Chapter 6
Hidden Markov Models

Chaochun Wei

Spring 2019

Contents

- Reading materials
- Introduction to Hidden Markov Model
  - Markov chains
  - Hidden Markov Models
  - Parameter estimation for HMMs

Reading

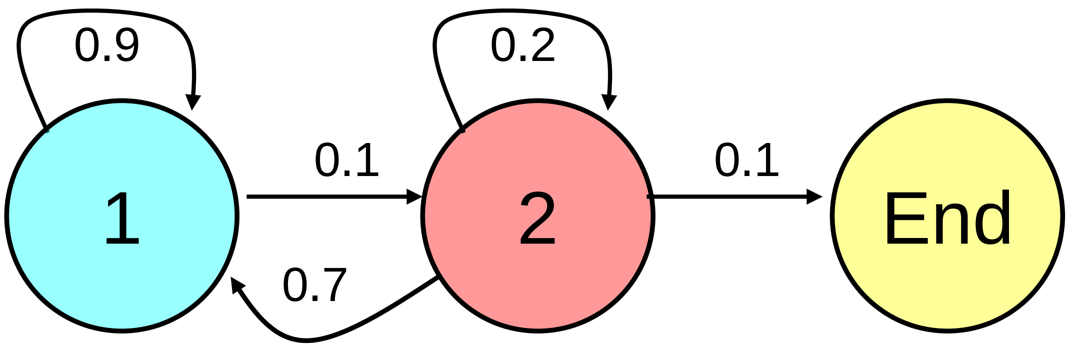
- Rabiner, L.(1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77 (2) 257-286
- Rabiner, L., and Juang, Biing-Hwang, (1993), Fundamentals of Speech Recognition, Prentice Hall.

Markov chain: a process that the current state depends on at most a limited number of previous states

- Weather
  - Sunny, Rain, Rain, Sunny, Cloudy, Cloudy,….
- Stock market index
  - Up, up, down, down, down, up, up, up, ….
- Girl/Boy friend's mood
  - High, low, low, high, high, high, …
- Genome sequence
  - ATGTTAGATATAACAGATAA
- Flip coins
  - HTTTHHHHHH

Hidden Markov Model

HMM for two biased coins flipping



$$e_1(H) = 0.8, e_1(T) = 0.2, e_2(H) = 0.3, e_2(T) = 0.7$$

TTHHTTHTTTTTHTHHHHHTHTH **Observed sequence x**

11221111122221111112222 **Hidden state sequence** π

$$\pi^* = \arg\max_{\pi} P(x, \pi)$$

Hidden Markov Model

- **Elements of an HMM (N, M, A, B, Init)**

1. **N: number of states in the model**
   - $S=\{S_1, S_2, \ldots, S_N\}$, and the state at time t is q_t.
2. **M: alphabet size (the number of observation symbols)**
   - $V=\{v_1, v_2, \ldots, v_M\}$
3. **A: state transition probability distribution**
   - $A=\{a_{ij}\}$ where $a_{ij}=P[q_{t+1}=S_j|q_t=S_i]$, $1 \le i,j \le N$
4. **E: emission probability**
   - $E=\{e_j(k)\}$ (observation symbols probability distribution in state j), where $e_j(k)=P[v_k \text{ at } t \mid q_t = S_j\}$, $1 \le j \le N$, $1 \le k \le M$
5. **Init: initial state probability**
   - $Init=\{I_i\}$, where $I_i=P[q_1=S_i]$, $1 \le i \le N$.

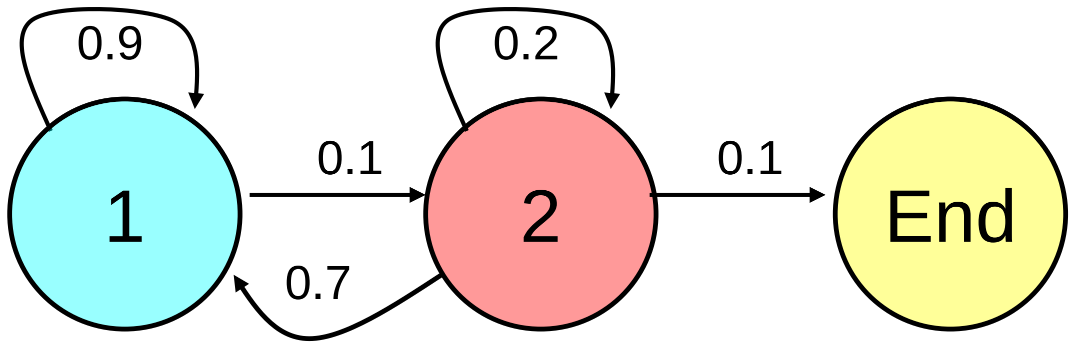
HMM is a generative model

- HMM can be used as a generator to produce an observation sequence $O=O_1O_2\ldots O_T$, where each O_t is one of the symbols from V, and T is the number of observations in the sequence.

1. Choose an initial state $q_1=S_i$ according to Init;
2. Set t=1;
3. Choose $O_t=v_k$ according to $e_i(k)$ (the symbol probability distribution in state S_i);
4. Transit to a new state $q_{t+1}=S_j$ according to a_{ij};
5. Set t=t+1; return to step 3 if t<T; otherwise terminate the procedure.

