# CS 364 COMPUTATIONAL BIOLOGY

Sara Mathieson
Haverford College

# Outline

Lab 8 due Thursday
Review: in-class on Thursday
Midterm 2: in-class on Tuesday
Office hours: TODAY
2:30-3:30pm in Zubrow

- Finish Forward-Backward algorithm
- Posterior decoding and posterior mean
- Parameter estimation
- Baum-Welch algorithm (EM for HMM)

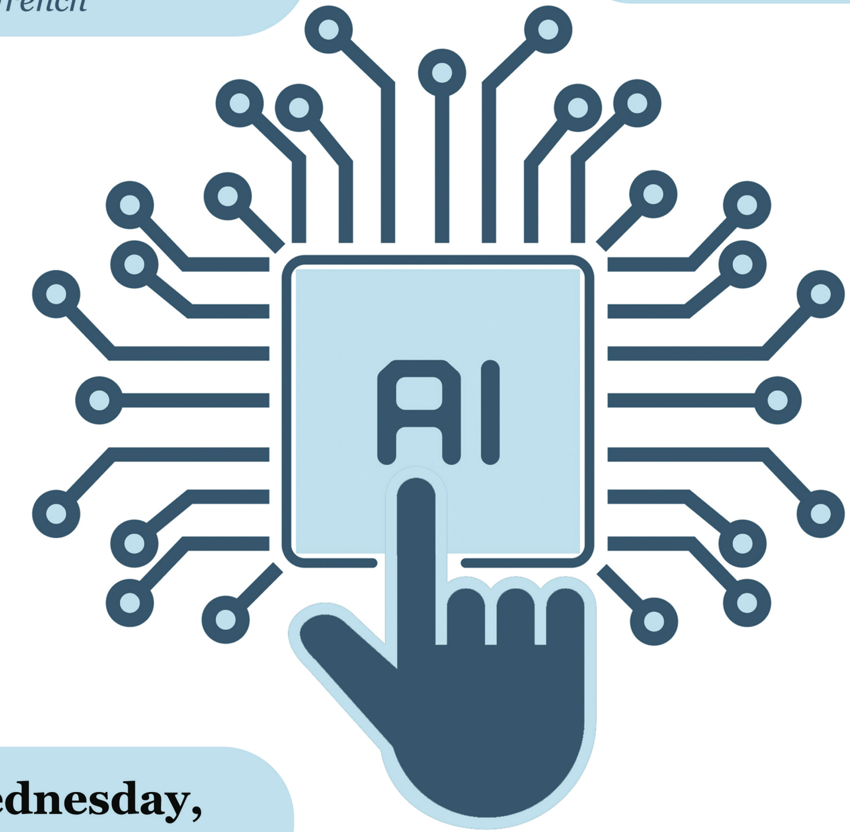
# AI Uses in the Classroom panel

**Benjamin Le**
*Associate Provost for Faculty Development, Professor of Psychology*

**Sara Mathieson**
*Associate Professor of Computer Science, Coordinator of Scientific Computing*

**Patrick Kelly '25**
*Political Science and French*

**Pranav Rane '25**
*Computer Science*




**Wednesday, November 20**
**4:30pm in Lutnick 232**

HAVERFORD
LIBRARIES

# More HMM examples

# HMMs in practice

## Example 1: part-of-speech tagging

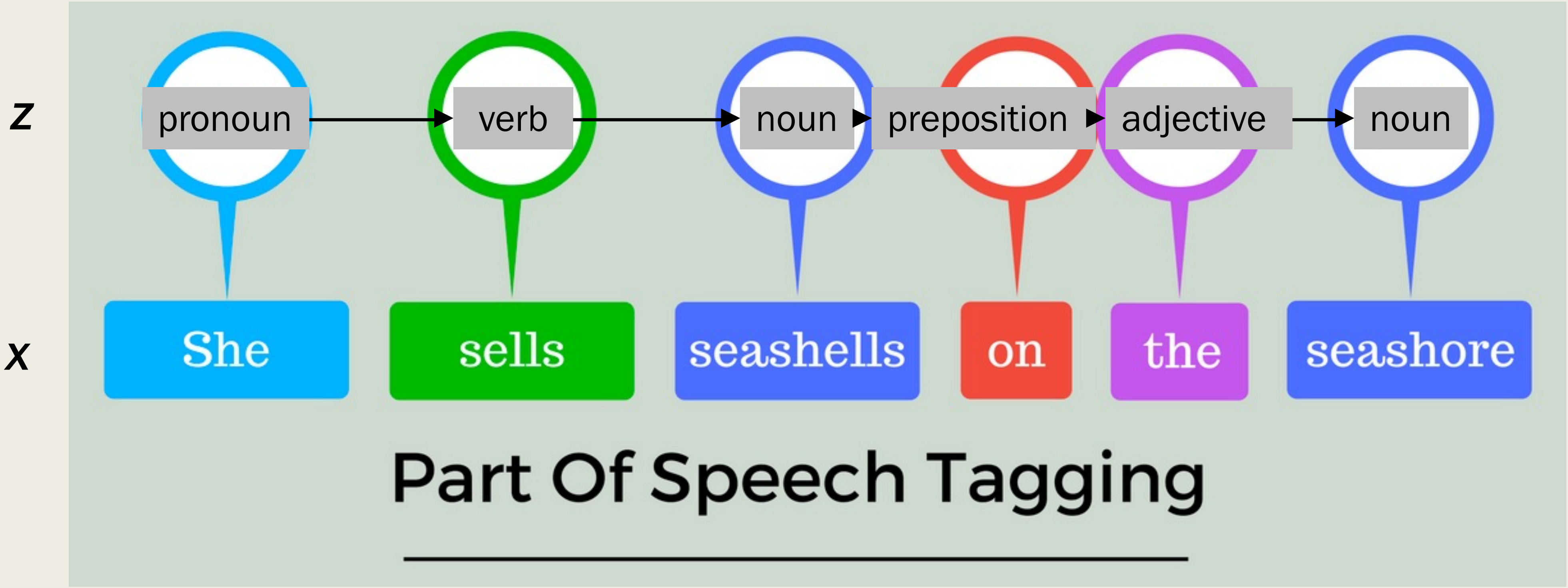



Image: modified from http://nlpforhackers.io/training-pos-tagger/

# HMMs in practice

## Example 2: multiple sequence alignment

**Hidden Markov models of biological primary sequence information**

**(multiple sequence alignments/protein modeling/adaptive algorithms/sequence classification)**

PIERRE BALDI*†, YVES CHAUVIN‡§, TIM HUNKAPILLER*¶, AND MARCELLA A. MCCLURE‖**

HMM architecture that models matches, insertions, and deletions

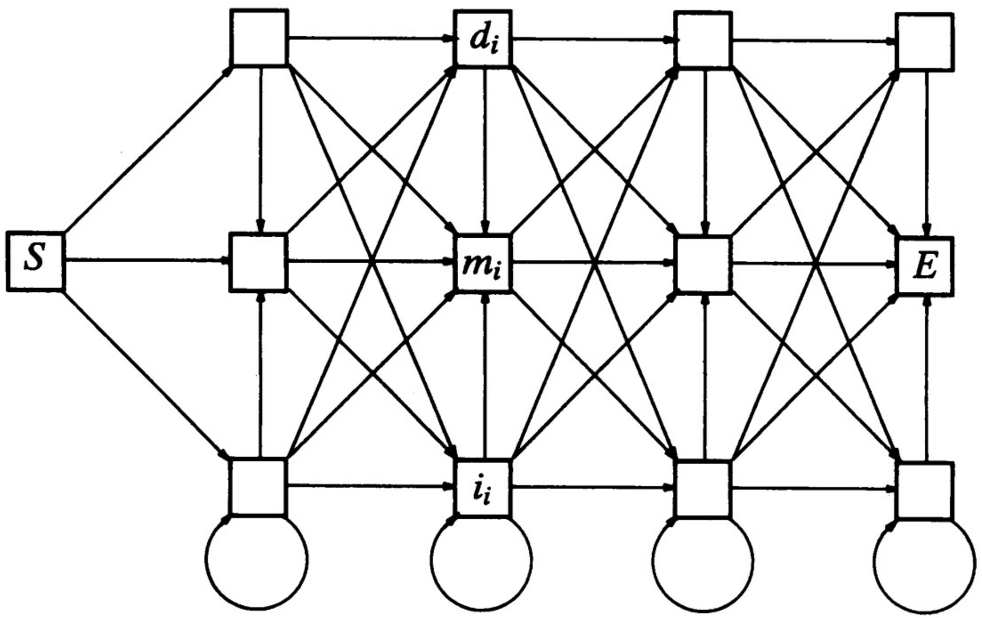



FIG. 1. HMM architecture. $S$ and $E$ are the start and end states. Sequence of main states $m_i$ is the backbone. Side states $d_i$ (resp. $i_i$) correspond to deletions (resp. insertions).

