IE 598 Inference in Graphical Models

Mid-term Quiz

Fall 2012

This quiz is an hour and 30 minutes long. There are 2 problems with 5 sub-problems in total.

Problem 1

Consider a stochastic process that transitions among a finite set of states $s_1, \ldots, s_k$ over time steps $i = 1, \ldots, N$. The random variables $X_1, \ldots, X_N$ representing the state of the system at each time step are generated as follows:

- Sample the initial state $X_1 = s$ from an initial distribution p_1, and set $i := 1$.
- Repeat the following:
  - Sample a duration d from a duration distribution p_D over the integers $\{1, \ldots, M\}$, where M is the maximum duration.
  - Remain in the current state s for the next d time steps, i.e., set

$$X_i := X_{i+1} := \ldots := X_{i+d-1} := s$$

  - Sample a successor state s' from a transition distribution $p_T(\cdot|s)$ over the other states $s' \neq s$ (so there are no sef-transitions).
  - Assign $i := i + d$ and $s := s'$.

This process continues indefinitely, but we only observe the first N time steps. You need not worry about the end of the sequence to do any of the problems. As an example calculation with this model, the probability of the sample state sequence $s_1, s_1, s_1, s_2, s_3, s_3$ is

$$p_1(s_1)p_D(3)p_T(s_2|s_1)p_D(1)p_T(s_3|s_2) \sum_{2 \geq d \leq M} p_D(d) \ .$$

Finally, we do not directly observe the X_i's, but instead observe emissions y_i at each step sampled from a distribution $p_{Y_i|X_i}(y_i|x_i)$.

(a) For this part only, suppose $M = 2$, and $p_D(d) = \begin{cases} 0.6 & \text{for } d = 1 \\ 0.4 & \text{for } d = 2 \end{cases}$, and each X_i takes on a value from an alphabet $\{a, b\}$. Draw a minimal directed I-map for the first five time steps using the variables $(X_1, \ldots, X_5, Y_1, \ldots, Y_5)$. Explain why none of the edges can be removed.
[Note: you do not need to solve part (a) in order to solve part (b) and (c).]

(b) This process can be converted to an HMM using an *augmented state representation*. In particular, the states of this HMM will correspond to pairs (x, t), where x is a state in the original system, and t represents the time elapsed in that state. For instance, the state sequence $s_1, s_1, s_1, s_2, s_3, s_3$ would be represented as $(s_1, 1), (s_1, 2), (s_1, 3), (s_2, 1), (s_3, 1), (s_3, 2)$. the transition and emission distribution for the HMM take the forms

$$\tilde{p}_{X_{i+1},T_{i+1}|X_i,T_i}(x_{i+1}, t_{i+1}|x_i, t_i) = \begin{cases} \phi(x_i, x_{i+1}, t_i) & \text{if } t_{i+1} = 1 \text{ and } x_{i+1} \neq x_i \\ \xi(x_i, t_i) & \text{if } t_{i+1} = t_i + 1 \text{ and } x_{i+1} = x_i \\ 0 & \text{otherwise} \end{cases}$$

and $\tilde{p}_{Y_i|X_i,T_i}(y_i|x_i, t_i)$, respectively. Express $\phi(x_i, x_{i+1}, t_i)$, $\xi(x_i, t_i)$, and $\tilde{p}_{Y_i|X_i,T_i}(y_i|x_i, t_i)$ in terms of parameters p_1, p_D, p_T, $p_{Y_i|X_i}$, k, N, and M of the original model.

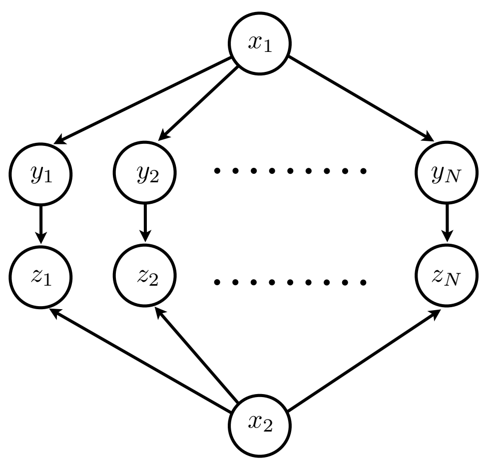
(*c*) We wish to compute the marginal probability for the final state X_N given the observations $Y_1, \dots, Y_N$. If we naively apply the sum-product algorithm to the construction in part (*b*), the computational complexity is $O(Nk^2M^2)$. Show that by exploiting additional structure in the model, it is possible to reduce the complexity to $O(N(k^2+kM))$. In particular, give the corresponding rules for computing the forward messages $\nu_{i+1\to i+2}(x_{i+1}, t_{i+1})$ from the previous message $\nu_{i\to i+1}(x_i, t_i)$. Do not worry about the beginning or the end of the sequence and restrict your attention to $2 \le i \le N-1$.
[Hint: substitute your solution from part (*b*) into the standard update rule for HMM messages and simplify as much as possible.]
[Note: If you cannot fully solve this part of the problem, you can receive substantial partial credit by constructing an algorithm with complexity $O(Nk^2M)$.]

