

BMI/CS 576 Fall 2008

Homework #3

Prof. Colin Dewey

Due Wednesday, October 28th, 2009 by 11:59pm

The goal of this assignment is to become more familiar with Markov chains and hidden Markov models.

To turn in your assignment, copy all relevant files to the directory:

`/u/medinfo/handin/bmi576/hw3/USERNAME`

where `USERNAME` is your account name for the BMI network.

1. An important property of a Markov chain is the probability distribution over the lengths of sequences it generates.

- (a) Give the distribution of lengths, $P(\ell)$, for sequences generated by the Markov chain below.

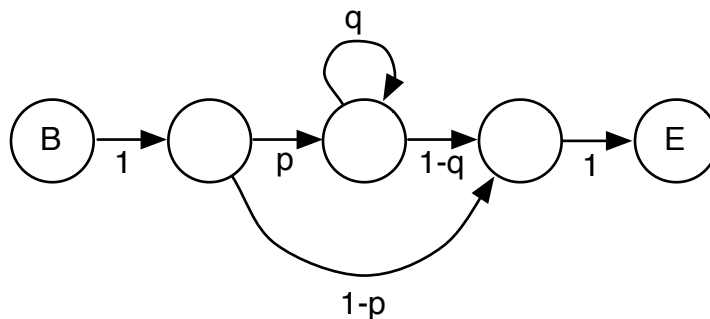


Figure 1: The state transition diagram for the Markov chain of problem 1(a). B and E are the silent begin and end states.

- (b) Give the state-transition diagram for a Markov chain with length distribution given by

$$P(\ell) = \begin{cases} 0 & \text{if } \ell = 0 \\ (1-p)(pq)^{(\ell-1)/2} & \text{if } \ell > 0, \text{ odd} \\ (1-q)p(pq)^{(\ell-2)/2} & \text{if } \ell > 0, \text{ even} \end{cases}$$

2. Given a hidden Markov model, M , with k states, briefly describe how to compute $P(x|M)$ (the probability of a sequence x given the model) with only $O(k)$ space.
3. The optimization problem of Problem 2 Homework 2 (partitioning a DNA sequence into segments that are predominantly composed of a single base) can be approximately modeled by a hidden Markov model.
 - (a) What are the observed data?
 - (b) What are the hidden states of the HMM?
 - (c) What do the emission parameters represent? How do they relate to the parameters b and c in the optimization problem?
 - (d) What do the transition parameters represent? How do they relate to the parameter a in the optimization problem?
 - (e) Which HMM algorithm should be used to partition a DNA sequence?
4. Given the simple HMM specified in Figure 2 and the observed sequence $x = \text{CGTCAG}$, calculate:
 - (a) $P(x|M)$ using the forward algorithm. Show your work by providing all values in the dynamic programming matrix.
 - (b) the most likely path of hidden states using the Viterbi algorithm. Show your work by providing all values in the dynamic programming matrix.
 - (c) $P(x|M)$ using the backward algorithm. Show your work by providing all values in the dynamic programming matrix.
 - (d) the posterior probabilities $P(\pi_i = 1|x, M)$ for $i = 1 \dots 6$, using your results for the forward and backward algorithms.

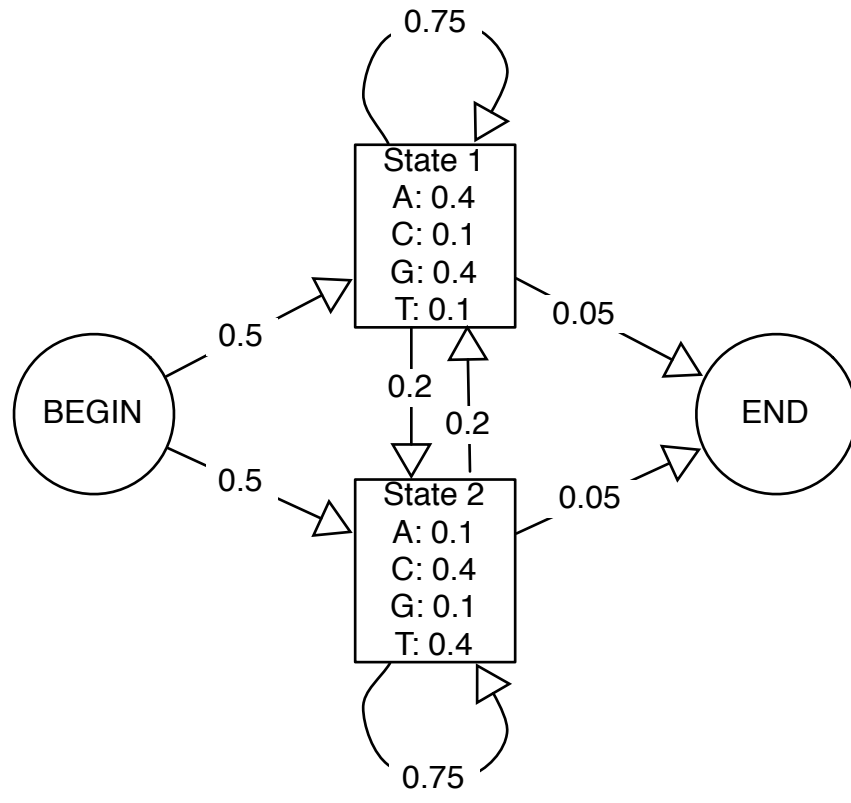


Figure 2: A simple HMM