Lecture 2: Expectation-Maximization and Hidden Markov Models

Main Algorithms with Applications in Bioinformatics

Grégory Nuel

Laboratoire Statistique et Génome
University of Evry, CNRS (8071), INRA (1152)
France

Bioinformatics and Comparative Genome Analysis, Institut Pasteur of Tunis from March 18 to April 6, 2007
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications

Main Algorithms with Applications in Bioinformatics by G. Nuel
Outline

1. **EM Algorithm**
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. **Markov Models**
   - Order 0
   - Order 1 and More
   - Model Selection

3. **Hidden Markov Models**
   - Motivation and Notations
   - EM for HMM
   - Applications
The Problem

The model (parameter $\theta$):

- $X = X_1, \ldots, X_\ell$ the **observed** variables – ex: sizes (m), nucleotides, 
  
- $S = S_1, \ldots, S_\ell$ the **hidden** variables – ex: sex (male/female), gene status, 

the likelihood $L(\theta \mid X, S)$ is computable but

$$L(\theta \mid X) = \sum_S L(\theta \mid X, S)$$

is **untractable** (the sum have at least $2^\ell$ terms).

$\Rightarrow$ How to get the MLE of $\theta$ **without** the hidden variables?
EM algorithm (Dempster et al. 1977)

1. \textbf{start with an initial value }\theta_0 \text{ (random or arbitrary), } i = 0
2. \textbf{while }L(\theta_i \mid X) \text{ has not converged do}
3. \textbf{\hspace{1em} }i = i + 1
4. \textbf{\hspace{1em}} // phase \textit{E} (Expectation)
5. \textbf{\hspace{2em}} compute \( P(S \mid X, \theta_{i-1}) \)
6. \textbf{\hspace{2em}} compute \( L(\theta \mid X) \trianglerighteq E L(\theta \mid X, S) \)
7. \textbf{\hspace{1em}} // phase \textit{M} (Maximization)
8. \textbf{\hspace{2em}} compute \( \theta_i = \text{argmax}_\theta L(\theta \mid X) \)

The sequence of the \( L(\theta_i \mid X) \) grows till convergence and the sequence \( \theta_i \) converges towards the MLE (as long as the algorithm is not stuck in a local extremum).
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
Gaussian Mixture

The mixture model

All $X_i$ and $S_i$ are independent and identically distributed:

- $\mathbb{P}(S_i = 0) = \pi_0$ and if $S_i = 0$ then $X_i \sim \mathcal{N}(\mu_0, \sigma_0^2)$
- $\mathbb{P}(S_i = 1) = \pi_1$ and if $S_i = 1$ then $X_i \sim \mathcal{N}(\mu_1, \sigma_1^2)$

Phase E for the Gaussian mixture model

$$p_i = \mathbb{P}(S_i = 1 \mid X_i, \theta) = \frac{\pi_1 f_1(X_i)}{\pi_0 f_0(X_i) + \pi_1 f_1(X_i)}$$

with $f_0$ and $f_1$ are the density of $\mathcal{N}(\mu_0, \sigma_0^2)$ and $\mathcal{N}(\mu_1, \sigma_1^2)$
EM for a Gaussian Mixture

Phase M for the Gaussian mixture model

\[
\pi_0 = \frac{\sum_i (1 - p_i)}{\ell} \quad \pi_1 = \frac{\sum_i p_i}{\ell}
\]

\[
\mu_0 = \frac{1}{\sum_i (1 - p_i)} \sum_i (1 - p_i) X_i \quad \mu_1 = \frac{1}{\sum_i p_i} \sum_i p_i X_i
\]

\[
\sigma_0 = \sqrt{\frac{1}{\sum_i (1 - p_i)} \sum_i (1 - p_i) X_i^2 - \mu_1^2}
\]

\[
\sigma_1 = \sqrt{\frac{1}{\sum_i p_i} \sum_i p_i X_i^2 - \mu_1^2}
\]

with \( p_i = \Pr(S_i = 1 \mid X_i, \theta) \)
Example \((\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5))\)

**Iteration 0: \(\text{theta}=(0.5,0.5,-1,1,1,1)\)**

The figure shows a plot of \(P(S_i=1 | \theta, X)\) against \(i\) for the first iteration of the EM algorithm.
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 1: \( \text{theta}=(0.2,0.8, -1.22, 1.54, 2.44, 1.12) \)
Example \((\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5))\)

Iteration 2: \(\theta = (0.22, 0.78, -1.13, 1.54, 2.5, 1.05)\)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 3: \( \theta = (0.23, 0.77, -1.03, 1.55, 2.55, 1) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 4: \( \theta = (0.25, 0.75, -0.93, 1.57, 2.59, 0.97) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

Iteration 5: \( \theta = (0.26, 0.74, -0.84, 1.59, 2.62, 0.94) \)
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

