HMM for Bioinformatics

Paweł Błażej
Department of Genomics, Faculty of Biotechnology,
blazej@smorfland.uni.wroc.pl

13 marca 2019

Conventions and notations

1. $\mathcal{X}, \mathcal{Y}$ – spaces;

Conventions and notations

1. $\mathcal{X}, \mathcal{Y}$ – spaces;
2. $|\mathcal{X}| = L$ – size of an alphabet;

Conventions and notations

1. $\mathcal{X}, \mathcal{Y}$ – spaces;
2. $|\mathcal{X}| = L$ – size of an alphabet;
3. X, Y discrete random variables ($X : \Omega \to \mathcal{X}$);

Conventions and notations

1. $\mathcal{X}, \mathcal{Y}$ – spaces;
2. $|\mathcal{X}| = L$ – size of an alphabet;
3. X, Y discrete random variables ($X : \ \Omega \rightarrow \mathcal{X}$);
4. $P_X(x_i)$ probability.

Conventions and notations

1. $\mathcal{X}, \mathcal{Y}$ – spaces;
2. $|\mathcal{X}| = L$ – size of an alphabet;
3. X, Y discrete random variables ($X : \ \Omega \rightarrow \mathcal{X}$);
4. $P_X(x_i)$ probability.
5. probability functions p, f

EM – algorithm

An algorithm for finding the maximum likelihood estimate of the parameters of the finite mixture.

Mixtures

Let us suppose that $\{(X_1, Y_1), \ldots, (X_n, Y_n)\} = \{(X_l, Y_l)\}_{l=1}^{n}$ is a sequence of independent (pairs of) random variables with the same distribution and for every l and $j = 1, 2, \ldots L$ we have:

$$\alpha_j = P(X_l = x_j)$$

and for any $y \in \mathcal{Y}$

$$p(y|\phi_j) = P(Y_l = y|X_l = x_j).$$

We also take θ as the parameter with:

$$\theta = (\alpha_1, \alpha_2, \ldots \alpha_L; \phi_1, \phi_2, \ldots, \phi_L).$$

We think initially about two data sequences $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and $\mathbf{x} = (x_{j_1}, x_{j_2}, \ldots x_{j_n})$. The assumption of pairwise independence means that:

$$p(\mathbf{x}, \mathbf{y}|\theta) = \Pi_{l=1}^{n} P(Y_l = y_l, X_l = x_{j_l}|\theta) = \Pi_{l=1}^{n} p(y_l|\phi_{j_l}) \cdot \alpha_{j_l}.$$

Mixtures – properties

We think initially about two data sequences $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and $\mathbf{x} = (x_{j_1}, x_{j_2}, \ldots x_{j_n})$. The assumption of pairwise independence means that:

$$p(\mathbf{x}, \mathbf{y}|\theta) = \Pi_{l=1}^{n} P(Y_l = y_l, X_l = x_{j_l}|\theta) = \Pi_{l=1}^{n} p(y_l|\phi_{j_l}) \cdot \alpha_{j_l}.$$

Remark

The goal is to estimate θ in a situation where the sequence of $\mathbf{x}$ is hidden.

Mixtures – notations

Using the rules for computing marginal distributions, we get for any (X_l, Y_l)

$$f(y|\theta) = \sum_{j=1}^{L} P(X_l = x_j, Y = y|\theta)$$

Mixtures – notations

Using the rules for computing marginal distributions, we get for any (X_l, Y_l)

$$f(y|\theta) = \sum_{j=1}^{L} P(X_l = x_j, Y = y|\theta)$$

so that

$$f(y|\theta) = \sum_{j=1}^{L} p(y|\phi_j)\alpha_j.$$

where $f(y|\theta)$ is a *finite mixture* whereas $\{\alpha_j\}_{j=1}^{L}$ is called *mixing distribution*.

Likelihood and missing information

The likelihood function for $\mathbf{y}$ with relation to θ is:

$$p(\mathbf{y}|\theta) = f(y_1|\theta) \cdot f(y_1|\theta) \cdot \ldots \cdot f(y_n|\theta).$$

The maximum likelihood estimate is

$$\hat{\theta} = argmax_\theta p(\mathbf{y}|\theta),$$

Likelihood and missing information

The likelihood function for $\mathbf{y}$ with relation to θ is:

$$p(\mathbf{y}|\theta) = f(y_1|\theta) \cdot f(y_1|\theta) \cdot \ldots \cdot f(y_n|\theta).$$

The maximum likelihood estimate is

$$\hat{\theta} = argmax_{\theta} p(\mathbf{y}|\theta),$$

but how to include information about “hidden” $\mathbf{x}$?

We start with the posterior probability

$$p(\mathbf{x}|\mathbf{y},\theta) = \frac{p(\mathbf{x},\mathbf{y}|\theta)}{p(\mathbf{y}|\theta)}.$$

missing information – posterior probability

We start with the posterior probability

$$p(\mathbf{x}|\mathbf{y},\theta) = \frac{p(\mathbf{x},\mathbf{y}|\theta)}{p(\mathbf{y}|\theta)}.$$

In other words

$$p(\mathbf{x}|\mathbf{y},\theta) = \Pi_{l=1}^{n} \frac{p(y_l|\phi_{j_l})\alpha_{j_l}}{f(y_l|\theta)}.$$

Therefore, we get

$$log(p(\mathbf{y}|\theta)) = \sum_{l=1}^{n} log(p(y_l|\phi_{j_l}))\alpha_{j_l} - log(p(\mathbf{x}|\mathbf{y},\theta)).$$

Therefore, we get

$$log(p(\mathbf{y}|\theta)) = \sum_{l=1}^{n} log(p(y_l|\phi_{j_l}))\alpha_{j_l} - log(p(\mathbf{x}|\mathbf{y}, \theta)).$$

Remark

We shal continue by giving a lower bound for $log(p(\mathbf{y}|\theta))$.