HMM is a generative model

HMM for two biased coins flipping



$$e_1(H)=0.8, e_1(T)=0.2, e_2(H)=0.3, e_2(T)=0.7$$

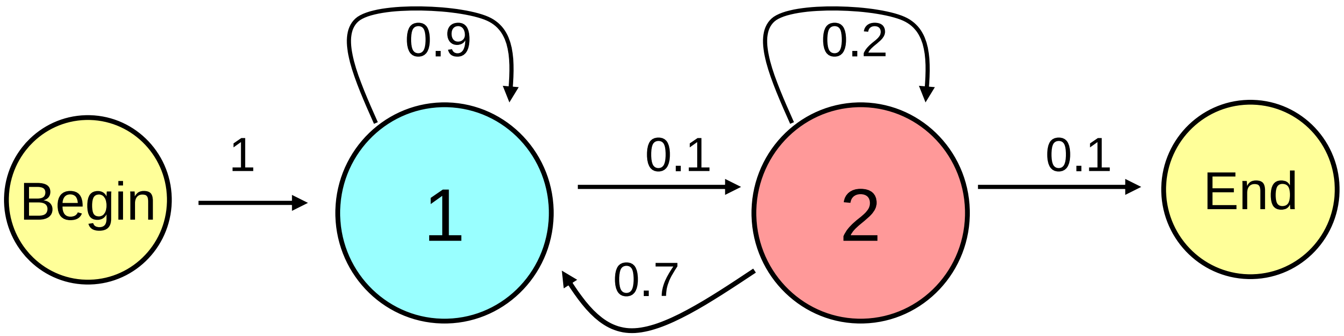
TTHHTTHTTTTTHTHHHHHTHTH **Observed sequence x**

11221111222211111112222 **Hidden state sequence π**

$$P(x,\pi\,|\,\lambda)=Init_{\pi_0} * e_{\pi_0}(x(0)) * \prod_{0\le i\le T}(a_{\pi_i\pi_{i+1}} e_{\pi_{i+1}}(x(i)))$$

HMM is a generative model

HMM for two biased coins flipping



$$e_1(H) = 0.8, e_1(T) = 0.2, e_2(H) = 0.3, e_2(T) = 0.7$$

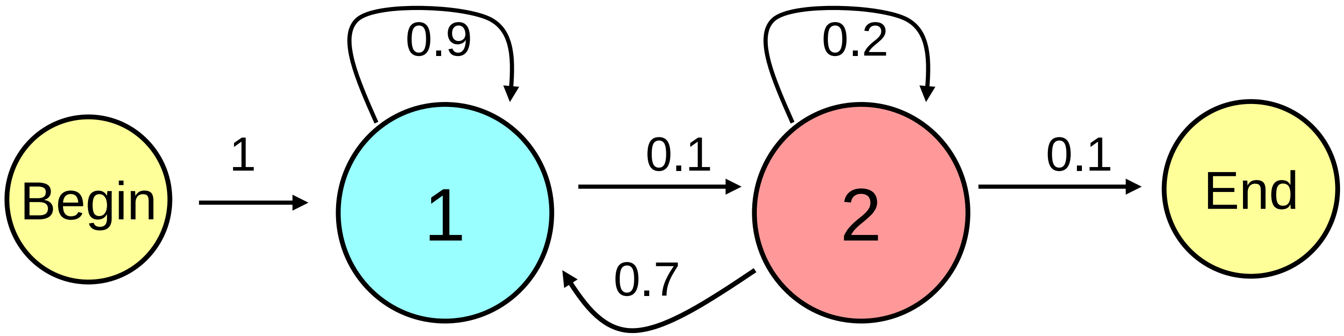
TTHHT **Observed sequence x**

11221 **Hidden state sequence** π

$$P(x, \pi \mid \lambda) = ?$$

HMM is a generative model

HMM for two biased coins flipping



$$e_1(H)=0.8, e_1(T)=0.2, e_2(H)=0.3, e_2(T)=0.7$$

TTHHT **Observed sequence x**

11221 **Hidden state sequence** π

$$P(x,\pi \mid \lambda) = Init_{\pi_0} * e_{\pi_0}(x(0)) * \prod_{0\le i\le T} (a_{\pi_i \pi_{i+1}} e_{\pi_{i+1}}(x(i)))$$

$$= 1 * e_1(T) * (a_{11} e_1(T)) * (a_{12} e_2(\text{H})) * (a_{22} e_2(\text{H})) * (a_{21} e_1(T))$$