Example ($\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)$)

iteration 6: theta=(0.28,0.72,-0.74,1.62,2.64,0.92)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 7: \( \theta = (0.29, 0.71, -0.66, 1.65, 2.67, 0.9) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 8: \( \theta = (0.3, 0.7, -0.58, 1.68, 2.69, 0.88) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 9: \( \theta = (0.31, 0.69, -0.5, 1.7, 2.71, 0.86) \)
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 10: \( \theta = (0.32, 0.68, -0.43, 1.72, 2.73, 0.85) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 11: \( \theta = (0.33, 0.67, -0.36, 1.75, 2.75, 0.84) \)
**EM Algorithm**

**Markov Models**

**Hidden Markov Models**

**The Principle**

**A Detailed Example**

**Application: Finding Patterns in Biopolymers**

**Main Algorithms with Applications in Bioinformatics** by G. Nuel

**Lecture 2: Expectation-Maximization and Hidden Markov Models**

---

**EM iterations**

**Example** \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

**Iteration 12:** \( \theta = (0.34, 0.66, -0.3, 1.76, 2.76, 0.82) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 13: \( \theta = (0.35, 0.65, -0.24, 1.78, 2.78, 0.81) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 14: \( \theta = (0.36, 0.64, -0.18, 1.8, 2.79, 0.79) \)
EM Algorithm
Markov Models
Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 15: \( \theta = (0.37, 0.63, -0.12, 1.81, 2.81, 0.78) \)
EM Algorithm
Markov Models
Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 16: \( \theta = (0.38, 0.62, -0.07, 1.83, 2.82, 0.76) \)
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

Main Algorithms with Applications in Bioinformatics by G. Nuel

Lecture 2: Expectation-Maximization and Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 17: \( \theta = (0.4, 0.6, -0.02, 1.84, 2.84, 0.75) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 18: \( \theta = (0.41, 0.59, 0.03, 1.86, 2.85, 0.73) \)
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

---

EM iterations

Example \( (\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)) \)

iteration 19: \( \theta = (0.42, 0.58, 0.08, 1.87, 2.86, 0.71) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

Iteration 20: \( \theta = (0.43, 0.57, 0.13, 1.88, 2.88, 0.7) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 21: \( \theta = (0.44, 0.56, 0.18, 1.9, 2.89, 0.68) \)
Example ($\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)$)

Iteration 22: $\theta = (0.45, 0.55, 0.23, 1.91, 2.9, 0.66)$
EM Algorithm
Markov Models
Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 23: \( \theta = (0.45, 0.55, 0.27, 1.92, 2.91, 0.64) \)
Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 24: \( \theta = (0.46, 0.54, 0.32, 1.93, 2.92, 0.63) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 25: \( \theta = (0.47, 0.53, 0.36, 1.95, 2.92, 0.61) \)
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 26: \( \theta = (0.48, 0.52, 0.39, 1.96, 2.93, 0.6) \)

P(Si=1 | theta,X)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

Iteration 27: \( \theta = (0.49, 0.51, 0.43, 1.96, 2.93, 0.58) \)

\( P(S_i = 1 | \theta, X) \)
**Example** \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

**Iteration 28:** \( \theta = (0.49, 0.51, 0.45, 1.97, 2.94, 0.57) \)
Example \( (\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)) \)

iteration 29: \( \theta = (0.5, 0.5, 0.48, 1.98, 2.94, 0.57) \)
Example ($\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)$)

iteration 30: $\theta=(0.5, 0.5, 0.5, 1.98, 2.94, 0.56)$
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 31: \( \theta = (0.51, 0.49, 0.51, 1.98, 2.94, 0.56) \)

Main Algorithms with Applications in Bioinformatics by G. Nuel
Lecture 2: Expectation-Maximization and Hidden Markov Models
EM Algorithm
Markov Models
Hidden Markov Models

A Detailed Example
Application: Finding Patterns in Biopolymers

Main Algorithms with Applications in Bioinformatics by G. Nuel
Lecture 2: Expectation-Maximization and Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 32: \( \theta = (0.51, 0.49, 0.53, 1.99, 2.95, 0.55) \)

Graph showing the iterative process of the EM algorithm.
Example \( (\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)) \)

iteration 33: \( \theta = (0.51, 0.49, 0.54, 1.99, 2.95, 0.55) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 34: \( \theta = (0.52, 0.48, 0.55, 1.99, 2.95, 0.55) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 35: \( \theta = (0.52, 0.48, 0.55, 1.99, 2.95, 0.54) \)
Example \((\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5))\)

iteration 36: \(\theta=(0.52,0.48,0.56,1.99,2.95,0.54)\)
EM Algorithm
Markov Models
Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

Iteration 37: \( \theta = (0.52, 0.48, 0.56, 1.99, 2.95, 0.54) \)
Example ($\theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5)$)

iteration 38: $\theta=(0.52, 0.48, 0.57, 1.99, 2.95, 0.54)$
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 39: \( \theta = (0.52, 0.48, 0.57, 1.99, 2.95, 0.54) \)
EM Algorithm
Markov Models
Hidden Markov Models

Main Algorithms with Applications in Bioinformatics by G. Nuel

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 40: \( \theta = (0.52, 0.48, 0.57, 2, 2.95, 0.54) \)
EM Algorithm
Markov Models
Hidden Markov Models

