HMM for Bioinformatics

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

27 marca 2019

Learning of Markov chains

1. Maximum likelihood;

Learning of Markov chains

1. Maximum likelihood;
2. Bayesian estimation;

Transition probability matrix

Let $\underline{\Theta}$ be the transition probability matrix

$$\underline{\Theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \dots & \theta_{1L} \\ \theta_{21} & \theta_{22} & x_{23} & \dots & \theta_{2L} \\ \hdotsfor{5} \\ \theta_{L1} & \theta_{L2} & \theta_{L3} & \dots & \theta_{LL} \end{bmatrix}$$

Transition probability matrix

Let $\underline{\Theta}$ be the transition probability matrix

$$\underline{\Theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \dots & \theta_{1L} \\ \theta_{21} & \theta_{22} & x_{23} & \dots & \theta_{2L} \\ \dots & \dots & \dots & \dots & \dots \\ \theta_{L1} & \theta_{L2} & \theta_{L3} & \dots & \theta_{LL} \end{bmatrix}$$

We are concerned with estimating the model $\underline{\Theta}$ in the family of probabilistic models $p(x|\underline{\Theta})$ for a (training) sequence $\mathbf{X}$ of $n+1$ symbols in S^{n+1}

Markov chains

$$p(x|\underline{\Theta}) = P(X_0 = j_0, \ldots, X_n = j_n|\underline{\Theta}) = \pi_{j_0}(0)\prod_{i=1}^{n}\theta_{j_{i-1}|j_i}$$

Markov chains

$$p(x|\underline{\Theta}) = P(X_0 = j_0, \ldots, X_n = j_n|\underline{\Theta}) = \pi_{j_0}(0)\prod_{i=1}^{n}\theta_{j_{i-1}|j_i}$$

Therefore, we have at most $L^2 - L$ trabsition parameters and the $J - 1$ initial probabilities to estimate using the data $\mathbf{X}$

The conditional likelihood function

$$L(\underline{\Theta}) = \prod_{i=1}^{n} \theta_{j_{i-1}|j_i}$$

The conditional likelihood function

$$L(\underline{\Theta}) = \prod_{i=1}^{n} \theta_{j_{i-1}|j_i}$$

The corresponding log likelihood function is

$$\mathcal{L}(\underline{\Theta}) = \sum_{i=1}^{n} ln\theta_{j_{i-1}|j_i}$$

ML of the transition probability matrix

Let us introduce $n_{i|j}$ as a number of l such that

$$1 \leqslant l \leqslant n, \; j_{l-1} = i, \; j_l = j$$

ML of the transition probability matrix

Let us introduce $n_{i|j}$ as a number of l such that

$$1 \leqslant l \leqslant n, \; j_{l-1} = i, \; j_l = j$$

Using the frequency counts $n_{i|j}$ we can write the likelihood function as

$$L(\underline{\Theta}) = \prod_{i=1}^{L} \prod_{j=1}^{n} \theta_{i|j}^{n_{i|j}}.$$

ML of the transition probability matrix

Let us introduce $n_{i|j}$ as a number of l such that

$$1 \leqslant l \leqslant n,\ j_{l-1} = i,\ j_l = j$$

Using the frequency counts $n_{i|j}$ we can write the likelihood function as

$$L(\underline{\Theta}) = \prod_{i=1}^{L} \prod_{j=1}^{n} \theta_{i|j}^{n_{i|j}}.$$

The log likelihood will be

$$\mathcal{L}(\underline{\Theta}) = \sum_{i=1}^{L} \sum_{j=1}^{n} n_{i|j} log(\theta_{i|j}).$$

ML of the transition probability matrix

Let us introduce n_i as the number of l such that:

$$0 \leqslant l \leqslant n-1,\ j_l = i$$

Proposition

The maximum likelihood estimate $\widehat{\theta_{i|j}}$ of $\theta_{i|j}$ is

$$\widehat{\theta_{i|j}} = \frac{n_{i|j}}{n_i},$$

for all i, j.