$$= 1 * 0.2 * (0.9 * 0.2) * (0.1 * 0.3) * (0.2 * 0.3) * (0.7 * 0.2)$$

Hidden Markov Model

- HMM： λ={A, B,Init}

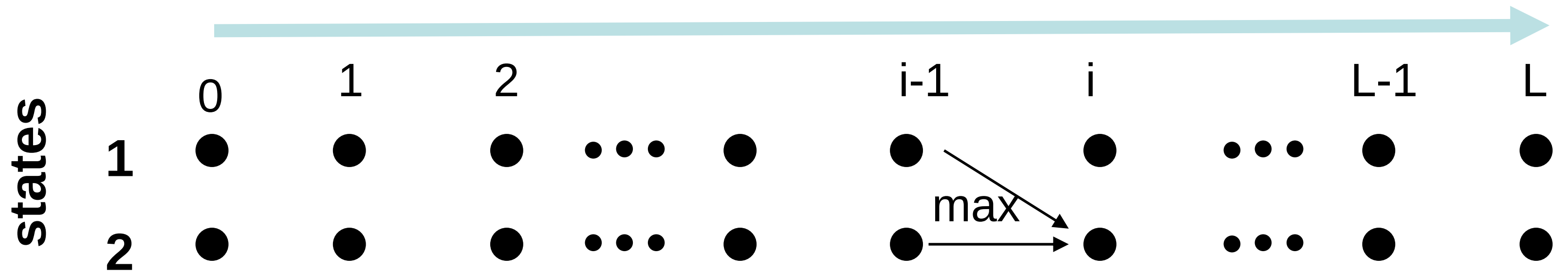
- **Three basic problems** for HMMs

1. Given the observation sequence $O=O_1O_2\ldots O_T$, and a model λ={A, B, Init}, how to compute P(O| λ)?
2. Given the observation sequence $O=O_1O_2\ldots O_T$, and a model λ={A, B, Init}, how to choose a corresponding state sequence $Q=q_1q_2\ldots q_T$, which is optimal in some meaningful sense..
3. How to estimate model parameters λ={A, B, Init} to maximize P(O| λ).

Hidden Markov Model

- HMM： λ={A, B, Init}
- Three basic problems for HMMs
  1. From the observation sequence and the model to a joint probability;
  2. Find the best hidden state sequence;
  3. Optimize the model parameters;

Most Probable Path and Viterbi Algorithm



Let $f_l(i) = \max_{\{\pi_0,\ldots,\pi_{i-1}\}} (\Pr(x_0,\ldots,x_{i-1},x_i,\pi_0,\ldots,\pi_{i-1},\pi_i = l))$

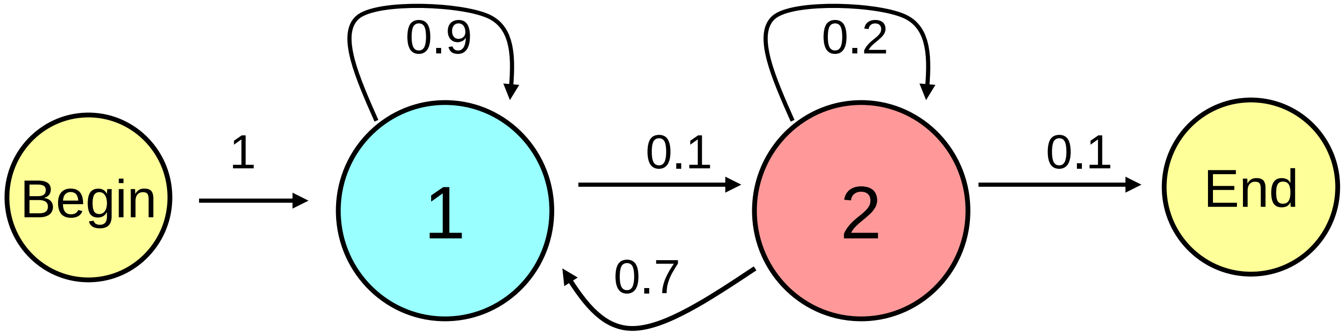
Recursion (i=1…L)

$$f_l(i) = e_l(x_i)\max_k(f_k(i-1)a_{kl});$$
$$ptr_i(l) = \arg\max_k(f_k(i-1)a_{kl}).$$

Time complexity $O(N^2L)$ space complexity $O(NL)$

Solution to problem 2: prob of best state sequence

Viterbi for the HMM for two biased coins flipping



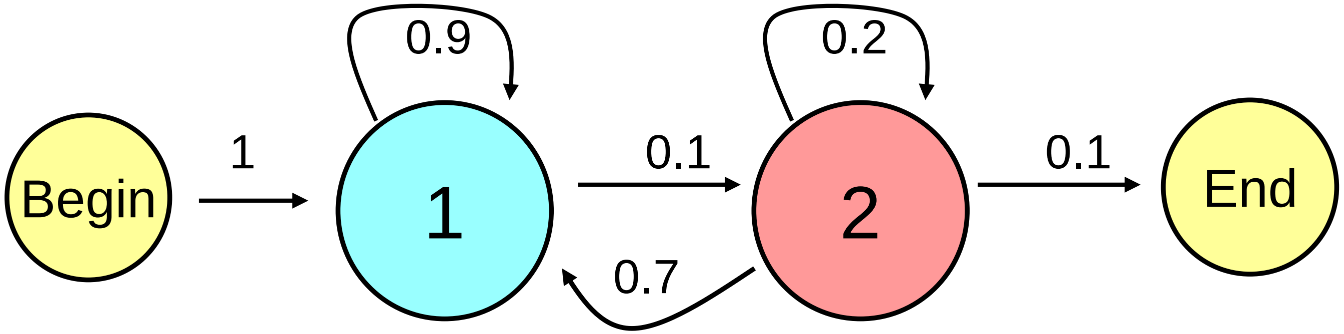
$$e_1(H) = 0.8, e_1(T) = 0.2, e_2(H) = 0.3, e_2(T) = 0.7$$

