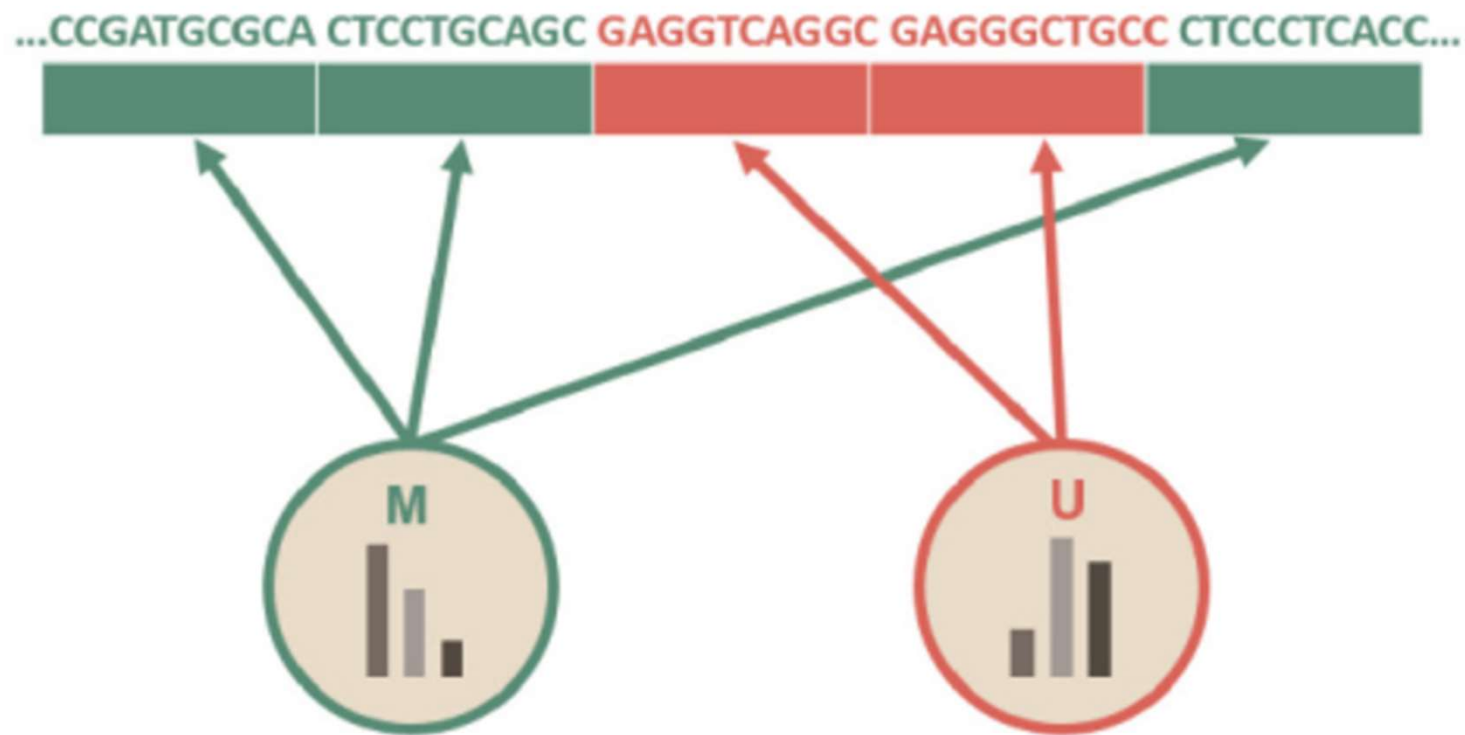


How does the rates of evolution change over a sequence?



Other examples of finding a pattern in the sequence: Metilation

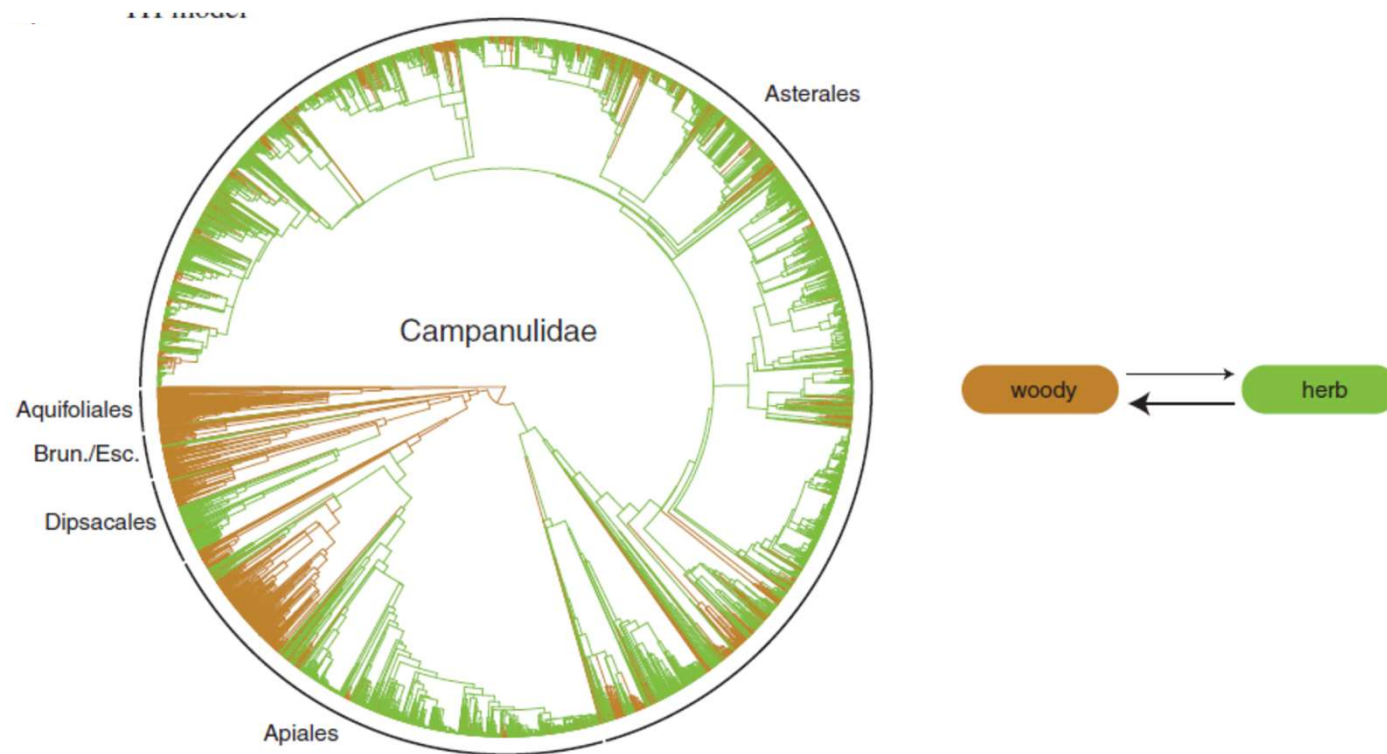
DNA Metilation



How does a phenotype change over a phylogeny?

Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character:
The Evolution of Plant Habit in Campanulid Angiosperms

JEREMY M. BEAULIEU^{1,*}, BRIAN C. O'MEARA², AND MICHAEL J. DONOGHUE¹



How to identify a pattern of characters?

2) To which rate of evolution each nucleotide belongs to?

Hidden because the state (Fast, Intermediate, Slow) is unknown

ATGATTTCAAAAAGGACCATCATTAGGA

Observed categories

FFFIIIIISSS...

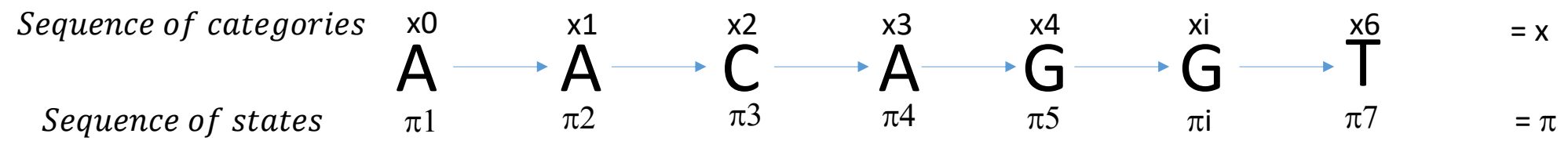
Hidden states

Fast (F), Intermediate (I) or Slow (S) is hidden (in fact, this is what we want to estimate!)

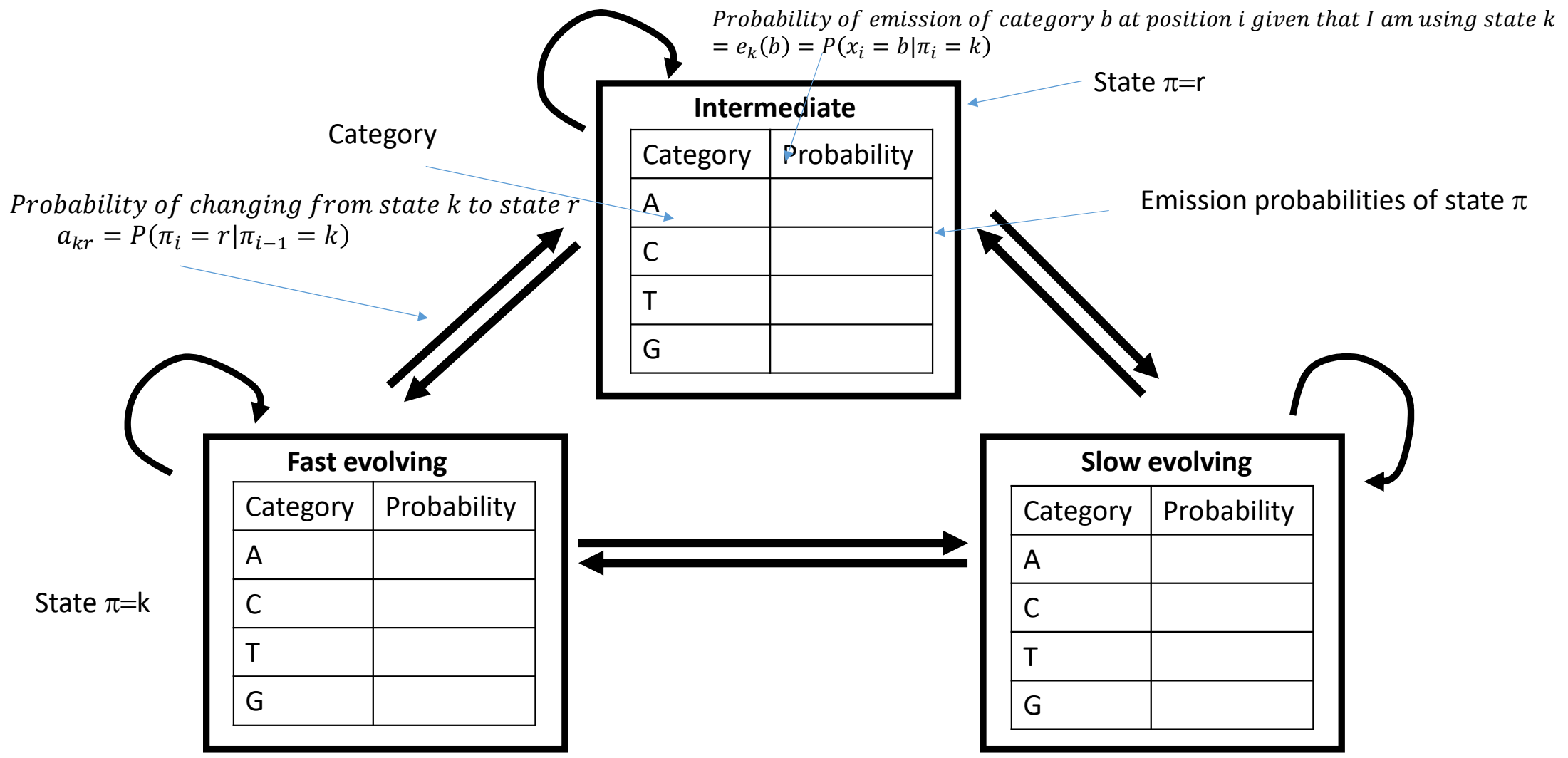
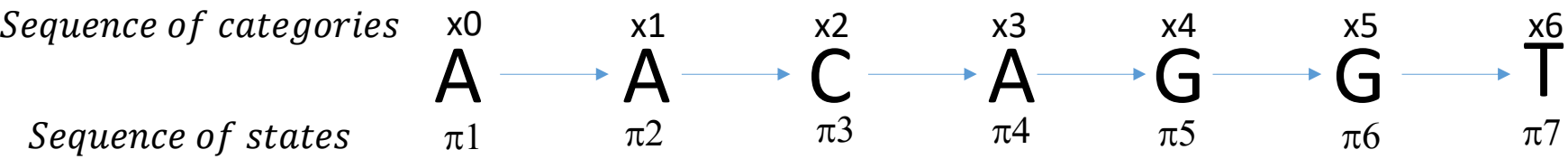
What am I saying when I use a HMM?