Multiple sequence alignment results on proteins

```
                                                1          2                   3                4
   ..............................AAAAAAAAAAAAAAAAAA.BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBCCCCCCCC..
V HAHU         -----------------VLSPA-DKTNVKAAWGKVGAHAgEYGAE-A---L--E-R--MFL----SFPTTKTYFP
T HAOR         -----------------MLTDA-EKKEVTALWGKAAGHGeEYGAE-A---L--E-R--LFQ----AFPTTKTYFS
V HADK         -----------------VLSAA-DKTNVKGVFSKIGGHAeEYGAE-T---L--E-R--MFI----AYPQTKTYFP
V HBHU         -----------------VHLTP-EEKSAVTALWGKVNVD-EVGGE-A---L--G-R--LLV----VYPWTQRFFE
T HBOR         -----------------VHLSG-GEKSAVTNLWGKVNIN-ELGGE-A---L--G-R--LLV----VYPWTQRFFE
V HBDK         -----------------VHWTA-EEKQLITGLWGKVNVA-DCGAE-A---L--A-R--LLI----VYPWTQRFFA
V MYHU         -----------------GLSDG-EWQLVLNVWGKVEADIpGHGQE-V---L--I-R--LFK----GHPETLEKFD
V MYOR         -----------------GLSDG-EWQLVLKVWGKVEGDLpGHGQE-V---L--I-R--LFK----THPETLEKFD
V IGLOB        mkffavlalcivgaias-PLTA-DEASLVQSSWKAVSHN-EV--E-I---L--A-A--VFA----AYPDIQNKFS
V GPYL         -----------------GVLTDvQVALVKSSFEEFNANIpKN---t-hr--ffT-L--VLE----IAPGAKDLF-
V GPUGNI       -----------------ALTEK-QEALLKQSWEVLKQNIpAH---s-l-------R--LFAliieAAPESKYVF-
V GGZLB        -----------------ML-D----QQTINIIKATVPVLkEHGVT-Ittt-f--y-knLFA----KHPEVRPLFD
   ..........................................................*..................*.....*.
   .................................................................................I....
```

# Forward-Backward, posterior decoding, and posterior mean

# Goal: compute the posterior probability of being in state *k* at step *i*

- ***Posterior probability***: probability of an "unknown" given observed (known) data

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

# Goal: compute the posterior probability of being in state *k* at step *i*

- ***Posterior probability***: probability of an “unknown” given observed (known) data

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

Likelihood of data given an unknown

- Aside: we can rewrite the numerator to include a prior (in a Bayesian setting)

$$= \frac{P(z_i = k) \cdot P(\vec{x}|z_i = k)}{P(\vec{x})}$$

Prior

Evidence (actual observed data)

# Goal: compute the posterior probability of being in state *k* at step *i*

Focus on the numerator

- ***Posterior probability***: probability of an "unknown" given observed (known) data

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

Likelihood of data given an unknown

- Aside: we can rewrite the numerator to include a prior (in a Bayesian setting)

$$= \frac{P(z_i = k) \cdot P(\vec{x}|z_i = k)}{P(\vec{x})}$$

Prior

Evidence (actual observed data)

# Goal: compute the posterior probability of being in state *k* at step *i*

$$P(\vec{x}, z_i = k) = P(x_1, \cdots x_i, z_i = k, x_{i+1}, \cdots x_L)$$

Break up emitted sequence around step *i*

# Goal: compute the posterior probability of being in state *k* at step *i*

$$P(\vec{x}, z_i = k) = P(x_1, \cdots x_i, z_i = k, x_{i+1}, \cdots x_L)$$

Break up emitted sequence around step *i*

$$= P(x_1, \cdots x_i, z_i = k) \cdot P(x_{i+1}, \cdots x_L | x_1, \cdots x_i, z_i = k)$$

Use conditional probability

# Goal: compute the posterior probability of being in state *k* at step *i*

$$P(\vec{x}, z_i = k) = P(x_1, \cdots x_i, z_i = k, x_{i+1}, \cdots x_L)$$

Break up emitted sequence around step *i*

$$= P(x_1, \cdots x_i, z_i = k) \cdot P(x_{i+1}, \cdots x_L | x_1, \cdots x_i, z_i = k)$$

Use conditional probability

$$= P(x_1, \cdots x_i, z_i = k) \cdot P(x_{i+1}, \cdots x_L | z_i = k)$$

Use Markov property

# Goal: compute the posterior probability of being in state *k* at step *i*

$$P(\vec{x}, z_i = k) = P(x_1, \cdots x_i, z_i = k, x_{i+1}, \cdots x_L)$$

Break up emitted sequence around step *i*

$$= P(x_1, \cdots x_i, z_i = k) \cdot P(x_{i+1}, \cdots x_L | x_1, \cdots x_i, z_i = k)$$

Use conditional probability

$$= P(x_1, \cdots x_i, z_i = k) \cdot P(x_{i+1}, \cdots x_L | z_i = k)$$

Use Markov property

$$= f_k(i) \cdot b_k(i)$$

Define these two pieces as the forward and backward probabilities

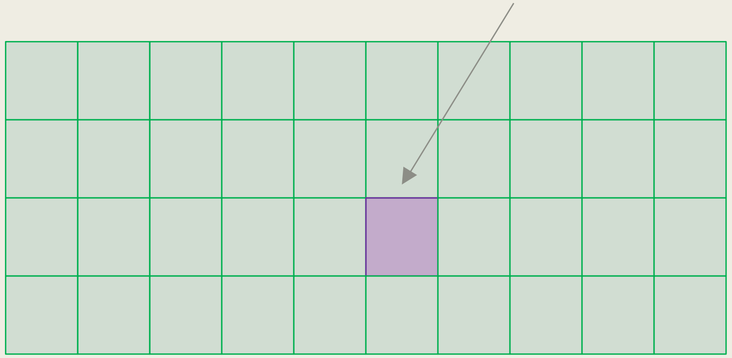
Forward

Backward

# Forward-backward Algorithm

- **Input:** observed sequence x=($x_1$,$x_2$,…,$x_L$) and transition/emission probabilities (***a*** and ***e*** matrices)
- **Output:** posterior probability of being in each hidden state at each time point $P(Z_i=k|x)$

What is the probability of being in this state, conditional on the observations?



Want to compute $P(Z_i=k|x_1,x_2 \ldots x_L) = P(Z_i=k, x_1,x_2 \ldots x_L) / P(x_1,x_2 \ldots x_L)$

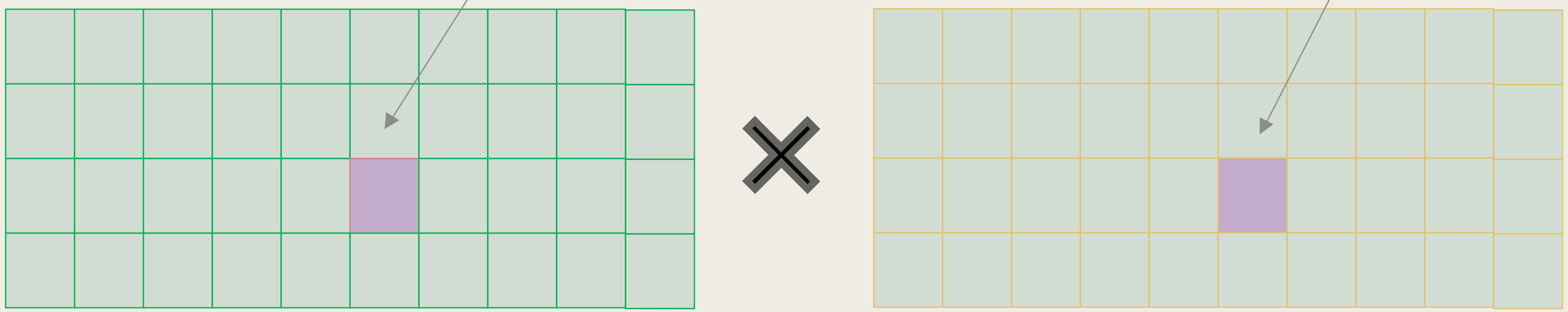
# Forward-backward Algorithm

Forward matrix $f_K(i)$

Backward matrix $b_k(i)$

What is the probability of being in this state, conditional on all the earlier observations

What is the probability of all the later observations, conditional on being in this state



Want to compute $P(Z_i=k \mid x_1,x_2 \ldots x_L) = P(Z_i=k, x_1,x_2 \ldots x_L) / P(x_1,x_2 \ldots x_L)$

