Genome 540 discussion

March 4th, 2025
Joe Min

Genome Sciences
UNIVERSITY OF WASHINGTON

Agenda

Homework 8 🥵

Homework 8

Homework 8 overview

Part 1:

- Make a 2-state HMM with given initial parameters

Part 2:

- Use the Baum-Welch method to find improved parameters for the HMM

Homework 8: part 1

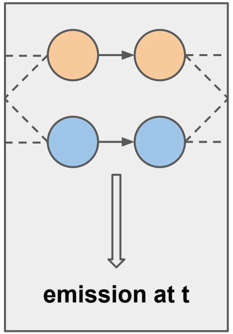
Making the initial HMM: definitions

π_i: initial probability of **state i**

a$_{ij}$: transition probability from **state i** to **state j**

b$_i$**(N)**: emission probability of **nucleotide N** from **state i**

t: the position of emission for the given node



notation borrowed from the Shen tutorial (also on website)

Making the initial HMM

State 1 is AT-rich

State 2 is GC-rich

Emission probabilities

	A	C	G	T
State 1	$b_1(A) = 0.3$	$b_1(C) = 0.2$	$b_1(G) = 0.2$	$b_1(T) = 0.3$
State 2	$b_2(A) = 0.15$	$b_2(C) = 0.35$	$b_2(G) = 0.35$	$b_2(T) = 0.15$

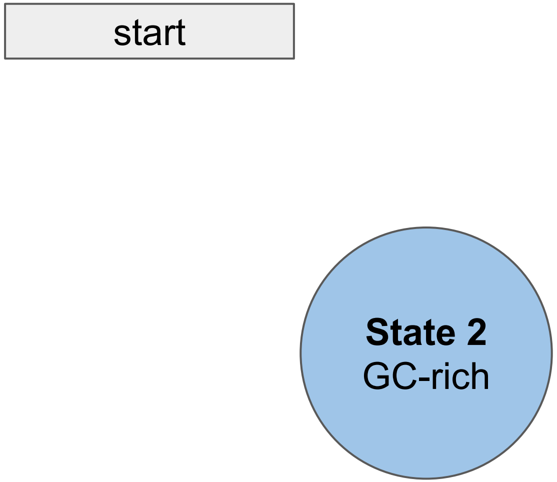
Initiation probabilities

- π_1: 0.996
- π_2: 0.004

Transition probabilities

	to State 1	to State 2
From State 1	$a_{11} = 0.999$	$a_{12} = 0.001$
From State 2	$a_{21} = 0.01$	$a_{22} = 0.99$

Visualizing the HMM

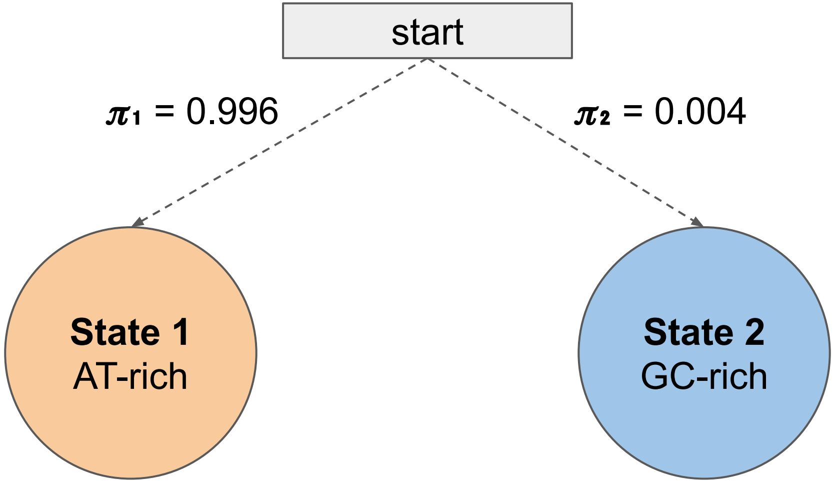


start

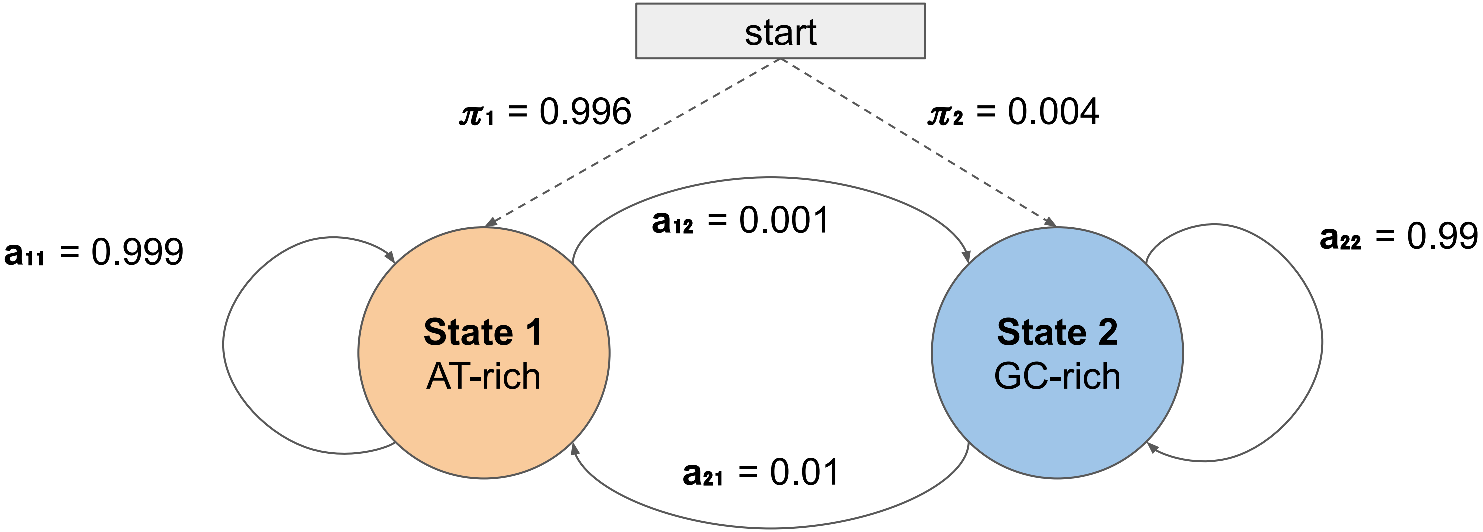
State 1
AT-rich

State 2
GC-rich

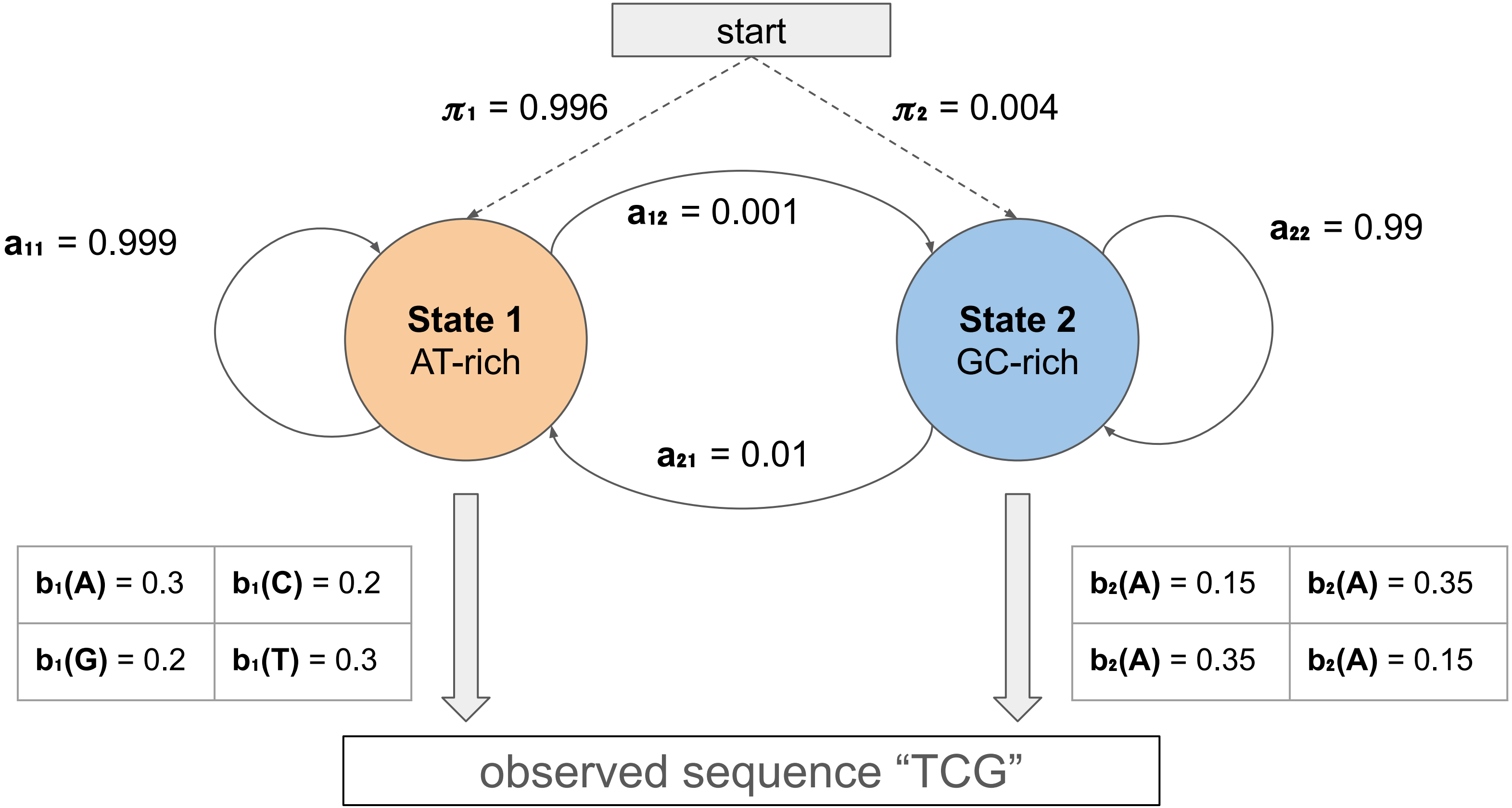
Visualizing the HMM



Visualizing the HMM



Visualizing the HMM



start

π_1 = 0.996

π_2 = 0.004

a_{11} = 0.999

a_{12} = 0.001

a_{22} = 0.99

a_{21} = 0.01

State 1
AT-rich

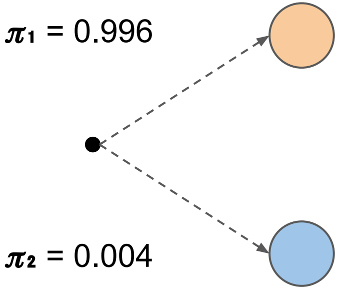
State 2
GC-rich

b_1(A) = 0.3	**b_1(C)** = 0.2
b_1(G) = 0.2	**b_1(T)** = 0.3

b_2(A) = 0.15	**b_2(A)** = 0.35
b_2(A) = 0.35	**b_2(A)** = 0.15

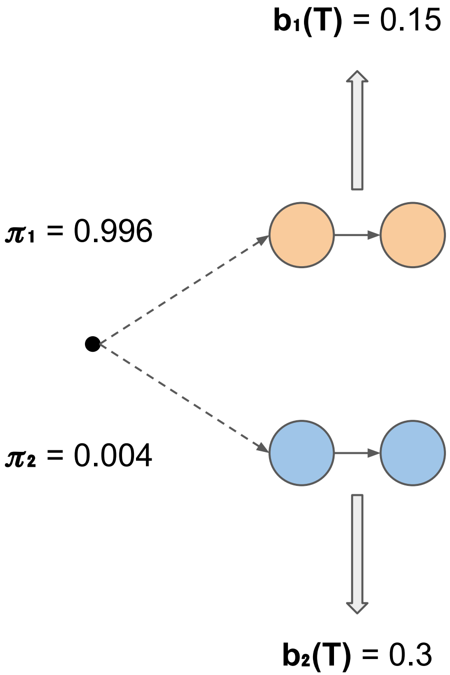
observed sequence “TCG”

