Course organization

- Introduction (Week 1-2)
  - Course introduction
  - A brief introduction to molecular biology
  - A brief introduction to sequence comparison
- Part I: Algorithms for Sequence Analysis (Week 3 - 11)
  - Chapter 1-3, Models and theories
    » Probability theory and Statistics (Week 4)
    » Algorithm complexity analysis   (Week 5)
    » Classic algorithms   (Week 6)
    » Lab: Linux and Perl
  - Chapter 4, Sequence alignment (week 7)
  - Chapter 5, Hidden Markov Models (week 8)
  - Chapter 6. Multiple sequence alignment (week 10)
  - Chapter 7. Motif finding (week 11)
  - Chapter 8. Sequence binning (week 11)
- Part II: Algorithms for Network Biology (Week 12 - 16)

Chapter 5
Hidden Markov Models

Chaochun Wei
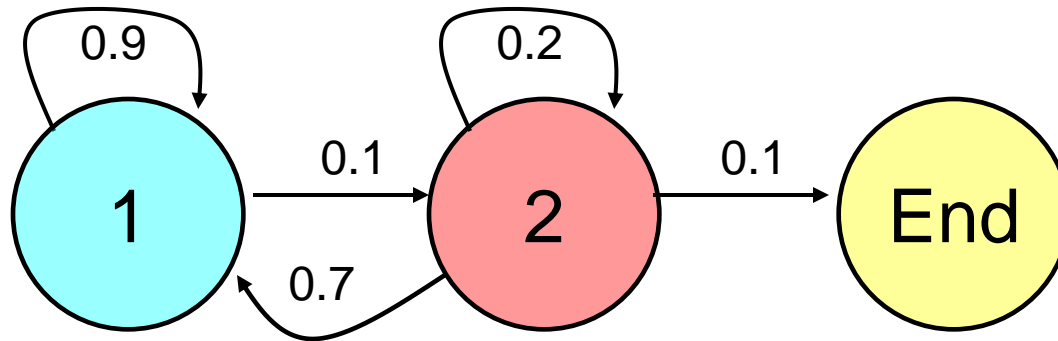
Fall 2014

Contents

- Reading materials
- Introduction to Hidden Markov Model
  - Markov chains
  - Hidden Markov Models
  - Three problems of HMMs
    - Calculate the probability from observations and the model
    - 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.

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

HMM is a generative model

- HMM can be used as a generator to produce an observation sequence $O=O_1O_2...O_T$, 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
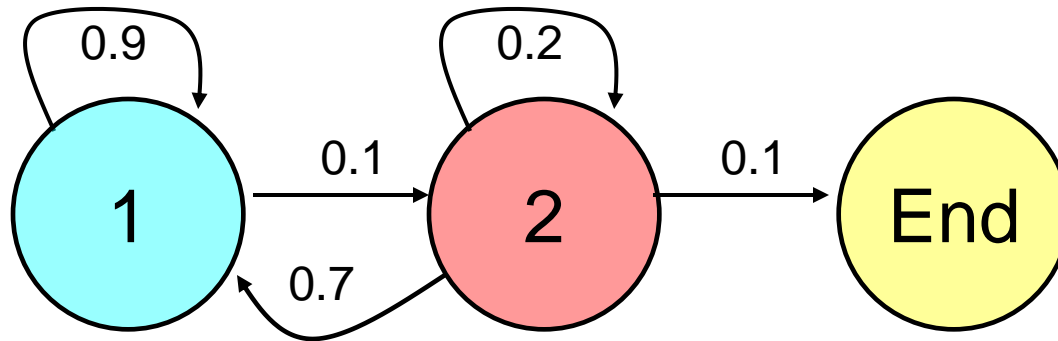
HMM for two biased coins flipping

$$e_1(H) = 0.8, e_1(T) = 0.2, e_2(H) = 0.3, e_2(T) = 0.7$$
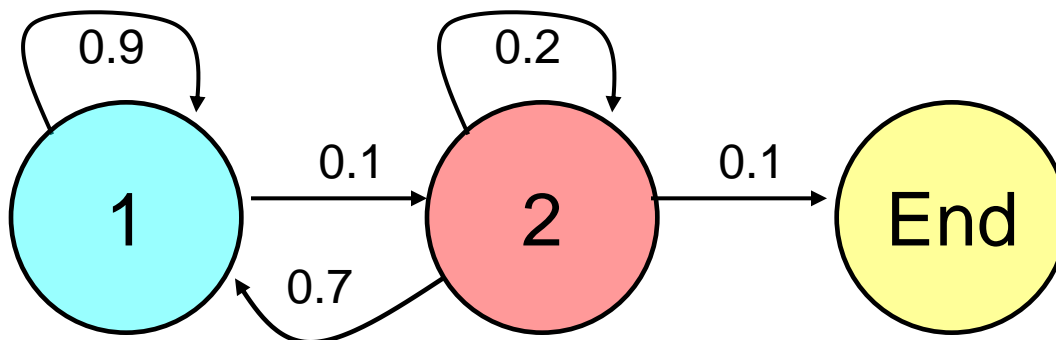
TTHHTTHTTTTTHTHHHHHTHTH   **Observed sequence x**