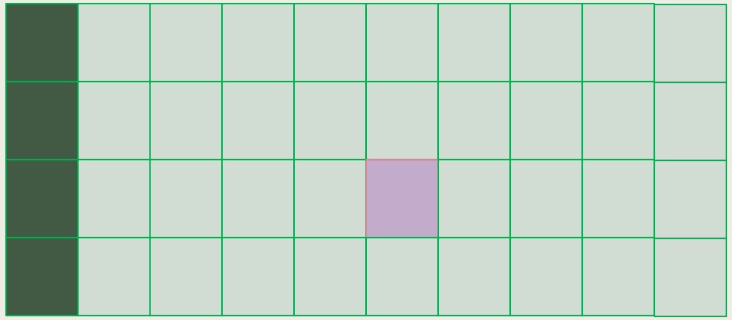
But we can break this up: $P(Z_i=k, x_1,x_2 \ldots x_L) = \boxed{P(Z_i=k, x_1,x_2 \ldots x_i)}\boxed{P(x_{i+1},x_2 \ldots x_L \mid Z_i=k)}$

Probability of being in state k at time i, given the sequence up to time i: Forward algorithm

Probability of the sequence after time i, given that you are in state k at time i: Backward algorithm

# Forward Algorithm

1. $f_k(1) = \pi_k e_k(x_1)$



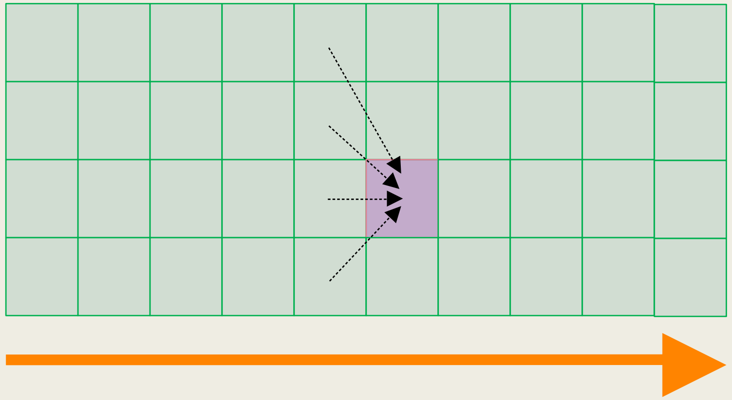
$P(Z_i=k, x_1,x_2 \ldots x_i)$

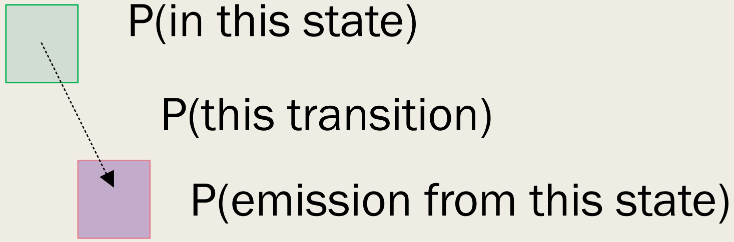
Probability of being in state k at time i, given the sequence up to time i: Forward algorithm

# Forward Algorithm

1. $f_k(1) = \pi_k e_k(x_1)$

2 $f_k(i) = \Sigma_{j=1}^{K} f_j(i-1) a_{jk} e_k(x_i)$






$\boxed{P(Z_i=k, x_1, x_2 \ldots x_i)}$

Probability of being in state k at time i, given the sequence up to time i: Forward algorithm

# Backward Algorithm

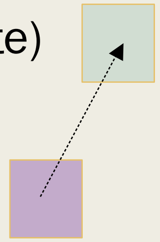
1 $b_k(K) = 1$

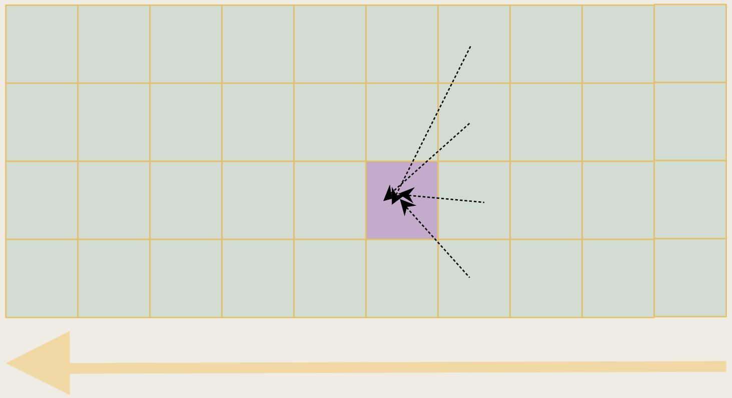
2 $b_k(i) = \Sigma_{j=1}^{K} b_j(i+1) a_{ij} e_j(x_{i+1})$

P(emission)

P(this state)

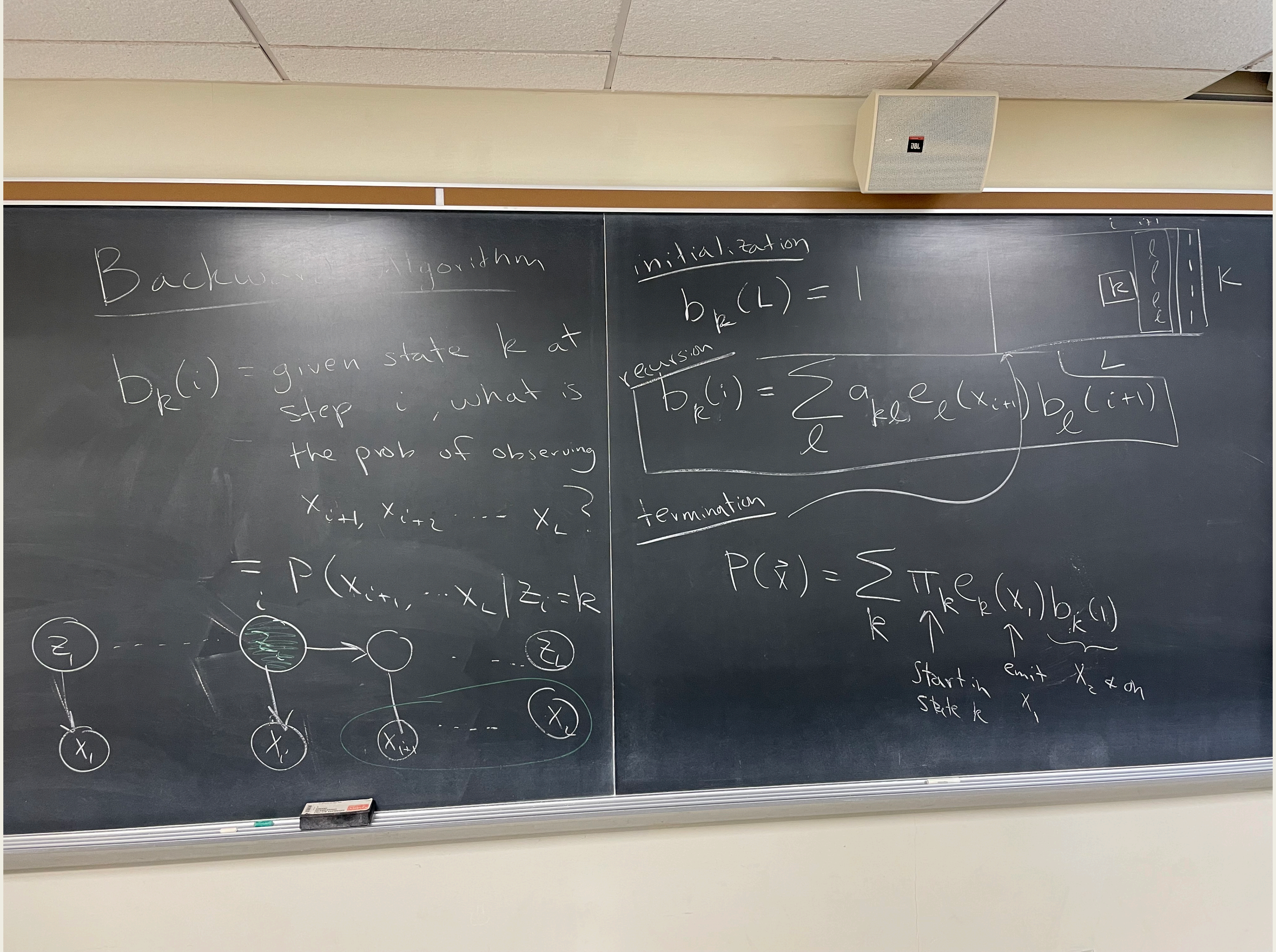
P(transition)





$P(x_{i+1}, x_2 \ldots x_L | Z_i = k)$

Probability of the sequence after time i, given that you are in state k at time i: Backward algorithm


Backward Algorithm
$b_k(i)$ = given state k at step i, what is the prob of observing $x_{i+1}, x_{i+2} \dots x_L$?
$= P(x_{i+1}, \dots x_L \mid z_i = k)$
$z_1$
$z_L$
$x_1$
$x_i$
$x_{i+1}$
$x_L$
initialization
$b_k(L) = 1$
recursion
$b_k(i) = \sum_\ell a_{k\ell} e_\ell(x_{i+1}) b_\ell(i+1)$
termination
$P(\vec{x}) = \sum_k \pi_k e_k(x_1) b_k(1)$
Start in state k
emit $x_1$
$x_2$ & on
k
K
L