Linearized visualization for specific sequence



observed sequence “TCG”

Linearized visualization for specific sequence



$\mathbf{b_1(T)} = 0.15$

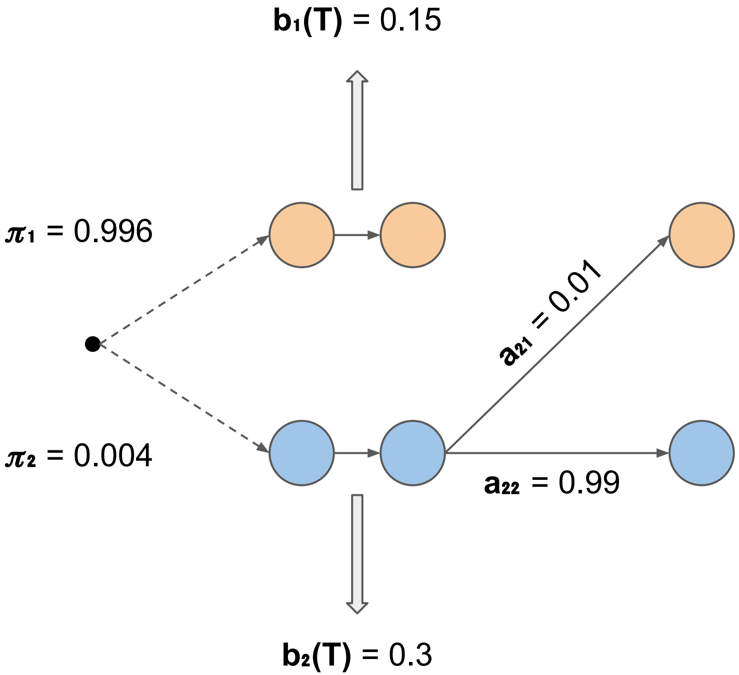
$\pi_1 = 0.996$

$\pi_2 = 0.004$

$\mathbf{b_2(T)} = 0.3$

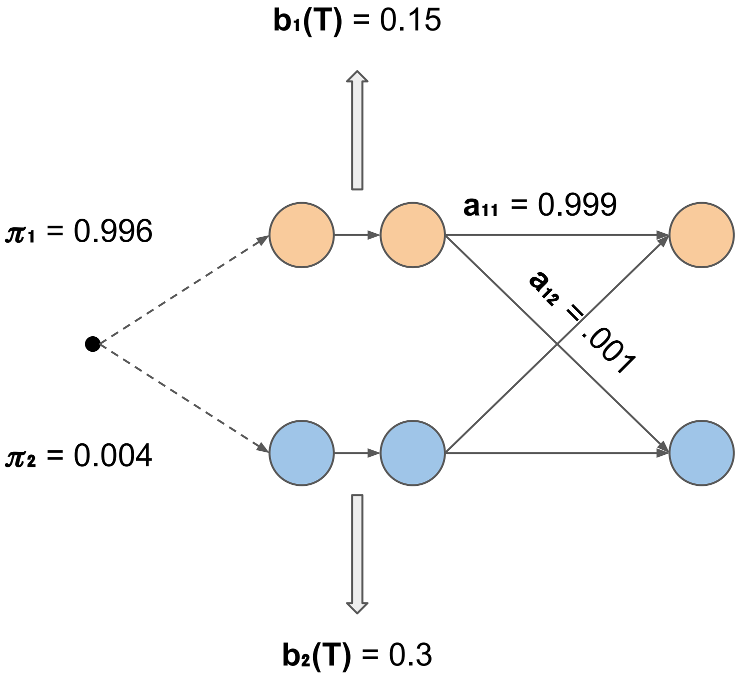
observed sequence “TCG”

Linearized visualization for specific sequence



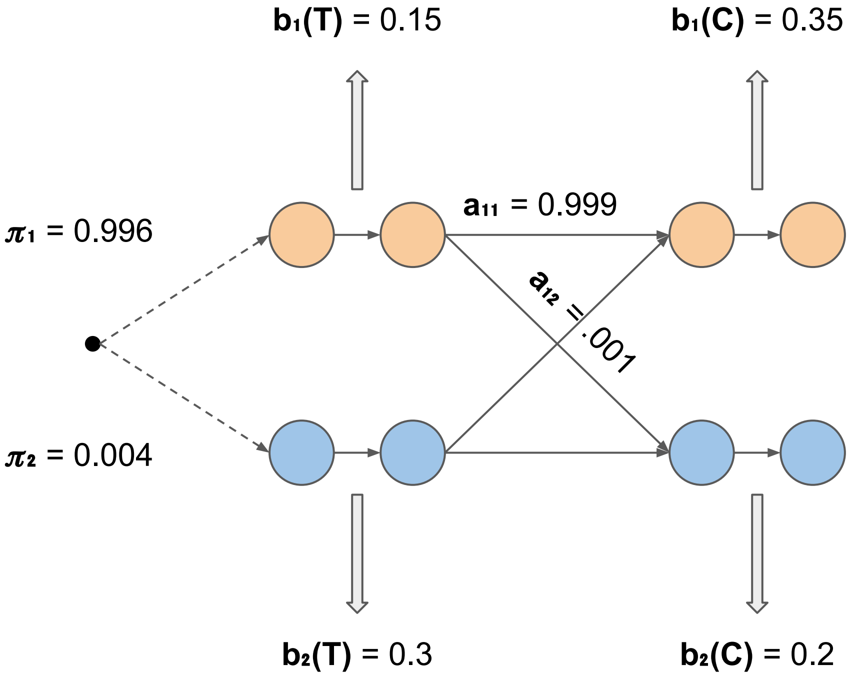
observed sequence “TCG”

Linearized visualization for specific sequence



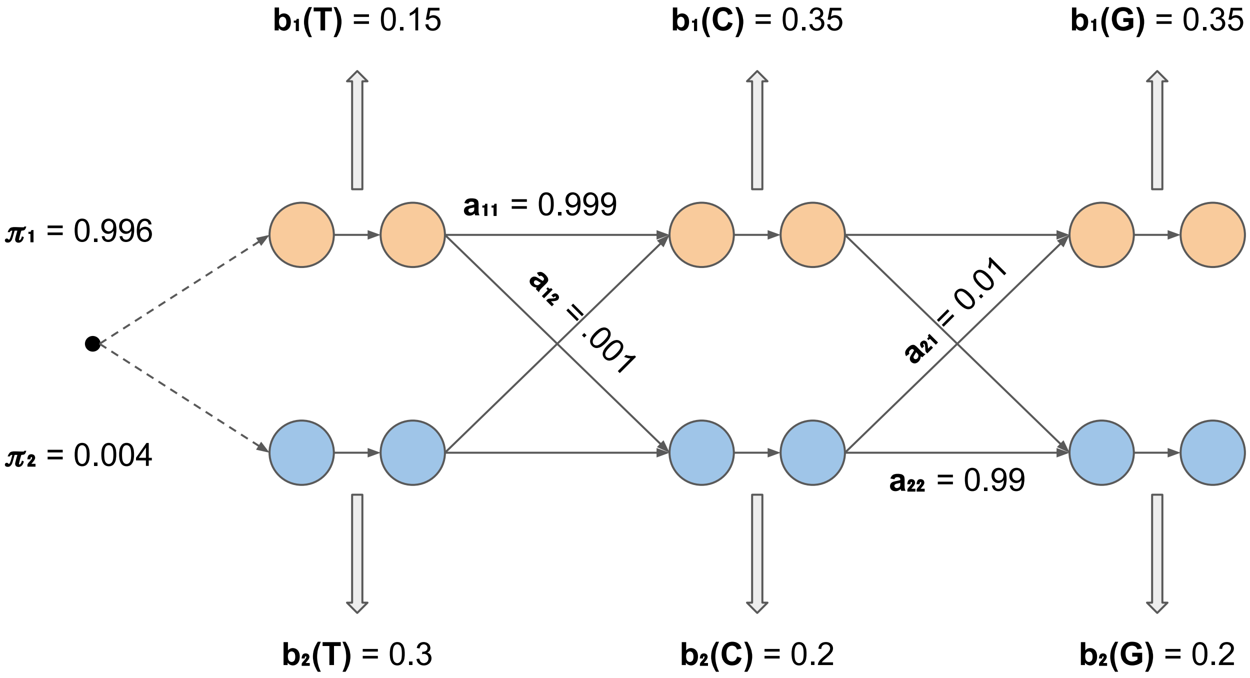
observed sequence “TCG”