The EM algorithm – Quasi-log likelihood

Let us suppose that we have obtained an approximation $\theta^{(t)}$ to the estimate $\hat{\theta}$ with

$$\theta^{(t)} = (\alpha_1^{(t)}, \ldots, \alpha_L^{(t)}; \phi_1^{(t)}, \ldots, \phi_L^{(t)})$$

Let us suppose that we have obtained an approximation $\theta^{(t)}$ to the estimate $\hat{\theta}$ with

$$\theta^{(t)} = (\alpha_1^{(t)}, \dots, \alpha_L^{(t)}; \phi_1^{(t)}, \dots, \phi_L^{(t)})$$

Remark

The general idea is to improve $\theta^{(t)}$ so as to get closer to $\hat{\theta}$.

The EM algorithm – Quasi-log likelihood

It is clear that :)

$$log(p(\mathbf{y}|\theta)) = \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \theta^{(t)}) log(p(\mathbf{x}, \mathbf{y}|\theta)) - \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \theta^{(t)}) log(p(\mathbf{x}|\mathbf{y}, \theta)).$$

We introduce also the auxiliary function

$$Q(\theta|\theta^{(t)}) = \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \theta^{(t)}) log(p(\mathbf{x}, \mathbf{y}|\theta))$$

and let us consider:

$$log(p(\mathbf{y}|\theta)) - log(p(\mathbf{y}|\theta^{(t)})).$$

It is easy to see :)

$$log(p(\mathbf{y}|\theta)) - log(p(\mathbf{y}|\theta^{(t)})) \geqslant Q(\theta|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t)})$$

Therefore, if we determine

$$\theta^{(t+1)} = argmax_{\theta} Q(\theta|\theta^{t})$$

we have found an estimate $\theta^{(t+1)}$ such that

$$log(p(\mathbf{y}|\theta^{(t+1)})) \geqslant log(p(\mathbf{y}|\theta^{(t)}))$$

and we have improved on $\theta^{(t)}$ in the sense of increased likelihood.

Step E and step M

1. **Start:** An estimate $\theta^{(t)}$ given by

$$\theta^{(t)} = (\alpha_1^{(t)}, \dots, \alpha_L^{(t)}; \phi_1^{(t)}, \dots, \phi_L^{(t)})$$

Step E and step M

1. **Start:** An estimate $\theta^{(t)}$ given by

$$\theta^{(t)} = (\alpha_1^{(t)}, \ldots, \alpha_L^{(t)}; \phi_1^{(t)}, \ldots, \phi_L^{(t)})$$

2. **Step E:** Calculate the conditional expectation:

$$Q(\theta|\theta^{(t)}) = \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \theta^{(t)}) log(p(\mathbf{x}, \mathbf{y}|\theta))$$

Step E and step M

1. **Start:** An estimate $\theta^{(t)}$ given by

$$\theta^{(t)} = (\alpha_1^{(t)}, \ldots, \alpha_L^{(t)}; \phi_1^{(t)}, \ldots, \phi_L^{(t)})$$

2. **Step E:** Calculate the conditional expectation:

$$Q(\theta|\theta^{(t)}) = \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}, \theta^{(t)}) log(p(\mathbf{x}, \mathbf{y}|\theta))$$

3. **Step M:** Determine $\theta^{(t+1)}$ by

$$\theta^{(t+1)} = argmax_{\theta} Q(\theta|\theta^{(t)}).$$

Let $\theta^{(t+1)} \rightarrow \theta^t$ and repeat from step **E**.

Improtant Questions

1. Is this really an algorithm?

Improtant Questions

1. Is this really an algorithm?
2. Does it converge ?

Improtant Questions

1. Is this really an algorithm?
2. Does it converge ?
3. Does this converge to a global/local maximum of the likelihood function?

An explicit form of step M

$$Q(\theta|\theta^{(t)}) = \sum_{j_l}^{L} \dots \sum_{j_n}^{L} \Pi_{l=1}^{n} \frac{p(y_l|\phi_{j_l}^{(t)})\alpha_{j_l}^{(t)}}{f(y_l|\theta^{(t)})} \cdot log \Pi_{l=1}^{n} p(y_l|\phi_{j_l}) \cdot \alpha_{j_l}$$

An explicit form of step M

$$Q(\theta|\theta^{(t)}) = \sum_{j_l}^{L} \dots \sum_{j_n}^{L} \Pi_{l=1}^{n} \frac{p(y_l|\phi_{j_l}^{(t)})\alpha_{j_l}^{(t)}}{f(y_l|\theta^{(t)})} \cdot log \Pi_{l=1}^{n} p(y_l|\phi_{j_l}) \cdot \alpha_{j_l}$$

A simple calculations shows that

$$Q(\theta|\theta^{(t)}) = \sum_{j=1}^{L} \sum_{l=1}^{n} log(p(y_l|\phi_j)\alpha_j) \frac{p(y_l|\phi_j^{(t)})\alpha_j^{(t)}}{f(y_l|\theta^{(t)})}$$

An explicit form of step M

Therefore:

$$\alpha_j^{(t+1)} = \frac{1}{n}\sum_{l=1}^{n} \frac{p(y_l|\phi^{(t)}{}_j)\alpha_j^{(t)}}{f(y_l|\theta^{(t)})}$$

and

$$\phi_j^{(t+1)} = argmax_{\phi_j} \sum_{l=1}^{n} log(p(y_l|\phi_j)) \frac{p(y_l|\phi_j^{(t)})\alpha_j^{t}}{f(y_l|\theta^{(t)})}.$$