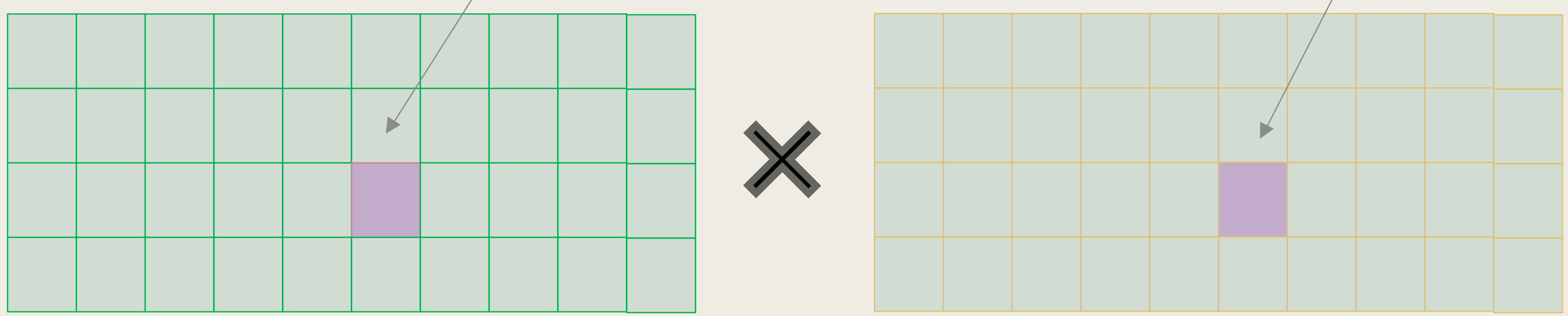
# Forward-backward Algorithm

Forward matrix $f_K(i)$

Backward matrix $b_k(i)$

What is the probability of being in this state, conditional on all the earlier observations

What is the probability of all the later observations, conditional on being in this state



Want to compute $P(Z_i=k \mid x_1,x_2 \ldots x_L) = P(Z_i=k,\ x_1,x_2 \ldots x_L\ ) \ /\ P(x_1,x_2 \ldots x_L)$

But we can break this up: $P(Z_i=k,\ x_1,x_2 \ldots x_L\ ) = \boxed{P(Z_i=k,\ x_1,x_2 \ldots x_i\ )}\boxed{P(x_{i+1},x_2 \ldots x_L \mid Z_i=k)}$

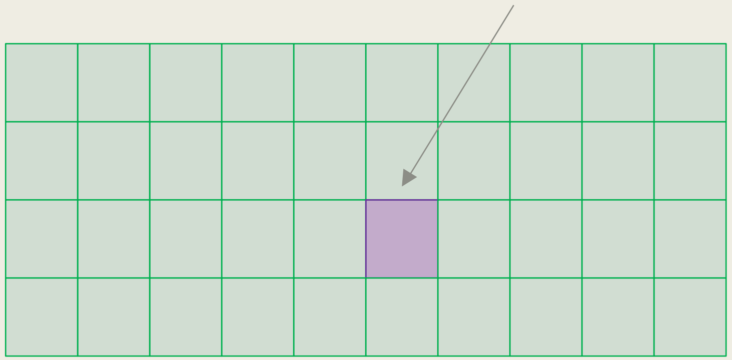
Probability of being in state k at time i, given the sequence up to time i: Forward algorithm

Probability of the sequence after time i, given that you are in state k at time i: Backward algorithm

# Forward-backward Algorithm

- **Input:** observed sequence $x=(x_1,x_2,\ldots,x_L)$ and transition/emission probabilities (***a*** and ***e*** matrices)
- **Output:** posterior probability of being in each hidden state at each time point $P(Z_i=k|x)$

What is the probability of being in this state, conditional on the observations?



Want to compute $P(Z_i=k|x_1,x_2 \ldots x_L) = P(Z_i=k, x_1,x_2 \ldots x_L) / P(x_1,x_2 \ldots x_L)$

What is the runtime of the forward-backward algorithm?

# Posterior decoding and posterior mean

Now we can complete the posterior probability:

$$P(z_i = k|\vec{x}) = \frac{f_k(i) \cdot b_k(i)}{P(\vec{x})}$$

Add last column of the forward DP table to get *P*(***x***):

$$P(\vec{x}) = \sum_k f_k(L)$$

# Posterior decoding and posterior mean

Now we can complete the posterior probability:

$$P(z_i = k|\vec{x}) = \frac{f_k(i) \cdot b_k(i)}{P(\vec{x})}$$

Add last column of the forward DP table to get *P(***x***)*:

$$P(\vec{x}) = \sum_k f_k(L)$$

<u>Posterior decoding</u>:

- Create a *K*x*L* table for the posterior probabilities
- Take the max over each column to find the posterior decoding

$$\hat{z}_i = \arg\max_k P(z_i = k|\vec{x})$$

# Posterior decoding and posterior mean

Now we can complete the posterior probability:

$$P(z_i = k|\vec{x}) = \frac{f_k(i) \cdot b_k(i)}{P(\vec{x})}$$

Add last column of the forward DP table to get *P(**x**)*:

$$P(\vec{x}) = \sum_k f_k(L)$$

Posterior decoding:

- Create a *K*x*L* table for the posterior probabilities
- Take the max over each column to find the posterior decoding

$$\hat{z}_i = \arg\max_k P(z_i = k|\vec{x})$$

$$\bar{g}_i = \sum P(z_i = k|\vec{x}) \cdot g(k)$$

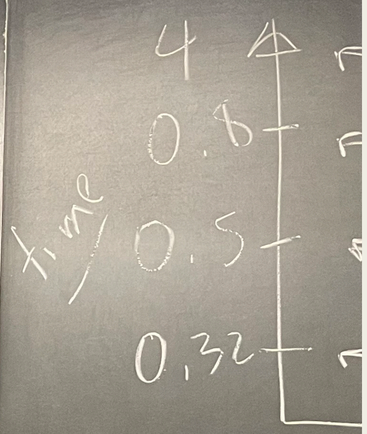
Posterior mean:

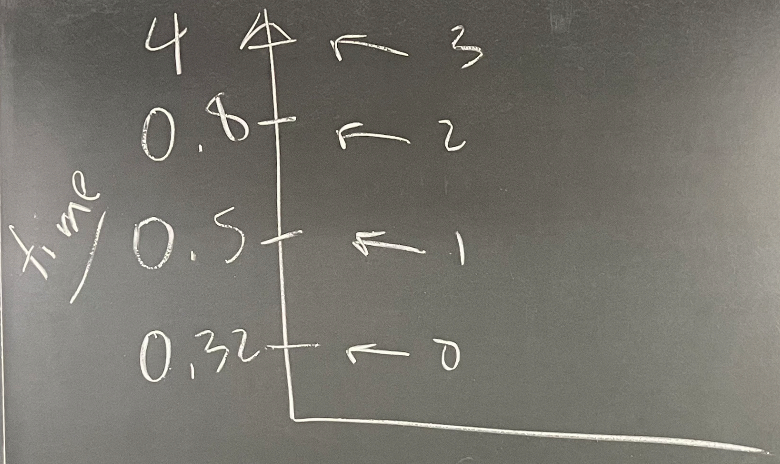
- The sum of each column in the posterior probability table should be 1
- Multiply these probabilities by the value of each state *g(k)* to get the mean

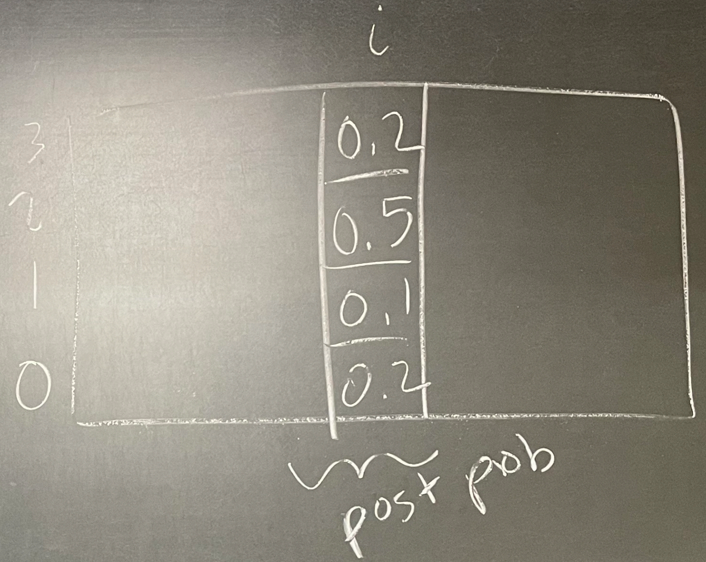
together $P(\vec{x}, z_i = k) = f_k(i)\, b_k(i)$

posterior: $\boxed{P(z_i = k \mid \vec{x}) = \dfrac{f_k(i)\, b_k(i)}{P(\vec{x})}}$

posterior decoding $\boxed{\hat{z}_i = \operatorname*{argmax}_k P(z_i = k \mid \vec{x})} \boxed{\forall\, i}$