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 41: \( \theta = (0.52, 0.48, 0.57, 2, 2.95, 0.54) \)
EM Algorithm
Markov Models
Hidden Markov Models

The Principle
A Detailed Example
Application: Finding Patterns in Biopolymers

EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 42: \( \theta = (0.52, 0.48, 0.57, 2, 2.95, 0.54) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 43: \( \theta = (0.52, 0.48, 0.57, 2, 2.95, 0.54) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 44: \( \theta = (0.52, 0.48, 0.57, 2, 2.95, 0.54) \)
EM iterations

Example \( \theta = (\pi_0, \pi_1, \mu_0, \sigma_0, \mu_1, \sigma_1) = (0.55, 0.45, 1, 2, 3, 0.5) \)

iteration 45: \( \theta = (0.52, 0.48, 0.57, 2, 2.95, 0.54) \)
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
The Problem

Given a set of $N$ biological sequences $Y_1, \ldots, Y_N$ we want to infer a pattern (or motif) associated to the dataset.

MEME approach (Bailey and Elkan, 1994)

We choose a length $h$ and perform

- split the sequences $Y_j$ in overlapping words of length $h$.
  Example: if $Y_1 = \text{atgct}a$ and $Y_2 = \text{aaat}a$ ($N = 2$) we get $X_1 = \text{atgc}$, $X_2 = \text{tgct}$, $X_3 = \text{gcta}$, $X_4 = \text{aat}a$ and $X_5 = \text{aat}$ a ($\ell = 5$).
- fit a mixture model on $X_1, \ldots, X_\ell$ using the EM algorithm

$\Rightarrow$ the parameters of the mixture model give the pattern’s profile.
The mixture model

\[ X = X_1 \ldots X_\ell \text{ with } X_i \in \mathcal{A}^h, \quad S = S_1 \ldots S_\ell \text{ with } S_i \in \{0, 1\}. \]

We assume that both \( X_i \) and \( S_i \) are independent and identically distributed with \( S_i \sim \mathcal{B}(\rho) \) and

\[
\begin{align*}
\mathbb{P}(X_i = a_1 \ldots a_h | S_i = 0) &= \mu_0(a_1) \ldots \mu_0(a_h) \\
\mathbb{P}(X_i = a_1 \ldots a_h | S_i = 1) &= \mu_1(a_1) \ldots \mu_h(a_h)
\end{align*}
\]

where \( \theta = (\rho, \mu_0, \mu_1, \ldots, \mu_h) \) is the parameter of the model.

Remarks

- independance assumption completely false
- the mixture model is very simple

⇒ MEME is a useful heuristic but HMM gives a more elegant solution to the problem.
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
### Definition (M00 model)

Let $X = X_1 \ldots X_\ell$ a random sequence over the (size $k$) alphabet $A$ (ex: $A = \{a, c, g, t\}$ for DNA). $X$ is under the M00 model if:

- all letters $X_i$ are independent and identically distributed
- $P(X_i = a) = 1/k$ for all $a \in A$ (uniform distribution)

### Remarks

- simple and easy to understand
- often implicitly used as the random model
- but absolutely unrealistic
## Example (In the HIV1 complete genome $\ell = 9718$)

<table>
<thead>
<tr>
<th>letter</th>
<th>a</th>
<th>c</th>
<th>g</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>expected under M00</td>
<td>2429.5</td>
<td>2429.5</td>
<td>2429.5</td>
<td>2429.5</td>
</tr>
<tr>
<td>observed</td>
<td>3411</td>
<td>1773</td>
<td>2370</td>
<td>2164</td>
</tr>
</tbody>
</table>

Pearson’s test gives $\chi^2(3) = 604.4$ which is highly significant.

## Definition (Shuffle model)

If $x = x_1 \ldots x_\ell$ is an observed sequence a random sequence $X = X_1 \ldots X_\ell$ is under the **shuffle model** if $X_i = x_{\sigma(i)}$ where $\sigma$ is drawn uniformly in the set of all permutations of $\{1, \ldots, \ell\}$.
Shuffling

**Algorithm**

- $x = x_1 \ldots x_\ell$ is the observed sequence

1. draw a sample $r_1, \ldots, r_\ell$ of random number ($\sim \mathcal{U}[0, 1]$)
2. (quick)sort the sample: $r_{\sigma(1)}, \ldots, r_{\sigma(\ell)}$
3. return $X = x_{\sigma(1)}, \ldots, x_{\sigma(\ell)}$

⇒ complexity is $O(\ell \log \ell)$ (quicksort)

**Example**

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>a</th>
<th>c</th>
<th>c</th>
<th>t</th>
<th>a</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>$r_i$</td>
<td>0.76</td>
<td>0.00</td>
<td>0.33</td>
<td>0.67</td>
<td>0.63</td>
<td>0.85</td>
</tr>
<tr>
<td>$r_{\sigma(i)}$</td>
<td>0.00</td>
<td>0.33</td>
<td>0.63</td>
<td>0.67</td>
<td>0.76</td>
<td>0.85</td>
</tr>
<tr>
<td>$\sigma(i)$</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>$X_i$</td>
<td>c</td>
<td>c</td>
<td>a</td>
<td>t</td>
<td>a</td>
<td>g</td>
</tr>
</tbody>
</table>
**Definition (M0 model)**