Linearized visualization for specific sequence



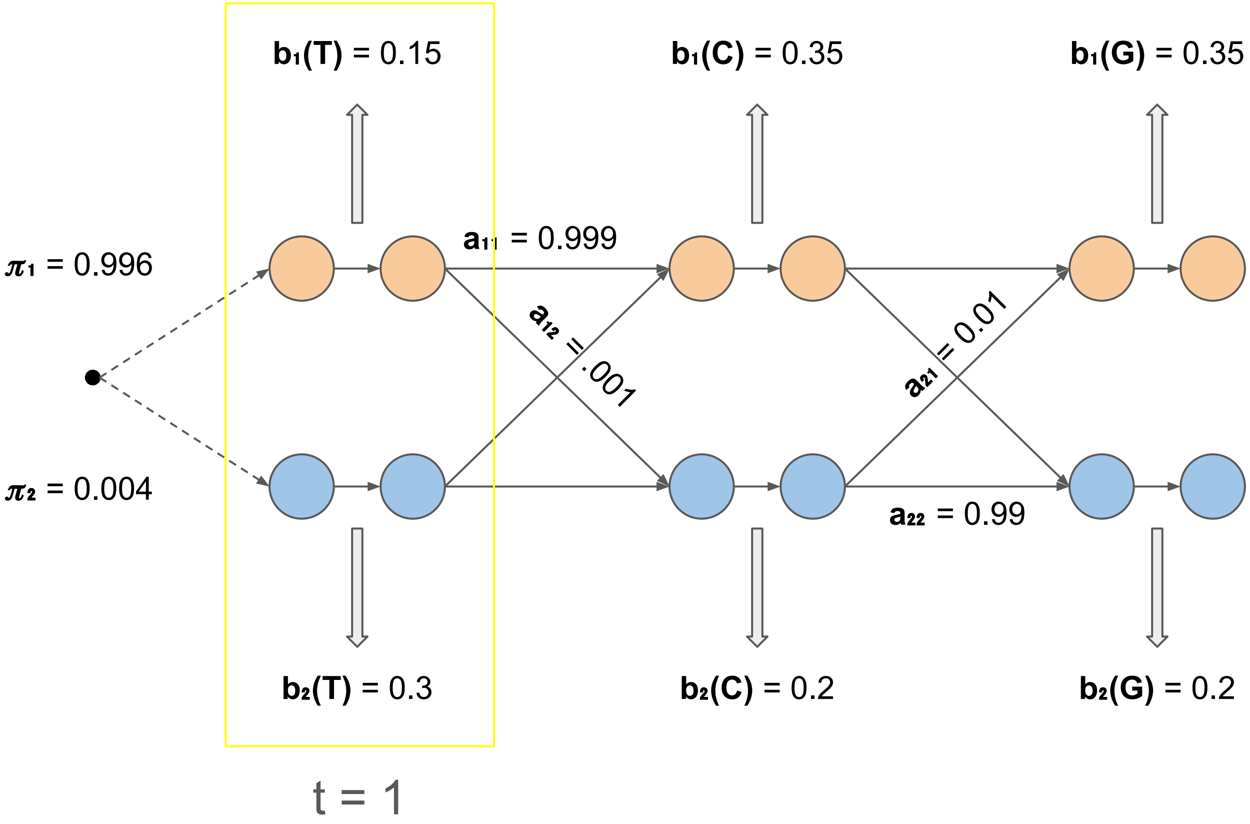
observed sequence “TCG”

Linearized visualization for specific sequence

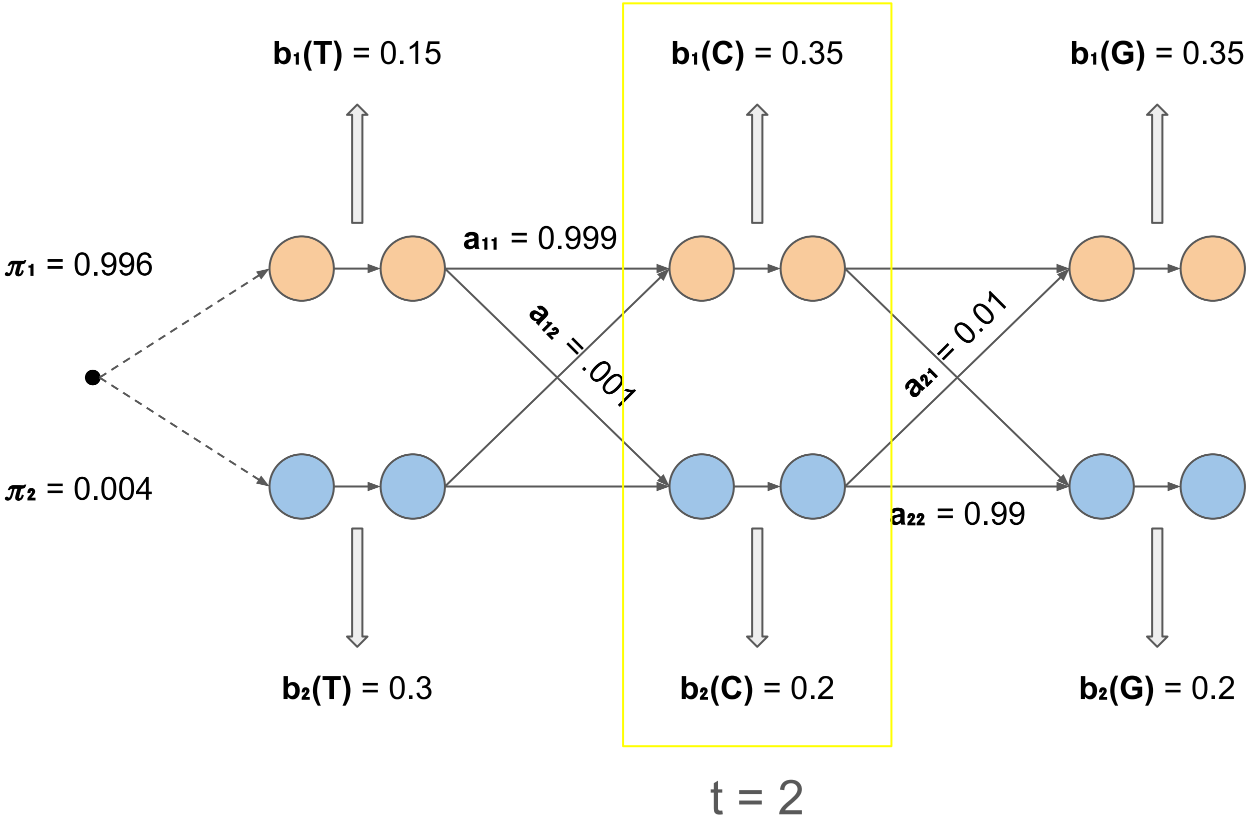


observed sequence “TCG”

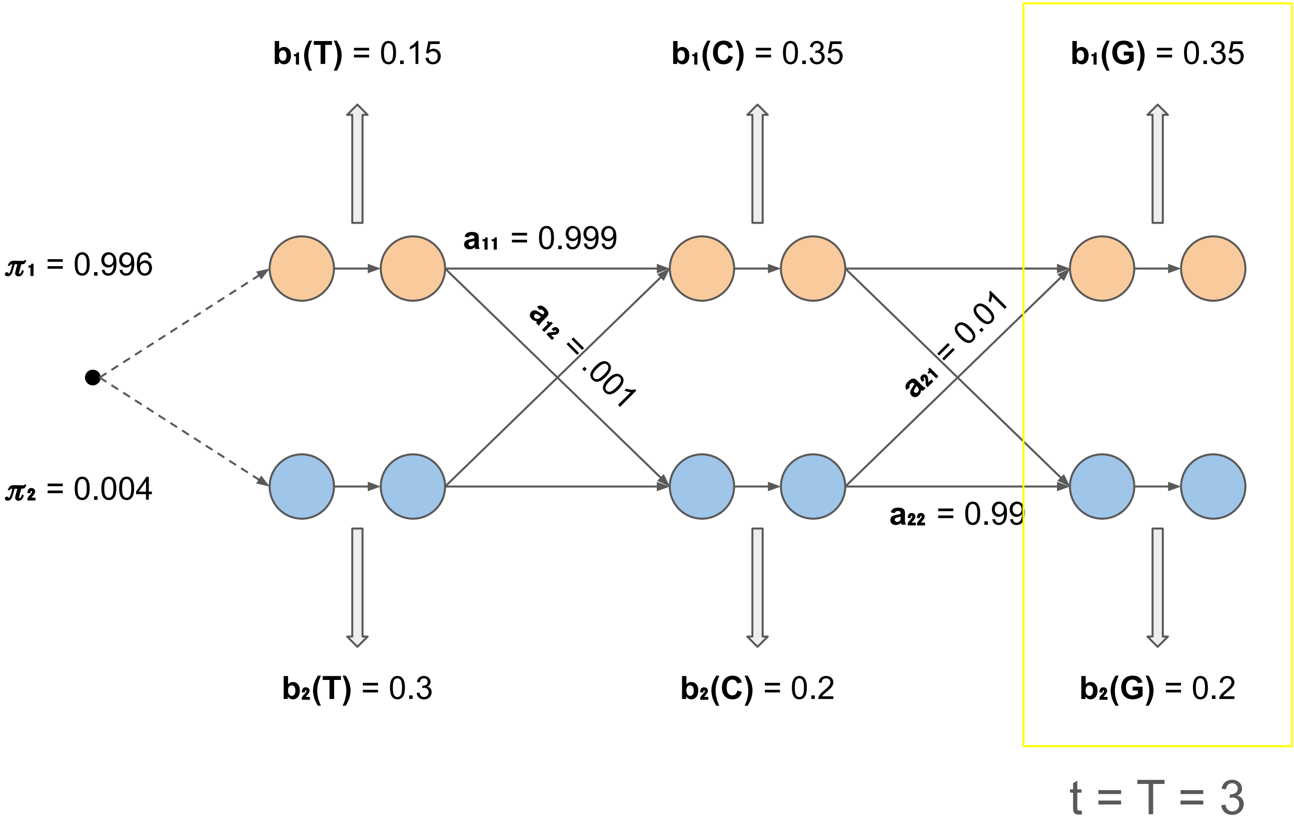
Linearized visualization for specific sequence



Linearized visualization for specific sequence



Linearized visualization for specific sequence



Baum-Welch example

Baum-Welch will iteratively adjust parameters of the HMM to maximize the likelihood of the observed sequence

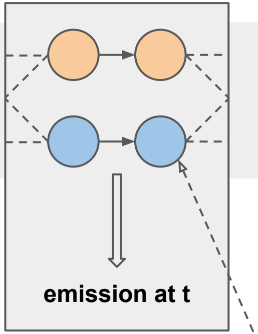
- First, calculate forward and backward variables using the forward/backward algorithm
- Using these values, update the HMM’s:
    - Emission probabilities
    - Transition probabilities

The forward part

Calculating forward values for all nodes is what we're used to! But with a twist:

- Use dynamic programming to find highest weight parent and weight of highest-weight path ending at each node
- The twist! For downstream reasons, keep things in the normal probability space (how??)

