Genome 540 Discussion

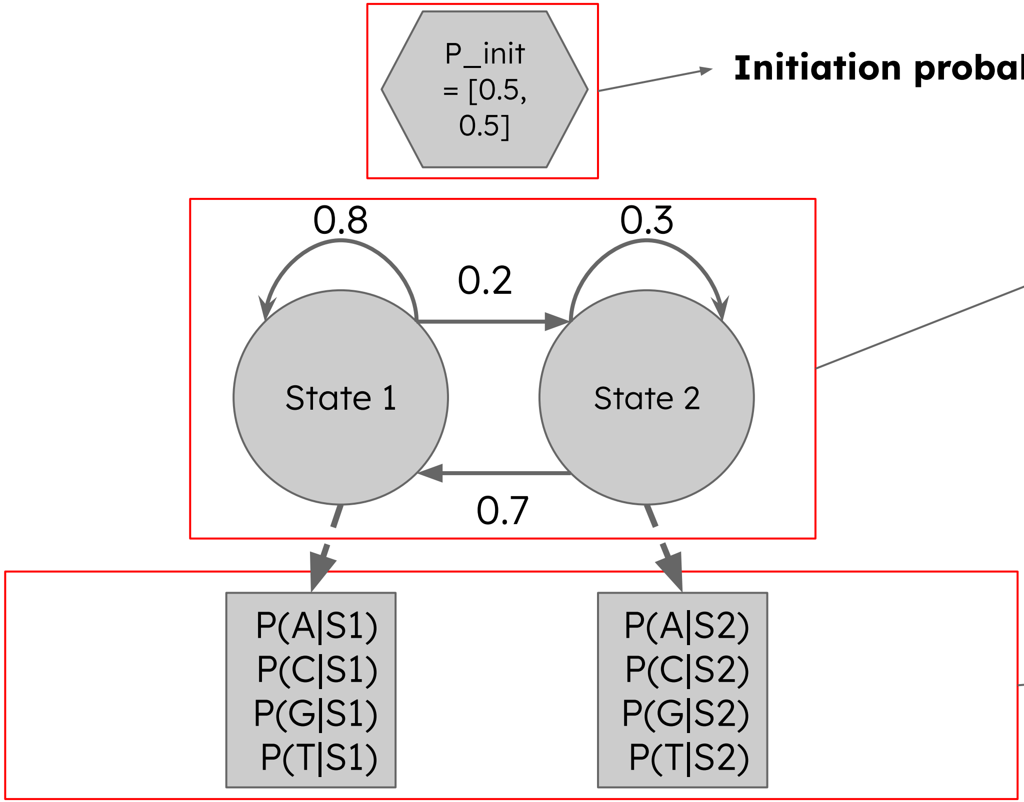
February 29th, 2024 - Happy Leap Day!

Clifford Rostomily

Genome Sciences
UNIVERSITY OF WASHINGTON

Assignment 8

Example - 2 state HMM for genomic sequences



Initiation probabilities

Transition Matrix

	1	2
1	0.8	0.2
2	0.7	0.3

Set of Hidden States

Emission probabilities of the observations

Some notation

$$S=\{S_1,S_2,\ldots,S_N\} \qquad q_t = \text{state at time } t$$

Set of states (Size N)

$$V=\{v_1,v_2,\ldots,v_m\}$$

Set of emitted symbols (vocabulary) (size M)

$$A=\{a_{ij}\} \qquad a_{ij}=P(q_{t+1}=S_j \mid q_t=S_i) \qquad 1\le i,j\le N$$

Transition matrix (N x N)

$$O=O_1 O_2 \ldots O_T \qquad O_t = \text{output symbol at time } t$$

Sequence of observed symbols (Length T)

$$B=\{b_j(k)\} \qquad \{b_j(k)\}=P(O_t=v_k \mid q_t=S_j) \qquad 1\le j\le N, 1\le k\le M$$