1122111112222111111112222   **Hidden state sequence** π

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

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

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

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

$$= 1 * 0.2 * (0.9 * 0.2) * (0.1 * 0.3) * (0.2 * 0.3) * (0.1 * 0.2) = 0.000001296$$

Hidden Markov Model

- HMM：λ={A, B,Init}

- Three basic problems for HMMs

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

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

  – Problem 3: how to estimate model parameters $\lambda=\{A, B, Init\}$ to maximize $P(O|\lambda)$.

Most Probable Path and Viterbi Algorithm
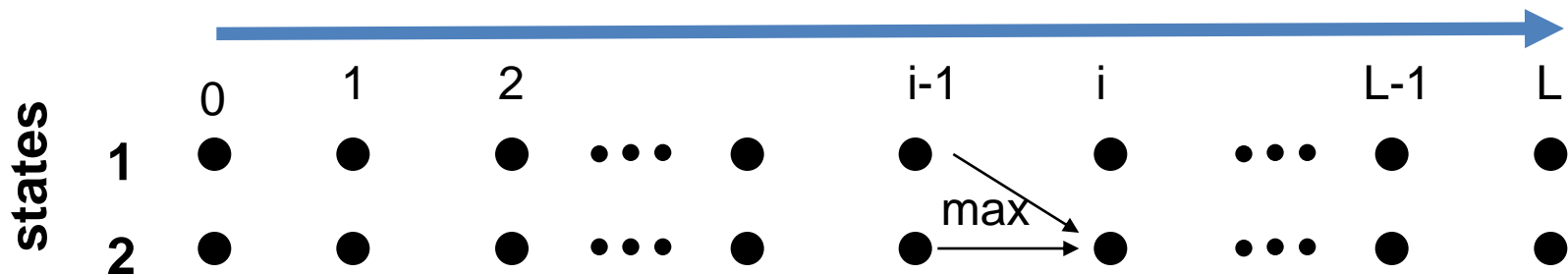
Let $$f_j(i) = \max_{\{\pi_0,...,\pi_{i-1}\}} (\Pr(x_0,...,x_{i-1},x_i,\pi_0,...,\pi_{i-1},\pi_i = j))$$

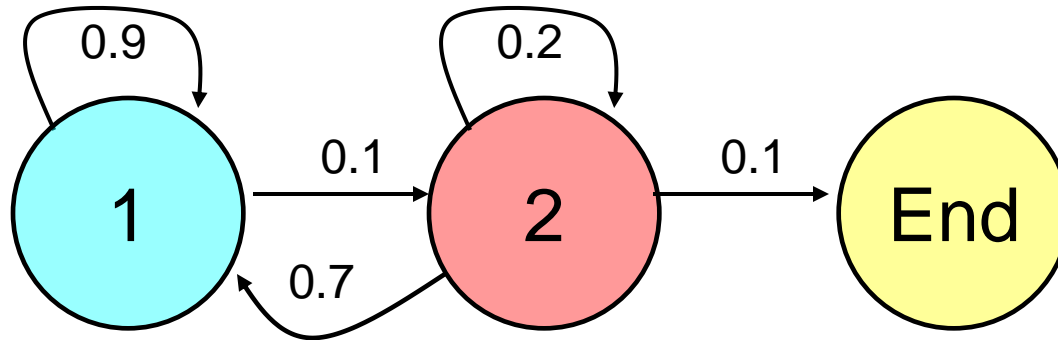Initialization (j=1…N) $$f_j(0) = \pi_j e_j(x_0)$$

Recursion (i=1…L)