The forward part: more definitions



post-emission node

t: the observation position for a given node

***α_t*(i)**: the highest probability of the partial observed sequence (i.e., observations from t=1 to t=t) for the post-emission node for **state i** at **time t** (unscaled)

c_t: the scaling factor at **time t**

***α^*_t*(i)**: the scaled α_t(i)

Notation discrepancies with the tutorial:

- $\ddot{a}_t$(i) = α_t(i) ⟵ redefinition from tutorial (eq10)
- $\hat{a}_t$(i) = α^*_t(i) ⟵ no "alpha hat" in slides, so use α^* instead

The forward part: scaling

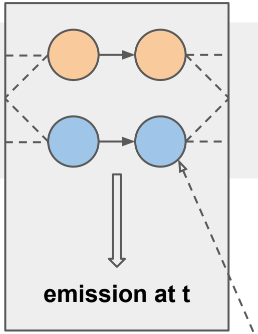
$\mathbf{c_t}$: the scaling factor at **time t**

- Note: independent of state! Because:

$$\mathbf{c_t} = \frac{1}{\sum_i \alpha_t(i)}$$

That is, $\mathbf{c_t}$ is the inverse of the sum of the α_t values across all states at **time t**

The forward part: more definitions



post-emission node

t: the observation position for a given node

$\boldsymbol{\alpha}_t$**(i)**: the highest probability of the partial observed sequence (i.e., observations from t=1 to t=t) for the post-emission node for **state i** at **time t** (unscaled)

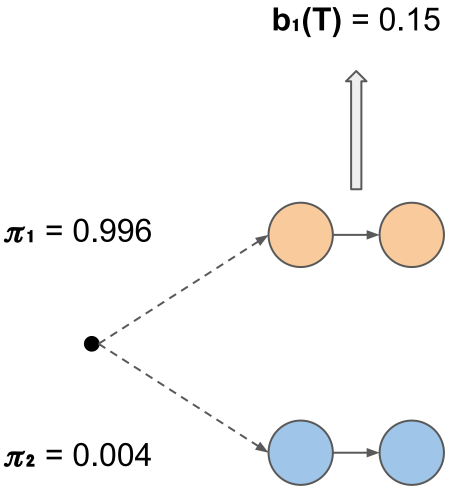
c$_t$: the scaling factor at **time t**

$\boldsymbol{\alpha}^*_t$**(i)**: $c_t\alpha_t(i)$ (scaled probability)

Notation discrepancies with the tutorial:

- $\ddot{a}_t(i) = \alpha_t(i)$ ⟵ redefinition from tutorial (eq10)
- $\hat{a}_t(i) = \alpha^*_t(i)$ ⟵ no "alpha hat" in slides, so use α^* instead

The forward part: calculating t=1



$\mathbf{b_1(T)} = 0.15$

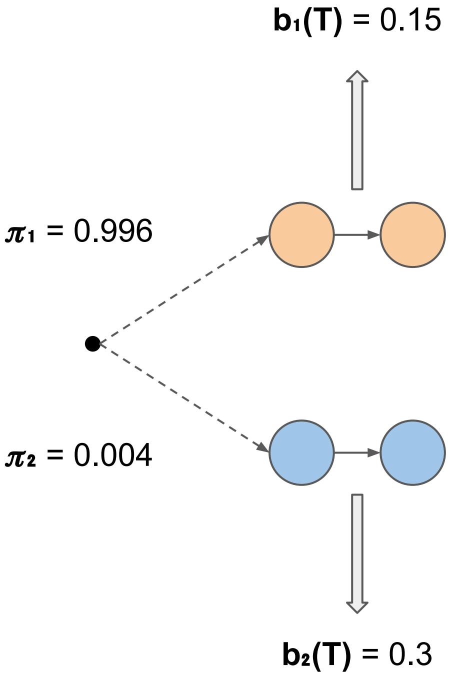
$\pi_1 = 0.996$

$\pi_2 = 0.004$

$\boldsymbol{\alpha_1(1)} = \pi_1 * b_1(T) = 0.996 * 0.15 = \mathbf{0.1494}$

observed sequence “TCG”

The forward part: calculating t=1



$\mathbf{b_1(T)}$ = 0.15

$\boldsymbol{\pi_1}$ = 0.996

$\boldsymbol{\pi_2}$ = 0.004

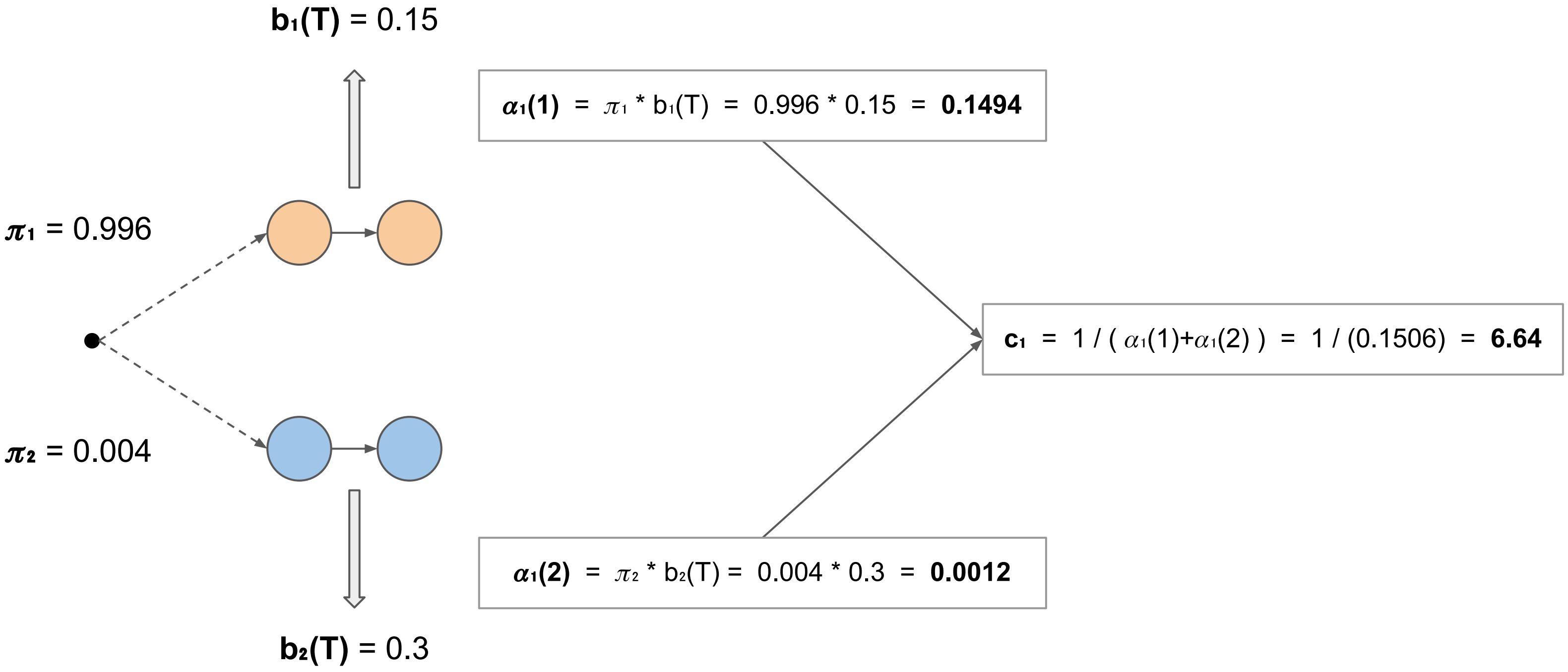
$\mathbf{b_2(T)}$ = 0.3

$\boldsymbol{\alpha_1(1)} = \pi_1 * b_1(T) = 0.996 * 0.15 = \mathbf{0.1494}$

$\boldsymbol{\alpha_1(2)} = \pi_2 * b_2(T) = 0.004 * 0.3 = \mathbf{0.0012}$

observed sequence “TCG”

The forward part: calculating t=1



$\mathbf{b_1(T)} = 0.15$

$\boldsymbol{\pi_1} = 0.996$

$\boldsymbol{\pi_2} = 0.004$

$\mathbf{b_2(T)} = 0.3$

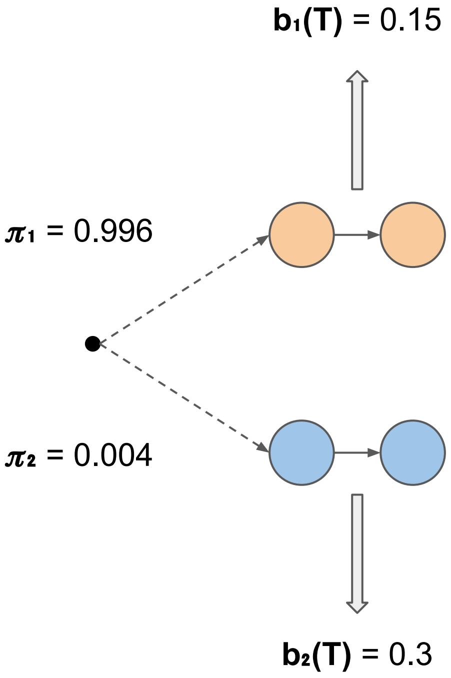
$\boldsymbol{\alpha_1(1)} = \pi_1 * b_1(T) = 0.996 * 0.15 = \mathbf{0.1494}$

$\boldsymbol{\alpha_1(2)} = \pi_2 * b_2(T) = 0.004 * 0.3 = \mathbf{0.0012}$

$\mathbf{c_1} = 1 / (\alpha_1(1)+\alpha_1(2)) = 1 / (0.1506) = \mathbf{6.64}$

observed sequence “TCG”

The forward part: calculating t=1



$\boldsymbol{\alpha_1(1)} = \pi_1 * b_1(T) = 0.996 * 0.15 = \mathbf{0.1494}$

$\boldsymbol{\alpha^*_1(1)} = c_1 * \alpha_1(1) = \mathbf{0.992}$

$\mathbf{c_1} = 1 / (\alpha_1(1)+\alpha_1(2)) = 1 / (0.1506) = \mathbf{6.64}$

$\boldsymbol{\alpha^*_1(2)} = c_1 * \alpha_1(2) = \mathbf{0.00797}$

$\boldsymbol{\alpha_1(2)} = \pi_2 * b_2(T) = 0.004 * 0.3 = \mathbf{0.0012}$

observed sequence “TCG”