TTHHT **Observed sequence x**

11221 **Hidden state sequence** π

	T	T	H	H	T
	0	1	2	3	4
1					
2					

Viterbi for the HMM for two biased coins flipping



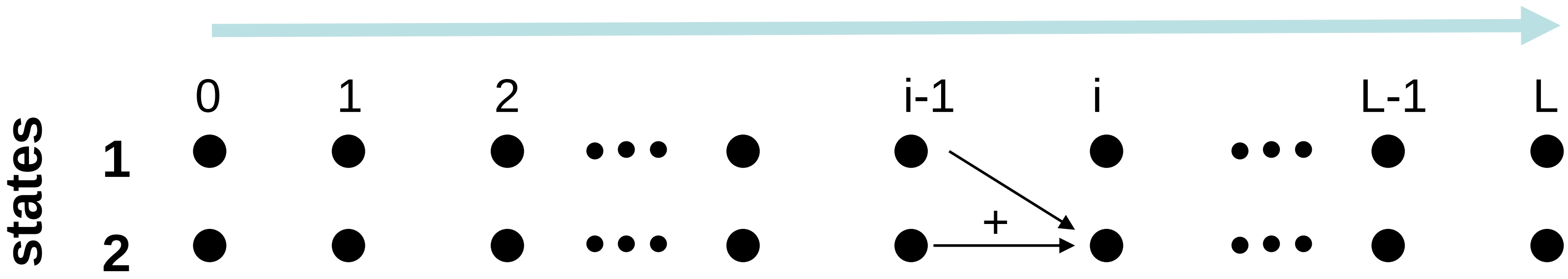
$$e_1(H)=0.8, e_1(T)=0.2, e_2(H)=0.3, e_2(T)=0.7$$

TTHHT **Observed sequence x**

11221 **Hidden state sequence** π

	T	T	H	H	T
	0	1	2	3	4
1	**0.2**	**Max0.2*(0.9 *0.2, 0) = 0.036**	**max0.8*(0.036* 0.9, 0.014*0.7) = 0.0259**	**Max 0.8 * (0.0259*0.9, 0.0108*0.7) = 0.018662**	**=max0.2*(0.018662* 0.9, 0.000777*0.7) = 0.003359**
2	0	Max0.7*(0.2 * 0.1, 0) = 0.014	Max0.3*(0.036* 0.1, 0.014*0.2) = 0.0108	Max0.3(0.0259* 0.1, 0.0108*0.2) = 0.000777	=max0.3(0.018662*0.1, 0.000777*0.7) = 0.0005599

Probability of All the Possible Paths and Forward Algorithm



Let $$f_l(i) = \Pr(x_0, \ldots, x_i, \pi_i = l)$$

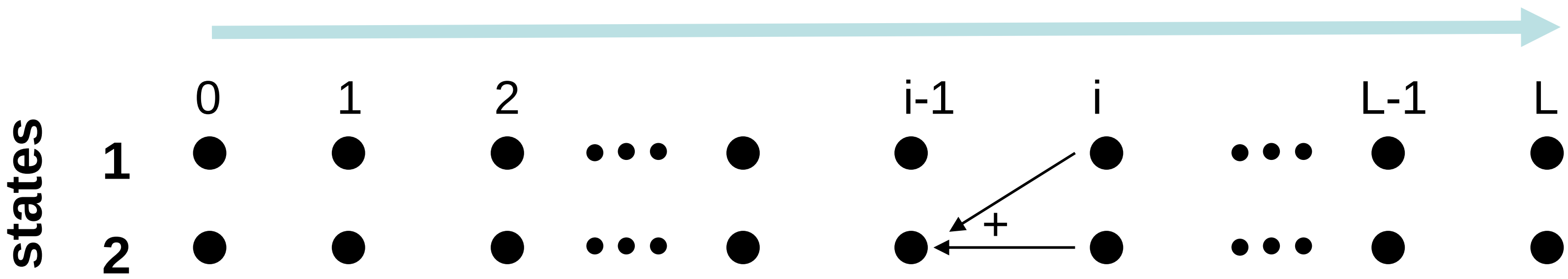
Initialization (i=1…L) $$f_0(i) = \pi_i e_i(x_0)$$

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

Probability of all the probable paths $$P(x) = \sum_\pi P(x, \pi) = \sum_k f_k(L)$$

Solution to problem 1

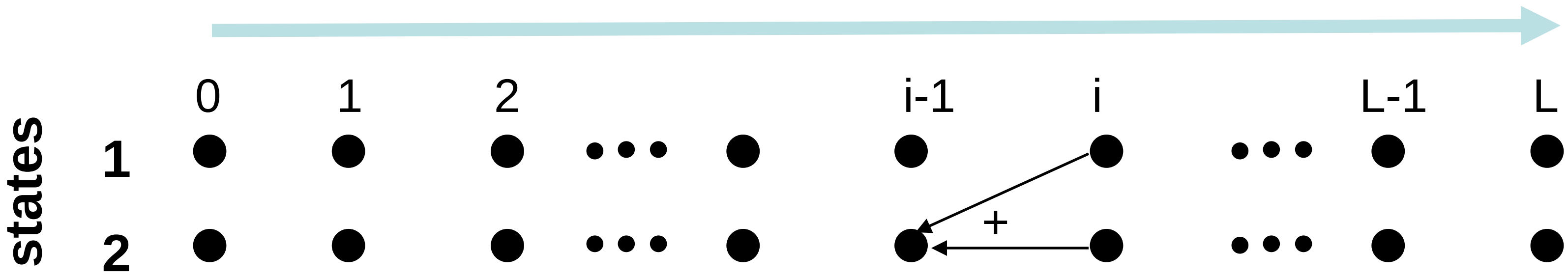
Posterior Probability and Forward and Backward Algorithm