Let $X = X_1 \ldots X_\ell$ a **random sequence** over the (size $k$) alphabet $\mathcal{A}$. $X$ is under the **M0 model** with parameter $\mu$ if:

- all letters $X_i$ are **independant** and identically distributed
- $P(X_i = a) = \mu(a)$ for all $a \in \mathcal{A}$

**Proposition**

The **log-likelihood** of the model considering $x = x_1 \ldots x_\ell$ is given by

$$L = \sum_{a \in \mathcal{A}} F_x(a) \log \mu(a)$$
Corollary

The Maximum Likelihood Estimator of $\mu$ is therefore given by

$$\hat{\mu}(a) = \frac{F_x(a)}{\ell} \quad \forall a \in \mathcal{A}$$

Example ($x = \text{acctag}$)

$$L = \log (\mu(a)\mu(c)\mu(c)\mu(t)\mu(a)\mu(g))$$

$$= \log \left( \mu(a)^2 \mu(c)^2 \mu(g) \mu(t) \right)$$

$$= 2 \log \mu(a) + 2 \log \mu(c) + \log \mu(g) + \log \mu(t)$$

$$\hat{\mu}(a) = \frac{2}{6} \quad \hat{\mu}(c) = \frac{2}{6} \quad \hat{\mu}(g) = \frac{1}{6} \quad \hat{\mu}(t) = \frac{1}{6}$$
EM Algorithm
Markov Models
Hidden Markov Models
Order 0
Order 1 and More
Model Selection

Drawing a sequence under M0

Algorithm ($\mu$ and $\ell$ are given)

1: compute $G$ the cumulative distribution of $\mu$
2: for all $i = 1 \ldots \ell$ do
3: draw a random number $r \sim \mathcal{U}[0, 1]$
4: $X_i$ is the smallest $a$ such as $G(a) > r$

$\Rightarrow$ complexity is $O(k \times \ell)$

Example

<table>
<thead>
<tr>
<th>$a$</th>
<th>a</th>
<th>c</th>
<th>g</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu(a)$</td>
<td>0.32</td>
<td>0.33</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>$G(a)$</td>
<td>0.32</td>
<td>0.65</td>
<td>0.82</td>
<td>1.00</td>
</tr>
<tr>
<td>$i$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$r$</td>
<td>0.76</td>
<td>0.00</td>
<td>0.33</td>
<td>0.67</td>
</tr>
<tr>
<td>$X_i$</td>
<td>g</td>
<td>a</td>
<td>c</td>
<td>g</td>
</tr>
</tbody>
</table>

Main Algorithms with Applications in Bioinformatics by G. Nuel
Lecture 2: Expectation-Maximization and Hidden Markov Models
$x = x_1 \ldots x_\ell$ and observed sequence and we compare the shuffle model to the M0 model (parameters are estimated on $x$):

- randomly shuffled sequences have exactly the same nucleotide frequencies than $x$
- random M0 sequences have on the average the same nucleotide frequencies than $x$
- shuffling is (slightly) slower than drawing under M0
- shuffling requires the original sequence
- shuffling is difficult (but not impossible) to extend to dinucleotide frequencies
- statistical properties of M0 are well known

$\Rightarrow$ no more shuffling from now
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
Taking into account dinucleotide frequencies

Example (In the HIV1 complete genome $\ell = 9718$ )

<table>
<thead>
<tr>
<th>observed</th>
<th>a</th>
<th>c</th>
<th>g</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1112</td>
<td>561</td>
<td>1024</td>
<td>713</td>
</tr>
<tr>
<td>c</td>
<td>795</td>
<td>413</td>
<td>95</td>
<td>470</td>
</tr>
<tr>
<td>g</td>
<td>820</td>
<td>457</td>
<td>661</td>
<td>432</td>
</tr>
<tr>
<td>t</td>
<td>684</td>
<td>342</td>
<td>590</td>
<td>548</td>
</tr>
</tbody>
</table>

$$\frac{F(ac)}{F(a.)} = \frac{561}{3410} = 16.45\% \quad \frac{F(cc)}{F(c.)} = \frac{95}{1773} = 5.36\%$$

$\Rightarrow$ we need to introduce some dependance
**M1 model**

**Definition (M1 model)**

$X = X_1 \ldots X_\ell$ is drawn according to the M1 model with starting distribution $\mu_1$ and transition matrix $\pi$ if:

- $\mathbb{P}(X_1 = a) = \mu_1(a)$ for all $a \in \mathcal{A}$
- $\mathbb{P}(X_i \mid X_1, \ldots, X_{i-1}) = \mathbb{P}(X_i \mid X_{i-1}) = \pi(X_{i-1}, X_i)$