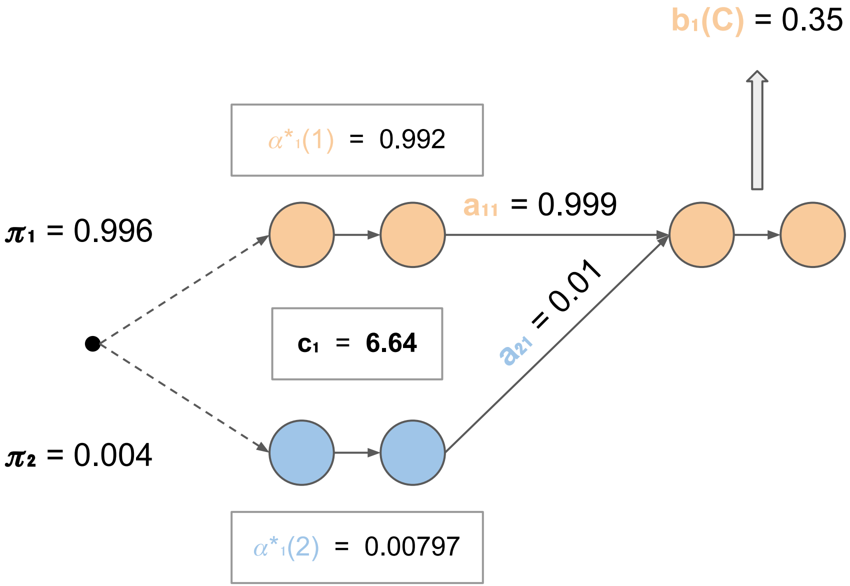
The forward part: t>1

Past t=1, we have to perform a sum to generate new values for $\alpha_t(i)$:

$$\alpha_{t+1}(j) = \textbf{[}\ \sum_i \alpha_t(i)*a_{ij}\ \textbf{]} * b_j(N_{+1})$$

This is the cumulative probability of all paths ending at **state j** at **time t+1** given the set of observed nucleotides

The forward part: t>1



$b_1(C) = 0.35$

$\alpha^*{}_1(1) = 0.992$

$a_{11} = 0.999$

$\pi_1 = 0.996$

$c_1 = \mathbf{6.64}$

$a_{21} = 0.01$

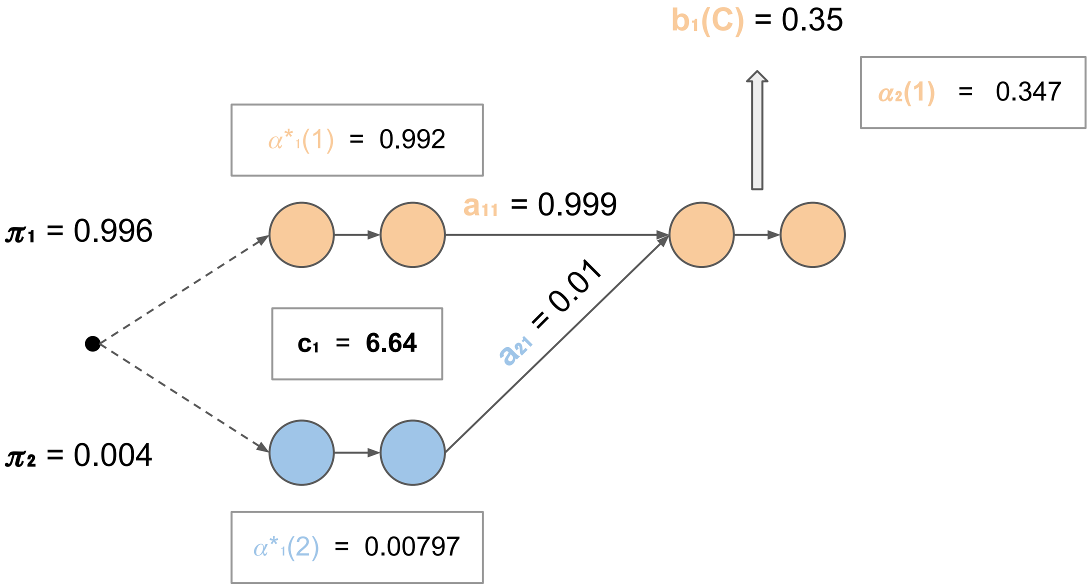
$\pi_2 = 0.004$

$\alpha^*{}_1(2) = 0.00797$

$$
\begin{aligned}
\boldsymbol{\alpha_2(1)} &= \text{sum}(\, a_{11} * \alpha_1{}^*(1) \,,\, a_{21} * \alpha_1{}^*(2)) * b_1(C) \\
&= \text{sum}(\, 0.9999 * 0.992 \,,\, 0.01 * 0.00797 \,) * 0.35 \\
&= 0.992 * 0.35 \\
&= \mathbf{0.347}
\end{aligned}
$$

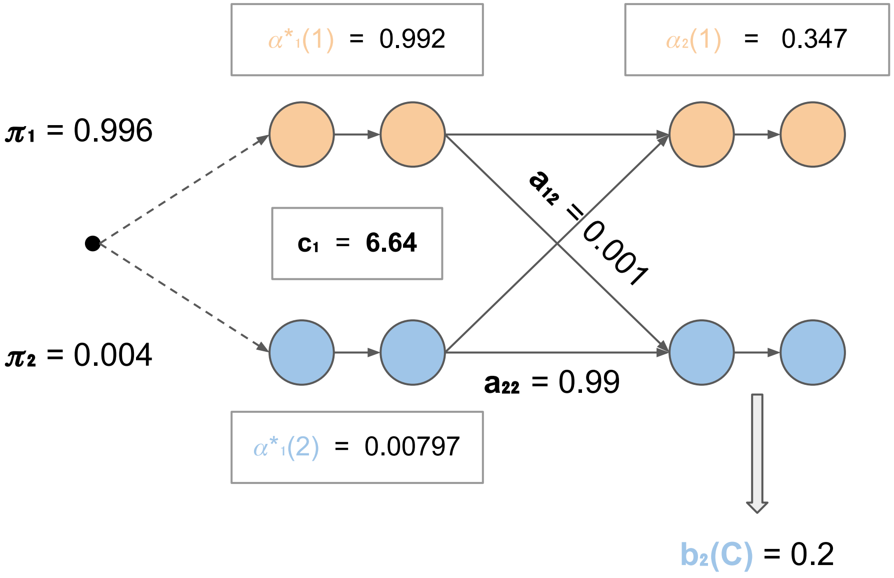
observed sequence “TCG”

The forward part: t>1



observed sequence “TCG”

The forward part: t>1



We can calculate $\alpha_2(2)$ the same as $\alpha_2(1)$

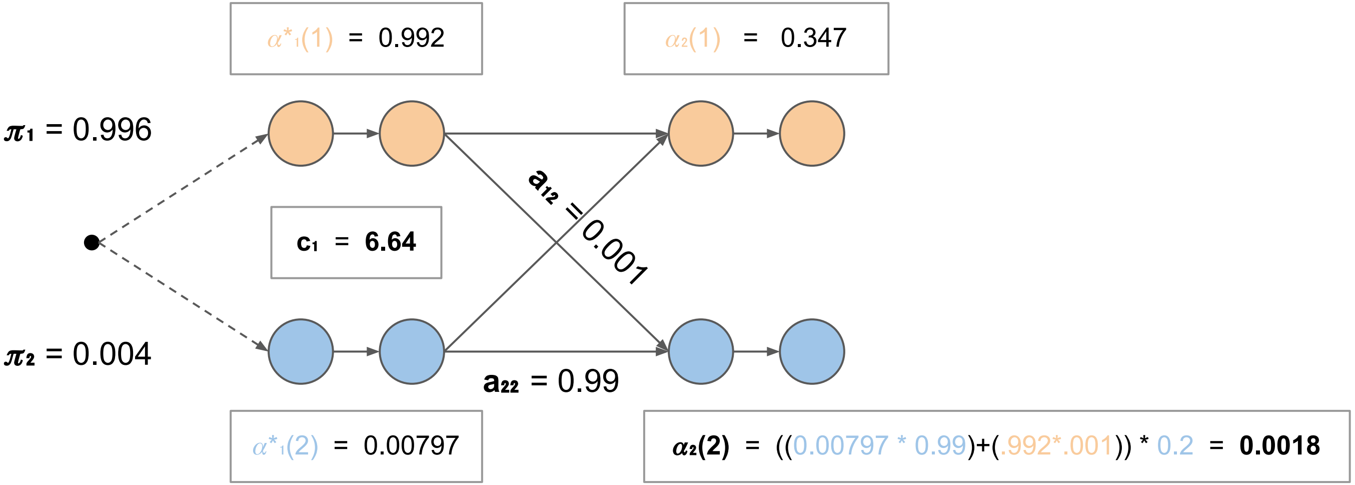
Then, we can calculate c_2

Which will allow us to calculate $\alpha^*_2(1)$ and $\alpha^*_2(1)$

Continue this procedure until we reach the end!

observed sequence “TCG”