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

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

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

Solution to problem 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

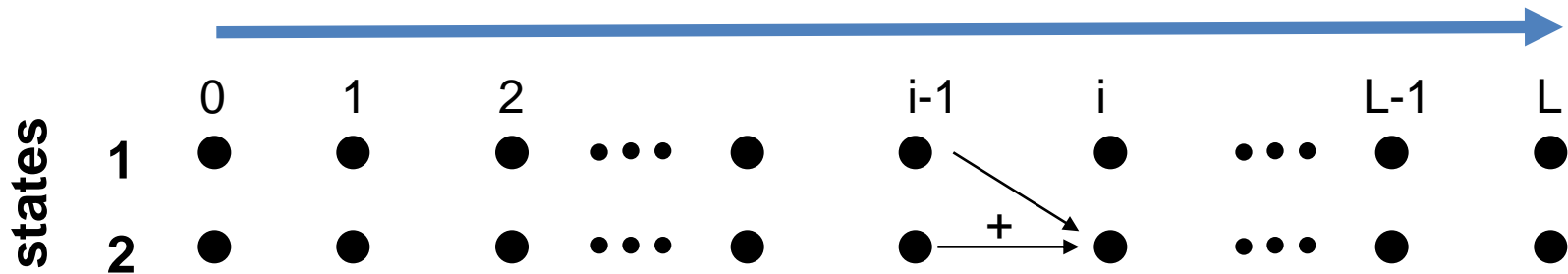|	T	T	H	H	T
	0	1	2	3	4
1	0.2	Max0.2*(0.9 *0.2, 0) = 0.036	max0.8*(0.036 *0.9, 0.014*0.7) = 0.0259	Max 0.8 * (0.0259*0.9, 0.0108*0.7) = 0.018662	=max0.2*(0.018662*0.9, 0.000777*0.7) = 0.003359
2	0	Max0.7*(0.2*0.1, 0) = 0.014	Max0.3*(0.036 *0.1, 0.014*0.2) = 0.0108	Max0.3(0.0259*0.1, 0.0108*0.2)=0.000777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

Probability of All the Possible Paths and Forward Algorithm

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

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

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

j = 1, .., N)

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

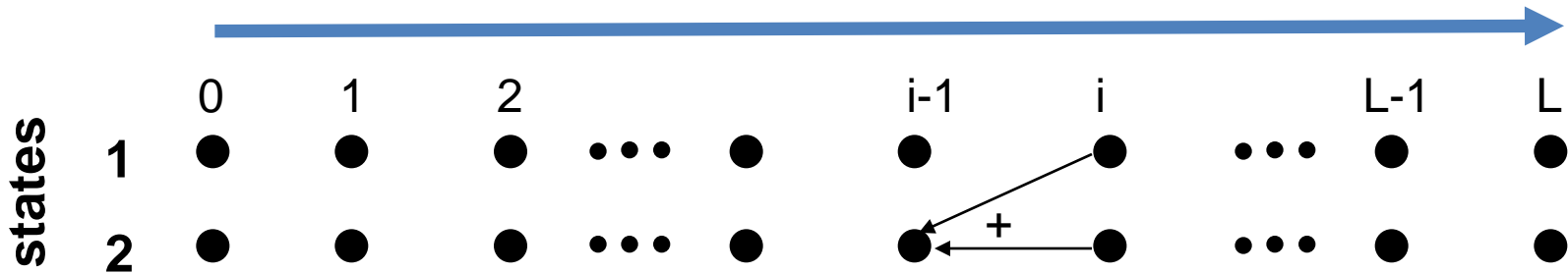Solution to problem 1

13

Posterior Probability and Forward and Backward Algorithm
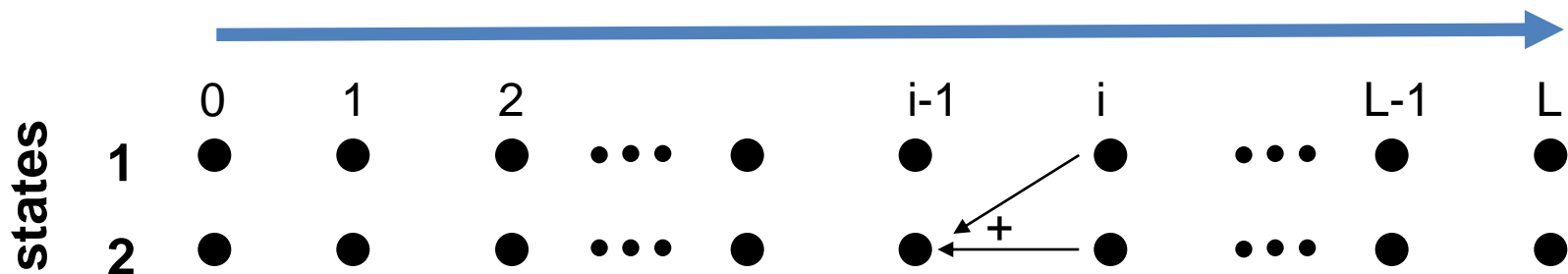
Posterior Probability