Problem 2

Consider random variables X_1, X_2, $Y_1, \dots, Y_N$, $Z_1, \dots, Z_N$ distributed according to

$$p_{X_1,X_2,Y,Z}(x_1, x_2, y, z) = p_{X_1}(x_1)p_{X_2}(x_2)\prod_{i=1}^{N}\left[p_{Y|X_1}(y_i|x_1)p_{Z|Y,X_2}(z_i|y_i, x_2)\right],$$

where $X_1, Y_1, \dots, Y_N, Z_1, \dots, Z_N$ take on values in $\{1, 2, \dots, K\}$ and X_2 instead takes on a value in $\{1, 2, \dots, N\}$. A minimal directed I-map for the distribution is as follows:



Assume throughout this problem that the complexity of table lookups for p_{X_1}, p_{X_2}, $p_{Y|X_1}$, and $p_{Z|Y,X_2}$ are $O(1)$.

(*a*) A Bayesian network represented by a directed acyclic graph can be turned into a Markov random field by *moralization*. The moralized counterpart of a directed acyclic graph is formed by connecting all pairs of nodes that have a common child, and then making all edges in the graph undirected. Draw the moral graph over random variables $X_1, X_2, Y_1, \dots, Y_N$ conditioned on $Z_1, \dots, Z_N$. In other words, find an undirected I-map for the distribution of random variables $X_1, X_2, Y_1, \dots, Y_N$ conditioned on $Z_1, \dots, Z_N$.

Provide a good elimination ordering for computing marginals of $X_1, X_2, Y_1, \dots, Y_N$ conditioned on $Z_1, \dots, Z_N$. For your elimination ordering, determine α and β such that complexity of computing $p_{X_1|Z_1,\dots,Z_N}$ using the associated elimination algorithm is $O(N^\alpha K^\beta)$.

(*b*) For the remainder of this problem, suppose that we also have the following *context-dependent* conditional independencies: Y_i is conditionally independent of Z_i given $X_2 = c$ for all $i \neq c$. For fixed $z_1, \ldots, z_N$, x_1, and c, show that

$$p_{Z_1,\ldots,Z_N|X_1,X_2}(z_1 \ldots, z_N|x_1, c) = \eta(x_1, c, z_c)\lambda(c, z_1, \ldots, z_{c-1}, z_{c+1}, \ldots, z_N)$$

for some function $\eta(x_1, c, z_c)$ that can be evaluated in $O(K)$ operations for fixed (x_1, c, z_c), and some function $\lambda(c, z_1, \ldots, z_N)$ that can be evaluated in $O(N)$ operations for fixed $(c, z_1, \ldots, z_N)$. Express $\eta(x_1, c, z_c)$ in terms of $p_{Y|X_1}$ and $p_{Z|Y,X_2}$, and $\lambda(c, z_1, \ldots, z_{c-1}, z_{c+1}, \ldots, z_N)$ in terms of $p_{Z|X_2}$.

Optional problem

The graph G is a perfect undirected map for some strictly positive distribution $\mu(x)$ over a set of random variables $x = (x_1, \ldots, x_n)$, each of which takes values in a discrete set $\mathcal{X}$. Choose some variable x_i and let x_A denote the rest of the variables in the model, i.e., $\{x_i, x_A\} = \{x_1, \ldots, x_N\}$. Construct the graph G' from G by removing the node x_i and all its edges. Let some value $c \in \mathcal{X}$ be given. Show that G' is not necessarily a perfect map for the conditional distribution $\mathbb{P}_{x_A|x_i}(\cdot|c)$ by giving a counterexample.