“My data is a mixture of k probability distributions, each determined by its own parameters”

“The probability of using one probability distribution or another depends on which one I used in the previous step”



SOME NOTATION



NOTATION HMM

State at position $i = \pi_i$

Sequence of states $= \pi$

Category at position $i = x_i$

Sequence of categories $= x$

$$P(x, \pi) = a_{0\pi} \prod_i^N e_{\pi_i}(x_i) a_{\pi_i \pi_{i+1}}$$

Initial state (“Probability that I started
the chain at state π ”)

Probability of changing from state k to state $r = a_{kr} = P(\pi_i = r | \pi_{i-1} = k)$

I am using state r

Given that I was in the previous position using state k

Transmission matrix

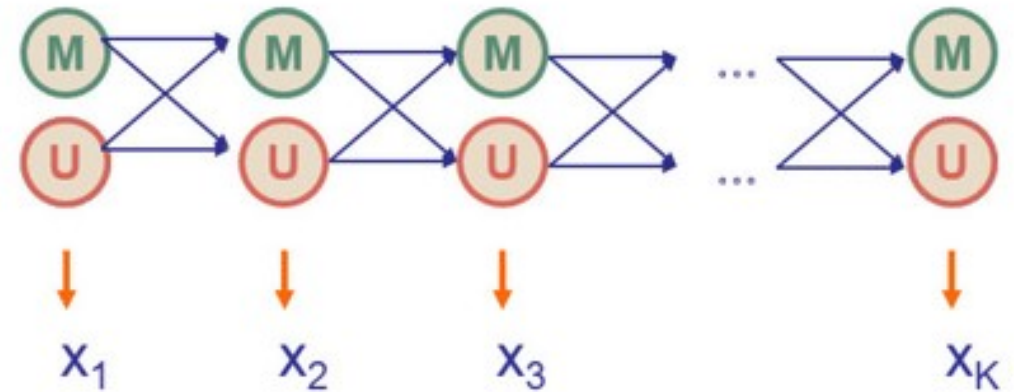
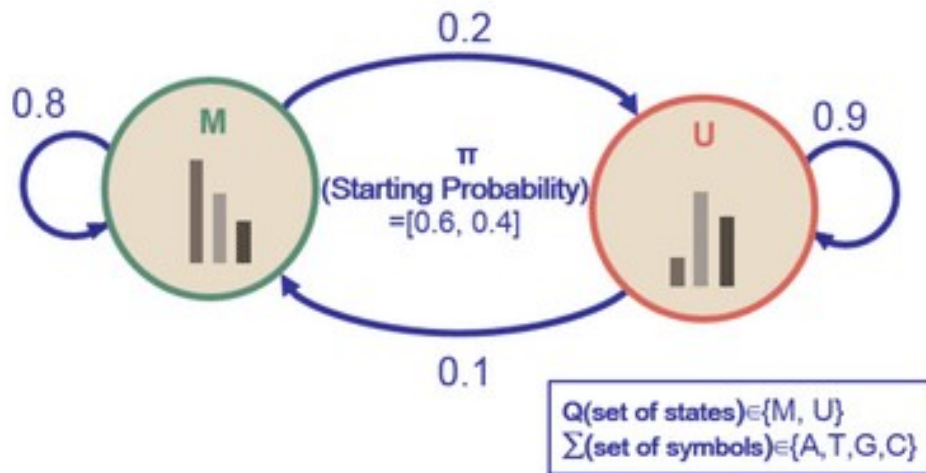
	Fast Evolving	Intermediate Evolving	Slow Evolving
Fast Evolving			
Intermediate Evolving			
Slow Evolving			

Emission Probability matrix for each possible state

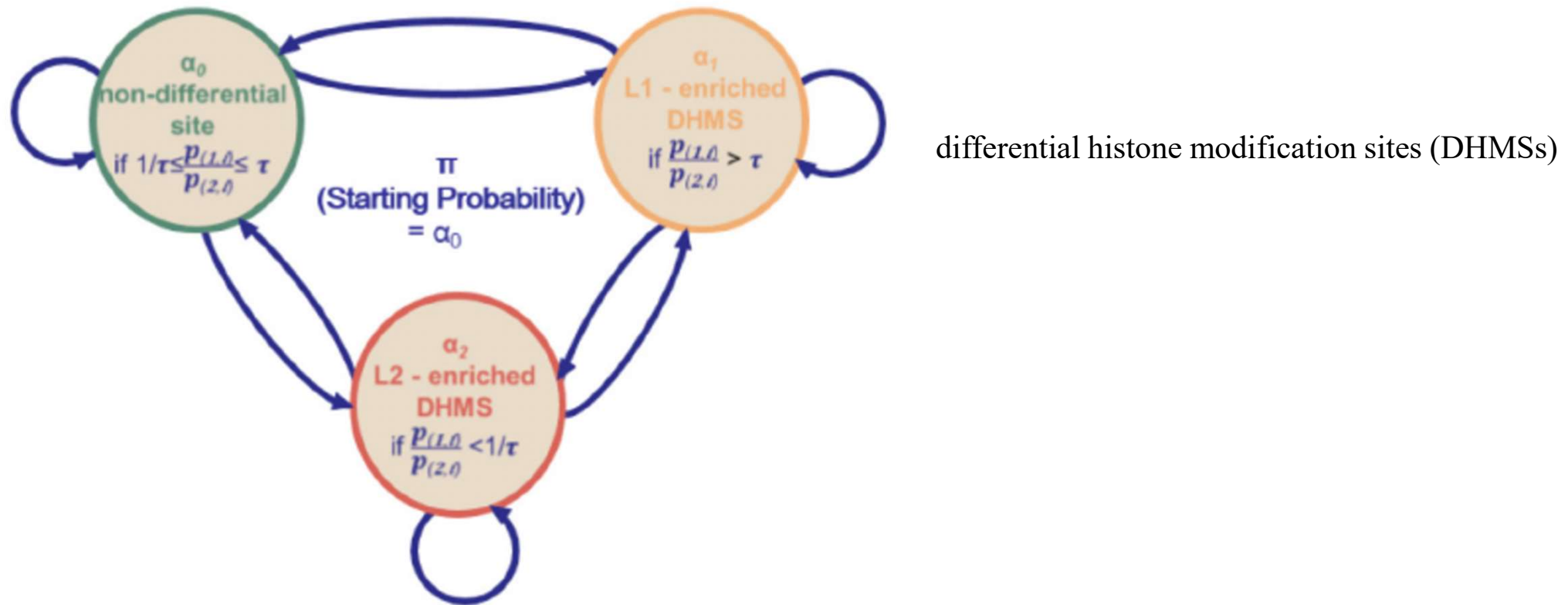
Fast evolving		Intermediate		Slow evolving	
Category	Probability	Category	Probability	Category	Probability
A		A		A	
C		C		C	
T		T		T	
G		G		G	

Probability of emission of category b at position i given that I am using state $k = e_k(b) = P(x_i = b | \pi_i = k)$

HMM: Multiple ways of modelling DNA Metilation

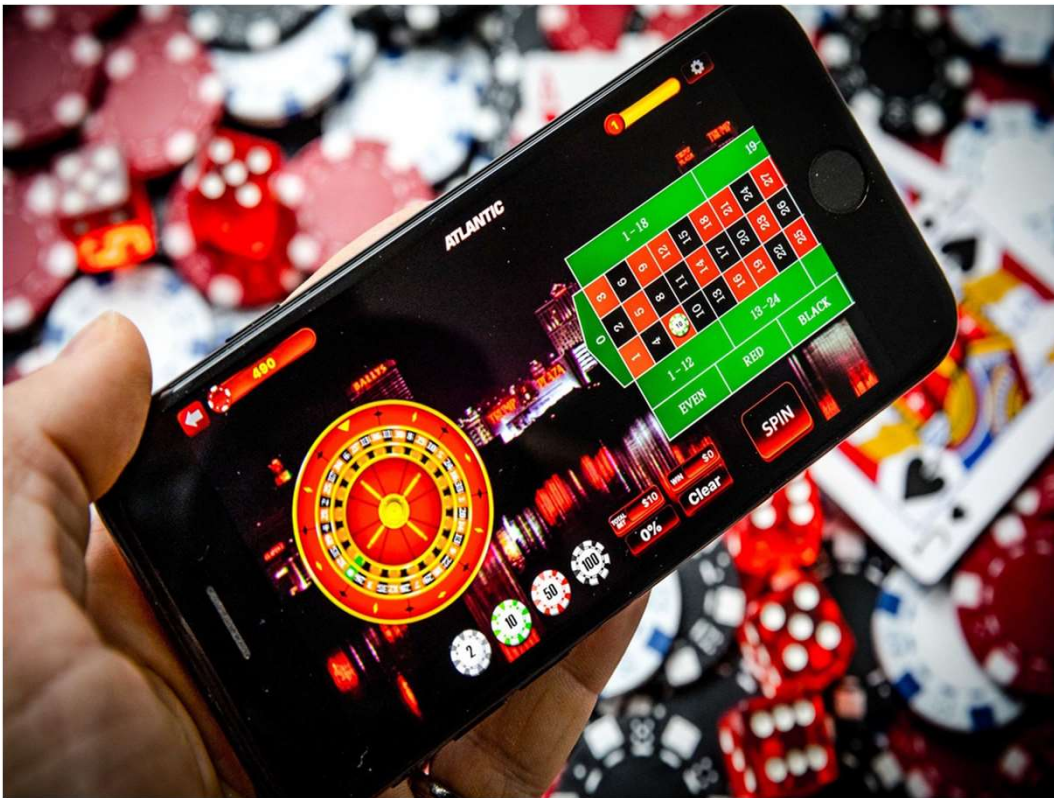


HMM: Multiple ways of modelling DNA Metilation



HMM: A simple working case

The occasionally dishonest casino problem

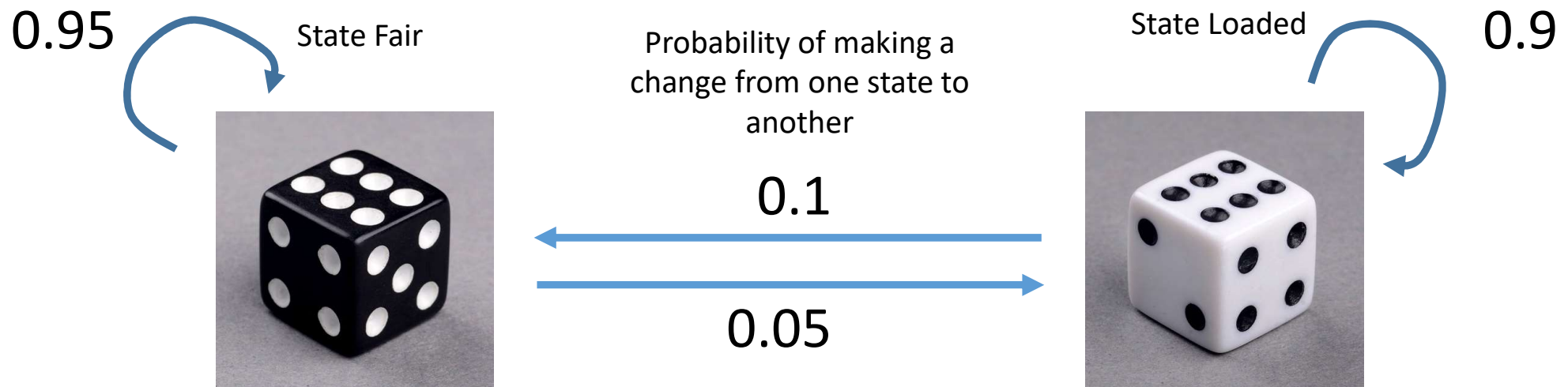


With some probability, the casino uses a dice that is **LOADED** so the number six (the bench wins) occurs more often ($P(6)=0.5$) than expected at random.

HOW CAN WE KNOW WHEN IT IS LOADED OF FAIR?

HMM

The occasionally dishonest casino problem.



Emission probabilities of fair dice

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

Fair (F) dice

Emission probabilities of loaded dice

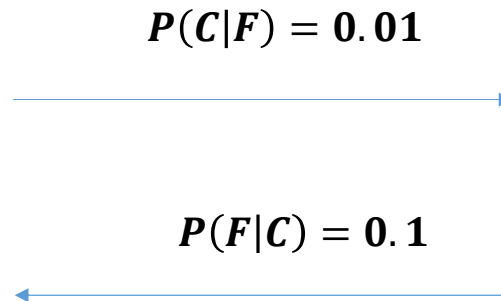
1	2	3	4	5	6
1/10	1/10	1/10	1/10	1/10	1/2

Loaded (L) dice

HMM:

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

x	$P(x F)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$



x	$P(x C)$
1	$\frac{1}{10}$
2	$\frac{1}{10}$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{1}{2}$

HMM: Probability of x given model?