Posterior Probability

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

Backward Algorithm



Let $b_k(i) = \Pr(x_{i+1}, x_{i+2} \ldots, x_L, \pi_i = k)$

Initialization (i=1…L) $$b_L(i) = 1, \quad 1 \le i \le N$$

Recursion (i=1…L) $$b_l(i) = \sum_k (a_{lk} e_k(x_i)) b_{l+1}(i),$$

$$l = L-1, L-2, \ldots, 0; 1 \le i \le N$$

Probability of all the probable paths $$P(x) = \sum_\pi P(x, \pi) = \sum_k b_k(0)$$

Optimize the model parameters

- HMM：λ={A, B, Init}
- With annotations
  - Maximum likely-hood
- Without annotations
  - EM algorithm

Optimize the model parameters

- **HMM：λ={A, B, Init}, Without annotations**

  Baum-Welch method (EM method)

Let $\xi_l(i, j) = P(\pi_l = i, \pi_{l+1} = j \mid x, \lambda)$

then $$\xi_l(i, j) = \frac{f_l(i) a_{ij} e_j(x_{l+1}) b_{l+1}(j)}{\sum_i^N \sum_j^N (f_l(i) a_{ij} e_j(x_{l+1}) b_{l+1}(j))}$$

Let $$\gamma_l(i) = \sum_{j=1}^N \xi_l(i, j)$$

then $\sum_{l=0}^N \gamma_l(i)$ = expected number of transitions from S_i

$\sum_{l=0}^N \xi_l(i, j)$ = expected number of transition S_i to S_j

Optimize the model parameters(2)

- **HMM：λ={A, B, Init}, Without annotations**

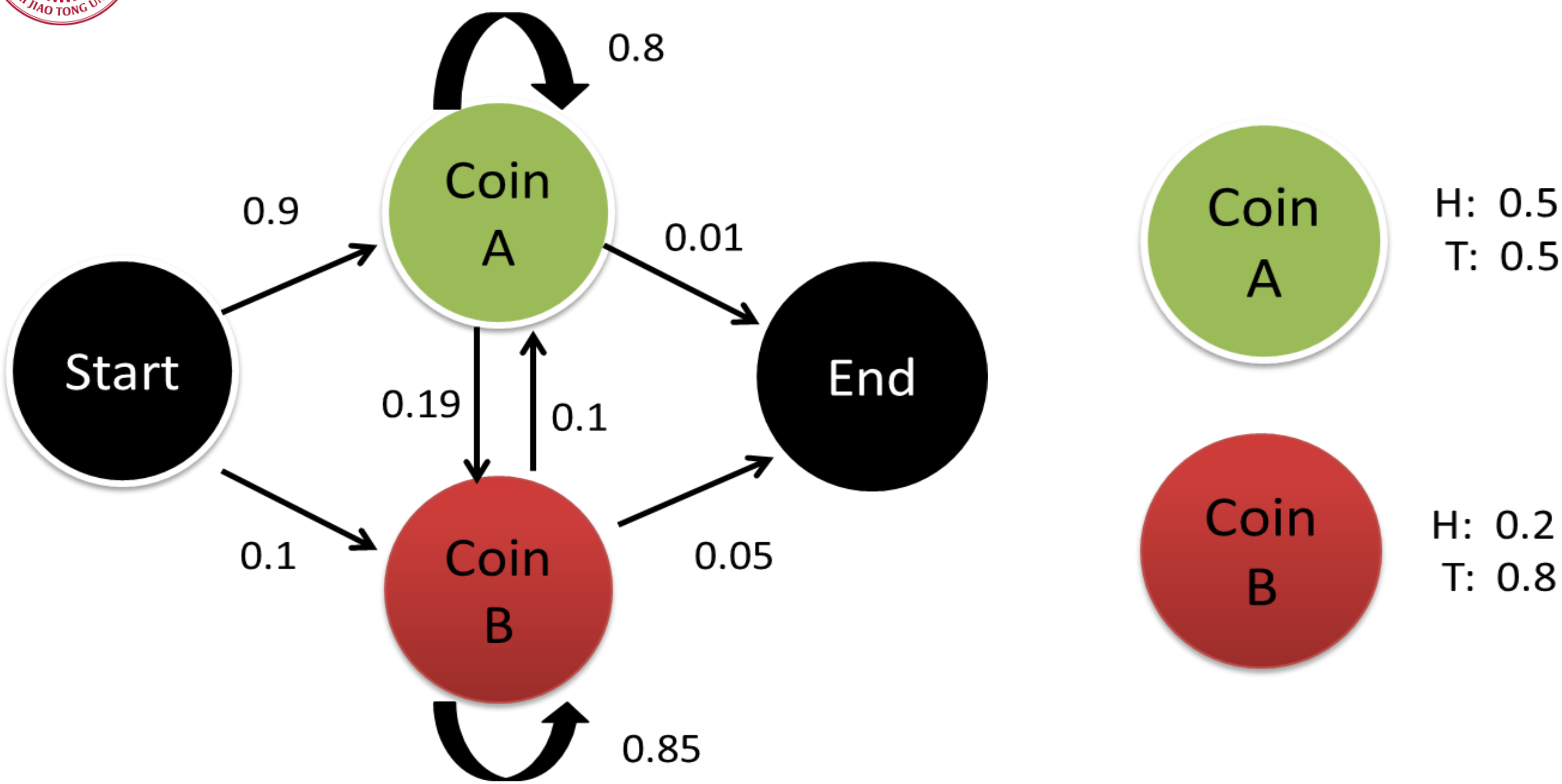
  Baum-Welch method (EM method)

