Assignment batch 3, Algorithmic Bioinformatics, Spring 2010

June 3, 2010

1 Implementation: MCMC-algorithm

In this assigment you are supposed to implement an MCMC algorithm. The model is the same as the model which has been described in part 2 of the implementation section of assignment 2. You can use the same data that you have received for assignment 2. Here the model is described again:

The sequences described below are circular and indices are counted modulo n, so $n = 0$. Consider the following probabilistic model with parameters $f_1, \ldots, f_n$ and $\lambda_1, \ldots, \lambda_n$, where f_i is a distribution over $\{1, \ldots, m\}$. A sequence $a_1, \ldots, a_n$ where $a_i \in \{1, \ldots, m\}$ is generated as follows: (1) a direction L or R is chosen for position i, the probability that L is chosen is λ_i and the probability for R is $1 - \lambda_i$ and (2) if i has direction L, then a_i is chosen according to f_{i-1} and otherwise according to f_i.

For the MCMC implementation, assume that (1) each distribution f_i gives probability $k/10$, where $k \in \{1, \ldots, 10\}$, to one number in $\{1, \ldots, m\}$ and distributes the rest of the probability mass equaly over the other numbers in $\{1, \ldots, m\}$ and (2) that each λ_i equals $1/k$ for some $k \in \{1, \ldots, 10\}$ (allow any such λ_i).

Contstruct an MCMC algorithm that estimates the posterior distribution over the paramters of the model, i.e., the different types of f_i and λ_i.

2 Problems

1. MCMC algorithms with a non-symmetric proposal distribution $\rho(a|b)$ converge to a stationary distribution ϕ if combined with an acceptance distribution defined by

$$\alpha(a|b) = \min(1, \frac{\phi(a)}{\phi(b)} \frac{\rho(b|a)}{\rho(a|b)}).$$

   Prove that this is true.

2. Described an MCMC algorithm that computes the posterior over HMMs with k layers after having observed a set of sequences F.