The forward part: t>1



$\alpha^*_1(1) = 0.992$

$\alpha_2(1) = 0.347$

$\pi_1 = 0.996$

$c_1 = 6.64$

$a_{12} = 0.001$

$\pi_2 = 0.004$

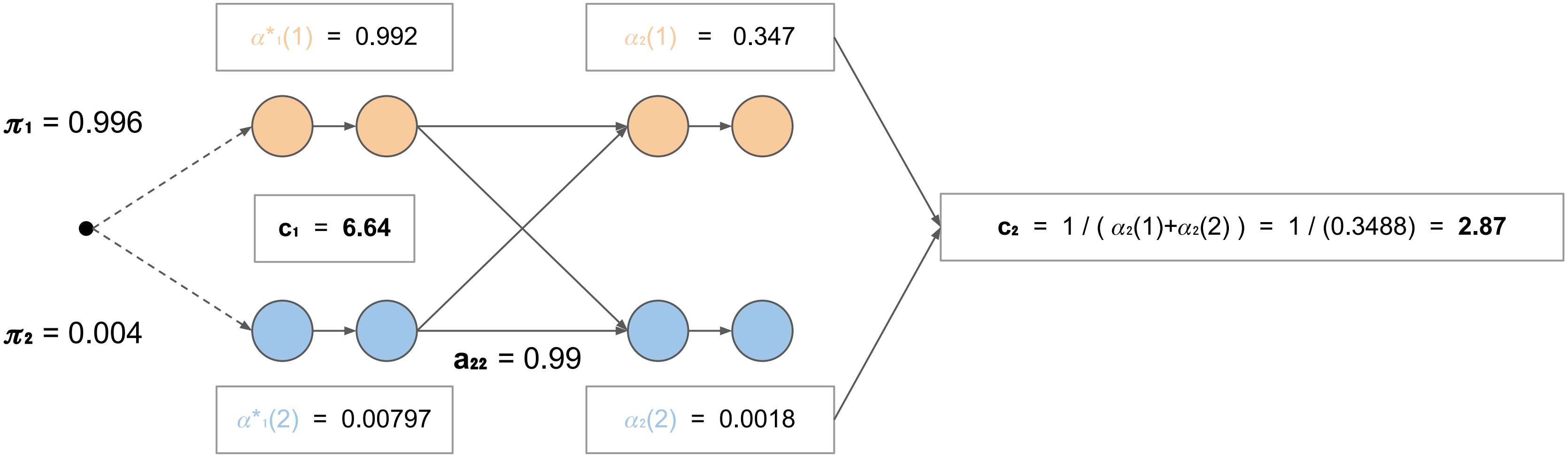
$a_{22} = 0.99$

$\alpha^*_1(2) = 0.00797$

$\alpha_2(2) = ((0.00797 * 0.99)+(.992*.001)) * 0.2 = 0.0018$

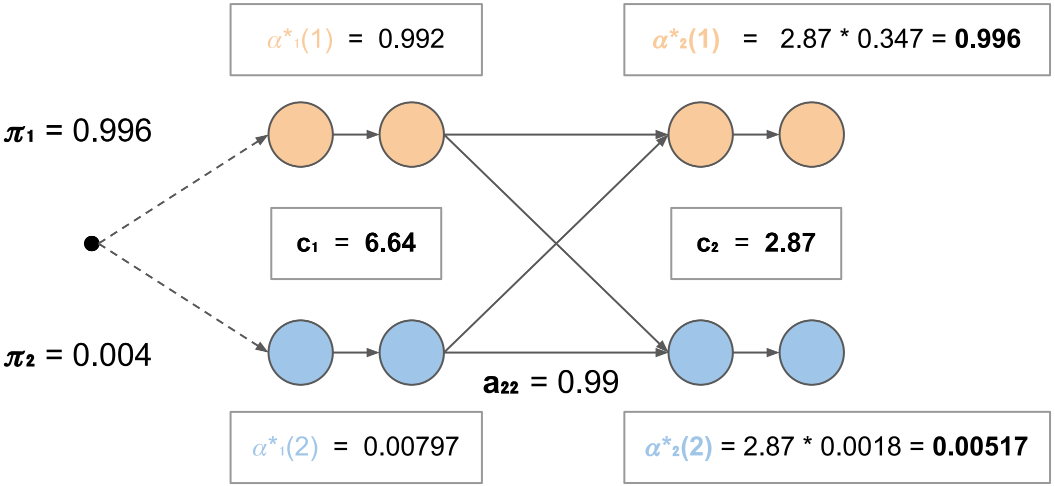
observed sequence “TCG”

The forward part: t>1



observed sequence “TCG”

The forward part: t>1



$\alpha^*_1(1)$ = 0.992

$\alpha^*_2(1)$ = 2.87 * 0.347 = **0.996**

π_1 = 0.996

c_1 = 6.64

c_2 = 2.87

π_2 = 0.004

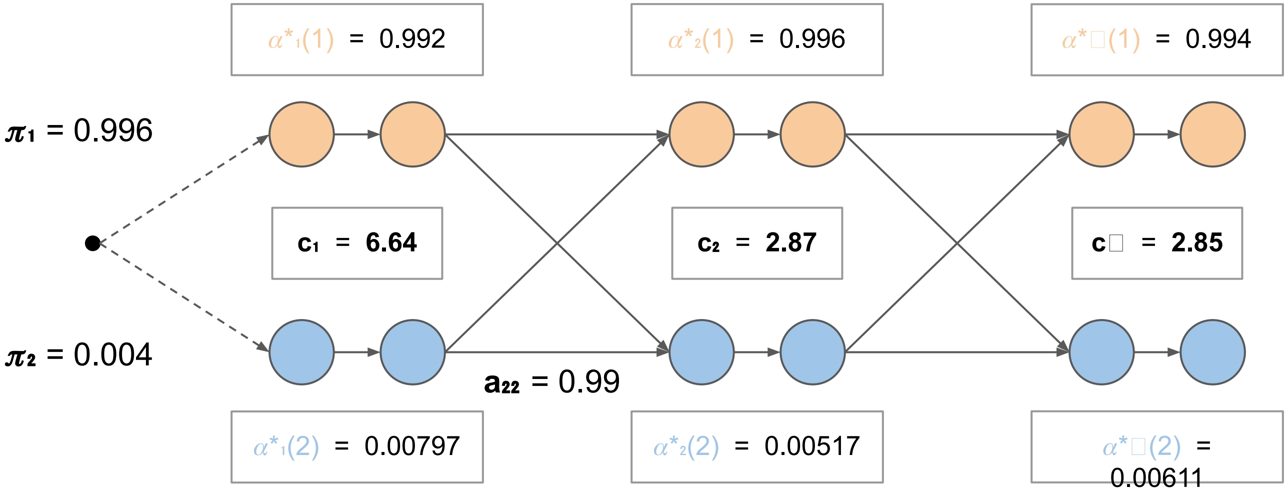
a_{22} = 0.99

$\alpha^*_1(2)$ = 0.00797

$\alpha^*_2(2)$ = 2.87 * 0.0018 = **0.00517**

observed sequence “TCG”

The forward part: t>1



$\alpha^*_1(1)$ = 0.992

$\alpha^*_2(1)$ = 0.996

$\alpha^*_□(1)$ = 0.994

π_1 = 0.996

c_1 = **6.64**

c_2 = **2.87**

$c_□$ = **2.85**

π_2 = 0.004

a_{22} = 0.99

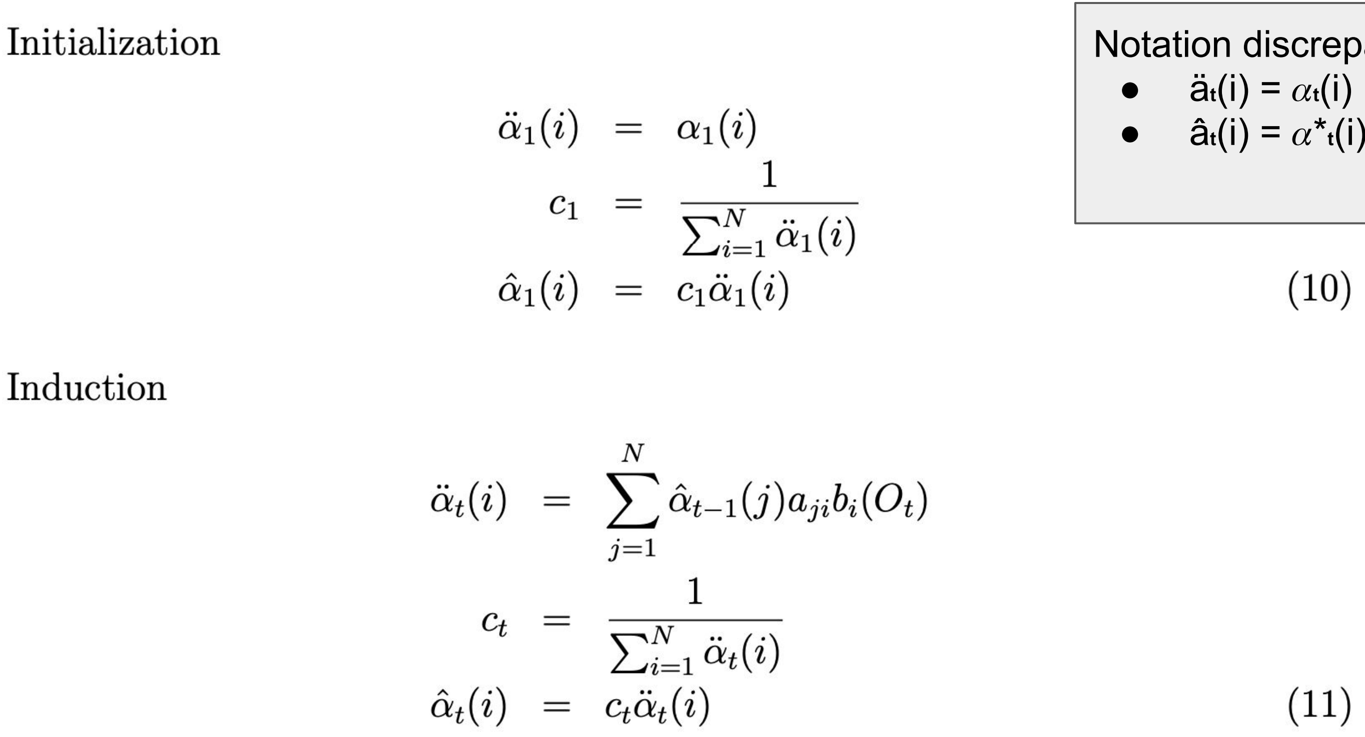
$\alpha^*_1(2)$ = 0.00797

$\alpha^*_2(2)$ = 0.00517

$\alpha^*_□(2)$ = 0.00611

observed sequence “TCG”

The forward part: summary from tutorial



Initialization

$$\begin{aligned}\ddot{\alpha}_1(i) &= \alpha_1(i)\\ c_1 &= \frac{1}{\sum_{i=1}^{N}\ddot{\alpha}_1(i)}\\ \hat{\alpha}_1(i) &= c_1\ddot{\alpha}_1(i)\end{aligned} \qquad (10)$$

Induction

$$\begin{aligned}\ddot{\alpha}_t(i) &= \sum_{j=1}^{N}\hat{\alpha}_{t-1}(j)a_{ji}b_i(O_t)\\ c_t &= \frac{1}{\sum_{i=1}^{N}\ddot{\alpha}_t(i)}\\ \hat{\alpha}_t(i) &= c_t\ddot{\alpha}_t(i)\end{aligned} \qquad (11)$$

Notation discrepancies with the tutorial:

- $\ddot{a}_t(i) = \alpha_t(i)$ ⟵ redefinition from tutorial (eq10)
- $\hat{a}_t(i) = \alpha^*_t(i)$ ⟵ no “alpha hat” in slides, so use α^* instead

https://www.cs.cmu.edu/~roni/11661/2017_fall_assignments/shen_tutorial.pdf

The forward part: summary from tutorial

also note we can prove that:

$$\hat{\alpha}_t(i) = \left(\prod_{\tau=1}^{t} c_\tau \right) \alpha_t(i). \tag{12}$$

Notation discrepancies with the tutorial:

- ä$_t$(i) = α_t(i) ← redefinition from tutorial (eq10)
- â$_t$(i) = α^*_t(i) ← no “alpha hat” in slides, so use α* instead

The backward part

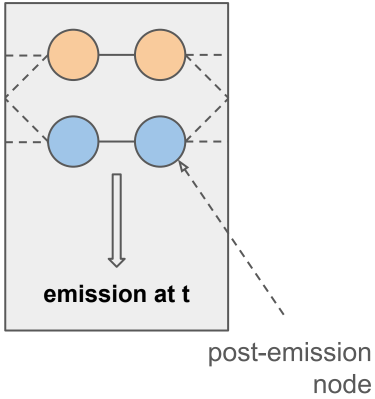
Similar to the forward part, but slightly different because of the ordering of the edges

Also, we get to reuse our forward-computed $\mathbf{c}_t$ values!