**Example (over the binary alphabet $\mathcal{A} = \{a, b\}$)**

$\mu_1 = \begin{pmatrix} 0.5 & 0.5 \end{pmatrix} \quad \pi = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}$

means that $\mathbb{P}(X_1 = a) = \mathbb{P}(X_1 = b) = 0.5$ and that

$\mathbb{P}(X_{i+1} = a \mid X_i = a) = 0.9 \quad \mathbb{P}(X_{i+1} = b \mid X_i = a) = 0.1$

$\mathbb{P}(X_{i+1} = a \mid X_i = b) = 0.3 \quad \mathbb{P}(X_{i+1} = b \mid X_i = b) = 0.7$
Likelihood, MLE

**Proposition**

The log-likelihood of the model considering $\mathbf{x} = x_1 \ldots x_\ell$ is given by

$$L = \log \mu_1(x_1) + \sum_{a,b \in A} F_x(a, b) \log \pi(a, b)$$

**Corollary**

The MLE of $\mu_1$ and $\pi$ are hence

$$\hat{\mu}_1(a) = \begin{cases} 1 & \text{if } a = x_1 \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \hat{\pi}(a, b) = \frac{F_x(ab)}{F_x(a.)} \quad \forall a, b \in A$$
A simple example

Example ($x = tgggaag \ldots$ is HIV1 $\ell = 9718$)

<table>
<thead>
<tr>
<th>observed</th>
<th>a</th>
<th>c</th>
<th>g</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1112</td>
<td>561</td>
<td>1024</td>
<td>713</td>
</tr>
<tr>
<td>c</td>
<td>795</td>
<td>413</td>
<td>95</td>
<td>470</td>
</tr>
<tr>
<td>g</td>
<td>820</td>
<td>457</td>
<td>661</td>
<td>432</td>
</tr>
<tr>
<td>t</td>
<td>684</td>
<td>342</td>
<td>590</td>
<td>548</td>
</tr>
</tbody>
</table>

$L = \log (\mu_1(t) \pi(t, g) \pi(g, g) \pi(g, a) \ldots) = \log \left(\mu_1(t) \pi(a, a)^{1112} \pi(a, c)^{561} \ldots\right) = \log \mu_1(t) + 1112 \log \pi(a, a) + 561 \log \pi(a, c) + \ldots$

$\hat{\pi}(a, a) = \frac{1112}{3410}$ \quad $\hat{\pi}(a, c) = \frac{561}{3410}$ \quad $\hat{\pi}(a, g) = \frac{1024}{3410}$ \quad $\ldots$
Extension to Mm

**Definition (Mm model)**

\(X = X_1 \ldots X_\ell\) is drawn according to the Mm model (\(m \geq 1\)) with starting distribution \(\mu_1\) and parameter \(\pi\) if:

- \(\mathbb{P}(X_1 = a_1, \ldots, X_m = a_m) = \mu_1(a_1, \ldots, a_m)\) for all \(a_i \in A\)
- \(\mathbb{P}(X_i \mid X_{i-m}, \ldots, X_{i-1}) = \pi(X_{i-m}, \ldots, X_{i-1}, X_i)\)

**Proposition**

The MLE of \(\pi\) is given by

\[
\hat{\pi}(a_1, \ldots, a_m, b) = \frac{F_X(a_1 \ldots a_mb)}{F_X(a_1 \ldots a_m.)} \quad \forall a_1, \ldots, a_m, b \in A
\]
Properties of the transition matrix

**Remark**

Back to the M1 model with transition matrix $\pi$:

- $\mathbb{P}(X_{i+1} = b \mid X_i = a) = \pi(a, b)$
- $\mathbb{P}(X_{i+2} = b \mid X_i = a) = \sum_{c \in A} \pi(a, c)\pi(c, b) = \pi^2(a, b)$
- $\ldots$
- $\mathbb{P}(X_{i+d} = b \mid X_i = a) = \pi^d(a, b)$

$\Rightarrow \mu_1 \times \pi^d$ gives the distribution of $X_{1+d}$
### Convergence of the marginal distribution

**Example (over the binary alphabet \( \mathcal{A} = \{a, b\} \))**

\[
\begin{align*}
\mu_1 &= \left( \begin{array}{cc}
0.0 & 1.0
\end{array} \right) \\
\pi &= \left( \begin{array}{cc}
0.9 & 0.1 \\
0.3 & 0.7
\end{array} \right)
\end{align*}
\]

| \( \mu_1 \times \pi^1 \) | 0.3 | 0.7 |
| \( \mu_1 \times \pi^2 \) | 0.48 | 0.52 |
| \( \mu_1 \times \pi^3 \) | 0.588 | 0.412 |
| \( \mu_1 \times \pi^4 \) | 0.6528 | 0.3472 |
| \( \mu_1 \times \pi^5 \) | 0.69168 | 0.30832 |
| \( \mu_1 \times \pi^6 \) | 0.715008 | 0.284992 |
| \( \mu_1 \times \pi^7 \) | 0.7290048 | 0.2709952 |
| \( \vdots \) | \( \vdots \) | \( \vdots \) |
| \( \mu_1 \times \pi^{40} \) | 0.75 | 0.25 |
| \( \mu_1 \times \pi^{41} \) | 0.75 | 0.25 |
**Definition (Ergodic Markov chain)**

A Markov chain with transition matrix $\pi$ is said **ergodic** if and only if it exist $\mu$ such as $\mu_1 \times \pi^d$ (quickly) converges towards $\mu$ for all $\mu_1$. $\mu$ is then called the **stationary distribution** of the Markov chain and verifies $\mu \times \pi = \mu$.