Posterior distribution of rows in the transition matrix

Let us assume that our uncertainty about the rows of Θ i.e.

$$\theta_i = (\theta_{i|1}, \ldots, \theta_{i|L})$$

is modelled by independent random variables that have their respective Dirichlet densities for $i = 1, 2, \ldots, L$. These we formulate as

$$Dir(\theta_i, \alpha_i q_{i|1}, \ldots, \alpha_i q_{i|L}) = \frac{\Gamma(\alpha_i)}{\prod_{j=1}^{L} \Gamma(\alpha_i q_{i|j})} \cdot \prod_{j=1}^{L} \theta_{i|j}^{\alpha_i q_{i|j} - 1}$$

where

$$\alpha_i > 0, \; q_{i|j} > 0, \; \sum_{j=1}^{L} q_{i|j} = 1.$$

The multivariate Dirichlet densitiy

$$\prod_{i=1}^{L} Dir(\theta_i, \alpha_i q_{i|1}, \ldots, \alpha_i q_{i|L})$$

The posterior distribution

$$p(\underline{\Theta}|\mathbf{X}) = \frac{\prod_{i=1}^{L} \frac{\Gamma(\alpha_i)}{\prod_{j=1}^{L} \Gamma(\alpha_i q_{i|j})} \prod_{j=1}^{L} \theta_{i|j}^{n_{i|j}+\alpha_i q_{i|j}-1}}{p(\mathbf{X})}$$

where $p(\mathbf{X})$ is the standarization that makes $p(\underline{\Theta}|\mathbf{X})$ a probability density.

Predictive probability

We may also pose a question: what is the probability of

$$P(X_{n+1} = j | X_n = i; \mathbf{X})?$$

Predictive probability

$$\widehat{P}_{ML}(X_{n+1} = j | X_n = i; \mathbf{X}) = \widehat{\theta_{i|j}}$$

Predictive probability

There are another way of addressing the stated question. we can consider

$$P^*(X_{n+1} = j | X_n = i; \mathbf{X}) = \int \theta_{i|j} p(\underline{\Theta} | \mathbf{X}) d\underline{\theta}$$

we get

$$P^*(X_{n+1} = j | X_n = i; \mathbf{X}) = \frac{n_{ij} + \alpha_i q_{i|j}}{n_i + \alpha_i}.$$

For any given $\mathbf{X} = (j_0, j_1, \ldots, j_n)$ we calculate $n_{i|j}$ and also let us define

$$n_{i|\cdot} = \sum_{j=1}^{L} n_{i|j}, \;\; n_{\cdot|j} = \sum_{i=1}^{L} n_{i|j}$$

for all i and j.

The Whittle Distribution

For any given $\mathbf{X} = (j_0, j_1, \ldots, j_n)$ we calculate $n_{i|j}$ and also let us define

$$n_{i|\cdot} = \sum_{j=1}^{L} n_{i|j}, \;\; n_{\cdot|j} = \sum_{i=1}^{L} n_{i|j}$$

for all i and j. Therefore,

$$n_{i|\cdot} - n_{\cdot|i} = \delta_{ij_0} - \delta_{ij_n}$$

and

$$\sum_{j=1}^{L} n_{i|j} = \sum_{i=1}^{L} n_{i|j} = n.$$

The Whittle Distribution

Let us define:

$$F = (n_{i|j})_{i=1,j=1}^{L,L}$$

is a $J \times J$ matrix of non-negative integers.

The Whittle Distribution – Proposition

Let F be an $J \times J$ matrix of non-negative integers $n_{i|j}$ such that

$$\sum_{i=1}^{L} \sum_{j=1}^{L} n_{i|j} = n$$

and

$$n_{i|\cdot} - n_{\cdot|i} = \delta_{iv} - \delta_{vi}$$

for some $u, v \in L$. Let $N_{u,v}^{n}(F)$ be the number of sequences $\mathbf{x} = (j_0, j_j, \ldots, j_n)$ having the frequency count F and satisfying $j_0 = u$, $j_n = v$. Then

$$N_{u,v}^{n}(F) = \frac{\prod_{i=1}^{L} n_{i|\cdot}!}{\prod_{i=1}^{L} \prod_{j=1}^{L} n_{i|j}!} \cdot F_{uv}^{*}.$$

F^*_{uv} is the (u, v)th cofactor of the matrix $F^* = (f_{ij})$. with the components

$$f^*_{ij} = \begin{cases} \delta_{ij} - \frac{n_{i|j}}{n_{i|\cdot}} & \text{if } n_{i|\cdot} > 0 \\ \delta_{ij} & \text{if } n_{i|\cdot} = 0. \end{cases}$$