Then, $\overline{Init}_i =$ expected frequency in S_i at time 0 = $\gamma_0(i)$

$$\overline{a}_{i,j} = \frac{\text{expected number of tranistions from } S_i \text{ to } S_j}{\text{expected number of tranistions from } S_i} = \frac{\sum_{l=0}^{L-1} \xi_l(i,j)}{\sum_{l=0}^{L-1} \gamma_l(i)}$$

$$\overline{e}_i(k) = \frac{\text{expected number of times in state i and oberving symbol } v_k}{\text{expected number of times in state i}} = \frac{\sum_{\substack{l=0 \\ s.t.x_t=v_k}}^{L} \gamma_l(i)}{\sum_{l=0}^{L} \gamma_l(i)}$$

One more example: Flipping two coins

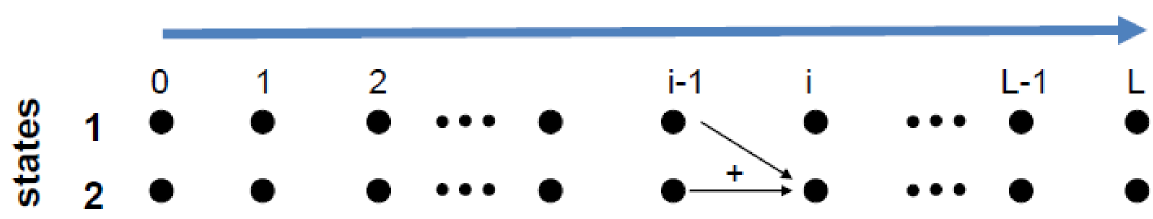


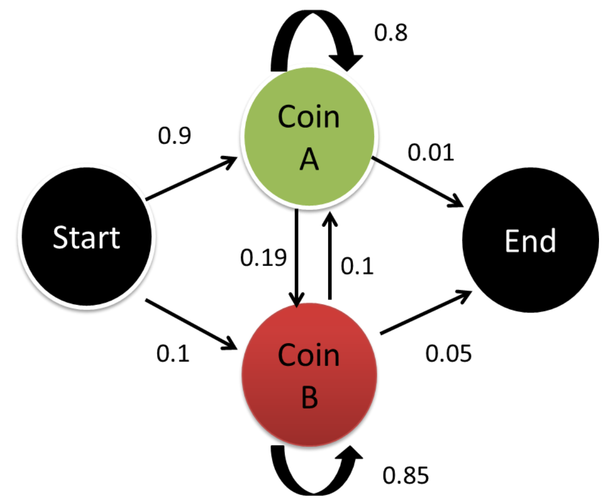
- **O= HHTHHTTTHT, P(O|λ)=?**

Problem 1: Given the observation sequence $O=O_1O_2\ldots O_T$, and a model λ={A, B, Init}, how to compute P(O| λ), the probability of the observation sequence given the model?

Forward Algorithms

	H	H	T	H	H	T	T	T	H	T	End
A	0.45	0.181	0.07342	0.03144	0.01306	0.05371	0.02345	0.00111	5.817e-4	2.580e-4	2.423e-5
B	0.02	0.0205	0.04145	0.00983	0.00286	0.00394	0.00349	0.00273	5.067e-4	4.329e-4	





Let $f_j(i) = \Pr(x_0, \ldots, x_i, \pi_i = j)$

Initialization (j=1…N) $f_j(0) = \pi_j e_j(x_0)$

Recursion (i=1…L; j = 1, .., N) $f_j(i) = e_j(x_i)\sum_k (f_k(i-1)a_{kj})$

Probability of all the probable paths $P(x) = \sum_\pi P(x,\pi) = \sum_k f_k(L)$

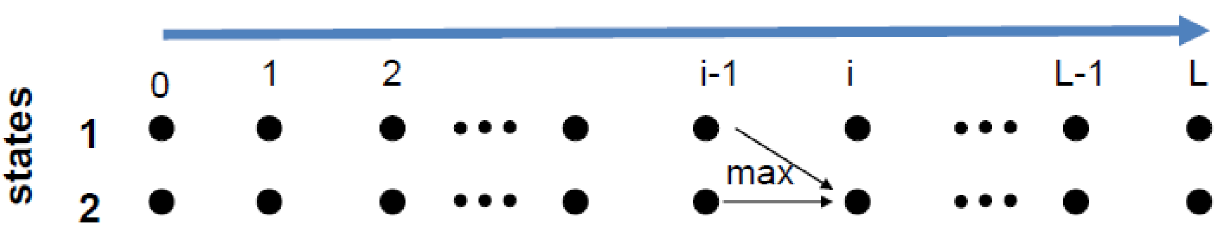
- **O= HHTHHTTTHT**
- **argmax(P(O, Q, λ))**