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

x	$P(x F)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

$$P(C|F) = 0.01$$

$$P(F|C) = 0.1$$

x	$P(x C)$
1	$\frac{1}{10}$
2	$\frac{1}{10}$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{1}{2}$

It refers to the values of emission and transmission that are used for the k considered states

HMM

Probability of a sequence x given that the states of the sequence are known and we have the HMM model

Imagine we observe the sequence x

$$x = \langle x_1, x_2, x_3 \rangle = \langle 6, 1, 6 \rangle$$

Assume that the prior probability of starting at one state or at the other is the same (0.5)

What would be the probability $P(x, \pi)$ if the state sequence was

$$\pi = \langle \pi_1, \pi_2, \pi_3 \rangle = \langle F, F, F \rangle$$

$$P(x, \pi) = a_{0\pi} \prod_i^N e_{\pi_i}(x_i) a_{\pi_i \pi_{i+1}}$$

$$P(x, \pi) = 0.5 * 0.95 * \frac{1}{6} * 0.95 * \frac{1}{6} * 0.95 * \frac{1}{6} = 0.001984664$$

Prior of starting at F

Prob of staying at state F

Emission prob of 6 at state F

HMM: Probability of x given model?

$$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$$

We know that in a MC it is just:

$$P(S | \text{Markov Chain}) = P(1)P(3 | 1)P(4 | 3) \dots$$

$$P(S | \text{Hidden Markov Chain}): P(1)P(3 | 1)P(4 | 3) \dots$$

In our previous
HMM notation

In a HMC
this chain
explodes!

$P(C)$

$P(F)$

$P(1 | C)$

$P(1 | F)$

{

{

$P(3 | F)P(F | C)$

$P(3 | C)P(C | C)$

$P(3 | F)P(F | F)$

$P(3 | C)P(C | F)$

$e_{\pi_i}(x_i)$ $a_{\pi_i \pi_{i+1}}$

HMM: Probability of x given model? Forward

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

	1	3	4	...
F	$0.5 * 1/6$			
L	$0.5 * 1/10$			

x	$P(x F)$		x	$P(x C)$
1	$\frac{1}{6}$	$\xrightarrow{P(C F) = 0.01}$	1	$\frac{1}{10}$
2	$\frac{1}{6}$		2	$\frac{1}{10}$
3	$\frac{1}{6}$		3	$\frac{1}{10}$
4	$\frac{1}{6}$		4	$\frac{1}{10}$
5	$\frac{1}{6}$	$\xleftarrow{P(F C) = 0.1}$	5	$\frac{1}{10}$
6	$\frac{1}{6}$		6	$\frac{1}{2}$

HMM: Probability of x given model? Forward

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

	1	3	4	...
F	$0.5 * 1/6$	$0.5 * 1/6 * 0.99 * 1/6$		
L	$0.5 * 1/10$			

Diagram illustrating the forward pass calculation for the HMM. The table shows the probability of the sequence x given the model, calculated using the forward pass. The first row shows the initial state probabilities for 'F' and 'L'. The second row shows the calculation for the first state (1) given the model, which is $0.5 * 1/6$. The third row shows the calculation for the second state (3) given the model, which is $0.5 * 1/6 * 0.99 * 1/6$. The diagram also shows the transition probabilities from state 1 to state 3, labeled 'F,F'.

x	$P(x F)$		x	$P(x C)$
1	$\frac{1}{6}$		1	$\frac{1}{10}$
2	$\frac{1}{6}$		2	$\frac{1}{10}$
3	$\frac{1}{6}$	$P(C F) = 0.01$	3	$\frac{1}{10}$
4	$\frac{1}{6}$		4	$\frac{1}{10}$
5	$\frac{1}{6}$	$P(F C) = 0.1$	5	$\frac{1}{10}$
6	$\frac{1}{6}$		6	$\frac{1}{2}$

Diagram illustrating the forward pass calculation for the HMM. The table shows the probability of the sequence x given the model, calculated using the forward pass. The first row shows the initial state probabilities for 'F' and 'L'. The second row shows the calculation for the first state (1) given the model, which is $0.5 * 1/6$. The third row shows the calculation for the second state (3) given the model, which is $0.5 * 1/6 * 0.99 * 1/6$. The diagram also shows the transition probabilities from state 1 to state 3, labeled 'F,F'.

HMM: Probability of x given model? Forward

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

	1	3	4	...
F	$0.5 * 1/6$	$0.5 * 1/6 * 0.99 * 1/6 +$ $0.5 * 1/10 * 0.1 * 1/6$		
L	$0.5 * 1/10$			

F,F

C,F

x	$P(x F)$		x	$P(x C)$
1	$\frac{1}{6}$	$P(C F) = 0.01$	1	$\frac{1}{10}$
2	$\frac{1}{6}$		2	$\frac{1}{10}$
3	$\frac{1}{6}$		3	$\frac{1}{10}$
4	$\frac{1}{6}$		4	$\frac{1}{10}$
5	$\frac{1}{6}$	$P(F C) = 0.1$	5	$\frac{1}{10}$
6	$\frac{1}{6}$		6	$\frac{1}{2}$

HMM: Probability of x given model? Forward

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

F,F

C,F

	1	3	4	...
F	$0.5 * 1/6$	$0.5 * 1/6 * 0.99 * 1/6 +$ $0.5 * 1/10 * 0.1 * 1/6$		
L	$0.5 * 1/10$	$0.5 * 1/10 * 0.9 * 1/10$		

C,C

x	$P(x F)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

$P(C|F) = 0.01$

$P(F|C) = 0.1$

x	$P(x C)$
1	$\frac{1}{10}$
2	$\frac{1}{10}$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{1}{2}$

HMM: Probability of x given model? Forward

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

	1	3	4	...
F	$0.5 * 1/6$	$0.5 * 1/6 * 0.99 * 1/6 +$ $0.5 * 1/10 * 0.1 * 1/6$		
L	$0.5 * 1/10$	$0.5 * 1/10 * 0.9 * 1/10 +$ $0.5 * 1/6 * 0.01 * 1/6$		

F,F
C,F
C,C
F,C

x	$P(x F)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

$P(C|F) = 0.01$

$P(F|C) = 0.1$

x	$P(x C)$
1	$\frac{1}{10}$
2	$\frac{1}{10}$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{1}{2}$

We have searched all the combinations

HMM: Probability of x given model? Forward

$P(1,3,4,5,1,2,6,6,3,6,6,2,1,2,4,5 | \text{HMM})$

	1	3	4	...
F	$0.5 * 1/6$	$0.5 * 1/6 * 0.99 * 1/6 +$ $0.5 * 1/10 * 0.1 * 1/6$		
L	$0.5 * 1/10$	$0.5 * 1/10 * 0.9 * 1/10 +$ $0.5 * 1/6 * 0.01 * 1/6$		

P1

P2

$P1 + P2 = P(S | \text{HMM})$

Some computational considerations....

- Each step implies multiplying a small number by a number between 0 and 1 (the probability of doing that step).
- We end having VERY small numbers, which ultimately will produce underflow
 - Do a log transformation (you can do this whenever you multiply probabilities)
 - Scale the probability at each step (you do this when you multiply and add probabilities)

HMM

- Scale the probability at each step (Rabiner 1989)

State	P(i)	P(i+1)
State1		
State2		
StateK		



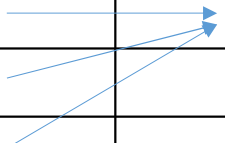
$$\hat{f}_l(i) = \frac{f_l(i)}{\prod_{j=1}^i s_j}$$

← Scaling factor


HMM

- Scale the probability at each step (Rabiner 1989)

State	$\hat{f}_l(i)$	$\hat{f}_l(i + 1)$
State1		
State2		
StateK		



New scaled variable $\hat{f}_l(i) = \frac{f_l(i)}{\prod_{j=1}^i s_j}$



HMM

- Scale the probability at each step (Rabiner 1989)

State	$\hat{f}_l(i)$	$\hat{f}_l(i + 1)$
State1		
State2		
StateK		

Probability of moving from state r to state l in $i+1$

$$\hat{f}_l(i + 1) = \frac{1}{s_{i+1}} e_l(x_{i+1}) \sum_{r=1}^K \hat{f}_r(i) a_{rl}$$

Scaled forward

Emission probability of category at x_{i+1} using state l

HMM

- Scale the probability at each step (Rabiner 1989)

State	$\hat{f}_l(i)$	$\hat{f}_l(i + 1)$
State1		
State2		
StateK		

If we impose that

Then s_{i+1} is

Total of the scaled value

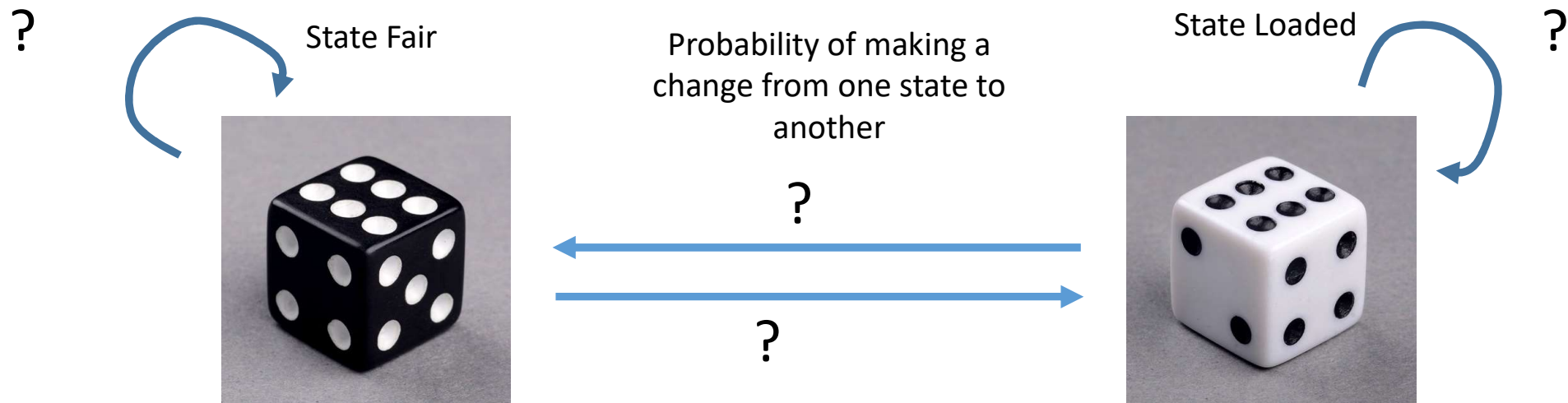
$$\sum_l \hat{f}_l(i) = 1$$

$$s_{i+1} = \sum_l e_l(x_{i+1}) \sum_k \hat{f}_k(i) a_{kl}$$

$$P(X|HMM) = \prod_i^L s_i$$

HMM

What happens when the parameters are not known?



Emission probabilities of fair dice

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

Fair (F) dice

Emission probabilities of loaded dice

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

Loaded (L) dice

HMM

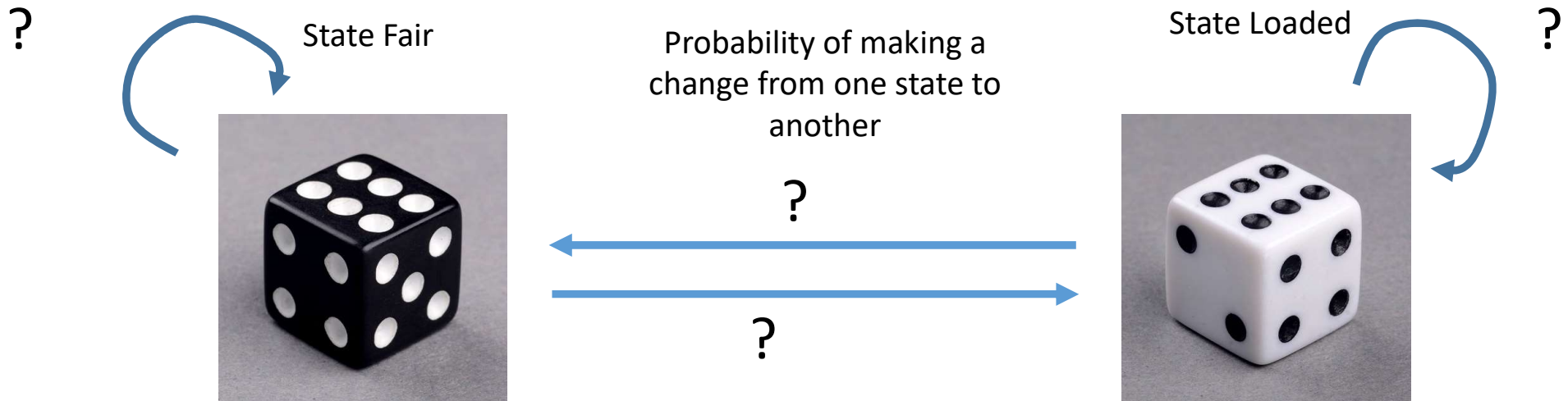
What happens when the parameters are not known?