**Definition (Stationary Markov chain)**

An ergodic Markov chain which starting distribution is the stationary one ($\mu$) is called a **stationary Markov chain**. For such a Markov chain $\mu \times \pi^d = \mu$ for all $d$ and hence all $X_i$ are distributed according to $\mu$ (but are *not* independant).
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
Remark

- the (log)likelihood $L$ of $M^m$ grows with $m$
- over a size $k$ alphabet, the $M^m$ model has $(k - 1)k^m$ free parameters

$\Rightarrow$ find a tradeoff between $m$ and the number of parameters

Definition (Penalized likelihood)

There are two common criteria of penalized likelihood:

- **AIC** = $-2L + 2K$
- **BIC** = $-L + K \log \ell$

where $K$ is the number of free parameters and $\ell$ the length of the data.
### Example *(Escherichia coli K12 $\ell = 4.6$Mb and HIV $\ell = 10$Kb)*

**With the AIC:**

<table>
<thead>
<tr>
<th>modèle</th>
<th>M00</th>
<th>M0</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV1</td>
<td>26.95</td>
<td>26.37</td>
<td>25.80</td>
<td><strong>25.68</strong></td>
<td>25.70</td>
<td>26.10</td>
<td>28.03</td>
<td>40.00</td>
<td>106.00</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>12863</td>
<td>12861</td>
<td>12743</td>
<td>12626</td>
<td>12546</td>
<td>12497</td>
<td>12456</td>
<td><strong>12435</strong></td>
<td>12443</td>
</tr>
</tbody>
</table>

**With the BIC:**

<table>
<thead>
<tr>
<th>modèle</th>
<th>M00</th>
<th>M0</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV</td>
<td>26.95</td>
<td>26.39</td>
<td><strong>25.89</strong></td>
<td>26.03</td>
<td>27.08</td>
<td>31.62</td>
<td>50.10</td>
<td>128.26</td>
<td>459.00</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>12863</td>
<td>12862</td>
<td>12743</td>
<td>12627</td>
<td>12548</td>
<td>12508</td>
<td><strong>12497</strong></td>
<td>12599</td>
<td>13099</td>
</tr>
</tbody>
</table>
Application to CDS Finding: Genmark

Genmark’s algorithm (Borodovsky and Mc Ininch, 1993)

1: list all ORFs of a genome
2: estimate the CDS Markov model (ex: $m = 4$) on the set of large enough ORFs
3: estimate the nonCDS Markov model (ex: $m = 2$) on the set of small enough ORFs
4: for all remaining ORFs do
5: compute $L_{CDS}$ of the CDS model on the ORF
6: compute $L_{nonCDS}$ of the nonCDS model on the ORF
7: if $\text{BIC}_{CDS} \geq \text{BIC}_{nonCDS}$ then
8: classify the ORF as a CDS
9: else
10: classify the ORF as a nonCDS
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
There is **multiple sources of heterogeneity** in biological sequences: gc content, genes, secondary structures, ...  

**Example (gc content in chromosome Chlorobium tepidum TLS)**

⇒ Let us assume that sequences are **homogeneous by segment**
Definition of a HMM

**Definition (M0M1 model)**

$X = X_1 \ldots X_\ell$, $X_i \in A$ (observed) and $S = S_1 \ldots S_\ell$, $S_i \in S$ (hidden) are both random sequences which distributions are given by

$$
\begin{align*}
P(S_i = t \mid S_{i-1} = s) &= \nu(s, t) \\
P(X_i = a \mid S_i = s) &= \mu_s(a)
\end{align*}
\quad \forall a \in A, \forall s, t \in S
$$

$S$ is hence M1 and $X$ is (conditionally to $S$) M0.

**Example (Simple example with $A = \{a, b\}$ and $S = \{1, 2\}$)**

$$
\nu = \begin{pmatrix}
1 - \varepsilon & \varepsilon \\
\eta & 1 - \eta
\end{pmatrix}
$$

$\mu_1(a) = p$ \quad $\mu_2(a) = q$

For example we can use $\varepsilon = \eta = 0.01$ and $p = 1 - q = 0.7$.
Remark

We denote \( \theta = \nu(s, s) \)

- \( \mathbb{P}(S_{i+1} = s \mid S_i = s) = \theta \)
- \( \mathbb{P}(S_{i+1} = s, S_{i+2} = s \mid S_i = s) = \theta^2 \)
- \( \ldots \)
- \( \mathbb{P}(S_{i+1} = s, \ldots, S_{i+d-1} = s \mid S_i = s) = \theta^{d-1} \)