Problem 2: Given the observation sequence $O=O_1O_2\ldots O_T$, and a model $\lambda=\{A, B, Init\}$, how to choose a corresponding state sequence $Q=q_1q_2\ldots q_T$, which is optimal in some meaningful sense..

Viterbi Algorithms

A -> A -> A -> A -> A -> B -> B -> B -> B -> B

	H	H	T	H	H	T	T	T	H	T	End
A	0.45	0.18	0.072	0.0288	0.0115	0.0046	1.843e-3	7.372e-4	2.949e-4	1.180e-4	4.680e-6
B	0.02	0.0171	0.0274	4.652e-3	1.109e-3	1.751e-3	1.191e-3	8.097e-4	1.376e-4	9.360e-5	



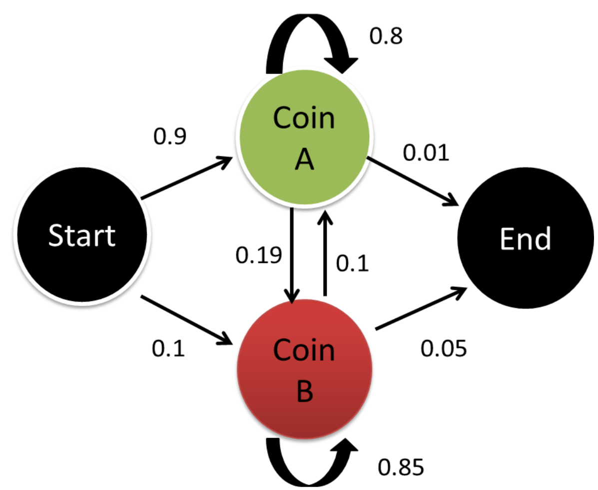
Let $f_j(i) = \max_{\{\pi_0,\ldots,\pi_{i-1}\}} (\Pr(x_0,\ldots,x_{i-1},x_i,\pi_0,\ldots,\pi_{i-1},\pi_i = j))$

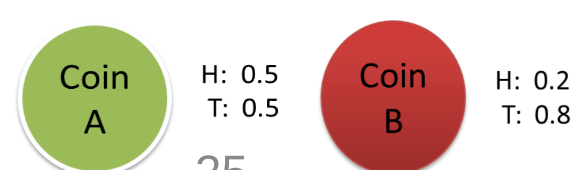
Initialization (j=1…N) $f_j(0) = \pi_j e_j(x_0)$

Recursion (i=1…L)