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

Backward Algorithm

Let $b_j(i) = \Pr(x_{i+1}, x_{i+2}..., x_L, \pi_i = j)$

Initialization (j=1…N) $b_j(L) = 1$
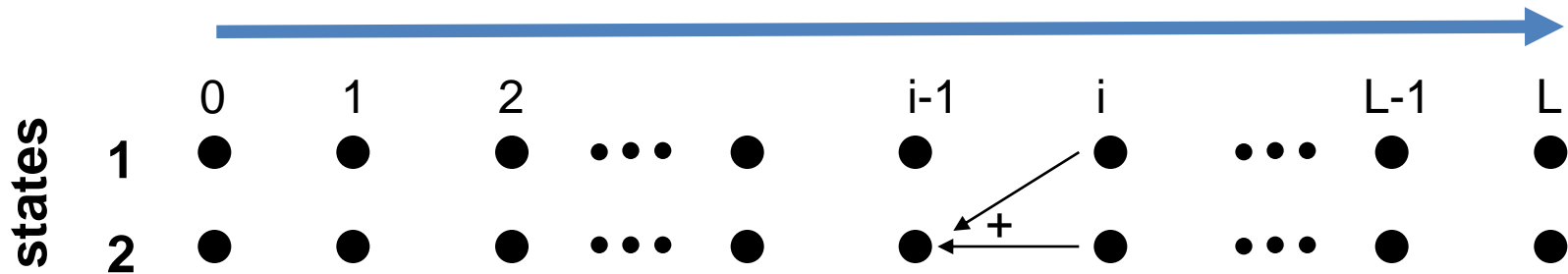
Recursion (i=L-1, L-2, …, 0, j=1, …, N) $b_j(i) = \sum_k (a_{jk} e_k(x_{i+1})) b_k(i+1)$

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

Posterior Probability and Forward and Backward Algorithm

Posterior Probability

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

$$= \frac{f_k(i) * b_k(i)}{\sum_k (f_k(i) * b_k(i))}$$

Problem 3: Optimize the model parameters from the observation

- HMM：λ={A, B, Init}
- With annotations
  - Maximum likely-hood ratio
- Without annotations
  - Baum-Welch algorithm (EM algorithm)

Baum-Welch algorithm (estimate model parameters)

- Goal：given the observation sequence data set, estimate the model parameter λ to maximize $P(O|\lambda)$.

- Algorithm:

  1. initialize the model λ_0,

  2. calculate the new model λ based on λ_0 and the observation sequences

  3. stop training if log $P(X|\lambda)$ - log($P(X|\lambda_0)$ < Delta

  4. otherwise, let $\lambda_0 = \lambda$, and go to step 2.

Baum-Welch method (EM method)

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

Let $\xi_t(i,j) = P(\pi_t = i, \pi_{t+1} = j \mid x, \lambda)$

then $\xi_t(i,j) = \dfrac{f_i(t)a_{ij}e_j(x_{t+1})b_j(t+1)}{\sum\limits_{i}^{N}\sum\limits_{j}^{N}(f_i(t)a_{ij}e_j(x_{t+1})b_j(t+1))}$

Let $\gamma_t(i) = \sum\limits_{j=1}^{N} \xi_t(i,j)$

then $\sum\limits_{t=0}^{L} \gamma_t(i)$ = expected number of transitions from S_i

$\sum\limits_{t=0}^{L} \xi_t(i,j)$ = expected number of transition S_i to S_j

Baum-Welch method (EM method) (2)

