

BMI/CS 776
Spring 2010
HW #1

(1) We can think of the MEME **OOPS** model as an HMM. Draw a diagram of the HMM structure for such a model for the case in which the motif width $W=3$. Show all states, transitions, and transition probabilities. Indicate which parameters, if any, are tied.

(2) Draw a diagram of the HMM structure that corresponds to the MEME **ZOOPS** model. As before, assume that the motif width $W=3$. Show all states, transitions, and transition probabilities. Indicate which parameters, if any, are tied.

(3) Consider using Gibbs sampling to find a motif of width $W=3$ in the following set of sequences:

ACATATTC
AGGCTATA
TCGTACCT
GCGGGGGT
CCTACGGC
AAGCGCAG
ATCTATAA

- (a) How large is the state space in this case? Explain how you determined this number.
- (b) Which state is more probable: the state in which the motif occurs in the first possible position in each sequence or the state in which the motif occurs in the last possible position in each sequence? Justify your answer by showing how the Gibbs sampling approach we discussed would calculate the probability of these two states.

(4) Draw the topology of a Markov model for DNA sequences that has the following length (i.e. duration) distribution:

$$P(l) = \begin{cases} 0 & \text{if } l \leq 0 \\ (l-1)\alpha^{l-2}(1-\alpha)^2 & \text{otherwise} \end{cases}$$

Show all states, transitions, and transition probabilities.

(5) Suppose you want to make a *motif logo* for a motif found in the sequences from question (3). Assume that this motif has width $W=3$ and that it occurs in the first position in each sequence. How tall will each character be in each position in this logo? Show your calculations.

(6) Suppose we are learning a CRM-modeling HMM, using the method of Noto and Craven. At some point in the search, suppose our model is a conjunction of two motifs, X and Y (i.e. X **AND** Y).

- a) Draw the HMM topology. Show all of the states and the transitions.
- b) Show how the HMM topology would be changed after applying the **OR** operator. Show the result for all possible applications of the **OR** operator.