$\forall \underline{\underline{i}}$

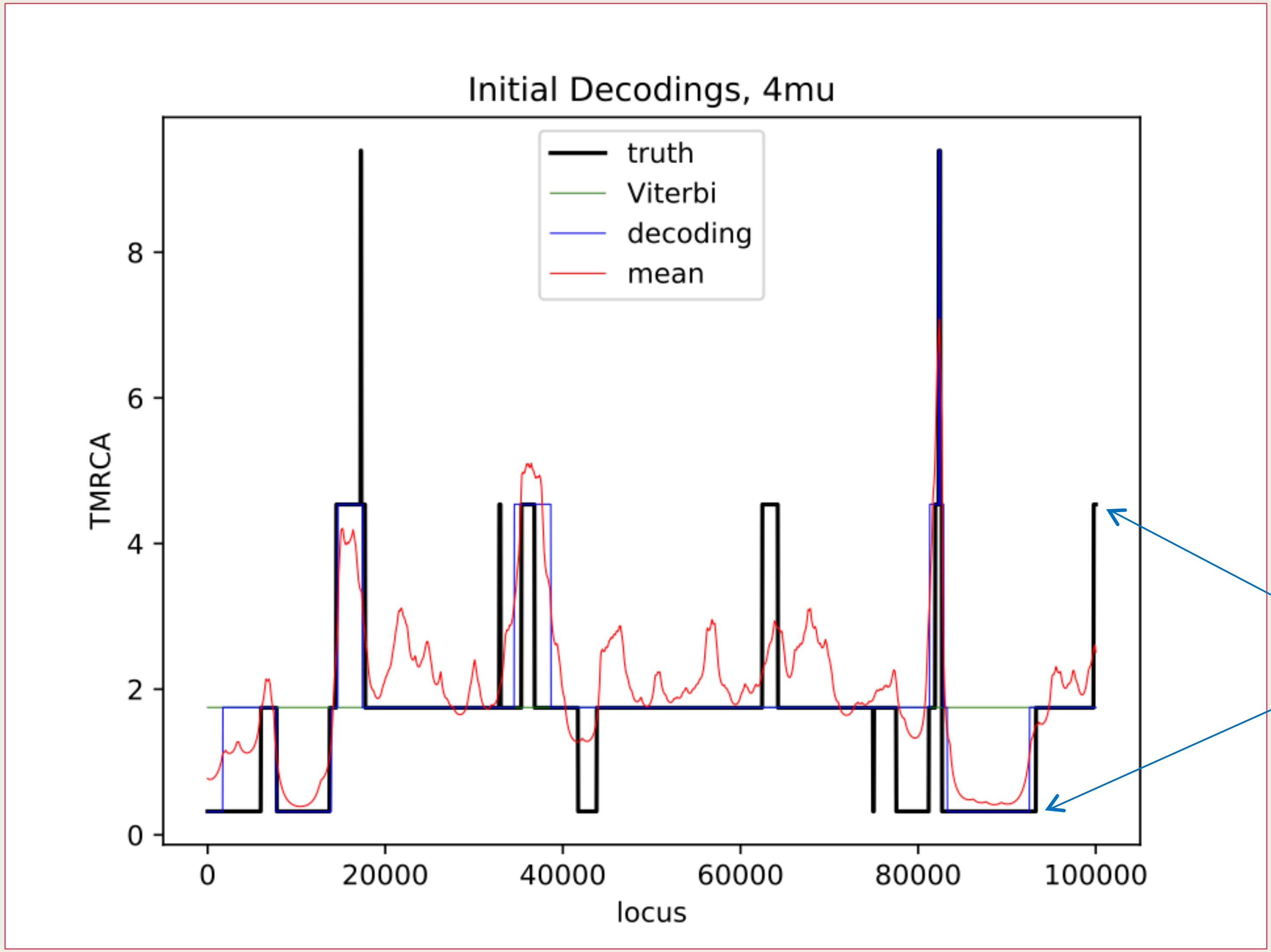
$$\bar{g}_i = \underbrace{0.2}_{P(g)}\underbrace{(0.32)}_{g} + 0.1(0.5) + 0.5(0.8) + 0.2(4)$$

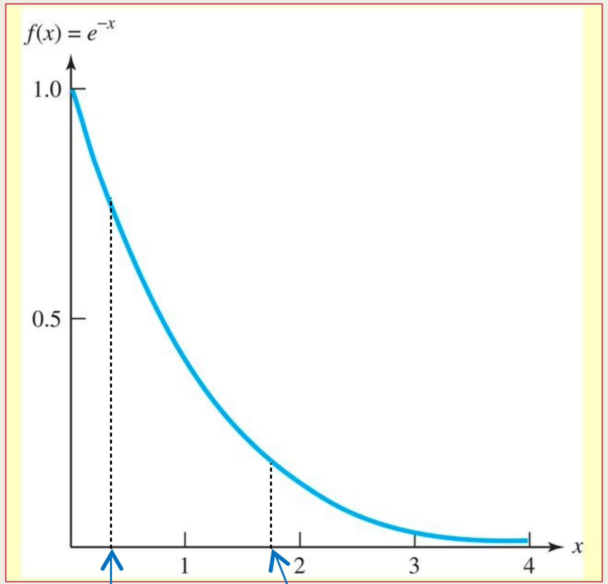
# Summary (Part 1 of Lab 8)

- Now we have 3 different ways to estimate the hidden states:
- 1) Viterbi traceback (from the max of last column)
- 2) Posterior decoding (max of posterior probability table from Forward-Backward algorithm)
- 3) Posterior mean (weighted average over each column of the posterior probability table)

- Note: the posterior mean already reflects the “meaning” of each hidden state. For the other two, we need to transform the state sequence (indices of hidden state) using a 1-1 mapping from state index to state value

# Lab 8 example: $T_{mrca}$ for $n$=2




g(0) = 0.32, g(1) = 1.75
g(2) = 4.54, g(3) = 9.40

Exponential distribution image: Modified from SlidePlayer

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on _____.

2) The value of $z_i$ only depends on _____.

3) What symbol would you use to denote the transition probability from state 2 to state 0?

4) What symbol would you use to denote the probability of emitting a 1 from state 3?

5) What symbol would you use to denote the probability of starting in state 2?

6) What is the runtime of the Viterbi, forward, backward algorithms?

7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\underline{\boxed{z_i}}$.

2) The value of $z_i$ only depends on ______.

3) What symbol would you use to denote the transition probability from state 2 to state 0?

4) What symbol would you use to denote the probability of emitting a 1 from state 3?

5) What symbol would you use to denote the probability of starting in state 2?

6) What is the runtime of the Viterbi, forward, backward algorithms?

7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\underline{\boxed{z_i}}$.
2) The value of $z_i$ only depends on $\underline{\boxed{z_{i-1}}}$.
3) What symbol would you use to denote the transition probability from state 2 to state 0?
4) What symbol would you use to denote the probability of emitting a 1 from state 3?
5) What symbol would you use to denote the probability of starting in state 2?
6) What is the runtime of the Viterbi, forward, backward algorithms?
7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\boxed{z_i}$.

2) The value of $z_i$ only depends on $\boxed{z_{i-1}}$.

3) What symbol would you use to denote the transition probability from state 2 to state 0? $\boxed{a_{2,0}}$

4) What symbol would you use to denote the probability of emitting a 1 from state 3?

5) What symbol would you use to denote the probability of starting in state 2?

6) What is the runtime of the Viterbi, forward, backward algorithms?

7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\boxed{z_i}$.
2) The value of $z_i$ only depends on $\boxed{z_{i-1}}$.
3) What symbol would you use to denote the transition probability from state 2 to state 0? $\boxed{a_{2,0}}$
4) What symbol would you use to denote the probability of emitting a 1 from state 3? $\boxed{e_3(1)}$
5) What symbol would you use to denote the probability of starting in state 2?
6) What is the runtime of the Viterbi, forward, backward algorithms?
7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\boxed{z_i}$.
2) The value of $z_i$ only depends on $\boxed{z_{i-1}}$.
3) What symbol would you use to denote the transition probability from state 2 to state 0? $\boxed{a_{2,0}}$
4) What symbol would you use to denote the probability of emitting a 1 from state 3? $\boxed{e_3(1)}$
5) What symbol would you use to denote the probability of starting in state 2? $\boxed{\pi_2}$
6) What is the runtime of the Viterbi, forward, backward algorithms?
7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\boxed{z_i}$.

2) The value of $z_i$ only depends on $\boxed{z_{i-1}}$.

3) What symbol would you use to denote the transition probability from state 2 to state 0? $\boxed{a_{2,0}}$

4) What symbol would you use to denote the probability of emitting a 1 from state 3? $\boxed{e_3(1)}$

5) What symbol would you use to denote the probability of starting in state 2? $\boxed{\pi_2}$

6) What is the runtime of the Viterbi, forward, backward algorithms?