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

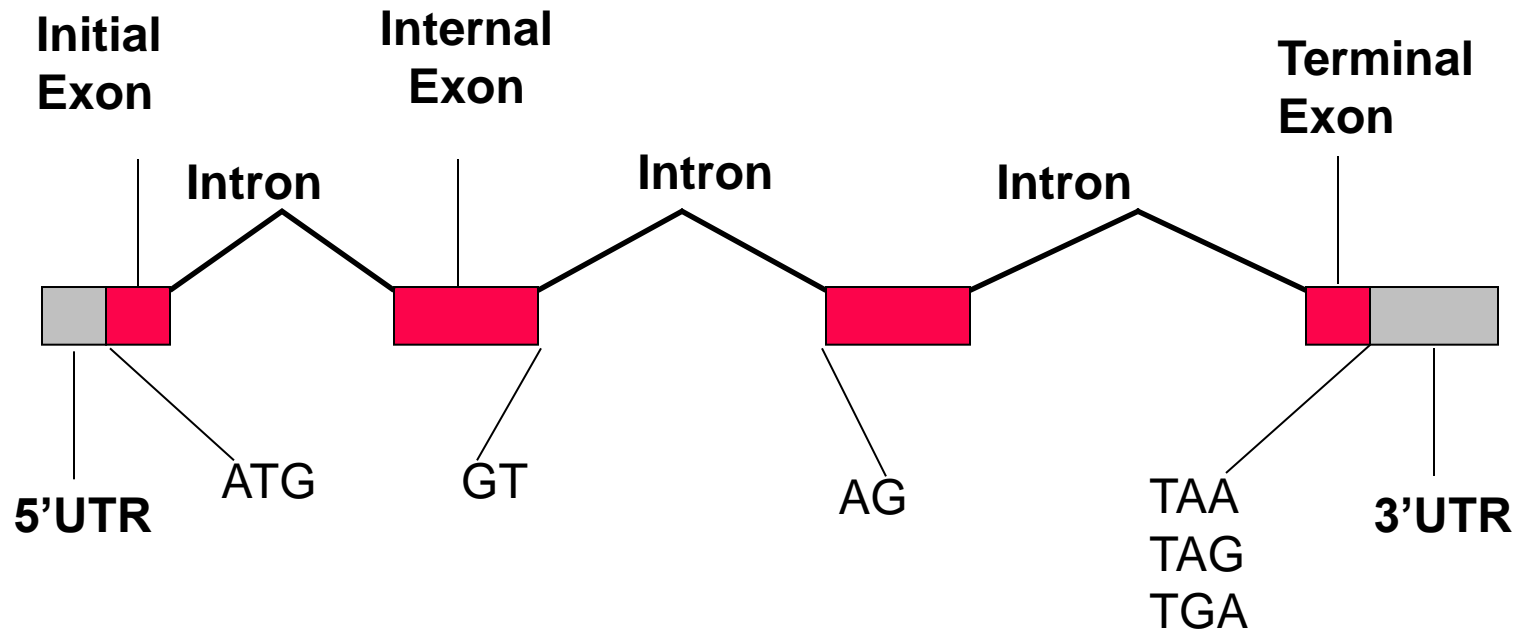Then, $\overline{Init_i} =$ expected frequency in S_i at time 0 = $\gamma_0(i)$

$$\overline{a_{i,j}} = \frac{\exp ected \quad number \quad of \quad tranistions \quad from \quad S_i \quad to \quad S_j}{\exp ected \quad number \quad of \quad tranistions \quad from \quad S_i}$$

$$= \frac{\sum_{t=0}^{L} \xi_t(i,j)}{\sum_{t=0}^{L} \gamma_t(i)}$$

$$\overline{e_i(k)} = \frac{\exp ected \quad number \quad of \quad times \quad in \quad state \quad j \quad and \, oberving \quad symbol \quad v_k}{\exp ected \quad number \quad of \quad times \quad in \quad state \quad j}$$

$$= \frac{\sum_{\substack{t=0 \\ s.t.x_t=v_k}}^{L} \gamma_t(i)}{\sum_{t=0}^{L} \gamma_t(i)}$$

Gene Structure

Initial Exon

Internal Exon

Terminal Exon

Intron   **Intron**   **Intron**

5'UTR   ATG   GT   AG   TAA   3'UTR
                           TAG
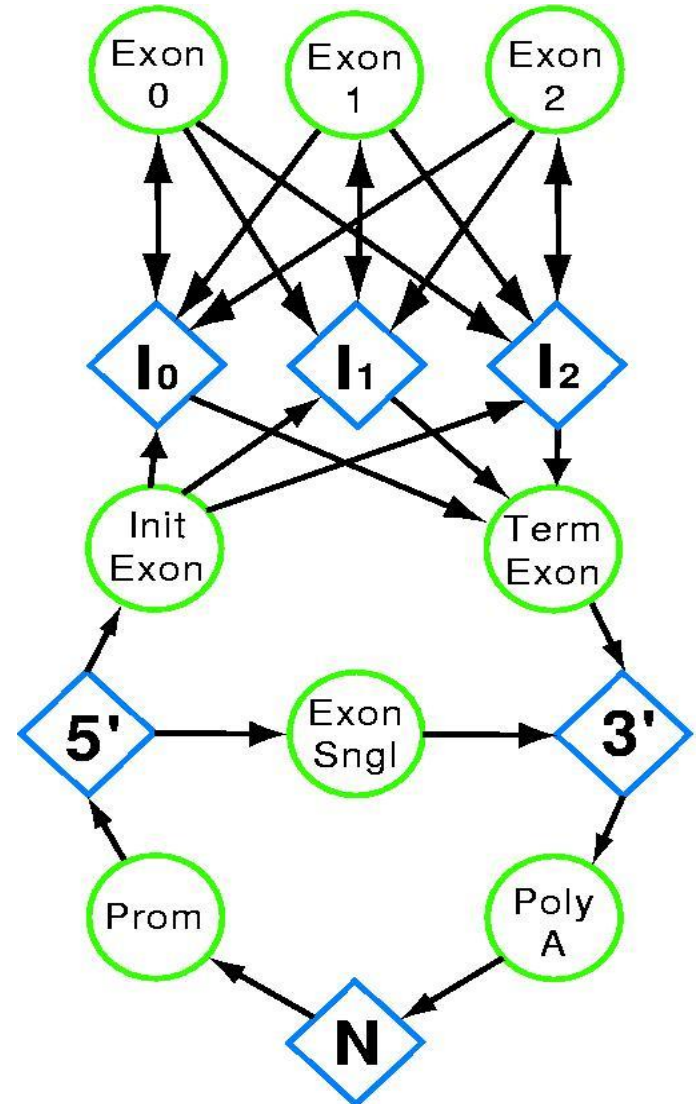   TGA