Emission probabilities

$$\pi=\{\pi_i\} \qquad \pi_i=P(q_1=S_i) \qquad 1\le i\le N$$

Initiation probabilities

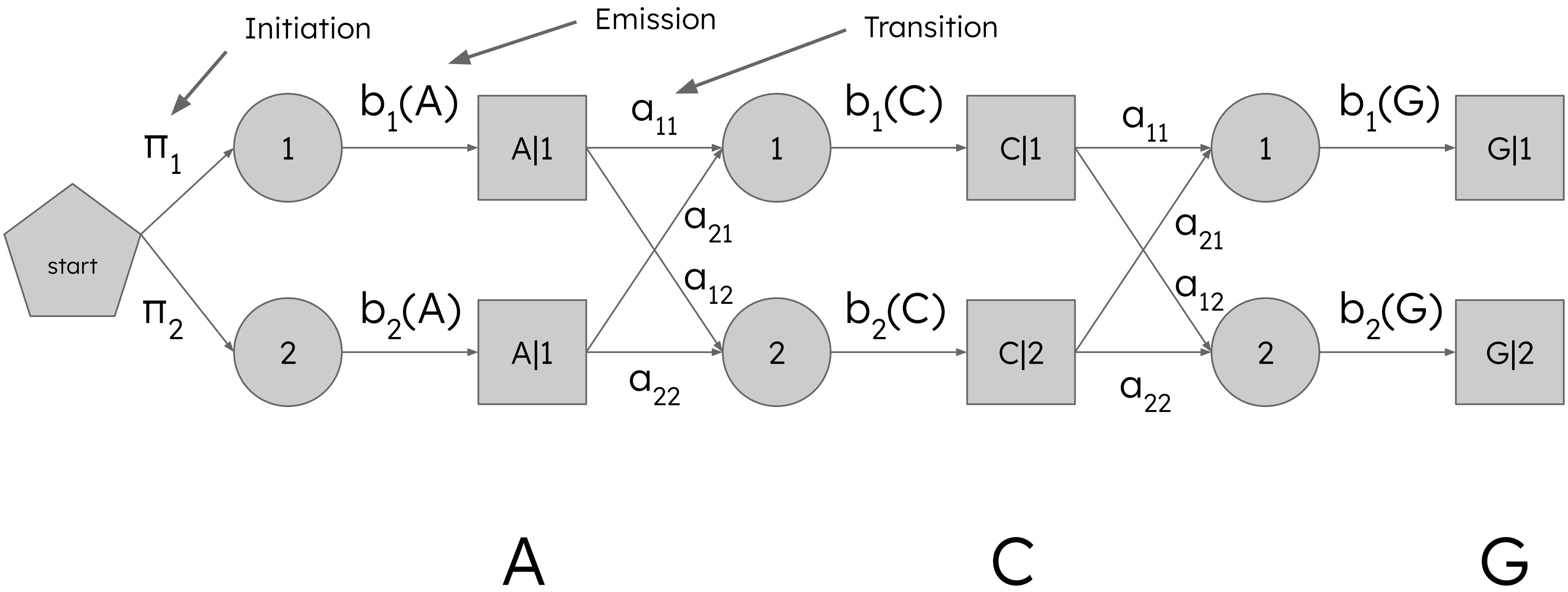
Baum Welch Steps

1. Compute (scaled) forward probabilities and scaling factors using the forward trellis
2. Compute (scaled) backward probabilities using the backward trellis
3. Compute updated parameter estimates
4. Repeat until convergence

Baum Welch Steps

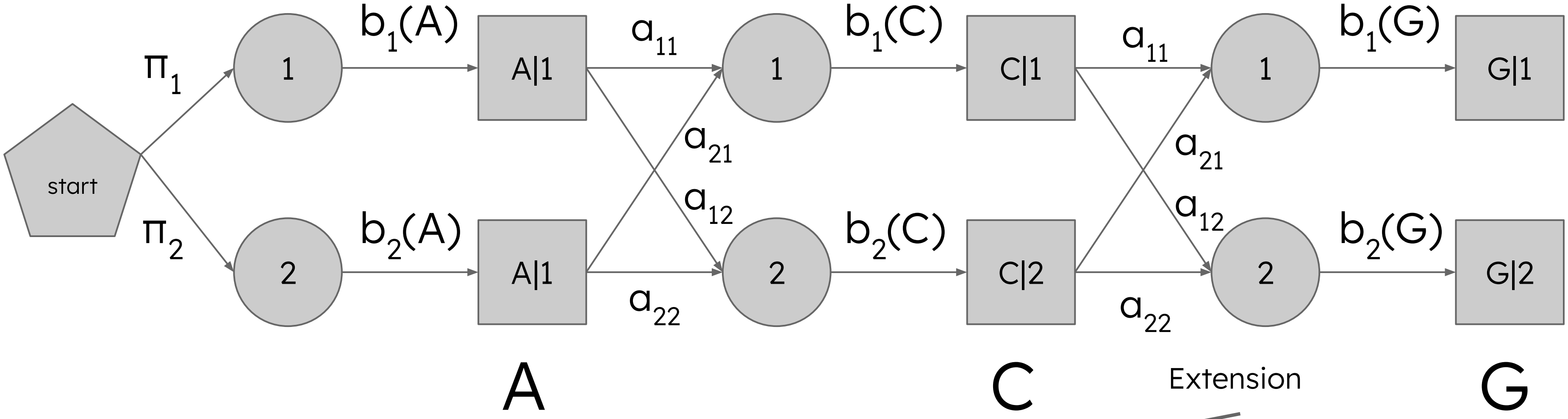
1. Compute (scaled) forward probabilities and scaling factors using the forward trellis
2. Compute (scaled) backward probabilities using the backward trellis
3. Compute updated parameter estimates
4. Repeat until convergence