   $O(K^2L)$, where $K$=number of hidden states, $L$=length of sequence

7) What is the runtime of obtaining the posterior decoding and posterior mean?

# HMM informal quiz: discuss with a partner

1) The value of $x_i$ only depends on $\boxed{z_i}$.

2) The value of $z_i$ only depends on $\boxed{z_{i-1}}$.

3) What symbol would you use to denote the transition probability from state 2 to state 0? $\boxed{a_{2,0}}$

4) What symbol would you use to denote the probability of emitting a 1 from state 3? $\boxed{e_3(1)}$

5) What symbol would you use to denote the probability of starting in state 2? $\boxed{\pi_2}$

6) What is the runtime of the Viterbi, forward, backward algorithms?

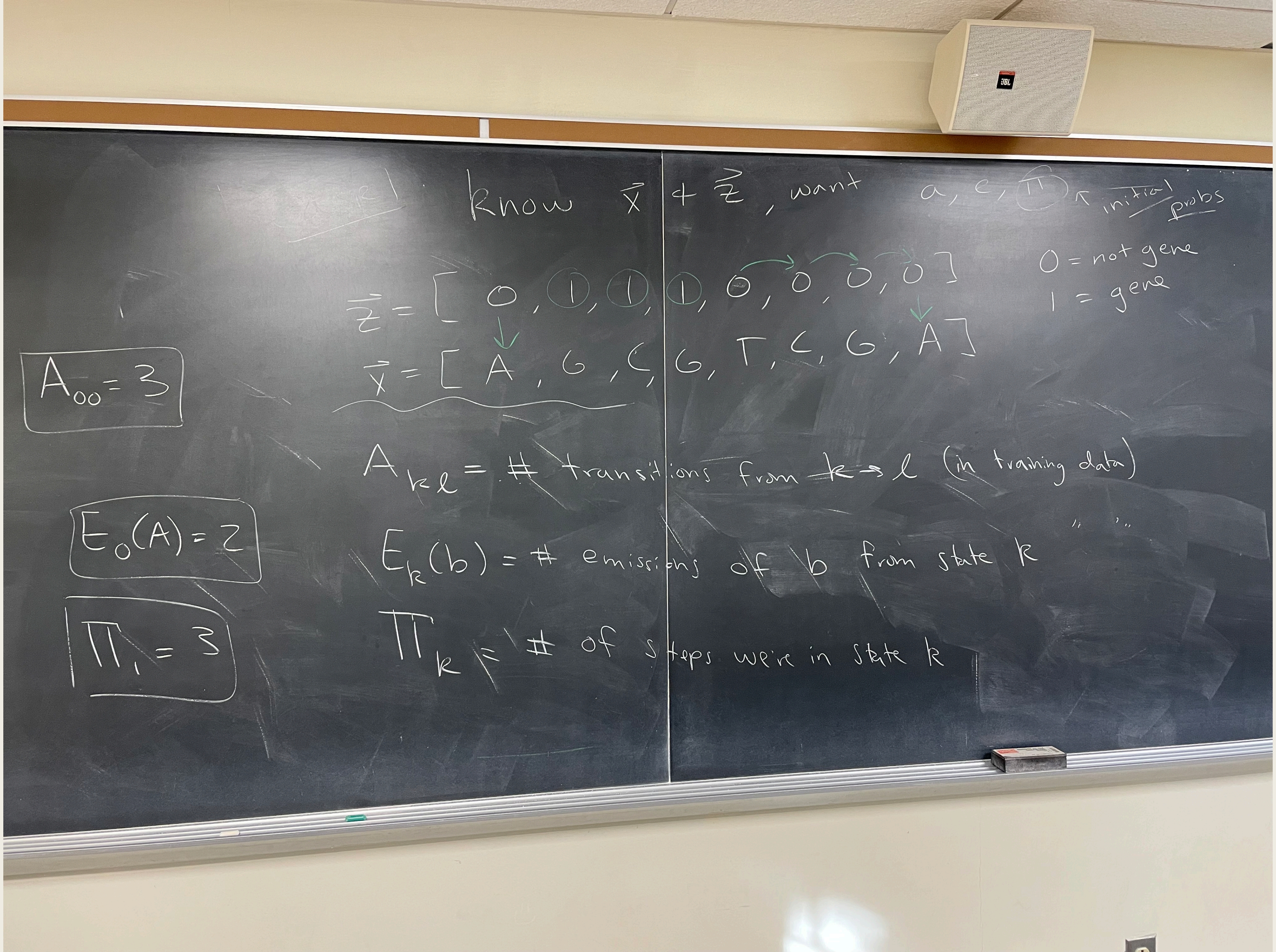
   *$O(K^2L)$,* where *K*=number of hidden states, *L*=length of sequence

7) What is the runtime of obtaining the posterior decoding and posterior mean?

   *$O(KL)$,* need to compute posterior probability for all *K* states, along the sequence

$$P(z_i = k|\vec{x}) = \frac{f_k(i) \cdot b_k(i)}{P(\vec{x})}$$

# Parameter Estimation for HMMs: Baum-Welch algorithm



know x + z, want a, e, π ← initial probs
0 = not gene
1 = gene
z = [0, 1, 1, 1, 0, 0, 0, 0]
x = [A, G, C, G, T, C, G, A]
A00 = 3
E0(A) = 2
π1 = 3
Akl = # transitions from k → l (in training data)
Ek(b) = # emissions of b from state k " "
πk = # of steps we're in state k

find prob

$L = 8$

$\rightarrow \boxed{\pi_0 = \frac{5}{8}, \ \pi_1 = \frac{3}{8}}$

initial

$$\pi_k = \frac{\Pi_k}{\sum_{k'} \Pi_{k'}}$$

$\rightarrow$ len of seq $L$

transition

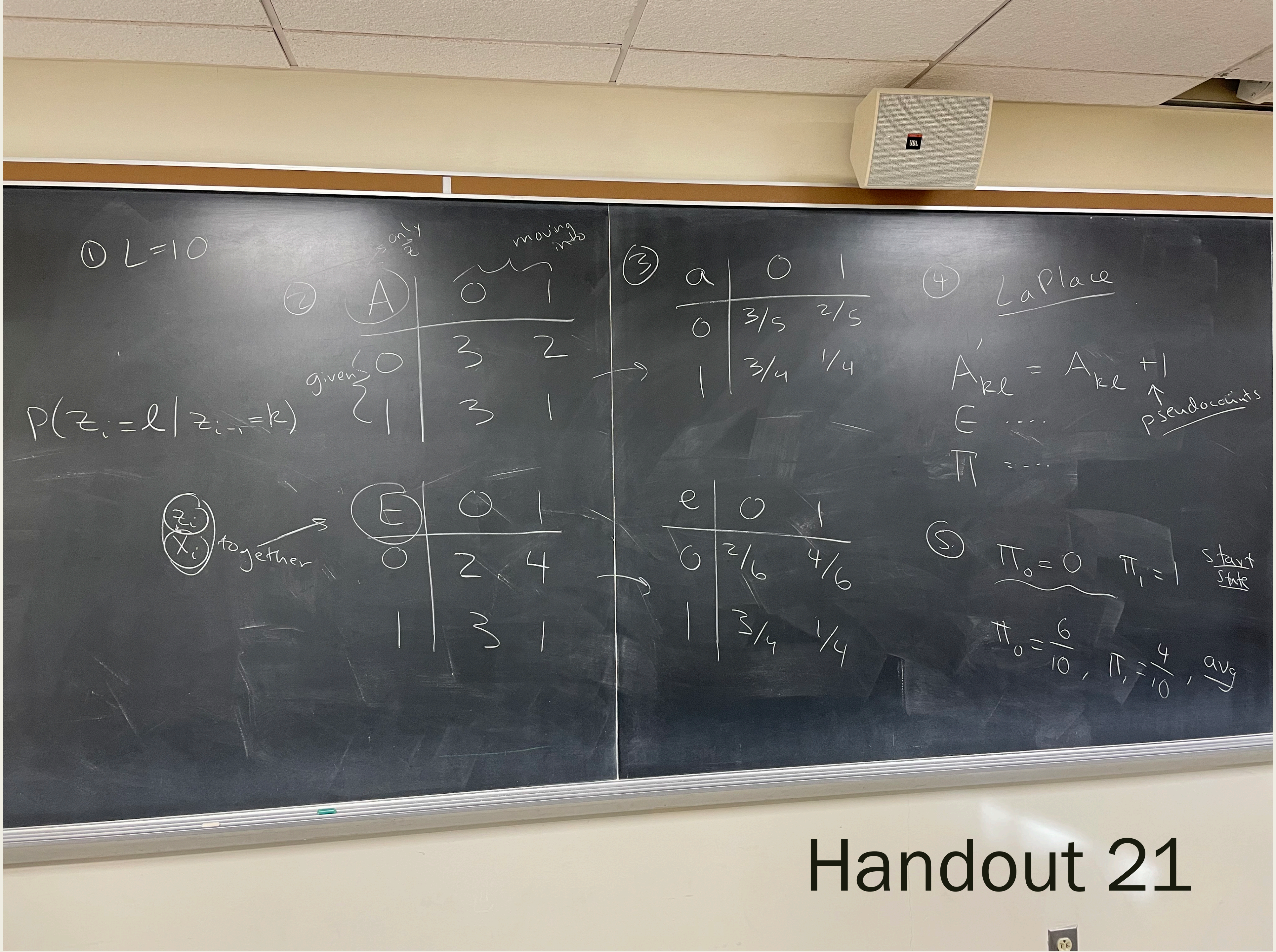
$$a_{kl} = \frac{A_{kl}}{\sum_{l'} A_{kl'}}$$