Additionally, initial probabilities are $\pi_1 = \pi_2 = 1$ because from post-emission nodes where **t**=T (the end), the only possible edge is to the end node

The backward part: more definitions

β_t(i): the highest probability of the partial observed sequence (this time, observations from t=t to t=T) for the post-emission node for **state i** at **time t** (unscaled)



β^*_t(i): the scaled version of β_t(i)

- Defined as: **c_t*β_t(i)**

Notation discrepancies with the tutorial:

- B^$_t$(i) = β^*_t(i) ⟵ no “beta hat” in slides, so use β* instead

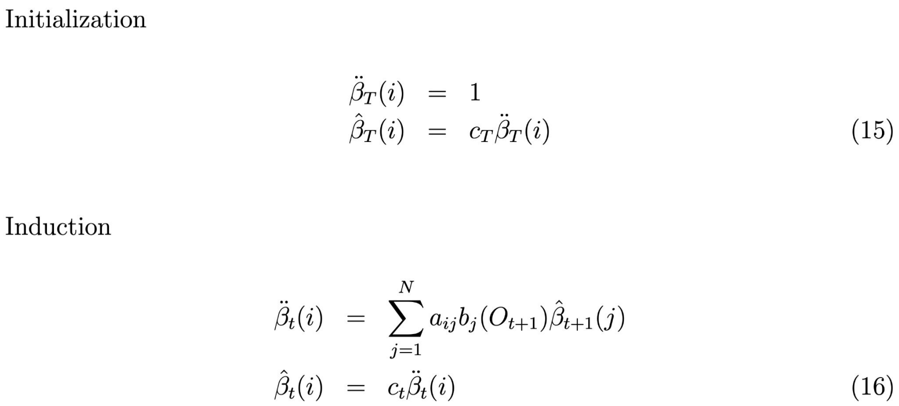
The backward part: calculating β*□(i)

For t = T:

- $\beta_T(i) = 1$
- $\beta^*_T(i) = c_T * \beta_T(i) = c_T$

For t < T, $\beta_t(i)$:

- $\beta_t(i) = \sum (a_{ij})(b_j)(\beta^*_{t+1}(j))$
- $\beta^*_t(i) = c_t * \beta_t(i)$



Initialization

$$\begin{aligned}\ddot{\beta}_T(i) &= 1 \\ \hat{\beta}_T(i) &= c_T\ddot{\beta}_T(i) \qquad (15)\end{aligned}$$

Induction

$$\begin{aligned}\ddot{\beta}_t(i) &= \sum_{j=1}^{N} a_{ij}b_j(O_{t+1})\hat{\beta}_{t+1}(j) \\ \hat{\beta}_t(i) &= c_t\ddot{\beta}_t(i) \qquad (16)\end{aligned}$$

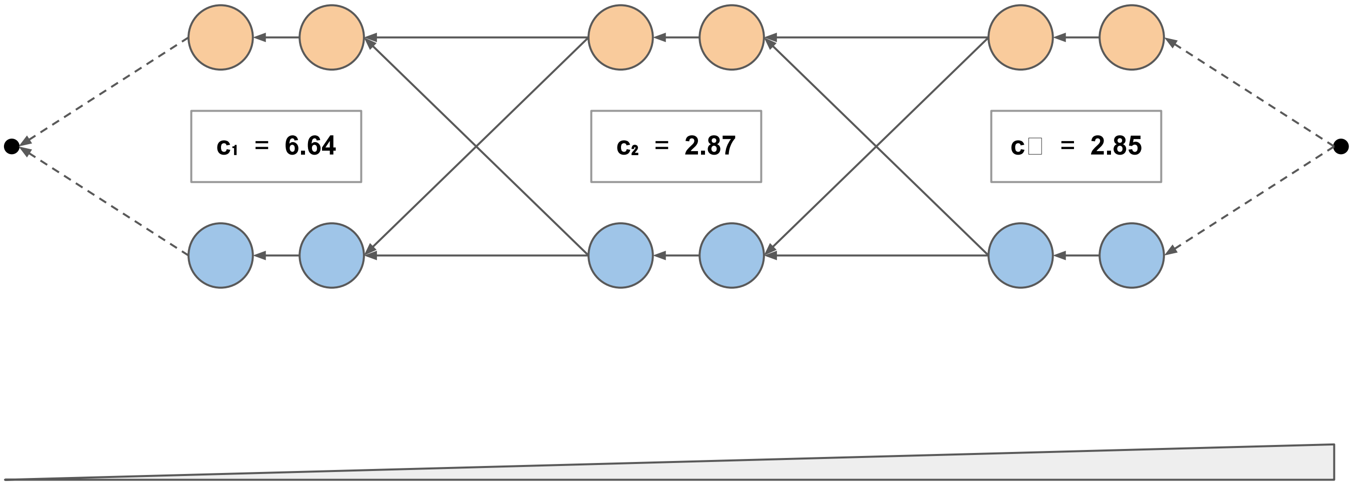
Notation discrepancies with the tutorial:

- B^$_t$(i) = $\beta^*_t(i)$ ← no “beta hat” in slides, so use β* instead

http://bozeman.genome.washington.edu/compbio/mbt599_2024/shen_tutorial.pdf

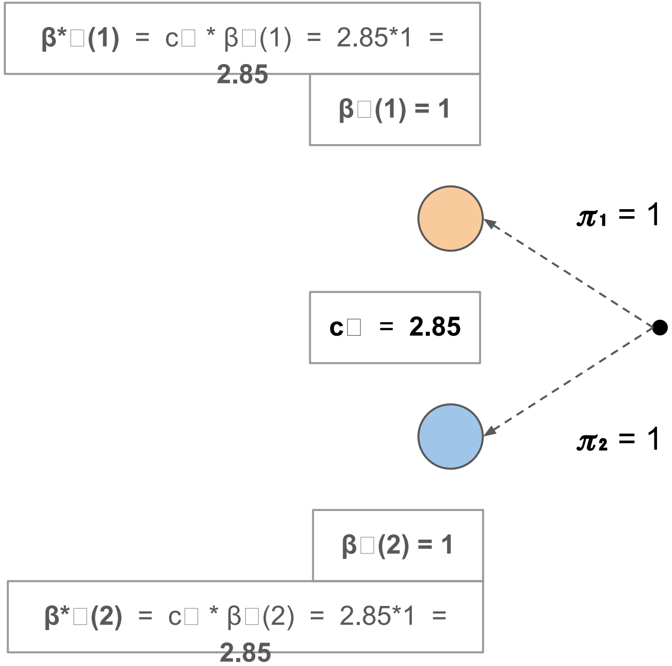
The backward part

First, reverse the graph:



same direction of increasing t →

The backward part: t=T



β*□(1) = c□ * β□(1) = 2.85*1 = **2.85**

β□(1) = 1

π_1 = 1

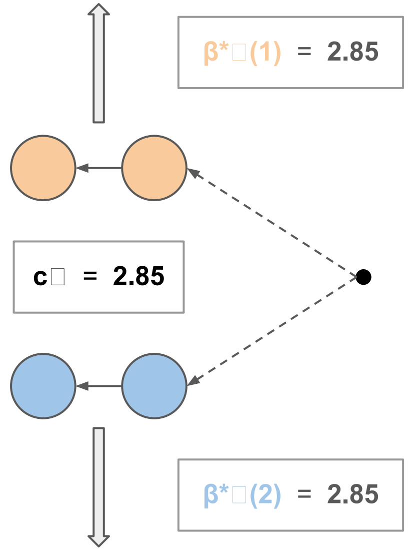
c□ = 2.85

π_2 = 1

β□(2) = 1

β*□(2) = c□ * β□(2) = 2.85*1 = **2.85**

The backward part: t<T



$b_1(G) = 0.35$

β*□(1) = 2.85

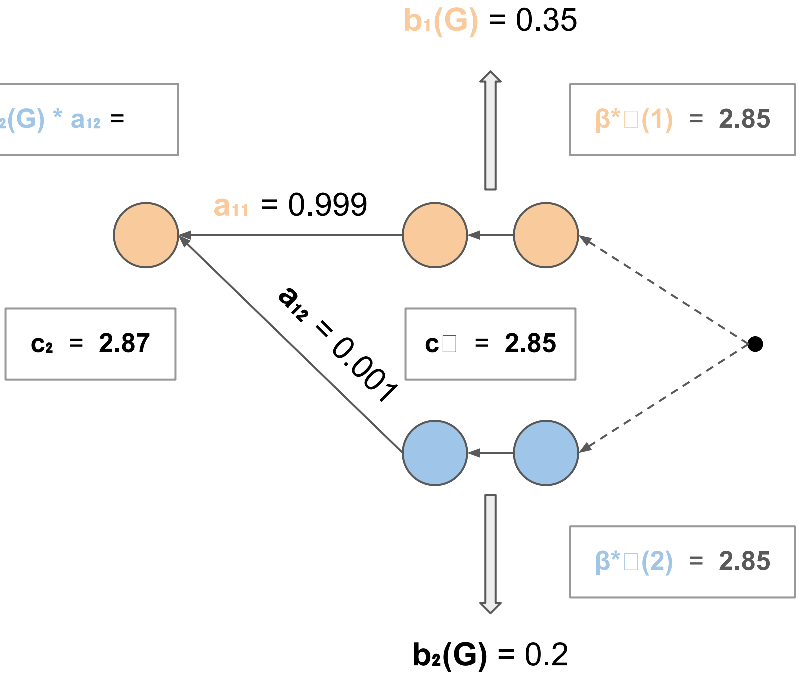
c□ = 2.85

β*□(2) = 2.85

$b_2(G) = 0.2$

The backward part: t<T

$\beta_2(2) = \beta^*\square(1) * b_1(G) * a_{11} + \beta^*\square(2) * b_2(G) * a_{12} =$ **0.997**