Forward trellis



Compute forward probabilities: $\alpha_t(i)$

$$\alpha_t(i) = P(O_1 O_2 \ldots O_T, q_t = S_i | \lambda)$$



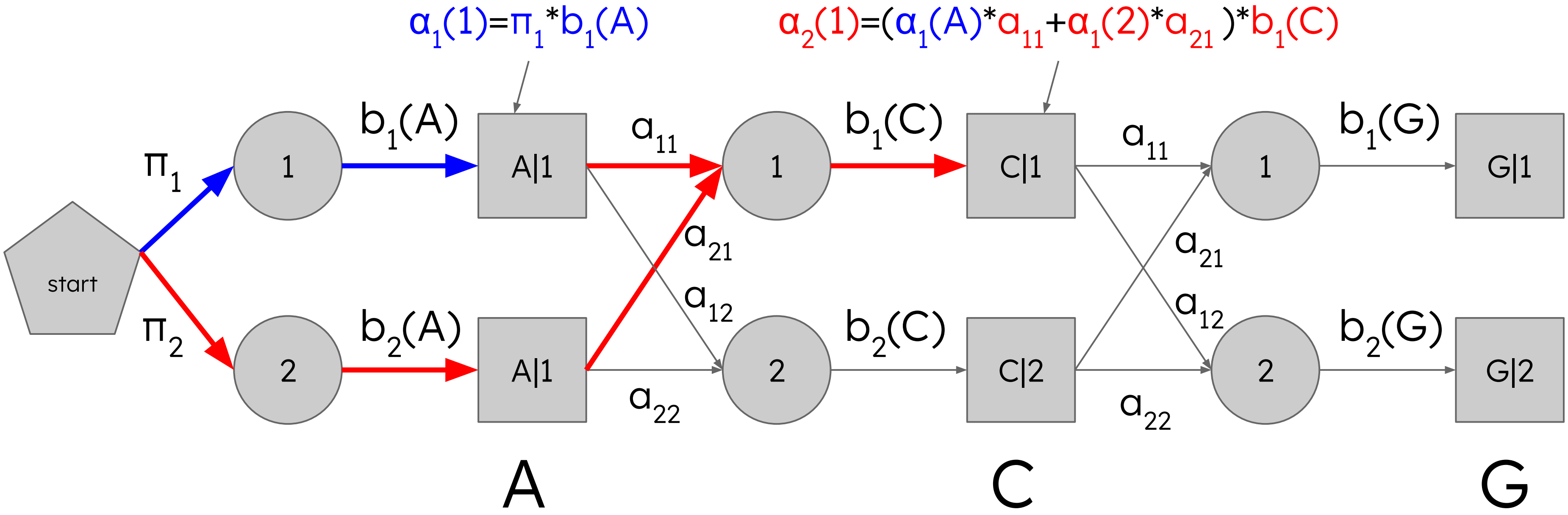
Initiation

$$\alpha_1(i) = \pi_i b_i(O_1)$$

Extension

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_t+1)$$

Example



$\alpha_1(1)=\pi_1*b_1(A)$
$\alpha_2(1)=(\alpha_1(A)*a_{11}+\alpha_1(2)*a_{21})*b_1(C)$
start
π_1
π_2
1
2
$b_1(A)$
$b_2(A)$
A|1
A|1
a_{11}
a_{21}
a_{12}
a_{22}
1
2
$b_1(C)$
$b_2(C)$
C|1
C|2
a_{11}
a_{21}
a_{12}
a_{22}
1
2
$b_1(G)$
$b_2(G)$
G|1
G|2
A
C
G