The length \( L \) of a run of \( s \) is hence distributed according to \( G(\theta) \):

\[
\mathbb{P}(L = d) = \theta^{d-1}(1 - \theta)
\]

\( \Rightarrow \) this consequence is a serious drawback of the HMM, it exists generalized HMM allowing to overcome this problem (but parameter estimation is difficult).
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
EM for HMM

The likelihood

\[
L(\theta \mid X, S) = \sum_{i=2}^{\ell} \sum_{s,t \in S} \mathbb{I}_{S_{i-1} = s, S_i = t} \log \nu(s, t) \\
+ \sum_{i=1}^{\ell} \sum_{s \in S} \sum_{a \in A} \mathbb{I}_{S_i = s} \mathbb{I}_{X_i = a} \log \mu_s(a)
\]

then if \( P_0 \) is a distribution of \( S \) we get

\[
E_0 L(\theta \mid X) = \sum_{i=2}^{\ell} \sum_{s,t \in S} P_0(S_{i-1} = s, S_i = t) \log \nu(s, t) \\
+ \sum_{i=1}^{\ell} \sum_{s \in S} \sum_{a \in A} P(S_i = s) \mathbb{I}_{X_i = a} \log \mu_s(a)
\]
Forward-Backward

\[ \alpha_i(s) = P(S_i = s \mid X_1^{i-1}) \]
\[ \varphi_i(s) = P(S_i = s \mid X_1^\ell) \]
\[ \beta_i(s) = P(S_i = s \mid X_1^i) \]
\[ \delta_i(s, t) = P(S_{i-1} = s, S_i = t \mid X_1^\ell) \]

1: // Forward
2: \beta_1(s) = \nu_0(s)\mu_s(X_1)/\text{normalization}
3: for \( i = 2 \ldots \ell \) do
4: \[ \alpha_i(s) = \sum_{t \in S} \beta_{i-1}(t)\nu(t, s) \]
5: \[ \beta_i(s) = \alpha_i(s)\nu(X_{i-1}, X_i)/\text{normalization} \]
6: // Backward
7: \varphi_\ell(s) = \beta_\ell(s)
8: for \( i = \ell \ldots 2 \) do
9: \[ \delta_i(s, t) = \frac{\nu(s,t)\beta_{i-1}(s)}{\alpha_i(t)}\varphi_i(t) \text{ and } \varphi_{i-1}(s) = \sum_{t \in S} \delta(s, t) \]
Phase M

Maximization

Using the current value of $\varphi(s)$ and $\delta(s, t)$ we get:

$$\hat{\nu}(s, t) = \frac{\sum_{i=2}^{\ell} \delta_i(s, t)}{\text{normalization}}$$

and

$$\hat{\mu}_s(a) = \frac{\sum_{i=1}^{\ell} \varphi_i(s) I_{X_i=a}}{\sum_{i=1}^{\ell} \varphi_i(s)}$$

⇒ only need the cumulative sums of $\varphi$ and $\delta$

Complexities

The overall complexities are $O(\ell)$ both in space and time.
The Problem

How to compute the best path $s^* = \arg\max_s P_\theta(S = s \mid X)$?

Proposition

$$Z_i(u) = \max_{s_1, \ldots, s_{i-1}} P_\theta(S_1 = s_1, \ldots, S_{i-1} = s_{i-1}, S_i = u \mid X)$$

then we have the following recurrence relation

$$Z_i(u) = \max_{t \in S} Z_{i-1}(t) \nu(t, u) \mu_u(X_i)$$

$\Rightarrow$ the Viterbi algorithm computes $s^*$ in $O(\ell)$. Caution: this is a simple but inaccurate replacement of Forward-Backward.
Example $(\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))$

- **Iteration 1:** $L = -902.262638$  $\theta = (0.1, 0.1, 0.8, 0.3)$
Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

Iteration 2: \( L = -898.514366 \) theta=(0.106,0.088,0.765,0.306)
Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 3: \(L = -889.139209\) \(\theta = (0.098, 0.084, 0.75, 0.312)\)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 4: \( L = -876.546028 \) \( \theta = (0.09, 0.08, 0.739, 0.314) \)
Example $(\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))$

iteration 5: $L = -862.270459$ \(\theta = (0.081, 0.074, 0.731, 0.315)\)
Step by Step

Example \( (\theta = (\epsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30)) \)

iteration 6: \( L = -847.035383 \) \( \theta = (0.073, 0.069, 0.724, 0.315) \)
Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

Iteration 7: \(L=-831.177299\) \(\theta=(0.065,0.063,0.718,0.316)\)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 8: L = -814.855156 theta = (0.057, 0.057, 0.713, 0.317)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 9: \( L = -798.168278 \) \( \theta = (0.05, 0.051, 0.708, 0.318) \)
Example \((\theta = (\epsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

Iteration 10: \(L = -781.238794\) \(\theta = (0.043, 0.045, 0.703, 0.32)\)
Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 11: \(L = -764.283293\) theta=(0.037,0.039,0.698,0.322)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 12: \( L = -747.677093 \), \( \text{theta} = (0.031, 0.033, 0.693, 0.325) \)
Step by Step