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

A Typical Human Gene Structure
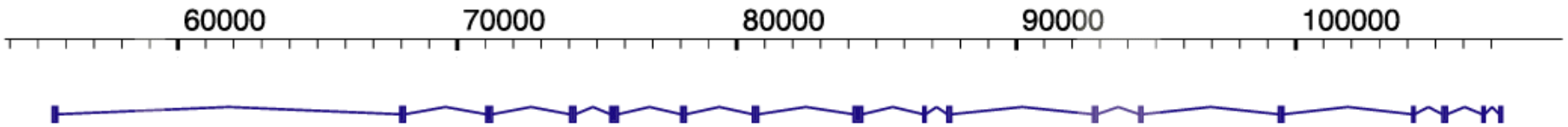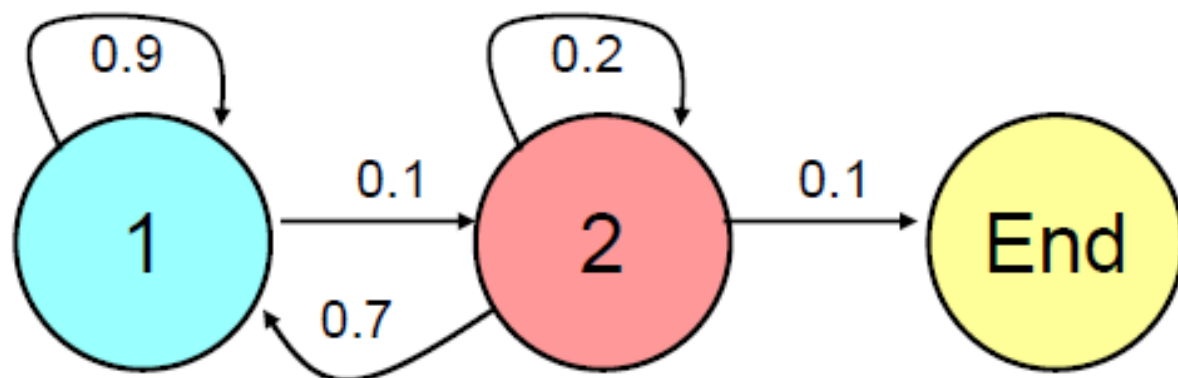
Gene Prediction Model

- **Generalized HMM**

- Each feature in a gene structure corresponds to one state.

- State-specific length models.

- State-specific sequence models

- Use Conservation information

23

Examples of

Hidden Markov Model (HMM)