Example... with scaling

2. $c_1 = \dfrac{1}{\alpha_1(1)+\alpha_1(2)}$

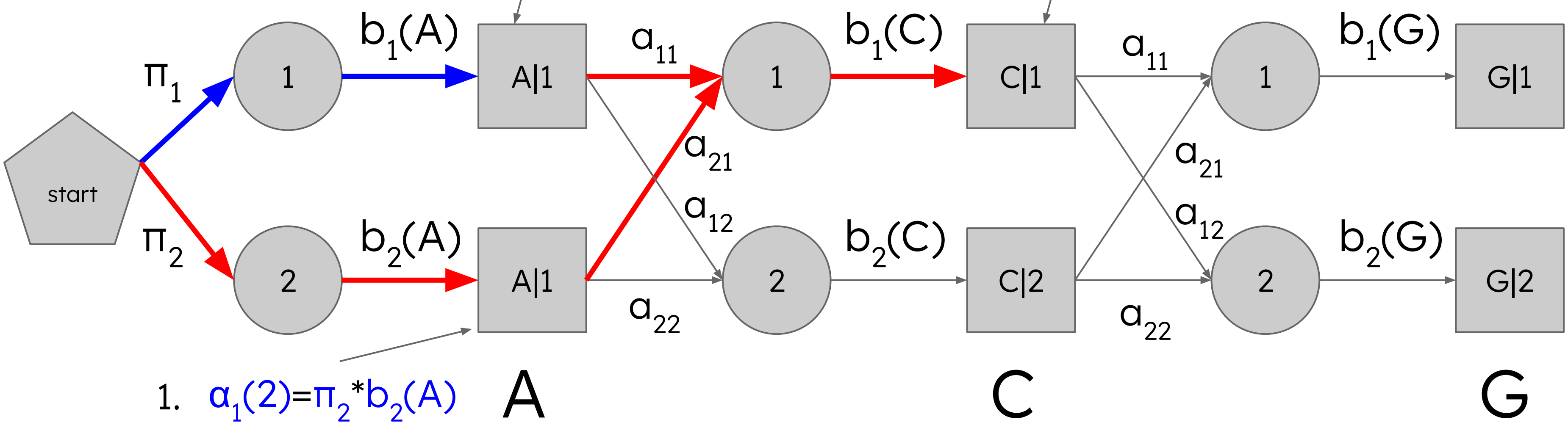
3. $\hat{\alpha}_1(1)=\pi_1{*}b_1(A){*}c_1$

5. Repeat 2-3 to get $\hat{\alpha}$

Scaling factor equation: $c_t = \dfrac{1}{\Sigma^{N}_{i=1}\ddot{\alpha}_t(i)}$

1. $\alpha_1(1)=\pi_1{*}b_1(A)$

4. $\ddot{\alpha}_2(A)=(\hat{\alpha}_1(1){*}a_{11}+\hat{\alpha}_1(2){*}a_{21}){*}b_1(C)$