1 We have a training dataset where we know the hidden state

x	1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
π	F,F,F,F,F,F,L,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,F,L,L

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,L,L



Emission probabilities of fair dice

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

Fair (F) dice

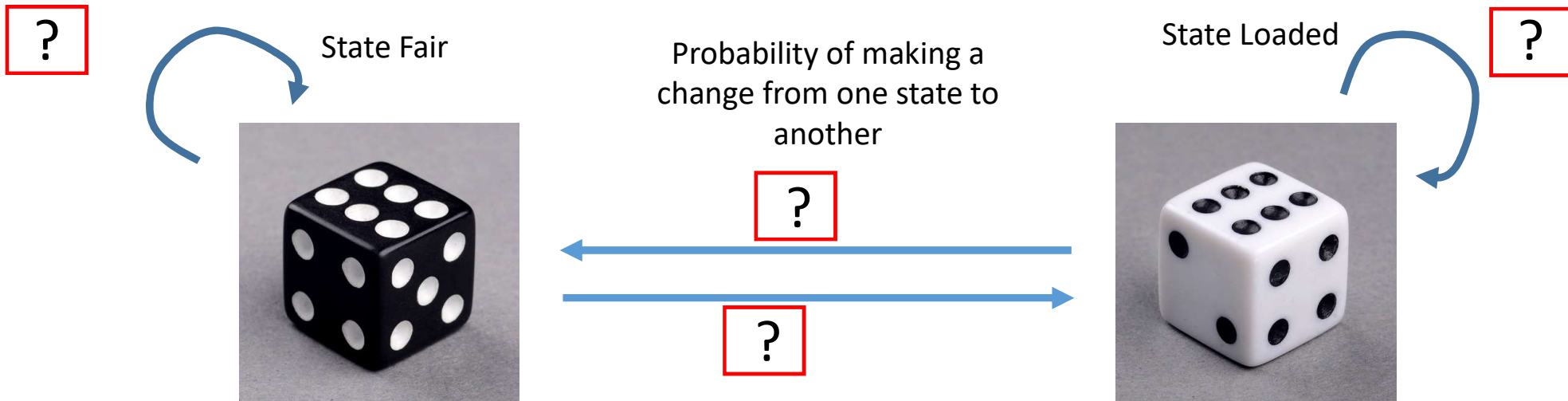
Emission probabilities of loaded dice

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

Loaded (L) dice

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L



Emission probabilities of fair dice

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

Fair (F) dice

Emission probabilities of charged dice

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

Loaded (L) dice

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L

How would you do it?

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L

Come from->Goes to	F	L
F		
L		

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

π F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,F,F,L,L

Come from->Goes to	F	L
F	1	
L		

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,L,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L

Come from->Goes to	F	L
F	2	
L		

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L

Come from->Goes to	F	L
F	α	β
L	γ	δ

a_{FF}

a_{FC}

a_{CF}

a_{CC}

?

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,L,L,L,L,L,L,L,L,F,F,F,F,F,F,F,F,F,F,F,F,L,L

Marginal of the row

Come from->Goes to	F	C	Total
F	α	β	$\alpha+\beta$
C	γ	δ	$\gamma+\delta$

Counts in cell row k, column l

Maximum likelihood estimates

$$a_{FF} = \frac{\alpha}{\alpha + \beta}$$

$$a_{FC} = \frac{\beta}{\alpha + \beta}$$

$$a_{kl} = \frac{A_{kl}}{\sum_{l'} A_{kl'}}$$

Row k
 "come from"

Column l
 "go to"

Counts in row k (Marginal)

HMM

Why the maximum likelihood can be computed like this?

HMM

For a given “I come from” state, we have two possible movements

“I stay at the state” with probability p

“I move to the other state” with probability $1-p$

We observe *“I stay at the state”* y observations out of n observations

Which is the statistical distribution that could model this process?

$$P(y|p; n) = \frac{n!}{y! (n - y)!} (p)^y (1 - p)^{n-y}$$

HMM

$$P(y|p; n) = \frac{n!}{y! (n - y)!} (p)^y (1 - p)^{n-y}$$

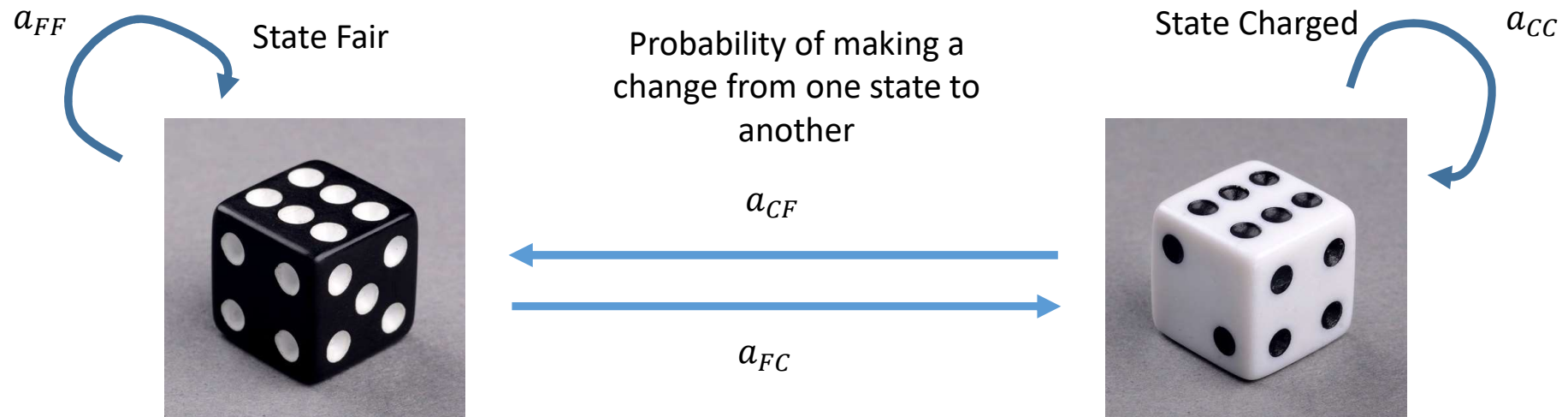
$$\log(L(p)) = \log(C) + y \log(p) + (n - y) \log(1 - p)$$

$$\frac{dL}{dp} = \frac{y}{p} - \frac{n-y}{1-p}$$

$$\frac{dL}{dp} = 0; p = \frac{y}{n}$$

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,C,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,F,F



Emission probabilities of fair dice

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

Fair (F) dice

Emission probabilities of charged dice

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

Loaded (L) dice

HMM

x	1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
π	F,F,F,F,F,F,C,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,F,F,F

Emission probabilities of fair dice

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

Fair (F) dice

Emission probabilities of loaded dice

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

Loaded (L) dice

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
 π F,F,F,F,F,F,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,F

How would you do it?

HMM

x	1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4
π	F,F,F,F,F,F,F,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,F,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,F,F

x	F
1	n_{1F}
2	n_{2F}
3	n_{3F}
4	n_{4F}
5	n_{5F}
6	n_{6F}

x	$P(x F)$
1	n_{1F}/N_F
2	n_{2F}/N_F
3	n_{3F}/N_F
4	n_{4F}/N_F
5	n_{5F}/N_F
6	n_{6F}/N_F

$$e_k(b) = \frac{E_k(b)}{\sum_{b'} E_k(b')}$$

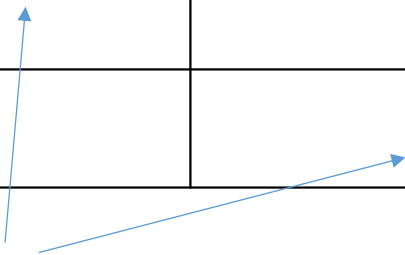
Can you identify the different variables?

HMM

Once the parameters have been estimated in training, they must be validated in a different dataset not used for training (replication)

Confusion matrix. In classification tasks, it allows to estimate how good is the model

Viterbi	F is predicted given HMM	C is predicted given HMM
F is true in replication		
C is true in replication		



I would like this to be as close as possible to 100%

Check also $P(S|HMM)$

HMM

The occasionally dishonest casino problem

Unknown paths

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

How to find HMM that maximizes $P(X | \text{HMM})$?

HMM

The occasionally dishonest casino problem

Unknown paths

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

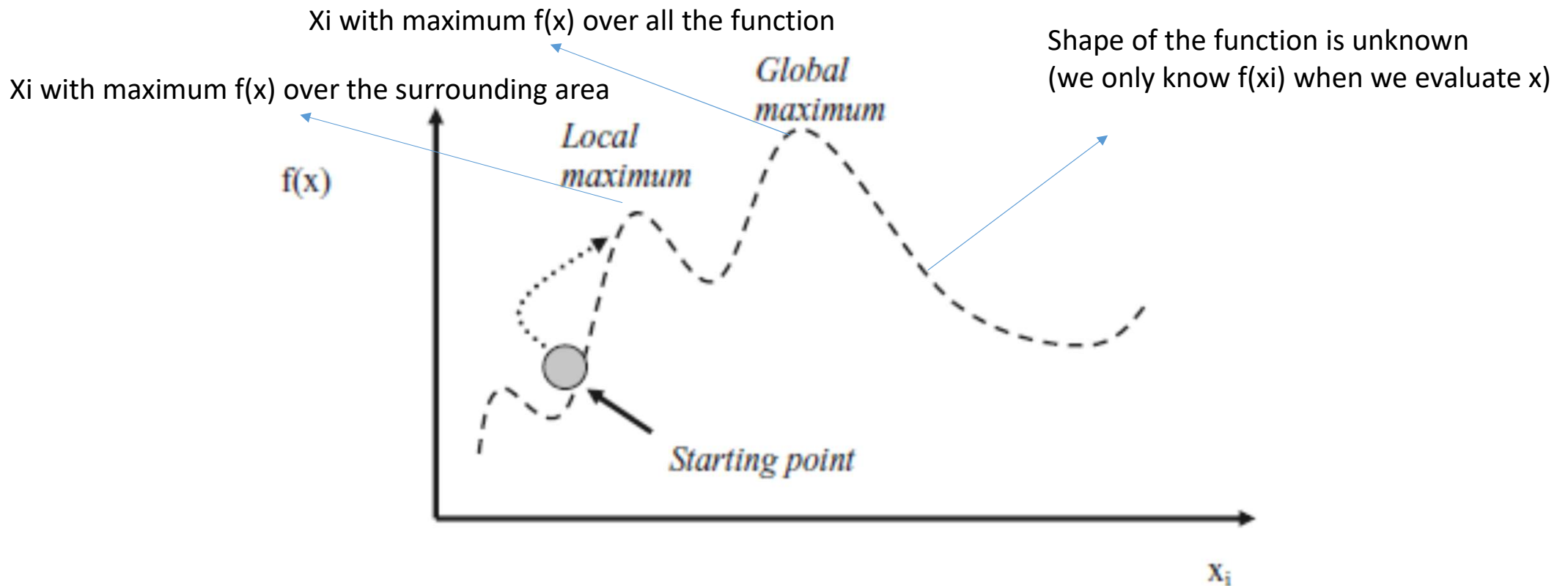
$P(x|\text{HMM}=a_{rk}; e_k(b))$

Likelihood of an observed sequence given a set of transition and emission probabilities from a given HMM, computed using the forward algorithm

How to find HMM that maximizes $P(X|\text{HMM})$?

HMM

How to find HMM that maximizes $P(X|HMM)$?



HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

```
Program Genetic Algorithm ()
{
  initialize population;
  evaluate population;
  while (termination conditions not reached)
  {
    select solutions for next population;
    perform crossover and mutation
    evaluate population;
  }
}
```

Classical Genetic algorithm

$$\text{chromosome} = [p_1, p_2, p_3, \dots, p_{N_{var}}]$$

Algorithm 3.1: Canonical Genetic Algorithm

Determine how the solution is to be encoded as a genotype and define the fitness function;
 Create an initial population of genotypes;
 Decode each genotype into a solution and calculate the fitness of each of the n solution candidates in the population;

repeat

Select n members from the current population of encodings (the *parents*) in order to create a mating pool;

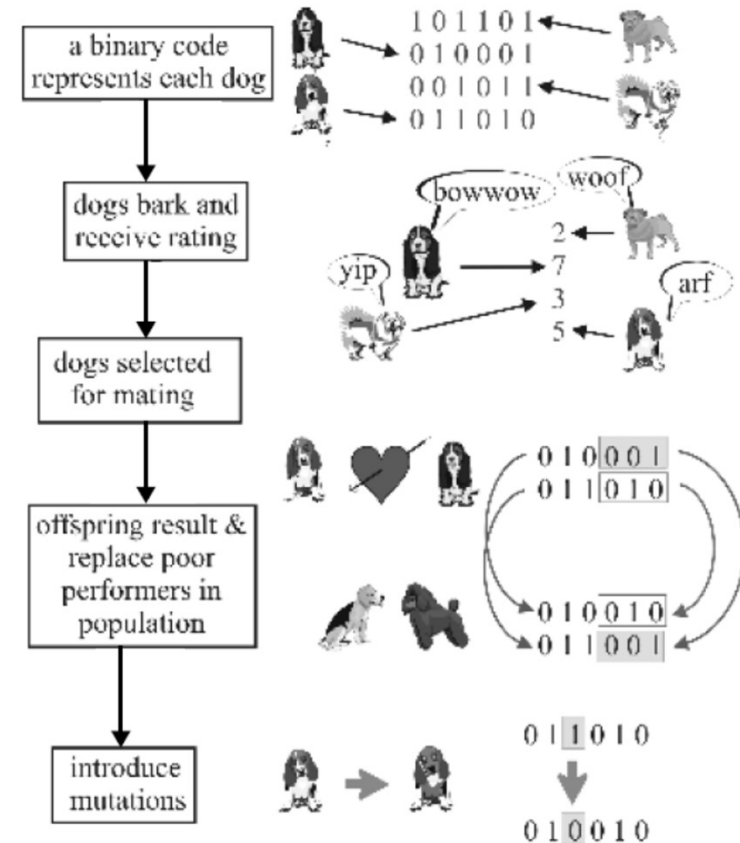
repeat

Select two parents randomly from the mating pool;
 With probability p_{cross} , perform a crossover process on the encodings of the selected parent solutions, to produce two new (*child*) solutions;
 Otherwise, crossover is not performed and the two children are simply copies of their parents;
 With probability p_{mut} , apply a mutation process to each element of the encodings of the two child solutions;

until n new child solutions have been created;

Replace the old population with the newly created one (this constitutes a generation);

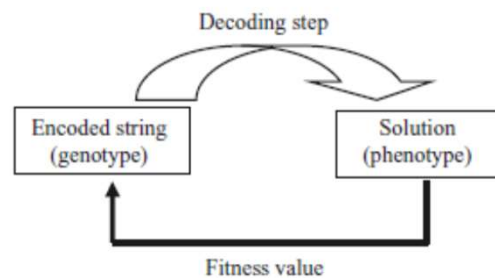
until terminating condition;



Classical Genetic algorithm

- What do we need?
 - A chromosome based representation of the solution that we are searching for:

$$chromosome = [p_1, p_2, p_3, \dots, p_{N_{var}}]$$



-3.12	23.11	3.93
-------	-------	------

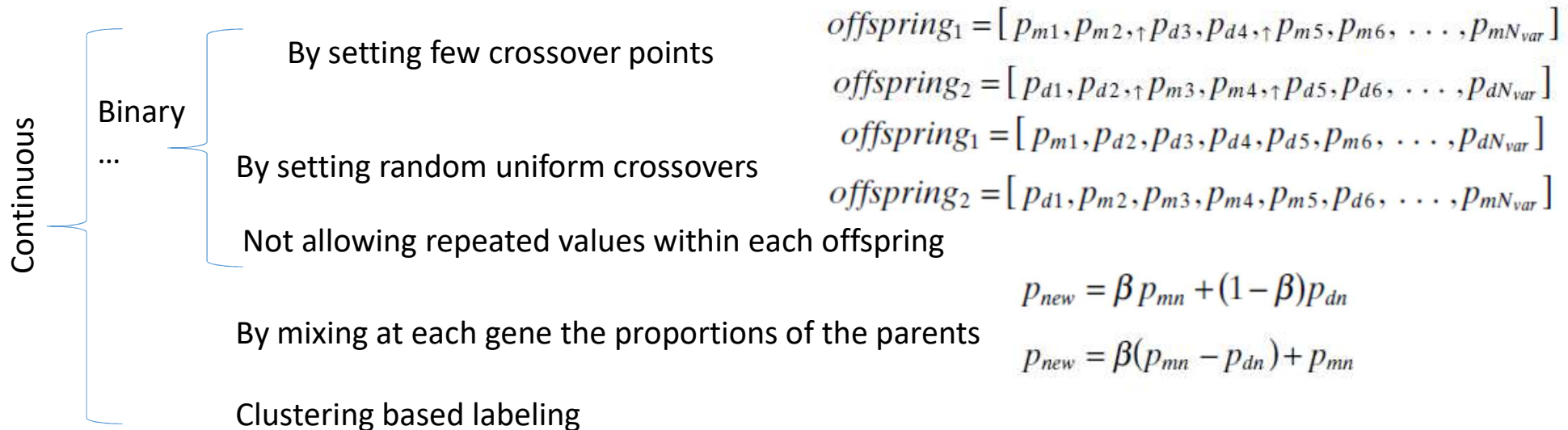
 \longrightarrow $y = -3.12 + 23.11x_1 + 3.93x_2$

Classical Genetic algorithm

- What do we need?
 - A way to generate new solutions by recombination from previous ones (Mating & Recombination):

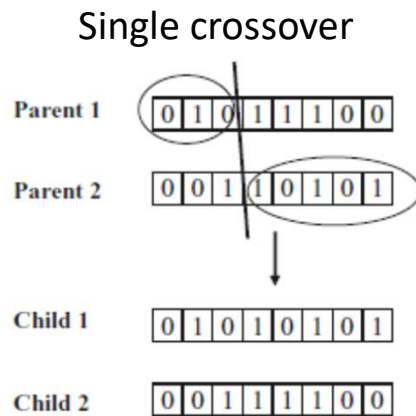
$$parent_1 = [p_{m1}, p_{m2}, p_{m3}, p_{m4}, p_{m5}, p_{m6}, \dots, p_{mN_{var}}]$$

$$parent_2 = [p_{d1}, p_{d2}, p_{d3}, p_{d4}, p_{d5}, p_{d6}, \dots, p_{dN_{var}}]$$

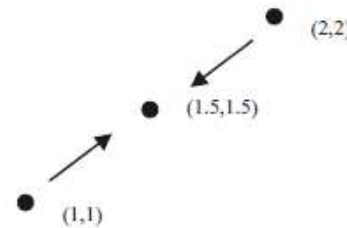


Classical Genetic algorithm

- What do we need?
 - A way to generate new solutions by recombination from previous ones (Mating & Recombination):



Real valued genotypes



Classical Genetic algorithm

- How to introduce new solutions (mutation)?

$$chromosome = [p_1, p_2, p_3, \dots, p_{N_{var}}]$$

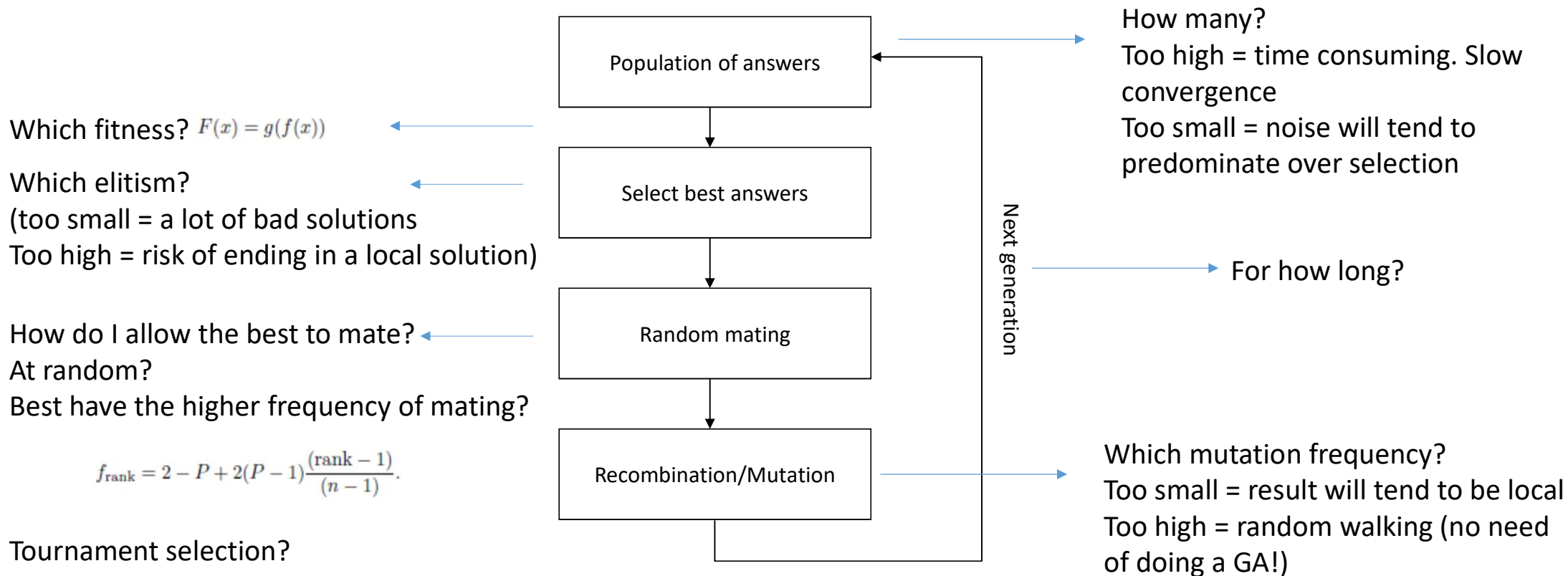


$f(px)$?

$$chromosome = [p_1, \mathbf{p_2}, p_3, \dots, p_{N_{var}}]$$

Classical Genetic algorithm

Other hyperparameters to consider

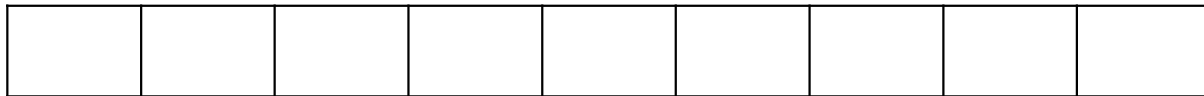


HMM

The occasionally dishonest casino problem

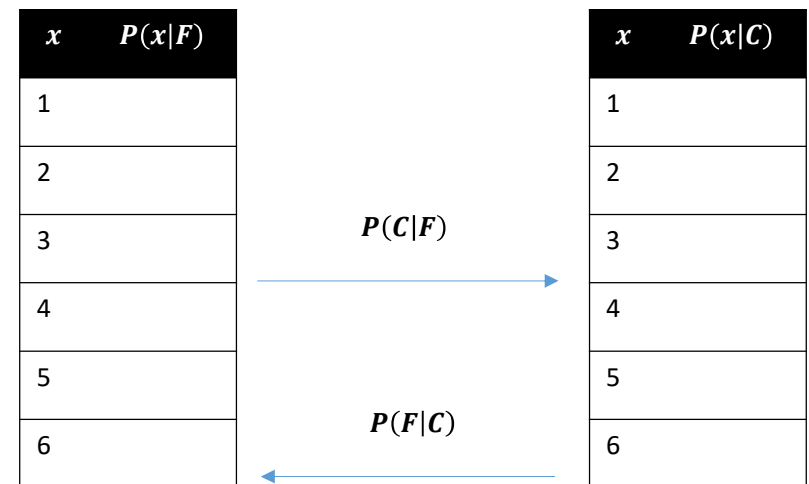
Optimization of a and e using The Genetic Algorithm

Chromosome structure



“Put in a vector the parameters you want to estimate”

Which are the parameters?

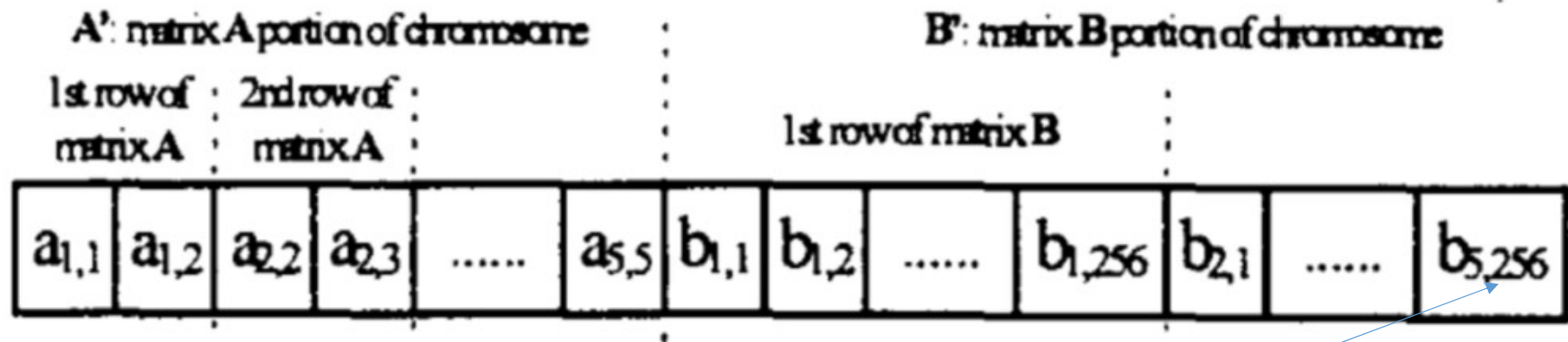


HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Coding of the matrices in a vector (=Chromosome in GA). Example



This is because in this case they are using characters for speech recognition problems

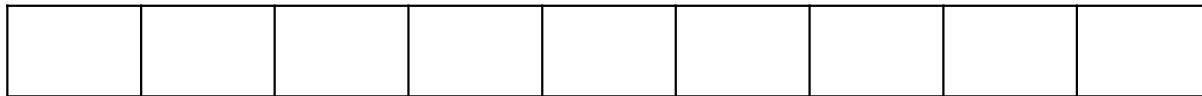
Optimization of HMM by a GA

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Chromosome structure



"Put in a vector the parameters you want to estimate"

"taking into account possible conditions"

All these variables must
add to one

Which are the parameters?

x	$P(x F)$
1	
2	
3	
4	
5	
6	

x	$P(x C)$
1	
2	
3	
4	
5	
6	

$P(C|F)$

$P(F|C)$

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Chromosome structure

--	--	--	--	--	--	--	--	--

"Put in a vector the parameters you want to estimate"

"taking into account possible conditions"

All these variables must
add to one

Which are the parameters?

x	$P(x F)$
1	
2	
3	
4	
5	
6	

x	$P(x C)$
1	
2	
3	
4	
5	
6	

$P(C|F)$

$P(F|C)$

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Chromosome structure

--	--	--	--	--	--	--	--	--

Which are the parameters?

