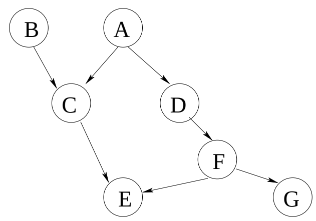
Probabilistic Reasoning in AI - Problem set 3

Due Wednesday, February 13, in class
This assignment is worth 8% of your grade

This assignment contains problems regarding exact and approximate inference methods (lectures 5-7)

1. [10 points] This problem is just practice for the exact inference algorithms we discussed. Consider the Bayes net below:



   - Suppose that we want to compute $P(E)$. Trace the variable elimination algorithm on this network, using the maximum cardinality search heuristic and starting with node A numbered as 1. If nodes are tied, break ties in alphabetic order. Show the cluster tree induced by variable elimination. Assuming all variables are Boolean, how many parameters are needed for each CPT in the tree?
   - Choose an ordering of the nodes that you think will be particularly bad. Explain in words why you chose the ordering. Perform variable elimination again, showing the cluster tree. How many parameters will the CPTs of the tree contain in this case?

2. [15 points] Assume that we have constructed a clique tree for a Bayesian network, and each clique has at most k nodes. Suppose we add an arc between two nodes in the graph. Give an upper bound on the maximum clique size in the clique tree for the new network. Justify your answer.

3. [35 points] In this problem you will show how the variable elimination algorithm can be changes to compute the Most Probable Explanation (MPE) hypothesis. Recall from lecture 5 that the most probable explanation is an assignment of values to all non-evidence variables X that maximizes $P(X|E = e)$ (where e is the evidence).

   (a) [10 points] Consider a chain of nodes linked sequentially:

   $$X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_n$$

   Without any evidence, show how to compute the probability of the most likely assignment to the chain, in time linear in the length of the chain. Hint: write down the expression for the probability, then consider an algorithm similar to variable elimination, but in which the summation is replaced by a max.

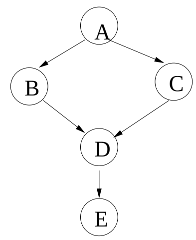
   (b) [10 points] Show how to compute the actual most likely assignment. Hint: this will require a second pass over the chain. You may wish to store intermediate results from the first pass.

   (c) [15 points] Consider an augmented chain, as shown below:

   $$\begin{array}{ccccccc} X_1 & \rightarrow & X_2 & \rightarrow & \cdots & \rightarrow & X_n \\ \downarrow & & \downarrow & & & & \downarrow \\ Y_1 & & Y_2 & & \cdots & & Y_n \end{array}$$

Assume that we have evidence $Y_i = y_i$ for every $i = 1, 2, \ldots n$, but none of the X_i is observed. Show how to compute the probability of the most likely assignment to X given $Y = y$. Then show how to compute the most likely assignment to X given $Y = y$. Hint: consider how your previous algorithms need to change.

4. [10 points] Consider the Bayes net below:

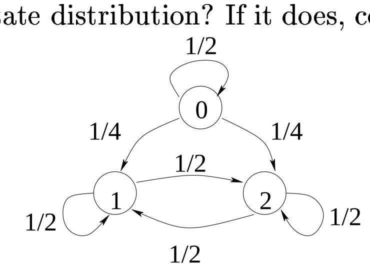


Assume $P(A) = 0.7$, $P(B = 1|A = 1) = 0.8$, $P(B = 1|A = 0) = 0.2$, $P(C = 1|A = 1) = 0.6$, $P(C = 1|A = 0) = 0.3$, $P(E = 1|D = 1) = 0.8$, $P(E = 1|D = 0) = 0.2$, and assume that D is a deterministic and of its parents. Suppose we want to compute $P(A = 1, E = 1|D = 1, C = 1)$ using likelihood weighting. Generate two samples using this methods and show their weights. Will likelihood weighting encounter any problems in this network? Justify your answer.

5. [10 points] Consider the Markov chain shown below. Compute the n-step transition probability matrix:

$$p_{ij}^{(n)} = P(s_n = j|s_0 = i), \forall i, j \in \{0, 1, 2\}$$

Does the chain have a steady-state distribution? If it does, compute it. Otherwise, justify why not.



6. [20 points] Gibbs sampling.

(a) [6 points] Show the equality from lecture 7, slide 16:

$$P(x_i|MB(X_i)) = \frac{P(x_i|Parents(X_i)) \prod_{j=1}^{k} P(Y_j|Parents(Y_j))}{\sum_{x_i'} P(x_i'|Parents(X_i)) \prod_{j=1}^{k} P(Y_j|Parents(Y_j))}$$

(b) [14 points] Consider the Bayes net shown below:

$$X \rightarrow Z \leftarrow Y$$

Assume X and Y are uniformly distributed, and Z is the deterministic exclusive or of X and Y. Show that Gibbs sampling on this structure with evidence $Z = 1$ will estimate $P(X = 1|Z = 1)$ as either 1 or 0. What happens if we make Z a slightly noisy exclusive or? E.g. Z is the exclusive or of X and Y with probability $1 - q$ and is chosen uniformly randomly with probability q. Explain this behavior in terms of what happens with the random walk generated by Gibbs sampling.