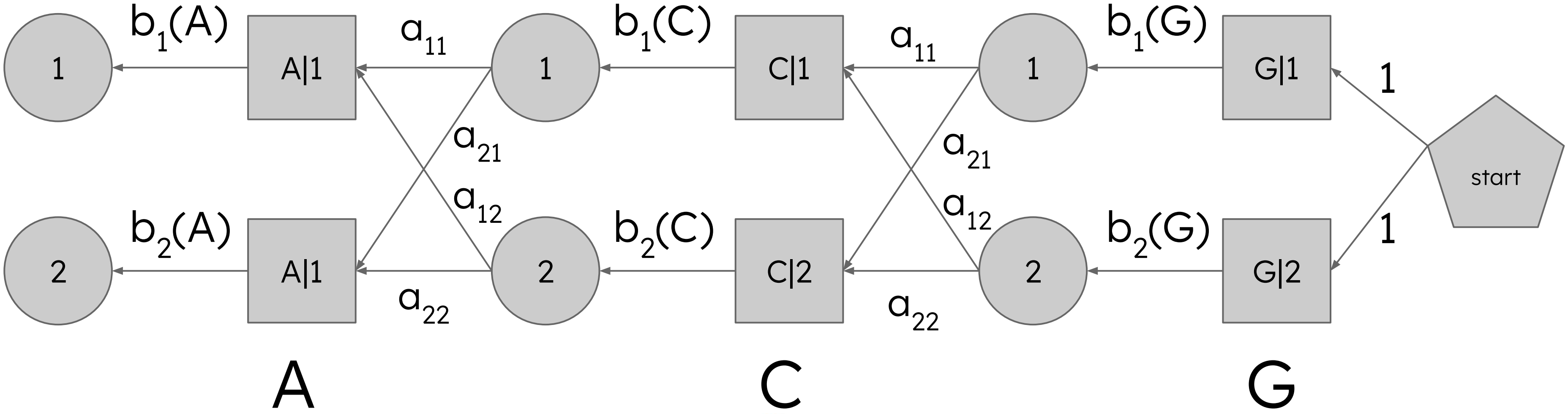


1. $\alpha_1(2)=\pi_2{*}b_2(A)$

Compute the backward probabilities: $\beta_t(i)$

$$\beta_t(i) = P(O_{t+1} O_{t+2} \cdots O_T | q_t = S_i, \lambda)$$

$$\beta_T(i) = 1$$

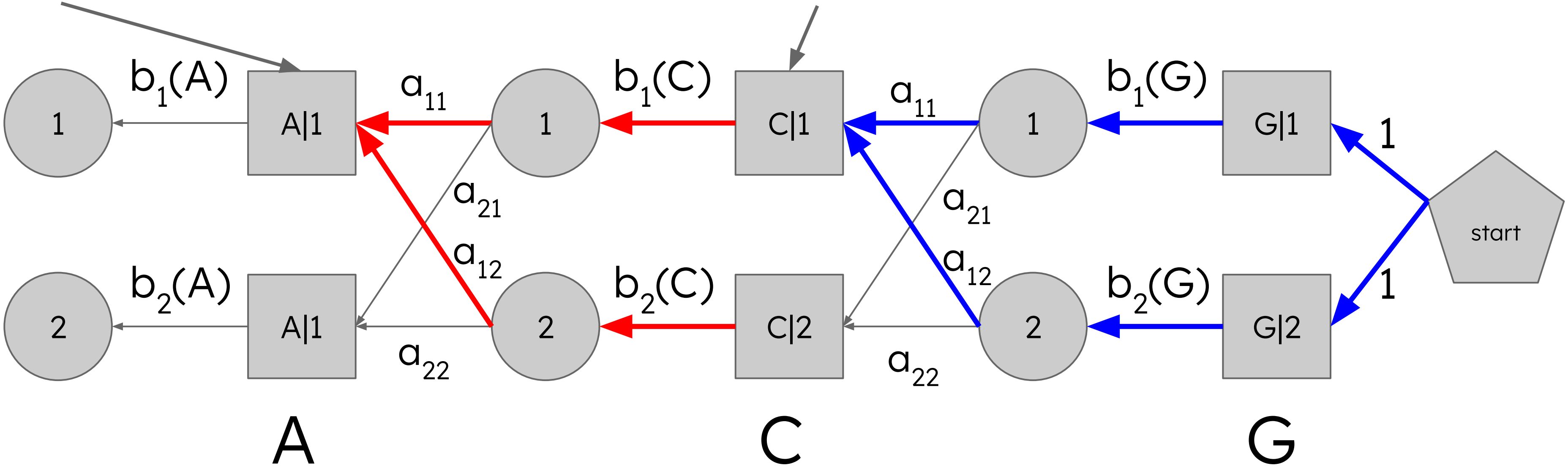


$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j).$$

Example

$\square_1(1) = (\square_2(1) * b_1(C) * a_{11} + \square_2(2) * b_2(C) * a_{12}) * c_1$

$\square_2(1) = (1 * b_1(G) * a_{11} + 1 * b_2(G) * a_{12}) * c_2$



*For scaling just multiply by the corresponding scaling factors from the forward probabilities

Calculating the updated probabilities

Once you have the forward and backward probabilities you can also calculate:

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

The probability of the observations given the model

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

The probability of being in a given state at a given position (time)

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

The probability of transitioning at a certain point

Calculating the updated probabilities

…and then you can compute the updated transition, initiation, and emission probabilities

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

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

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

But what about scaling?

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \cdot a_{ij} b_j(O_{t+1}) \cdot \hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \cdot \hat{\beta}_t(i) / c_t}$$

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

Avoiding vanishing probabilities

- These slides follow the following tutorial:
  - Shen Scaling Tutorial
- Alternatively you can skip scaling and work in log space. How to do this is described here:
  - Mann 2006

Reminders

- HW8 due this Sunday, 11:59pm
- Please have your name in the filename of your homework assignment and match the template