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

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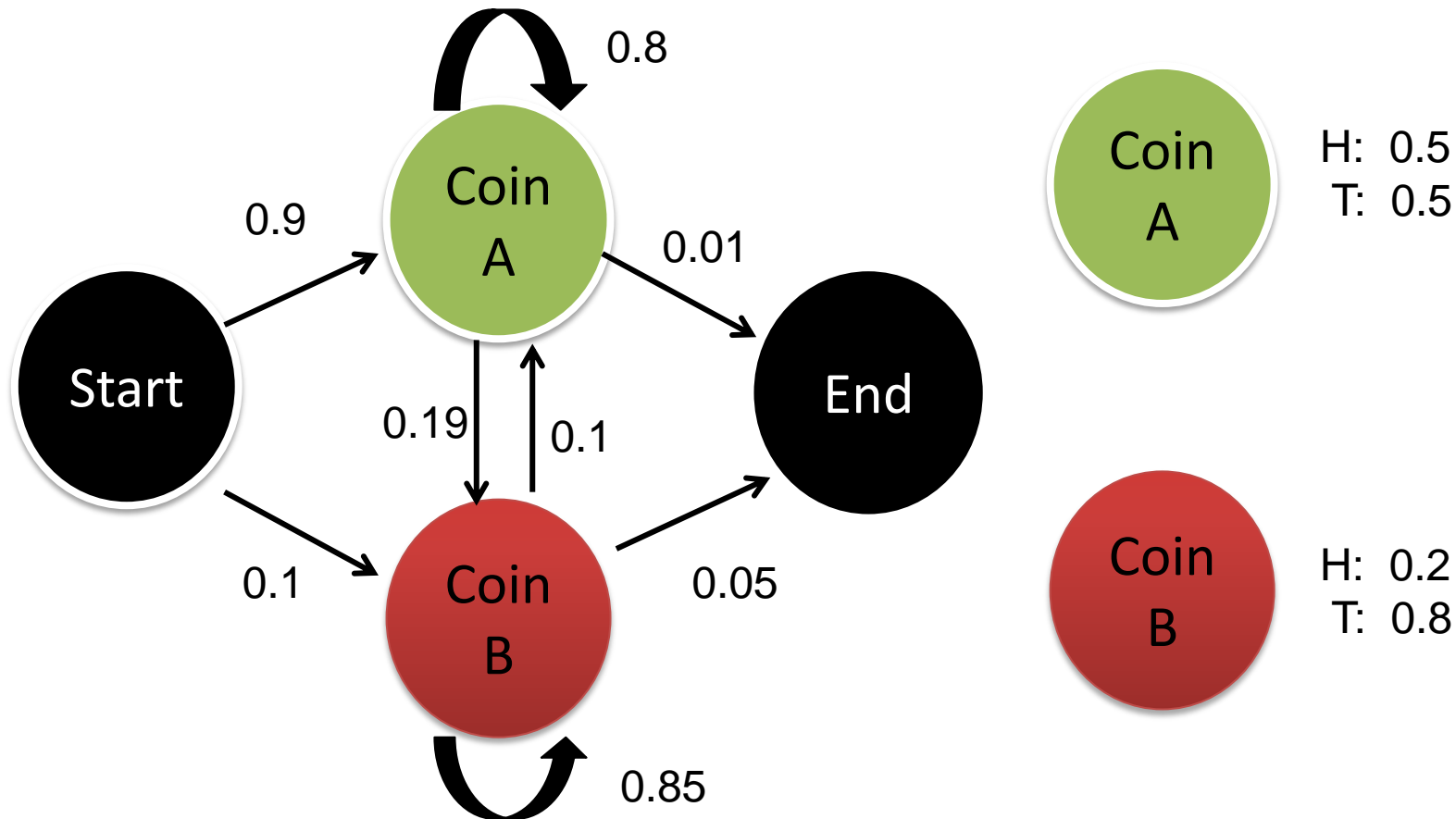
- Elements of an HMM (N, M, A, B, Init)
    1. N: number of states in the model
        - $S=\{S_1, S_2, ..., S_N\}$, and the state at time t is q_t.
    2. M: alphabet size (the number of observation symbols)
        - $V=\{v_1, v_2, ..., v_M\}$
    3. A: state transition probability distribution
 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

Hidden Markov Model

- HMM：λ={A, B,Init}

- Three basic problems for HMMs

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

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

  - Problem 3: how to estimate model parameters λ={A, B, Init} to maximize P(O| λ).
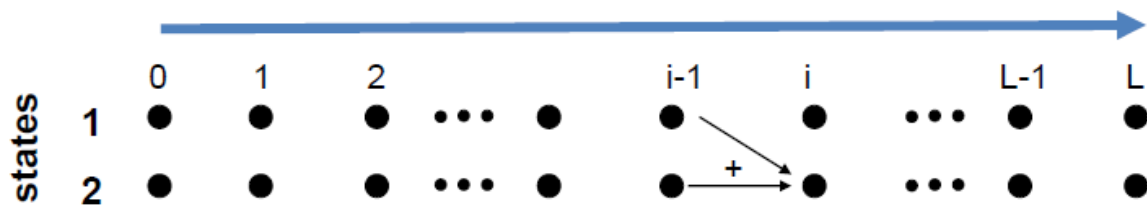
A case study

- Flipping two coins

Problem 1: Given the observation sequence $O = O_1 O_2 ... O_T$, and a model $\lambda = \{A, B, Init\}$, how to compute $P(O \mid \lambda)$, the probability of the observation sequence given the model?