$$f_j(i) = e_j(x_i)\max_k (f_k(i-1)a_{kj});$$

$$ptr_j(i) = \arg\max_k (f_k(i-1)a_{kj}).$$

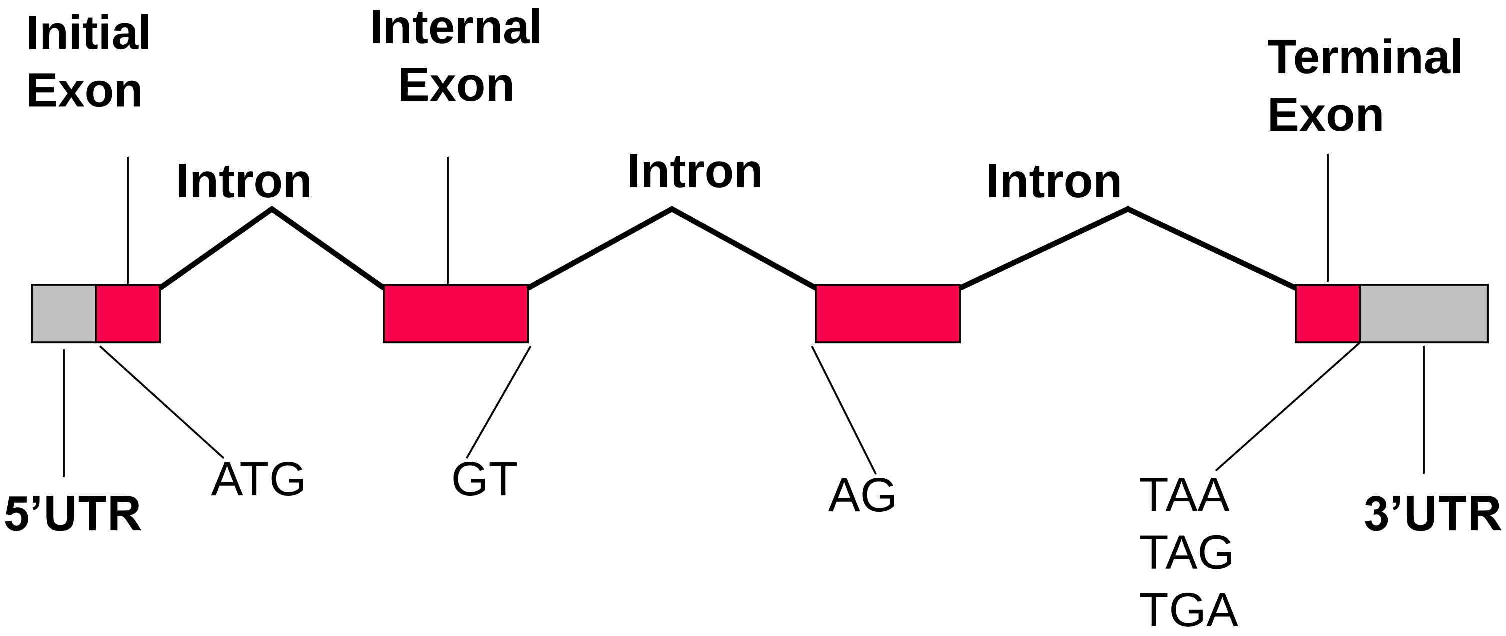




Problem 3. Model parameter estimation

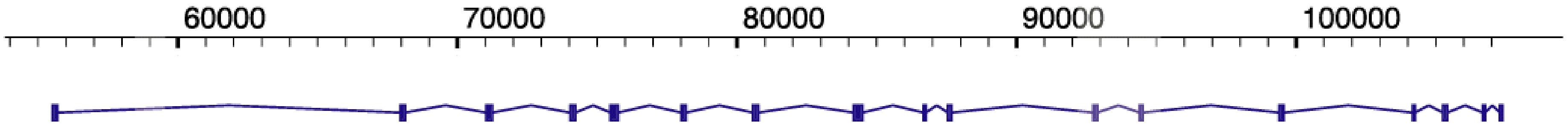
- See
  - Rabiner, L.(1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77 (2) 257-286
  - Rabiner, L., and Juang, Biing-Hwang, (1993), Fundamentals of Speech Recognition, Prentice Hall.

Gene Structure prediction with HMM



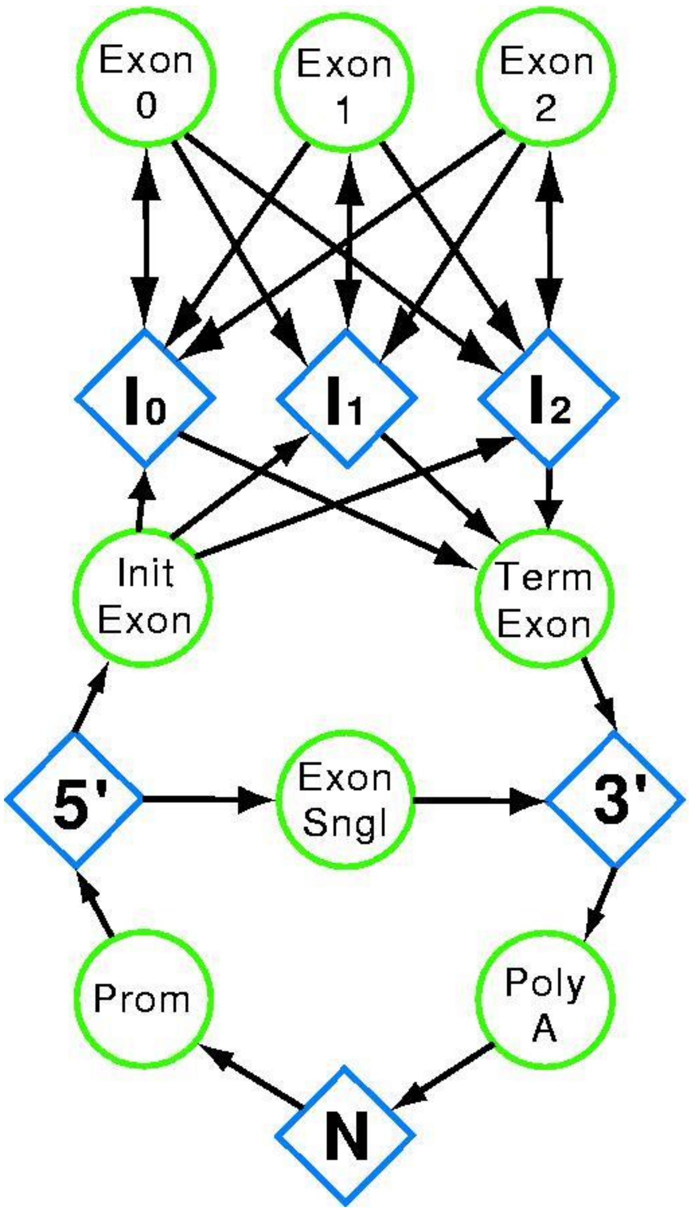
A gene is a highly structured region of DNA, it is a functional unit of inheritance.

A Typical Human Gene Structure



Gene Prediction Model

- HMM (27 states)
- Each state
  - For a gene structure
- State-specific models (Generalized HMM)
  - Length distribution
  - Sequence content



Another example:
Pair HMM for local alignment