$b_1(G) = 0.35$

$\beta^*\square(1) = 2.85$

$a_{11} = 0.999$

$a_{12} = 0.001$

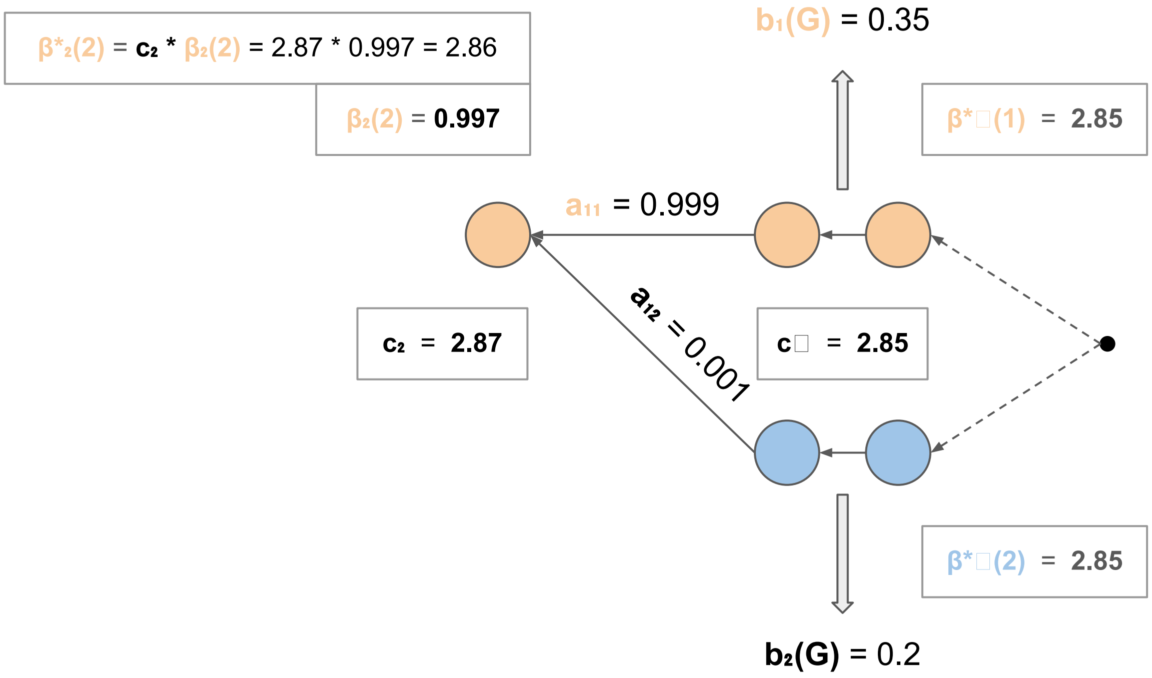
$c_2 = 2.87$

$c\square = 2.85$

$\beta^*\square(2) = 2.85$

$b_2(G) = 0.2$

The backward part: t<T



$\beta^*_2(2) = c_2 * \beta_2(2) = 2.87 * 0.997 = 2.86$

$\beta_2(2) = \mathbf{0.997}$

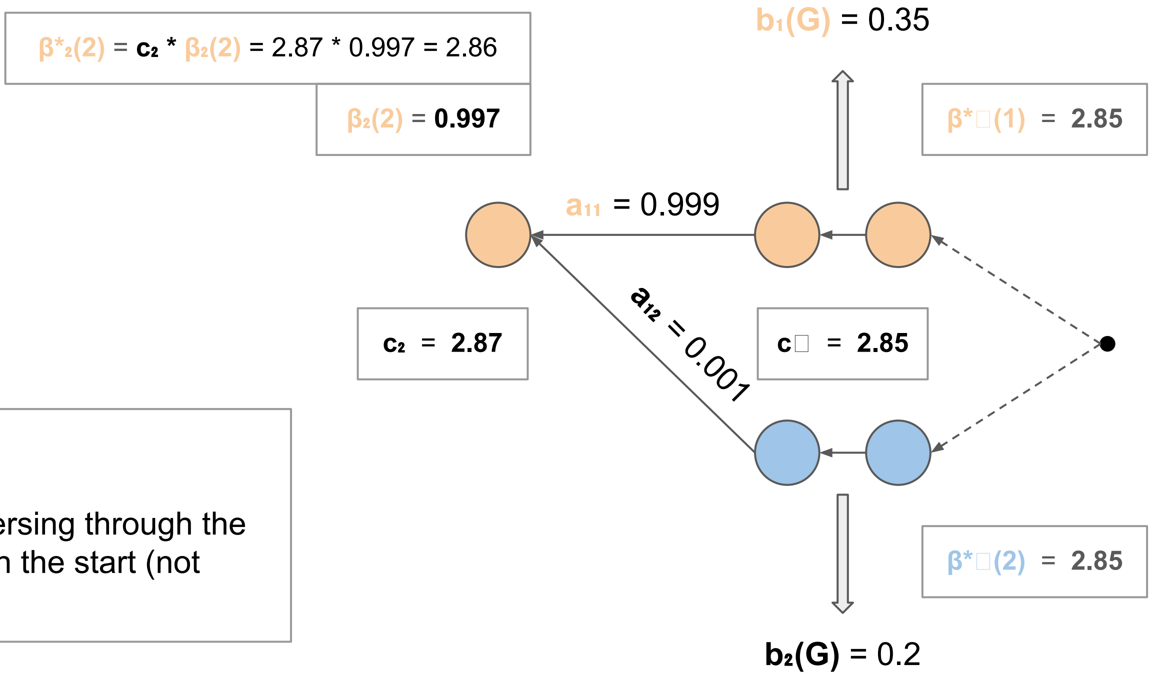
$c_2 = \mathbf{2.87}$

$c_{\square} = \mathbf{2.85}$

$\beta^*_{\square}(1) = \mathbf{2.85}$

$\beta^*_{\square}(2) = \mathbf{2.85}$

The backward part: t<T



$\beta^*_2(2)$ = c_2 * $\beta_2(2)$ = 2.87 * 0.997 = 2.86

$\beta_2(2)$ = **0.997**

$b_1(G)$ = 0.35

$\beta^*_{□}(1)$ = **2.85**

a_{11} = 0.999

a_{12} = 0.001

c_2 = **2.87**

$c_{□}$ = **2.85**

$\beta^*_{□}(2)$ = **2.85**

$b_2(G)$ = 0.2

…and so on!

From here, you can keep traversing through the backwards path until you reach the start (not shown here)

Homework 8: part 2

More variable definitions: unscaled

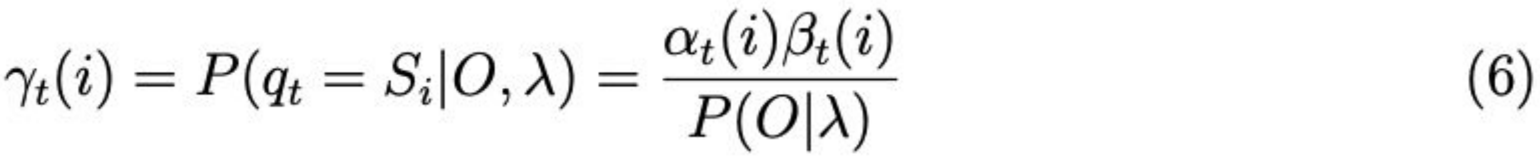
P(O | λ): Total probability of the **observed sequence O** given the **model λ**

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

$$P(O|\lambda) = \sum_{i=1}^{N} \pi_i \beta_1(i) b_i(O_1)$$

More variable definitions: unscaled

γ_t(i): the probability of being in **state i** at **time t** given the observed **sequence O** and the **model λ**



$$\gamma_t(i) = P(q_t = S_i|O, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} \qquad (6)$$

Somewhat intuitively, summing this term over all **t** is the expected number of times we are ever in **state i**

More variable definitions: unscaled

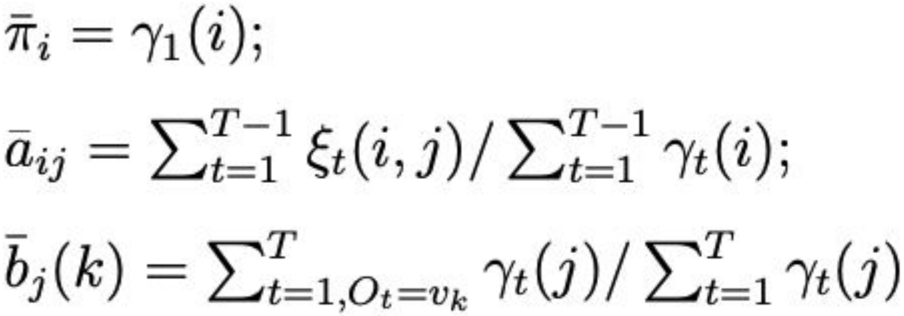
ξ_t**(i, j)**: the probability of being in **state i** at **time t** and **state j** at time **t+1** given the observed **sequence O** and the **model λ**

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \qquad (7)$$

Summing this term over all **t** (except T) is the expected number of times we transition from **state i** to **state j**

Baum-Welch parameter updates: unscaled

To calculate new parameters, you simply follow:



$$\bar{\pi}_i = \gamma_1(i);$$

$$\bar{a}_{ij} = \textstyle\sum_{t=1}^{T-1} \xi_t(i,j) / \sum_{t=1}^{T-1} \gamma_t(i);$$

$$\bar{b}_j(k) = \textstyle\sum_{t=1, O_t=v_k}^{T} \gamma_t(j) / \sum_{t=1}^{T} \gamma_t(j)$$

You can just keep alternating between this and the forward/backward calculations until the parameters converge!

Baum-Welch **SCALED**

The equations on the following slides basically just plug in the definitions we’ve seen for forward and backward variables, allowing us to use our newly computed scaled variables for the parameter update step

Baum-Welch **SCALED**

Log likelihood:

$$\log[P(O|\lambda)] = -\sum_{t=1}^{T} \log c_t.$$

Parameter updates:

$$\pi'_i = \frac{\hat{\alpha}_1(i)\hat{\beta}_1(i)}{\sum_j \hat{\alpha}_1(j)\hat{\beta}_1(j)}$$

$$\begin{aligned}\bar{b}_j(k) &= \frac{\sum_{t=1,O_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \\ &= \frac{\sum_{t=1,O_t=v_k}^{T} \hat{\alpha}_t(j) \cdot \hat{\beta}_t(j)/c_t}{\sum_{t=1}^{T} \hat{\alpha}_t(j) \cdot \hat{\beta}_t(j)/c_t}.\end{aligned}$$

$$\begin{aligned}\bar{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\ &= \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)} \\ &= \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i)/\mathbf{C}_t \cdot a_{ij} b_j(O_{t+1}) \cdot \hat{\beta}_{t+1}(j)/\mathbf{D}_{t+1}}{\sum_{t=1}^{T-1} \hat{\alpha}_t(i)/\mathbf{C}_t \cdot \hat{\beta}_t(i)/\mathbf{D}_t} \\ &= \frac{\left(\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \cdot a_{ij} b_j(O_{t+1}) \cdot \hat{\beta}_{t+1}(j)\right)/\mathbf{C}_T}{\left(\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \cdot \hat{\beta}_t(i)/c_t\right)/\mathbf{C}_T} \\ &= \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \cdot a_{ij} b_j(O_{t+1}) \cdot \hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \cdot \hat{\beta}_t(i)/c_t}.\end{aligned}$$

Office hours

Reminder:

Homework 8 is due Sunday, March 9th at 11:59pm!

Also remember:

Don’t round decimal values until the very end! This will affect all your calculations, especially here when dealing with a lot of decimal products