- O= HHTHHTTTHT
- $P(O \mid \lambda) = ?$

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) $\quad f_j(0) = \pi_j e_j(x_0)$

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

j = 1, .., N)

Probability of all the probable paths $\quad P(x) = \sum_\pi P(x, \pi) = \sum_k f_k(L)$
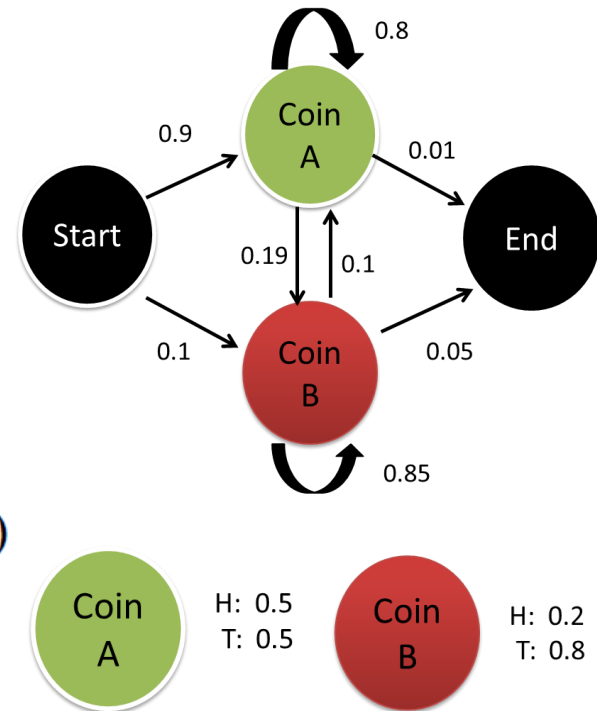
Problem 2: Given the observation sequence $O=O_1O_2...O_T$, and a model $\lambda=\{A, B, Init\}$, how to choose a corresponding state sequence $Q=q_1q_2...q_T$, which is optimal in some meaningful sense..

- O= HHTHHTTTHT
- $\underset{Q}{argmax}(P(O, Q, \lambda))$

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	
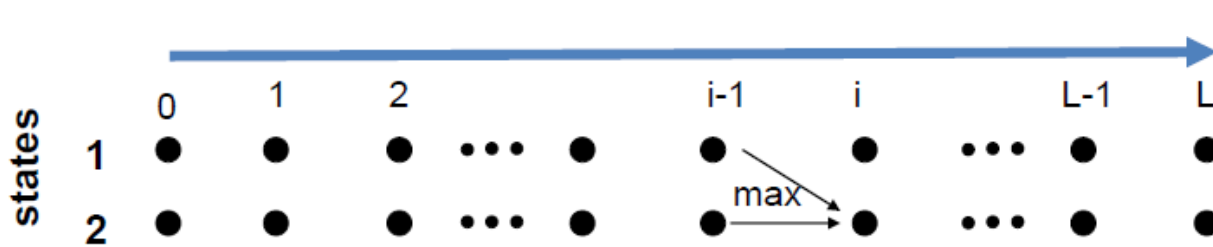
states

0 1 2 ... i-1 i ... L-1 L

1
2

max
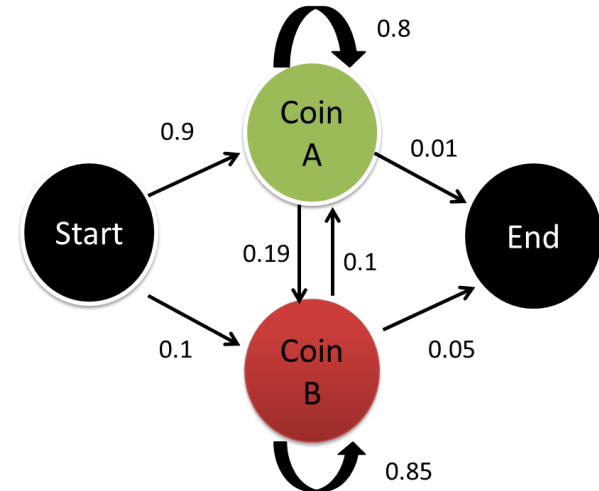
Let

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

Initialization (j=1…N)

$$f_j(0) = \pi_j e_j(x_0)$$

Recursion (i=1…L)

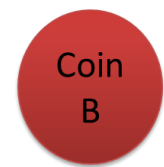$$f_j(i) = e_j(x_i) \max_k (f_k(i-1) a_{kj});$$

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

Start → Coin A : 0.9
Coin A self loop: 0.8
Coin A → End : 0.01
Coin A → Coin B : 0.19
Coin B → Coin A : 0.1
Start → Coin B : 0.1
Coin B → End : 0.05
Coin B self loop: 0.85

Coin A — H: 0.5 T: 0.5
Coin B — H: 0.2 T: 0.8

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.

Another example:
Pair HMM for local alignment