"Put in a vector the parameters you want to estimate"

"taking into account possible conditions"

All these variables must
add to one

States $P(S_j S_i)$	S_1	S_2	S_n
S_1			
S_2			
S_n			

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Chromosome structure



"Put in a vector the parameters you want to estimate"

"taking into account possible conditions"

How many independent
variables do we have
here?

Which are the parameters?

x	$P(x F)$
1	
2	
3	
4	
5	
6	

x	$P(x C)$
1	
2	
3	
4	
5	
6	

$P(C|F)$

$P(F|C)$

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Chromosome structure



“Put in a vector the parameters you want to estimate”

“taking into account possible conditions”

How many independent
variables do we have
here?

5 (because all the
variables must add to 1)

In statistical jargon we would say $n-1$ degrees of freedom

Which are the parameters?

x	$P(x F)$
1	
2	
3	
4	
5	
6	

x	$P(x C)$
1	
2	
3	
4	
5	
6	

$P(C|F)$

$P(F|C)$

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Chromosome structure



“Put in a vector the parameters you want to estimate”

“taking into account possible conditions”

Some subsets of parameters must add to 1 (we can exclude one)

All parameters cannot be greater than 1

Conditions of this particular problem

Find a way how to transform the variables so the conditions are fulfilled

HMM

Discrete probability function

$$P(i) = p_i$$

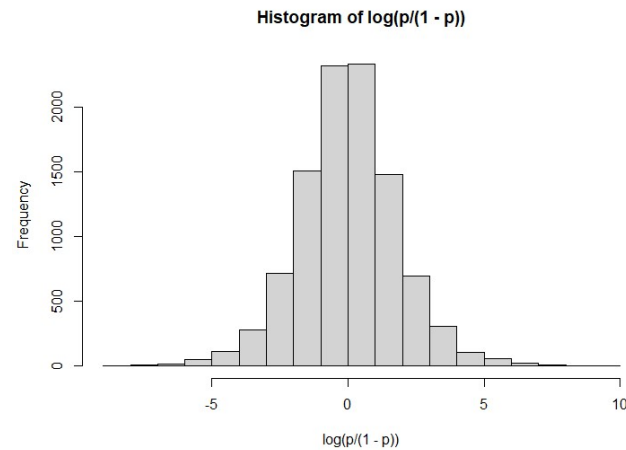
$$\sum_i^K p_i = 1$$

Logit transformation for a binary variable

$$\theta_i = \log \left(\frac{p_i}{1 - p_i} \right)$$

Which is the range?

Which is the range?



$$p \sim U(0,1)$$

Looks like a Normal distribution, but it is not

HMM

Discrete probability function

$$P(i) = p_i$$

$$\sum_i^K p_i = 1$$

Logit transformation for a binary variable

$$\theta_i = \log \left(\frac{p_i}{1 - p_i} \right)$$

How to estimate p_i ?

HMM

Discrete probability function

$$P(i) = p_i$$

$$\sum_i^K p_i = 1$$

*“Logit transformation of
multinomial distribution”*

$$p_i = \frac{e^{\theta_i}}{1 + \sum_j^{K-1} e^{\theta_j}}$$

$$p_K = 1 - \sum_j^{K-1} p_j$$

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

e_1	e_1	e_1	e_1	e_1	...			
θ_1	θ_2	θ_3	θ_4	θ_5				

Chromosome structure

We need to keep track to which table each parameter refers to (BUT IT IS NOT CODED IN THE chromosome)

Which are the parameters?

$$p_i = \frac{e^{\theta_i}}{1 + \sum_j^{K-1} e^{\theta_j}}$$

Decode the information of the chromosome

x	$P(x F)$
1	
2	
3	
4	
5	
6	

x	$P(x C)$
1	
2	
3	
4	
5	
6	

$P(C|F)$

$P(F|C)$

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Breakpoints

$\theta_{1,1}$	$\theta_{2,1}$	$\theta_{3,1}$	$\theta_{4,1}$	$\theta_{5,1}$	$\theta_{1,1}$	$\theta_{2,1}$	$\theta_{3,1}$...
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----

$\theta_{1,2}$	$\theta_{2,2}$	$\theta_{3,2}$	$\theta_{4,2}$	$\theta_{5,2}$	$\theta_{1,2}$	$\theta_{2,2}$	$\theta_{3,2}$...
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----

$\theta_{1,1}$	$\theta_{2,1}$	$\theta_{3,1}$	$\theta_{4,2}$	$\theta_{5,2}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{3,1}$...
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----

$\theta_{1,2}$	$\theta_{2,2}$	$\theta_{3,2}$	$\theta_{4,1}$	$\theta_{5,1}$	$\theta_{1,1}$	$\theta_{2,2}$	$\theta_{3,2}$...
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----

Parents

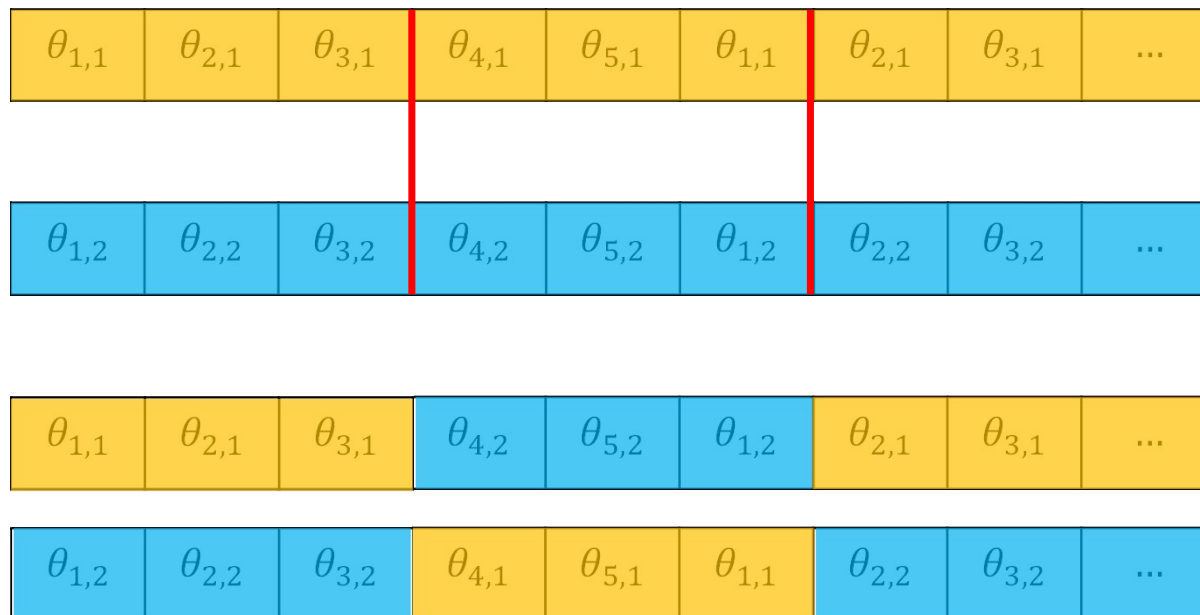
Offspring

HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Breakpoints



HMM

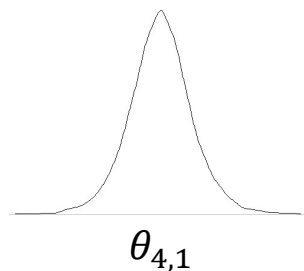
The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

$\theta_{1,2}$	$\theta_{2,2}$	$\theta_{3,2}$	$\theta_{4,1}$	$\theta_{5,1}$	$\theta_{1,1}$	$\theta_{2,2}$	$\theta_{3,2}$...
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----

Pick one of the variables at random

Mutation

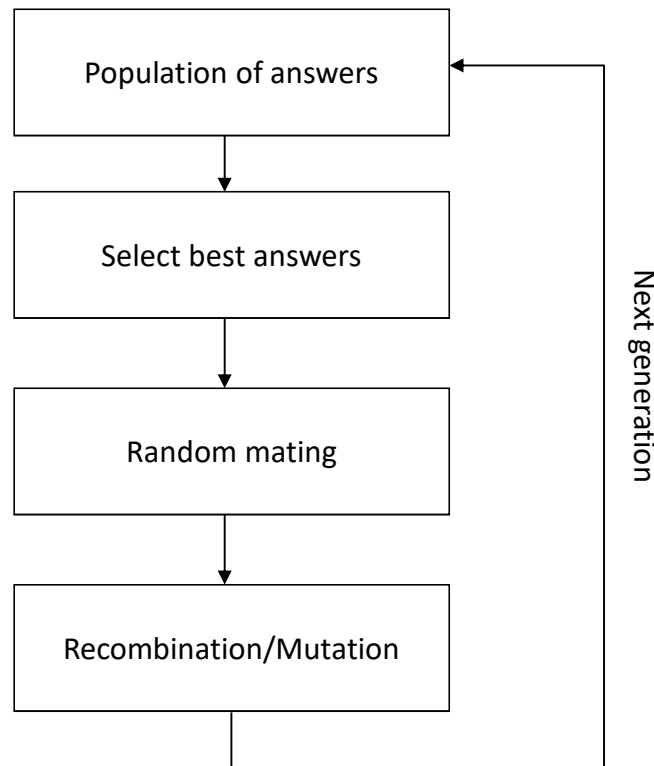


Propose a new $\theta_{4,1}^{t+1}$ using $N(\theta_{4,1}^t, sd)$

Classical Genetic algorithm

Other hyperparameters to consider

Which fitness?



HMM

The occasionally dishonest casino problem

Optimization of a and e using The Genetic Algorithm

Experimen	#1	#2	#3	#4	#5
Genetic Algorithm					
P_{same}	-4.9473	-3.5693	-3.2932	-3.0982	-4.2345
$P_{\text{different}}$	-7.4982	-8.9727	-8.6473	-8.5291	-9.1483
Forward-Backward Procedure					
P_{same}	-4.7359	-4.2125	-4.9843	-4.3908	-4.3876
$P_{\text{different}}$	-7.2714	-8.6137	-7.5914	-7.7634	-7.1007

Experimen	#6	#7	#8	#9	#10
Genetic Algorithm					
P_{same}	-3.3281	-4.1869	-4.2322	-4.3872	-3.1539
$P_{\text{different}}$	-7.5581	-7.6257	-8.6274	-8.7812	-8.3641
Forward-Backward Procedure					
P_{same}	-4.9811	-4.3481	-4.0567	-4.4860	-4.9251
$P_{\text{different}}$	-7.3825	-7.7351	-7.9328	-7.7514	-8.2254

Optimization of HMM by a GA

HMM

The occasionally dishonest casino problem

Baum–Welch algorithm

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$P(x|\text{HMM}=a_{rk}; e_k(b))$

Likelihood of an observed sequence given a set of **transition** and **emission** probabilities from a given HMM, computed using the forward algorithm

HMM

Imagine I could establish the probability of observing F,F given the sequence and the observed symbols and a proposed a_{kl}

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

π F,F,F,F,F,F,F,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,C,C,C,C,C,C,C,C,F,F,F,F,F,F,F,F,F,F,F,F,F,F

Which is the probability of observing this sequence given a transition a_{kl} probability and all the observed sequence?

$$P(\pi_i = k; \pi_{i+1} = l | x; HMM) = \frac{P(\pi_i = k | x_1 \dots x_i; HMM) P(x_i | \pi_i = k; HMM) P(\pi_{i+1} = l | \pi_i = k; HMM) P(x_{i+1} | \pi_{i+1} = l; HMM) P(\pi_{i+1} = l | x_{i+1} \dots x_L; HMM)}{P(x)}$$

Probability of observing state k at position i, state l at position i+1 given the observed sequence of symbols and a proposal of parameters

Emission probability of symbol x at position i using state k

Transition probability of state k to l


Emission probability of symbol x at position i+1 using state l

$$P(\pi_i = k; \pi_{i+1} = l | x; HMM) = \frac{P(\pi_i = k | x_1 \dots x_i; HMM) e_k(b = x_i) a_{kl} e_l(b = x_{i+1}) P(\pi_{i+1} = l | x_{i+1} \dots x_L; HMM)}{P(x)}$$

Knowing that, I would be able to get a new estimate of the expected number of times that I expect to see such combination of k and l and, given that, propose a new a_{kl} to maximize the likelihood of the data given the parameters.