$p(z_i = l \mid z_{i-1} = k)$

emission

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

$\Pi_k$



① L=10
only
moving into
②
A 0 1
given
0 3 2
1 3 1
P(z_i = l | z_{i-1} = k)
z_i
X_i
together
E 0 1
0 2 4
1 3 1
③
a 0 1
0 3/5 2/5
1 3/4 1/4
e 0 1
0 2/6 4/6
1 3/4 1/4
④ LaPlace
A'_kl = A_kl +1
pseudocounts
E ....
π ....
⑤ π_0 = 0 π_1 = 1
start state
π_0 = 6/10, π_1 = 4/10, avg


# Handout 21

# Parameter estimation for HMMs

$$A_{kl} = \#\text{ expected number of transitions between state } k \text{ and } l$$

$$E_k(b) = \#\text{ expected number of emissions of } b \text{ from state } k$$

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

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

# Parameter estimation for HMMs

$$A_{kl} = \#\text{ expected number of transitions between state } k \text{ and } l$$

$$E_k(b) = \#\text{ expected number of emissions of } b \text{ from state } k$$

- For either Case 1 (state sequence known) or Case 2 (state sequence unknown), we need a way of computing the expected number of times we use a specific transition or emission

- For Case 1, we can observe these counts directly

- For Case 2, we will obtain these expected counts by adding up the probability of using each transition and emission across the entire sequence

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

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

# What about the start state?

1) Count how many times we observe each state throughout the sequence, then normalize

2) See how many times (or how many expected times) we actually start in each state (this is assuming many independent sequences)

# Case 2: state sequence unknown

- Compute the expected transition and emission counts using the forward and backward probabilities:

$$P(z_i = k, z_{i+1} = l|\vec{x}) = \frac{f_k(i) \cdot a_{kl} \cdot e_l(x_{i+1}) \cdot b_l(i+1)}{P(\vec{x})}$$

$$A_{kl} = \sum_{i=1}^{L-1} P(z_i = k, z_{i+1} = l|\vec{x})$$

Expected transition counts

# Case 2: state sequence unknown

- Compute the expected transition and emission counts using the forward and backward probabilities:

$$P(z_i = k, z_{i+1} = l|\vec{x}) = \frac{f_k(i) \cdot a_{kl} \cdot e_l(x_{i+1}) \cdot b_l(i+1)}{P(\vec{x})}$$

$$A_{kl} = \sum_{i=1}^{L-1} P(z_i = k, z_{i+1} = l|\vec{x})$$

Expected transition counts

Expected emission counts

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

# Case 2: state sequence unknown

- Compute the expected transition and emission counts using the forward and backward probabilities:

$$P(z_i = k, z_{i+1} = l|\vec{x}) = \frac{f_k(i) \cdot a_{kl} \cdot e_l(x_{i+1}) \cdot b_l(i+1)}{P(\vec{x})}$$

$$A_{kl} = \sum_{i=1}^{L-1} P(z_i = k, z_{i+1} = l|\vec{x})$$

Expected transition counts

Expected emission counts

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

- We still need to normalize as before: $$a_{kl} = \frac{A_{kl}}{\sum_{l'} A_{kl'}}$$ $$e_k(b) = \frac{E_k(b)}{\sum_{b'} E_k(b')}$$

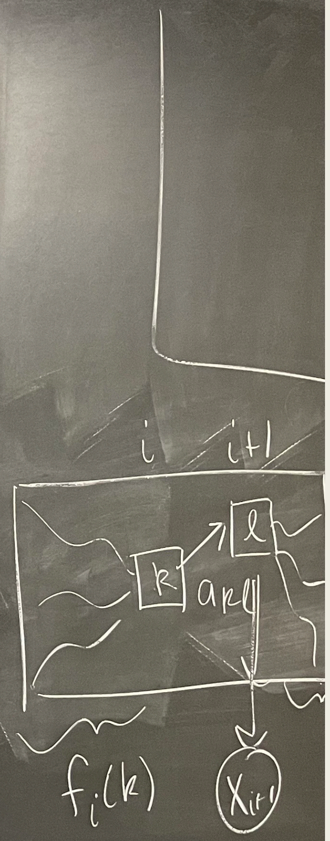
Case 2 only know $\vec{x}$ want: $\underbrace{a, e, \pi}_{\text{guess!}}$ + $\vec{z}$

$A_{kl}$ first

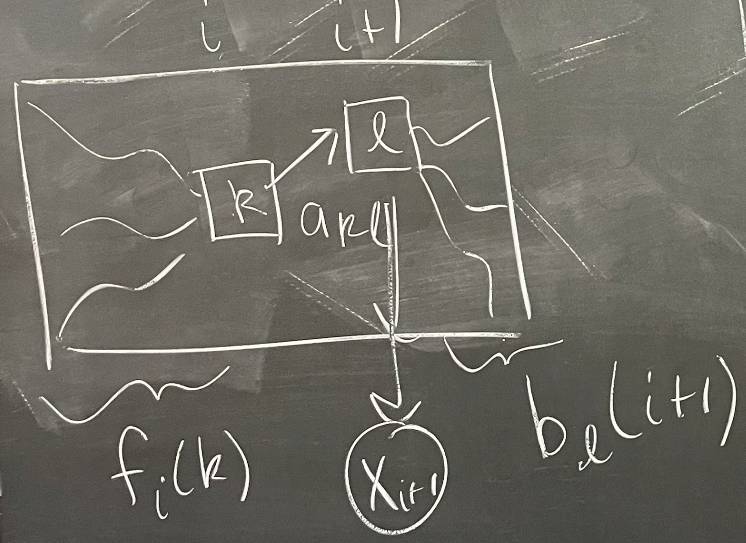
$$P(z_i = k, z_{i+1} = l \mid \vec{x}) = \frac{f_k(i)\, a_{kl}\, e_l(i+1)\, b_l(i+1)}{P(\vec{x})}$$

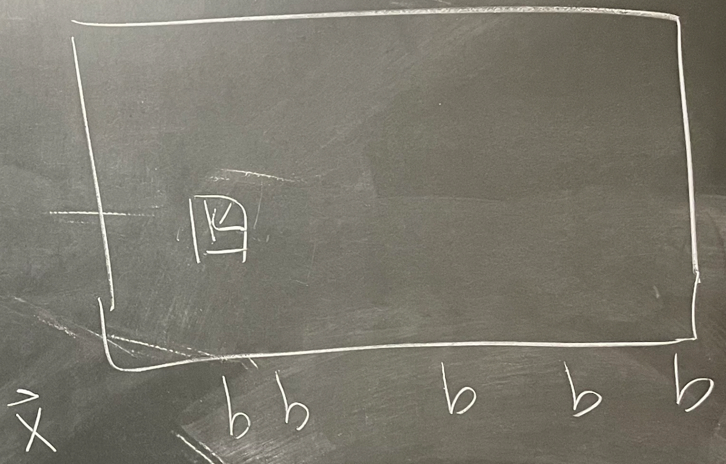
$$\boxed{A_{kl} = \sum_i P(z_i = k, z_{i+1} = l \mid \vec{x})}$$

"count"



$$E_k(b) = \sum_{\{i \mid x_i = b\}} \frac{f_k(i)\, b_k(i)}{P(\vec{x})}$$



# Baum-Welch Algorithm (1960's)

**Initialization**

Choose model parameters (***a*** & ***e*** matrices, $\boldsymbol{\pi}$ vector) arbitrarily

If we have some prior knowledge of the problem, use that to initialize parameters

# Baum-Welch Algorithm (1960's)

Initialization

Choose model parameters (***a*** & ***e*** matrices, $\boldsymbol{\pi}$ vector) arbitrarily

If we have some prior knowledge of the problem, use that to initialize parameters

Iteration

E-step

- Run forward algorithm to get $f_k(i)$ p. 59
- Run backward algorithm to get $b_k(i)$ p. 60
- Compute expected transition/emission counts $A_{kl}$, $E_k(b)$ Equations 3.20 & 3.21