Example \( \theta = (\epsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 13: \( L = -731.989376 \) \( \theta = (0.025, 0.027, 0.689, 0.328) \)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 14: \( L = -717.941178 \) \( \theta = (0.02, 0.022, 0.685, 0.331) \)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

Iteration 15: \( L = -706.24826 \) \( \theta = (0.016, 0.018, 0.681, 0.334) \)
Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 16: \( L = -697.38839 \) theta=(0.013,0.014,0.679,0.337)
Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 17: \( L = -691.391995 \) theta=(0.011,0.012,0.677,0.34)
Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

Iteration 18: \( L = -687.787817 \) \( \theta = (0.01, 0.01, 0.676, 0.342) \)
Step by Step

**Example ($\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30)$)**

*iteration 19: $L = -685.829208$  $\theta = (0.009, 0.009, 0.675, 0.343)$*
Example \( (\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30)) \)

iteration 20: \( L = -684.83443 \) \( \theta = (0.008, 0.009, 0.675, 0.344) \)
Example \((\theta = (\epsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 21: \(L=-684.347732\) \(\theta=(0.008,0.008,0.675,0.344)\)
Step by Step

Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 22: \(L = -684.113965\) theta=(0.008,0.008,0.675,0.345)
Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 23: \(L = -684.002654 \quad \text{theta} = (0.008, 0.008, 0.675, 0.345)\)
EM Algorithm
Markov Models
Hidden Markov Models

Motivation and Notations
EM for HMM
Applications

Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 24: \( L = -683.949864 \) \( \text{theta} = (0.008, 0.008, 0.675, 0.345) \)
Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 25: \( L = -683.924875 \) \( \theta = (0.008, 0.008, 0.675, 0.345) \)
Step by Step

Example \( \theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 26: \( L = -683.913057 \), \( \theta = (0.008, 0.008, 0.675, 0.345) \)
Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 27: \(L = -683.907469\) \(\theta = (0.008, 0.008, 0.675, 0.345)\)
Example \((\theta = (\varepsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30))\)

iteration 28: \(L = -683.904828\) \(\theta = (0.008, 0.008, 0.675, 0.345)\)
Step by Step

Example \( \theta = (\epsilon, \eta, p, q) = (0.01, 0.01, 0.70, 0.30) \)

iteration 29: \( L = -683.90358 \) \( \theta = (0.008, 0.008, 0.675, 0.345) \)
Outline

1. EM Algorithm
   - The Principle
   - A Detailed Example
   - Application: Finding Patterns in Biopolymers

2. Markov Models
   - Order 0
   - Order 1 and More
   - Model Selection

3. Hidden Markov Models
   - Motivation and Notations
   - EM for HMM
   - Applications
HMM profiles

structure of the hidden states

The distribution of each state gives the profile. We may accept either one or several patterns in each sequence (user choice)
HMM profiles

structure of the hidden states

we can also accept **insertions** or **deletions** in the pattern.
Finding Procaryote Genes with HMM

structure of the hidden states

- direct coding
- indirect coding
- non coding
Finding Prokaryote Genes with HMM

structure of the hidden states

Main Algorithms with Applications in Bioinformatics by G. Nuel

Lecture 2: Expectation-Maximization and Hidden Markov Models
Finding Procaryote Genes with HMM

structure of the hidden states

non coding

Main Algorithms with Applications in Bioinformatics by G. Nuel
Lecture 2: Expectation-Maximization and Hidden Markov Models
Finding Procaryote Genes with HMM

structure of the hidden states
Sample of results

Example (Software SHOW/MUGENE)
Summary

EM Algorithm

- suitable for many hidden data problems
- importance of the expression of likelihood
- application in pattern detection: MEME

Markov models

- widely used model in sequence modelization
- parameter estimation (MLE) model selection (BIC, …)
- application to CDS finding: Genmark

Hidden Markov Models

- Forward-Backward rather than Viterbi
- elegant way to take into account biological knowledge
- many applications: patterns (HMMER), genes, …
Summary

**EM Algorithm**
- suitable for many hidden data problems
- importance of the expression of likelihood
- application in pattern detection: MEME

**Markov models**
- widely used model in sequence modelization
- parameter estimation (MLE) model selection (BIC, ...)
- application to CDS finding: Genmark

**Hidden Markov Models**
- Forward-Backward rather than Viterbi
- elegant way to take into account biological knowledge
- many applications: patterns (HMMER), genes, ...
## Summary

### EM Algorithm
- suitable for many **hidden data** problems
- importance of the expression of **likelihood**
- application in pattern detection: **MEME**

### Markov models
- **widely used** model in sequence modelization
- parameter **estimation** (MLE) **model selection** (BIC, 
  . . .)
- application to CDS finding: **Genmark**

### Hidden Markov Models
- **Forward-Backward** rather than Viterbi
- elegant way to take into account **biological knowledge**
- many applications: **patterns** (HMMER), **genes**, 
  . . .