HMM for Bioinformatics

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

19 marca 2019

Markov chains – Introduction

1. $L = \{l_1, l_2, \ldots l_n\}$ – alphabet
2. $X_1, X_2, \ldots, X_n$ sequence of random variables assuming values in L;
3. The symbols $l_i \in L$ are called states .

Definition – Markov property

A sequence of random variables $\{X_n\}_{n=0}^{\infty}$ is called a markov chain, if for all $n \geqslant 0$ and $j_0, j_1, \ldots, j_n \in L$,

$$P(X_n = j_n | X_0 = j_0, \ldots, X_{n-1} = j_{n-1}) = P(X_n = j_n | X_{n-1} = j_{n-1})$$

The conditional probabilities

$$p_{ij} = P(X_n = j | X_{n-1} = i), \;\; n \geqslant 1, \;\; i,j \in L$$

are assumed to be independent of n and are called (stationary) one step transition probabilities. A Markov chain with stationary transition probabilities is called *homogeneous*

Transition probability matrix

$$\mathbf{P} = (p_{ij})_{i,j \in L}, \ \ \textit{where} \ \ p_{ij} \geqslant 0 \ \textit{and} \sum_{j \in L} p_{ij} = 1.$$

Transition probability matrix

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1L} \\ p_{21} & p_{22} & x_{23} & \cdots & p_{2L} \\ \hdotsfor{5} \\ p_{L1} & p_{L2} & p_{L3} & \cdots & p_{LL} \end{bmatrix}$$

Joint probaility distributions of Markov chains

$$P(X_n = j_n, X_0 = j_0, \ldots, X_{n-1} = j_{n-1}) =$$

$$p_{X_0} \cdot p_{j_0 j_1} \cdot p_{j_1 j_2} \cdot \ldots \cdot p_{j_{n-1} j_n}.$$

Joint probaility distributions of Markov chains

If $\{X_n\}$ is a homogeneous Markov chain, then

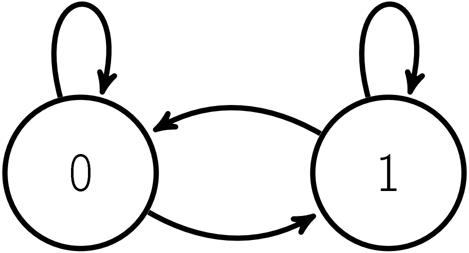
$$P(X_n = j_n, X_0 = j_0, \dots, X_{n-1} = j_{n-1}) = p_{X_0} \Pi_{l=1}^{n} p_{j_{l-1}j_l}$$

Examples – binary Markov chains

1. It is a sequential mechanism producing zeros (0) and ones (1);
2. It can be used modelling 'regulatory region' or 'not regulatory region';

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

Examples – binary Markov chains



Markov chains of kth order

Definition

A sequence of random variables $\{X_n\}_{n=0}^{\infty}$ is called a *kth* order Markov chain, if for all $n \geqslant 1$ and $j_0, j_1, \dots, j_n \in L$,

$$P(X_n = j_n | X_0 = j_0, \dots, X_{n-1} = j_{n-1}) =$$

$$P(X_n = j_n | X_{m-k} = j_{n-k}, \dots, X_{n-1} = j_{n-1})$$

for a positive integer k.