# Baum-Welch Algorithm (1960's)

**Initialization**

Choose model parameters (***a*** & ***e*** matrices, $\boldsymbol{\pi}$ vector) arbitrarily
If we have some prior knowledge of the problem, use that to initialize parameters

**Iteration**

E-step

- Run forward algorithm to get $f_k(i)$ p. 59
- Run backward algorithm to get $b_k(i)$ p. 60
- Compute expected transition/emission counts $A_{kl}$, $E_k(b)$ Equations 3.20 & 3.21

M-step

- Maximize likelihood $P(\mathbf{x})$, given expected counts
- Transition and emission: $\longrightarrow$

$$a_{kl} = \frac{A_{kl}}{\sum_{l'} A_{kl'}} \qquad e_k(b) = \frac{E_k(b)}{\sum_{b'} E_k(b')}$$

Equation 3.18

- Initial probability update for Lab 8:

$$\pi_k = \frac{f_k(1) \cdot b_k(1)}{P(\vec{x})}$$

# Baum-Welch Algorithm (1960's)

**Initialization**

Choose model parameters (***a*** & ***e*** matrices, $\boldsymbol{\pi}$ vector) arbitrarily
If we have some prior knowledge of the problem, use that to initialize parameters

**Iteration**

E-step
- Run forward algorithm to get $f_k(i)$ p. 59
- Run backward algorithm to get $b_k(i)$ p. 60
- Compute expected transition/emission counts $A_{kl}$, $E_k(b)$ Equations 3.20 & 3.21

M-step
- Maximize likelihood $P(\mathbf{x})$, given expected counts
- Transition and emission: $\longrightarrow$

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

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

Equation 3.18

- Initial probability update for Lab 8:

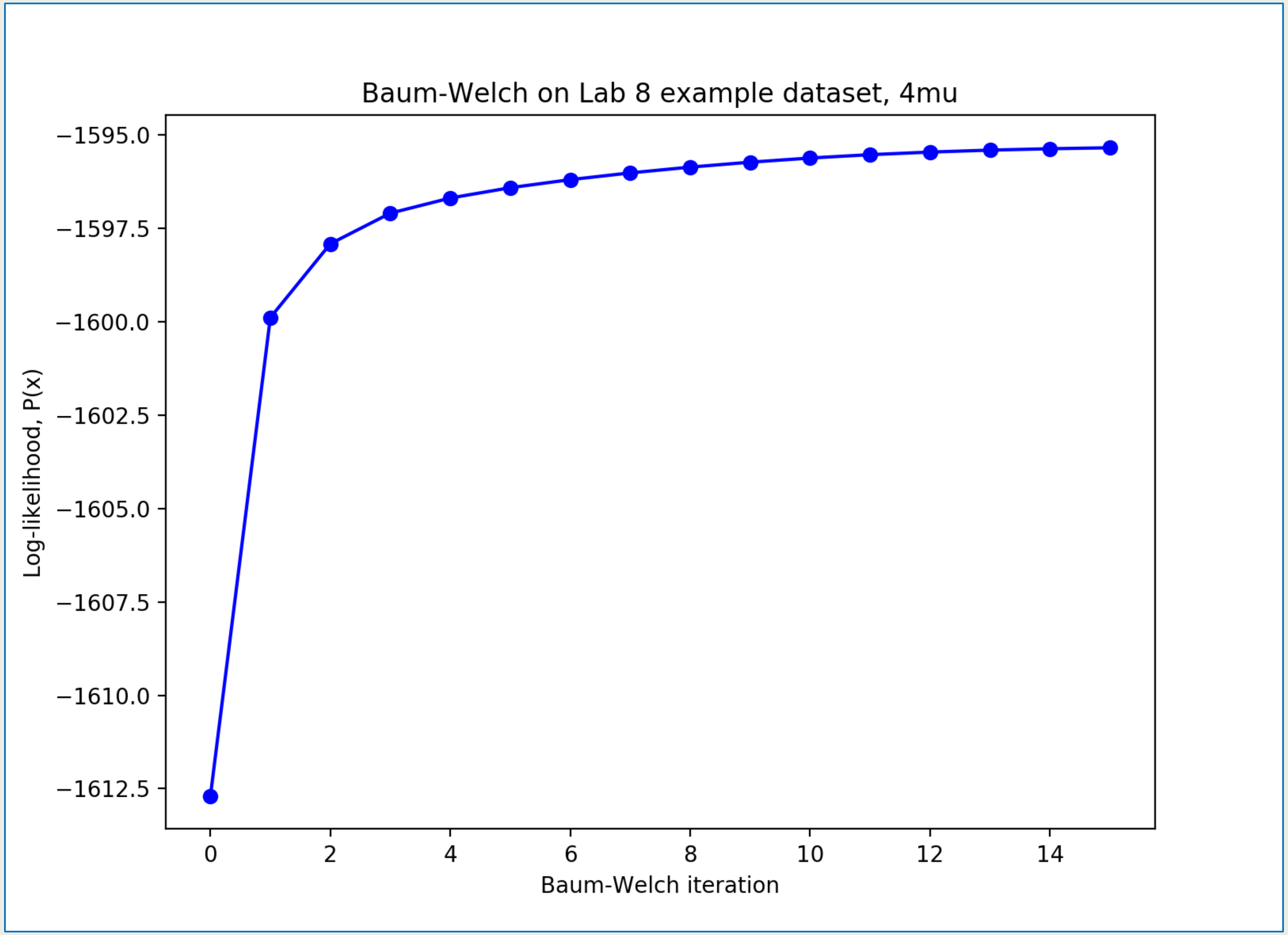
$$\pi_k = \frac{f_k(1) \cdot b_k(1)}{P(\vec{x})}$$

**Termination**

Stop when:
- Likelihood does not improve by much (choose threshold)
- Maximum number of iterations exceeded 15 iterations for us

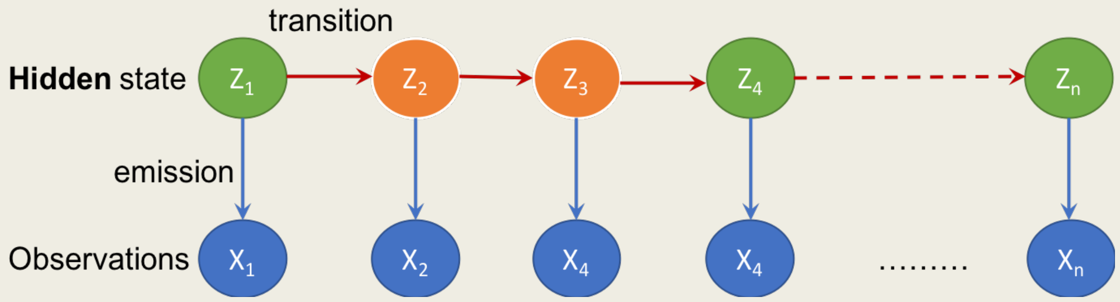
Note: in slides: $i$=1,2,...$L$
Python: $i$=0,1,...($L$-1)

# Log-likelihood improves with each Baum-Welch iteration



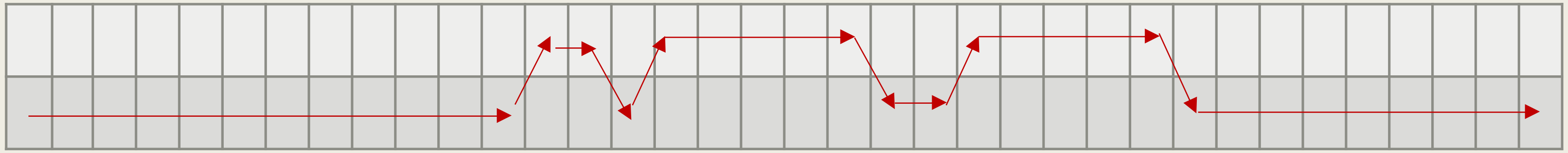
# Summary of HMMs

Structure of a Hidden Markov Model: Observations, hidden states, emissions, transitions

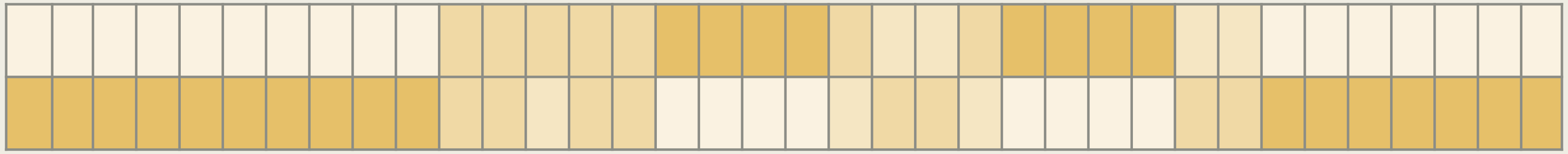



Three inference problems:

1) What is the most likely sequence of hidden states?: Viterbi algorithm



2) What is the most likely hidden state at each observation? Forward-Backward algorithm



3) What are the parameters? Baum-Welch algorithm