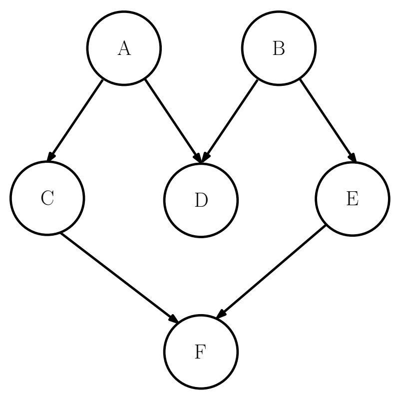
CSE 250A Assignment 3

This assignment is due at the start of class on Thursday October 25, 2012. Instructions are the same as for the first assignment. You must work in partnership with one other student, but you may keep the same partner or change partners, as you wish. *Acknowledgment: These questions are versions of ones written by Lawrence Saul.*

1. Ordering nodes

Consider the following Bayesian network:



Draw the minimal DAGs needed to represent the same joint distribution for the following node orderings:

(a) F, E, D, C, B, A

(b) D, A, B, C, E, F.

For each DAG, explain briefly which edges you include or omit, based on properties of conditional independence.

2. Clustering nodes

Consider a Bayesian network that is divided into layers numbered $j = 1$ to $j = k$ At each layer there are m binary nodes. There is an edge from every node at layer j to every node at layer $j + 1$, for $j = 1$ to $j = k - 1$.

(a) In total, how many free parameters are there in the CPTs for this network?

(b) Minimally, which nodes should be combined in order to make the network into an equivalent polytree?

(c) After combining nodes, how many free parameters are there in total in the CPTs for the polytree?

3. Intelligent inference

Consider a Bayesian network with nodes X_1 to X_T and edges $X_t \to X_{t+1}$ for $t < T$. Each node is a discrete random variable with n alternative values, that is $X_t \in \{1, 2, \ldots, n\}$. Suppose that the CPT for every node X_t for $t \geq 2$ is the same, so we have one fixed $n \times n$ transition matrix V with entries v_{ij}:

$$p(X_{t+1} = j | X_t = i) = v_{ij}.$$

(a) Show that the t-step transition probability $p(X_{t+1} = j | X_1 = i) = [V^t]_{ij}$.

(b) Devise a simple algorithm based on matrix-vector multiplication that computes $p(X_{t+1} = j | X_1 = i)$ in $O(n^2 t)$ time.

(c) Devise an alternative algorithm that performs the same computation in $O(n^3 \log_2 t)$ time.

4. Stochastic simulation

Consider a Bayesian network with n binary random variables $B_i \in \{0, 1\}$ and one integer random variable Z. There is an edge $B_i \to Z$ for each i. Let $f(B) = \sum_{i=1}^{n} 2^{i-1} B_i$ be the nonnegative integer whose binary representation is the string $B_n B_{n-1} \ldots B_2 B_1$. Suppose that the prior probability of each bit is $p(B_i = 1) = 0.5$ and that

$$p(Z | B_1, B_2, \ldots, B_n) = \frac{1 - \alpha}{1 + \alpha} \alpha^{|Z - f(B)|}$$

where $0 < \alpha < 1$ is a parameter measuring the amount of noise in the interpretation of the binary string. Larger α means greater noise.

(a) Show that the conditional distribution is normalized. That is, show that $\sum_z p(Z | B_1, B_2, \ldots, B_n) = 1$ where the sum is over all integers $z \in [-\infty, +\infty]$.

(b) Implement the method of likelihood weighting to estimate the probability $p(B_5 = 1 | Z = 20)$ for a network with $n = 9$ bits and noise level $\alpha = 0.01$. Hand in your source code, and a plot of the estimated probability as a function of the number of samples used in the method. You may code in the programming language of your choice, and use any software to plot the results. MATLAB works well for both.