HMM
i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4


$$P(\pi_i = k | x) = \frac{P(\pi_i = k, x)}{P(x)}$$

"I want to know the probability of state k at position i given that I have observed the whole sequence x "

HMM

i

x

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

$$P(\pi_i = k, x_1 \dots x_i)$$

$$P(x_{i+1} \dots x_L | \pi_i = k, x_1 \dots x_i)$$

$$P(\pi_i = k, x) = P(\pi_i = k, x_1 \dots x_i) P(x_{i+1} \dots x_L | \pi_i = k, x_1 \dots x_i)$$

HMM

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(\pi_i = k, x_1 \dots x_i)$$

$$P(x_{i+1} \dots x_L | \pi_i = k)$$

Because it only depends on the last element!

$$P(\pi_i = k, x) = P(\pi_i = k, x_1 \dots x_i) P(x_{i+1} \dots x_L | \pi_i = k)$$

HMM

i

x

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

$$P(\pi_i = k, x_1 \dots x_i)$$

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

$$P(\pi_i = k | x) = \frac{P(\pi_i = k)P(x | \pi_i = k)}{P(x)}$$

HMM

The occasionally dishonest casino problem

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

Imagine we have observed this part of the sequence

Imagine we know the parameters of the HMM (A, B)

$$P(X_1, X_2 \dots X_i, \pi_i = l | HMM)$$

“Which is the probability that I observe the sequence of symbols and at position i the state is l given the parameters of my HMM?”

HMM

The occasionally dishonest casino problem

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

Imagine we have observed this part of the sequence
Imagine we know the parameters of the HMM (A, B)

Transition probability matrix

Emission probability matrix

$$P(X_1, X_2 \dots X_i, \pi_i = l | HMM)$$

"Which is the probability that I observe the sequence of symbols and at position i the state is l given the parameters of my HMM?"

HMM

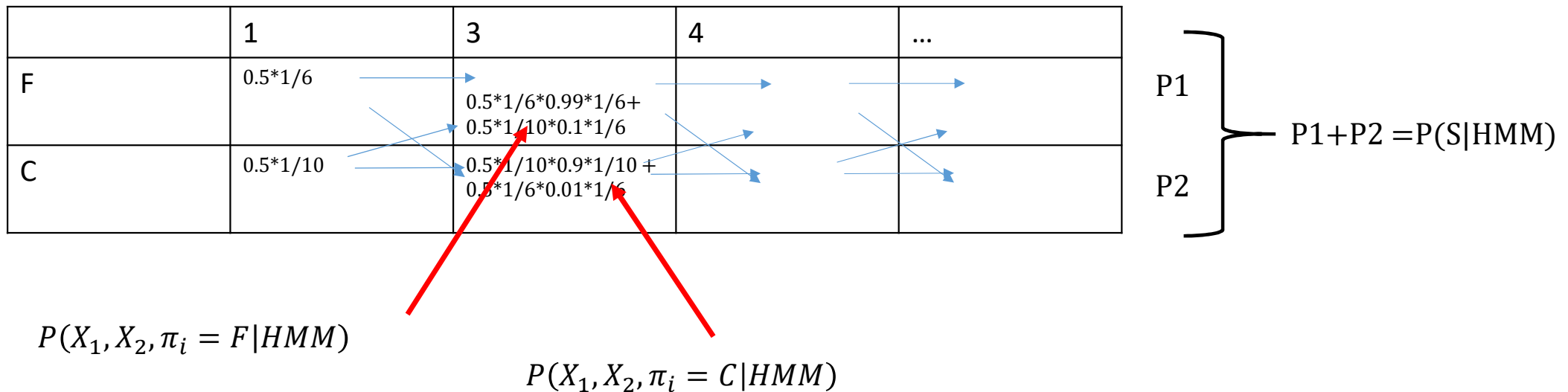
Forward algorithm: how to estimate the probability of a sequence ending in state k given a HMM

$$P(X_1, X_2 \dots X_i, \pi_i = l | HMM)$$

“Which is the probability that I observe the sequence of symbols and at position i the state is k given the parameters of my HMM?”

HMM

Forward algorithm: how to estimate the probability of a sequence ending in state k given a HMM



HMM

Forward algorithm: how to estimate the probability of a sequence ending in state k given a HMM

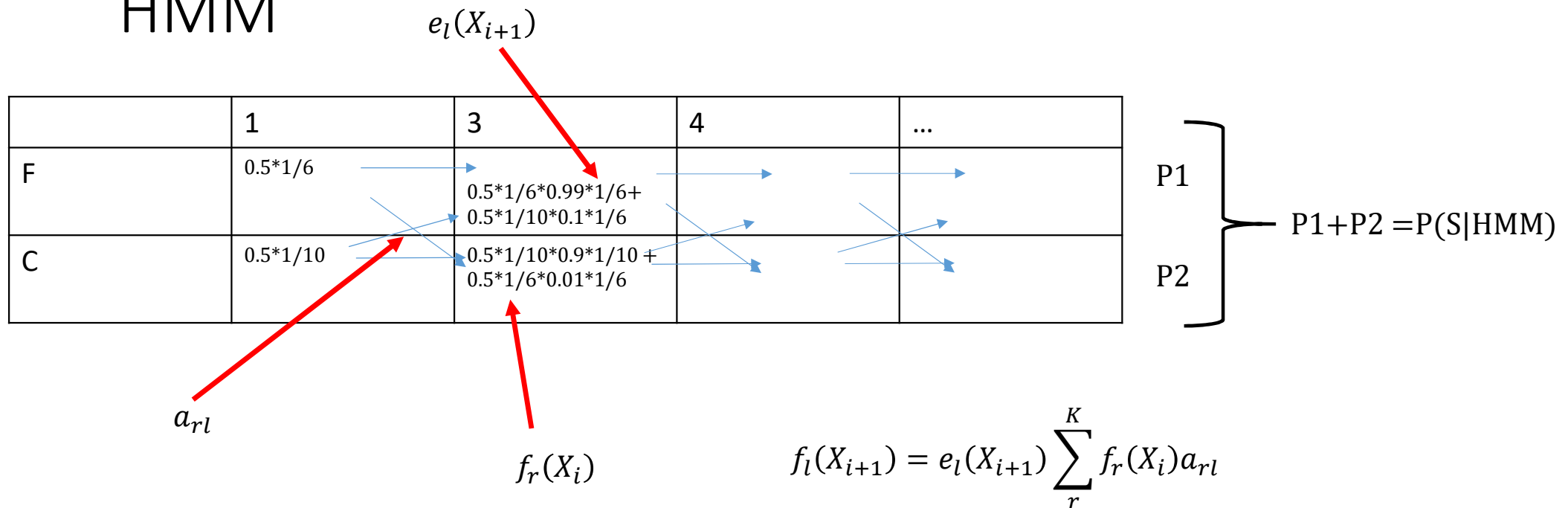
$$P(X_1, X_2 \dots X_{i+1}, \pi_{i+1} = l | HMM) = f_l(X_{i+1}) = e_l(X_{i+1}) \sum_r^K f_r(X_i) a_{rl}$$

Emission probability of X at time $i+1$ given that the state is l Probability of the sequence up to i , ending at i in state r Transition probability to move from state r to l

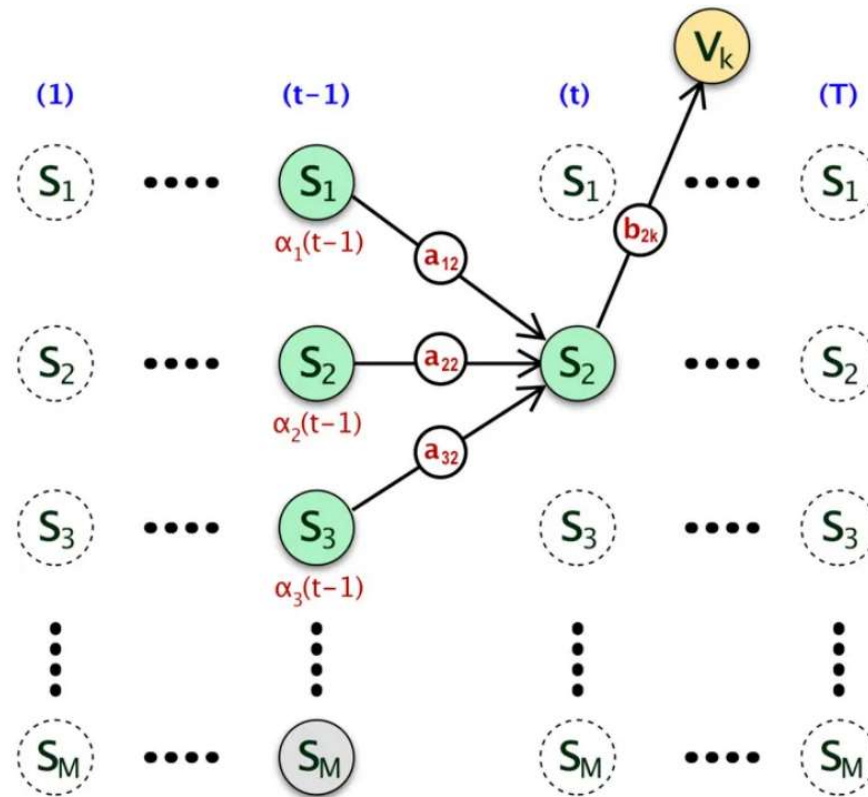
"Which is the probability that I observe the sequence of symbols and at position i the state is k given the parameters of my HMM?"

HMM

Forward algorithm: how to estimate the probability of a sequence ending in state k given a HMM



Another way of visualizing the forward algorithm



HMM

Forward algorithm: how to estimate the probability of a sequence ending in state k given a HMM. Pseudocode (From the book of Durbin)

Algorithm: Forward algorithm

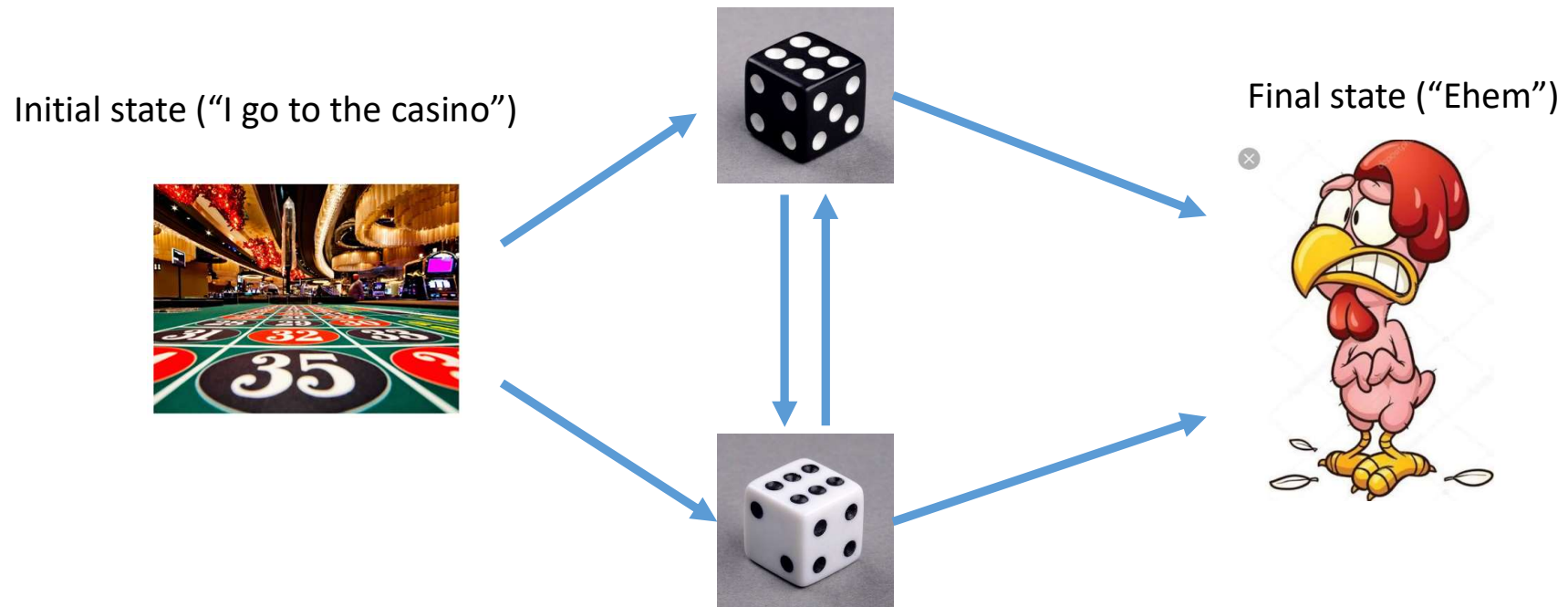
Initialisation ($i = 0$): $f_0(0) = 1, f_k(0) = 0$ for $k > 0$.

Recursion ($i = 1 \dots L$): $f_l(i) = e_l(x_i) \sum_k f_k(i-1) a_{kl}$.

Termination: $P(x) = \sum_k f_k(L) a_{k0}$.

HMM

In fact, there is an initial state and a final state

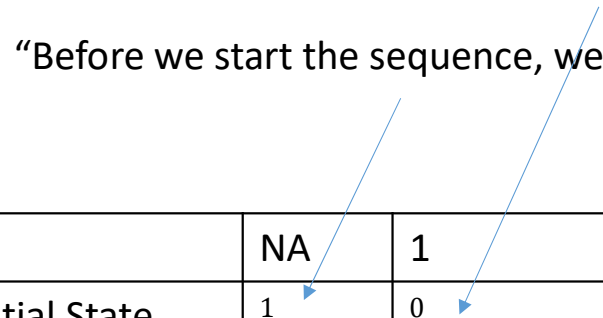


HMM

Forward algorithm: There are two other states!

“After we start the sequence, the probability of going to initial state is 0”

“Before we start the sequence, we are at the initial state”

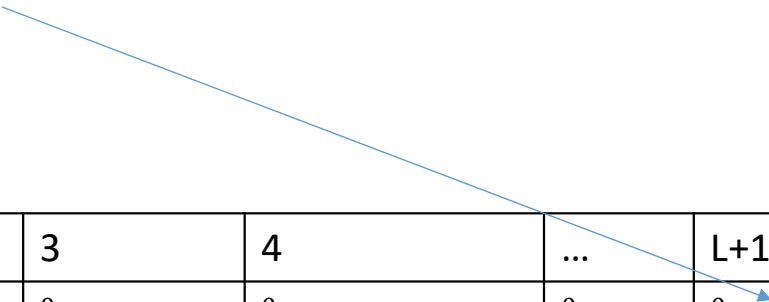


	NA	1	3	4	...	L+1
Initial State	1	0	0	0	0	0
F	0					0
C	0					0
Final State	0	0	0	0	0	$P(S HMM)$

HMM

Forward algorithm: There are two other states!

“After we finish the sequence, the probability of going to the final state is 1”



	NA	1	3	4	...	L+1
Initial State	1	0	0	0	0	0
F	0					0
C	0					0
Final State	0	0	0	0	0	$P(S HMM)$

HMM

i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

HMM

i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

	...	5	5	4
F				
C				

Backward algorithm: how to estimate the probability of a sequence given an initial k state

Backward algorithm

HMM
i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

	...	5	5	4
F				1
C				1


Starting at the last element, the entries in this column denote the probability moving to the end state from each state after generating the entire sequence.

Backward algorithm

HMM
i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

	...	5	5	4
F				1
C				1

$$a_{kr} e_r(X_i) b_k(i)$$

$$0.95 * \frac{1}{6} * 1$$

Backward algorithm

HMM
i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

	...	5	5	4
F			$0.95 * \frac{1}{6} * 1$	1
C				1

$$a_{kr} e_r(X_i) b_k(i)$$

$$0.05 * \frac{1}{6} * 1$$

Backward algorithm

HMM
i

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

	...	5	5	4
F			$0.95 * \frac{1}{6} * 1 + 0.05 * \frac{1}{6} * 1$	1
C				1

$$a_{kr} e_r(X_i) b_k(i)$$

$$0.05 * \frac{1}{6} * 1$$

Backward algorithm

HMM
i

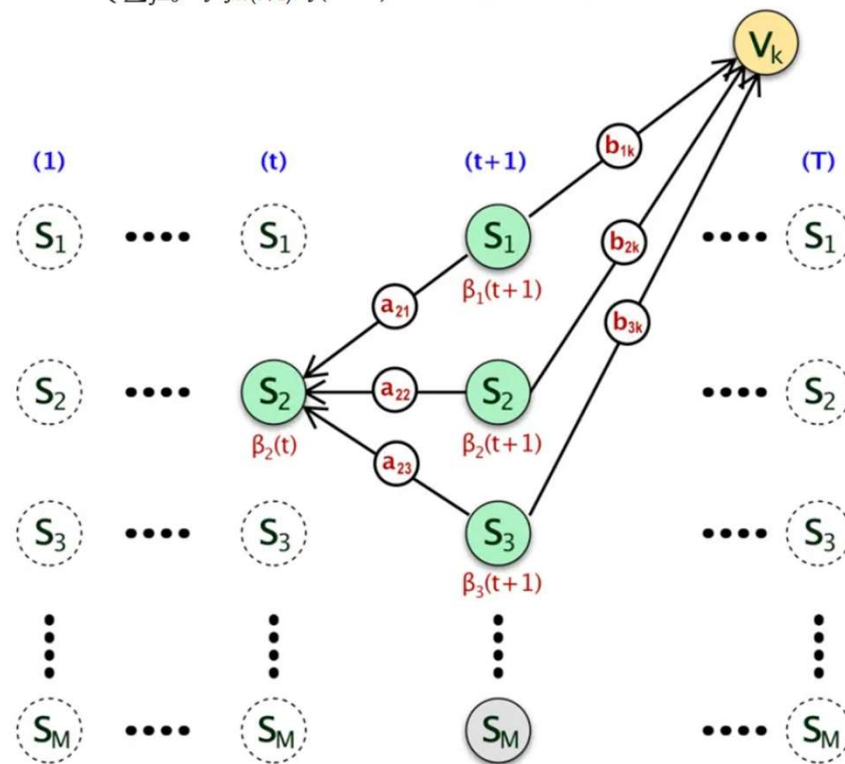
x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

$$P(x_{i+1} \dots x_L | \pi_i = k) = b_k(i)$$

	...	5	5	4
F			$0.95 * \frac{1}{6} * 1 + 0.05 * \frac{1}{6} * 1$	1
C			$0.9 * \frac{1}{10} * 1 + 0.1 * \frac{1}{10} * 1$	1

Another way of visualizing the backward algorithm

$$\beta_i(t) = \begin{cases} 1 & \text{when } t = T \\ \sum_{j=0}^M a_{ij} b_{jkv(t+1)} \beta_j(t+1) & \text{when } t \text{ less than } T \end{cases}$$

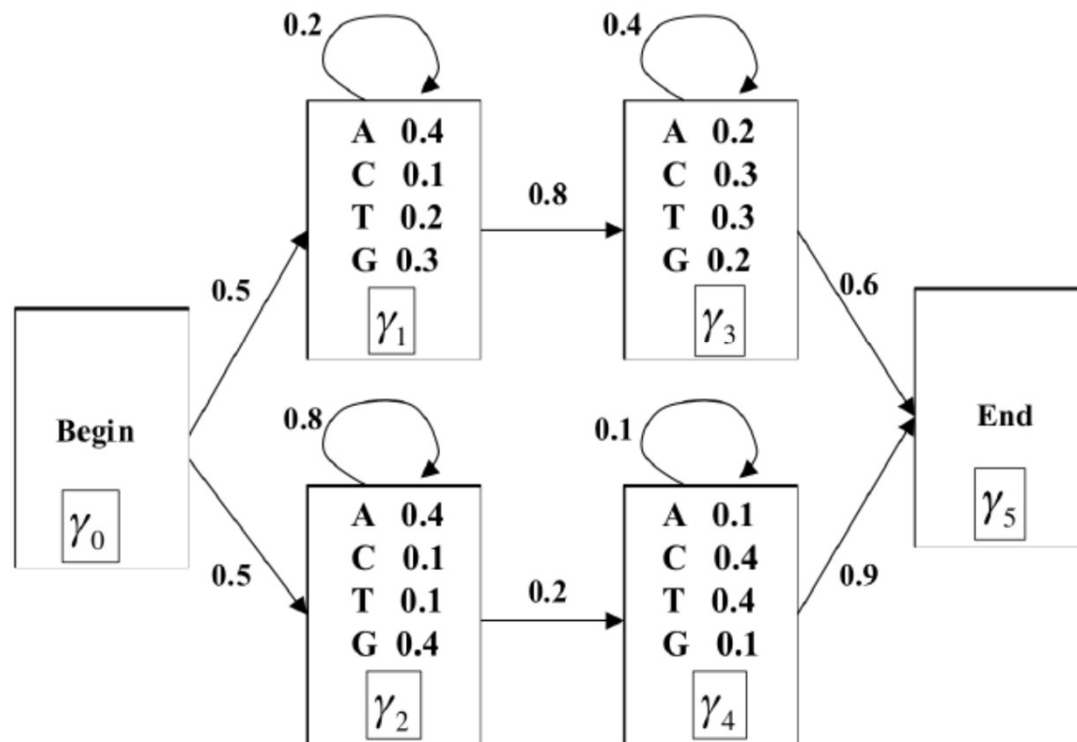


HMM

Backward algorithm

Remember: Even if we have not explicitly stated, beginning and end are also states

What is strange in this HMM compared to the ones we have been using?



HMM

Backward algorithm

If our subsequence was TAGA, which is the probability that we have ended at that each stage?

Time Step, t State, l	0	1 (T)	2 (A)	3 (G)	4 (A)
γ_0	-	-	-	-	0
γ_1	-	-	-	-	0
γ_2	-	-	-	-	0
γ_3	-	-	-	-	0.6
γ_4	-	-	-	-	0.9
γ_5	-	-	-	-	0

HMM

Backward algorithm

Algorithm: Backward algorithm

Initialisation ($i = L$): $b_k(L) = a_{k0}$ for all k .

Recursion ($i = L - 1, \dots, 1$): $b_k(i) = \sum_l a_{kl} e_l(x_{i+1}) b_l(i + 1)$.

Termination: $P(x) = \sum_l a_{0l} e_l(x_1) b_l(1)$.

HMM

The occasionally dishonest casino problem

Baum–Welch algorithm

Diagram illustrating the Baum–Welch algorithm equation, showing the probability of a state transition and emission:

$$P(\pi_i = k, \pi_{i+1} = l | X, HMM) = \frac{f_k(i) a_{kl} e_l(x_{i+1}) b_l(i+1)}{P(X)}$$

Annotations explaining the terms in the equation:

- Probability of ending at position i in state k* (points to $f_k(i)$)
- Probability that I change from k to l* (points to a_{kl})
- Probability of emission of symbol x at position $i+1$ using state l* (points to $e_l(x_{i+1})$)
- Probability of the sequence from $i+1$ to L if I start at $i+1$ at state l* (points to $b_l(i+1)$)
- Probability of a transition from state k at position i to state l in position $i+1$* (points to the denominator $P(X)$)

HMM

The occasionally dishonest casino problem

Baum–Welch algorithm

We can do this for each position. Adding over all the positions, we have an estimate of “*how often would occur a transition from state k to l if the parameter a_{kl} that we are using was the one that generated the data*”

$$A_{kl} = \sum_i^L P(\pi_i = k, \pi_{i+1} = l | X, HMM)$$

$$a_{kl}^{(t+1)} = \frac{A_{kl}^t}{\sum_{l'}^K A_{kl'}^t}$$

“Given that I come from state k , which is the probability of moving to any other state”

Posterior probability of $a_{kl}(t+1)$ given $a_{kl}(t)$ and the observed data

When the state was not hidden, this was the **maximum likelihood** estimate of the transition a_{kl} !

HMM

The occasionally dishonest casino problem

Baum–Welch algorithm

How can we estimate the emission probabilities?

x 1,4,5,1,3,2,1,6,6,4,1,6,6,6,6,6,2,3,4,5,1,4,3,6,2,1,3,5,6,6,6,6,6,6,1,2,3,1,1,1,1,2,3,4,5,5,4

Pick one of the symbols (for example, $b = 6$)

The index i of this position is 8

$$E_k(b) = \frac{1}{P(x)} \sum_{\{i|x_i = b\}} f_k(i) b_k(i)$$

← Probability that we are at state k in position i

For all the positions where we observe 6

New emission probabilities

$$e_k^{t+1}(b) = \frac{E_k(b)}{\sum_{b'} E_k(b')}$$

HMM

Baum-Welch algorithm

- 1) Propose a set of transition ($a_{..}$) and emission ($b_{.}(.)$) matrices
- 2) Until change of likelihood $P(X|HMM)$ is not smaller than threshold, do
- 3) Using the forward and backward algorithm, estimate the new $A_{..}$ and $B_{..}$ variables
- 4) Estimate the new $a_{..}$ and $b_{.}(.)$ matrices
- 5) Go to 2)

The Baum-Welch is a special case of the EM algorithm

- Two steps
 - Expectation: missing data are estimated given the observed data and current estimate of the model parameters.
 - Maximization: the likelihood function is maximized under the assumption that the missing data are known.

How does a phenotype change over a phylogeny?

Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character:
The Evolution of Plant Habit in Campanulid Angiosperms
JEREMY M. BEAULIEU^{1,*}, BRIAN C. O'MEARA², AND MICHAEL J. DONOGHUE¹

Model adequacy is a different concern than model fit: the latter determines which of a set of models is least bad (or “best”) for the data, whereas the former determines whether a single model adequately describes the data. For example, a model that says humans, chimps, and mice all diverged simultaneously 70 million years ago is a better fit to the data than one that puts that divergence 700 million years ago, but neither model adequately describes the data. Adequacy is often evaluated by simulating under the focal model to see if it generates data indistinguishable from the empirical data using one or more measures.