Problem Set 2

Conditional Independence and d-Separation in Bayesian Networks

Prof. Jim Rehg
CS 7636 Computational Perception
College of Computing
Georgia Institute of Technology

January 22, 2004

Solutions must be submitted at beginning of class on Friday, January 30, 2004. Hand-written answers are fine as long as they are legible and organized. Please show all work. I've tried to leave space on this handout for your solutions. This problem set is worth 7% of your total grade, distributed among the problems as shown.

For Problem 5 you will have to perform significant numeric calculations and will need to use a computer package for Bayesian networks. You can find many packages on the web and are free to use whichever one you would like. I recommend that you use Kevin Murphy's *BNT* package which runs under Matlab and is quite popular. It may be downloaded from `http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html`.

1. [2% of grade]
Let x, y, z be binary random variables.

(a) For each of the following joint distributions, determine whether x and z are *dependent*, showing your analysis:

i. $p(x, z) =$

	$x = 0$	$x = 1$
$z = 0$	0.02	0.08
$z = 1$	0.18	0.72

ii. $p(x, z) =$

	$x = 0$	$x = 1$
$z = 0$	0.025	0.125
$z = 1$	0.125	0.725

(b) Suppose $p(x, y, z)$ factors as $p(x, y, z) = p(y)p(x|y)p(z|y)$ and that $p(y) = [0.5 \quad 0.5]$. For each of the following parameterizations, identify all of the *independencies* in the joint distribution, showing the details of your analysis. (Be sure to consider lack of dependency resulting from a specific numerical parameterization, as well as any structural independencies).

i. $p(x|y) =$

	$y = 0$	$y = 1$
$x = 0$	0.2	0.1
$x = 1$	0.8	0.9

$p(z|y) =$

	$y = 0$	$y = 1$
$z = 0$	0.2	0.1
$z = 1$	0.8	0.9

ii. $p(x|y) =$

	$y = 0$	$y = 1$
$x = 0$	0.2	0.2
$x = 1$	0.8	0.8

$p(z|y) =$

	$y = 0$	$y = 1$
$z = 0$	0.1	0.1
$z = 1$	0.9	0.9

(c) Suppose $p(x, y, z)$ factorizes as $p(x, y, z) = p(x)p(y|x)p(z|x, y)$.

i. Draw this factorization as a directed graphical model.

ii. Characterize the dependencies and independencies.

iii. Calculate the total number of parameters (degrees of freedom) in the joint distribution. How would this number change if $p(x, y, z)$ factored as $p(x, y, z) = p(y)p(x|y)p(z|y)$?

iv. Find a specific numerical parameterization θ for this model with the property that $y \perp\!\!\!\perp x \mid z, \theta$.

(d) Suppose $p(x, y, z)$ factorizes as $p(x, y, z) = p(x)p(y|x)p(z|y)$. Find a specific numerical parameterization for this model with the properties:

i. $z \perp\!\!\!\perp x \mid y$

ii. $x = z$.

By $x = z$ we mean that when a sample (x, y, z) is drawn according to $p(x, y, z)$, it will always be the case that x and z have the same value.

2. [1% of grade]
Let $\mathcal{G}(V, E)$ be a directed acyclic graph with n vertices. Each node $j \in V$ represents a variable $x_j \in X$, establishing a one-to-one correspondence between V and X. Let $\pi_j = \{i \in V | (i, j) \in E\}$ denote the set of parents of node j. For each node x_i, define the potential function $f(x_i; x_{\pi_i})$ with the properties $f(x_i; x_{\pi_i}) \geq 0$ and $\sum_{x_i} f(x_i; x_{\pi_i}) = 1$. Prove the following:

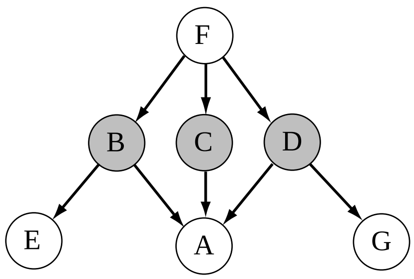
Theorem PS2-1 Let the joint distribution for X be defined by $p(X) = \prod_{i=1}^{n} f(x_i; x_{\pi_i})$. Then:

i. $p(X)$ is a valid probability distribution (it is non-negative and sums to one).

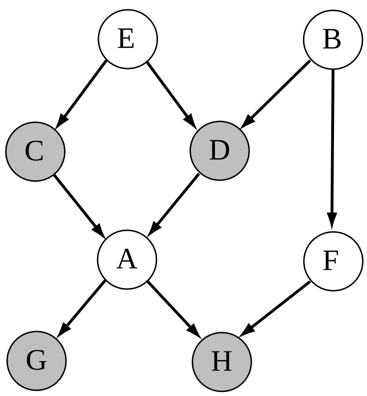
ii. $p(x_i | x_{\pi_i}) = f(x_i, x_{\pi_i})$ (the conditional probability of a node given its parents is exactly its potential function).

3. [1% of grade]

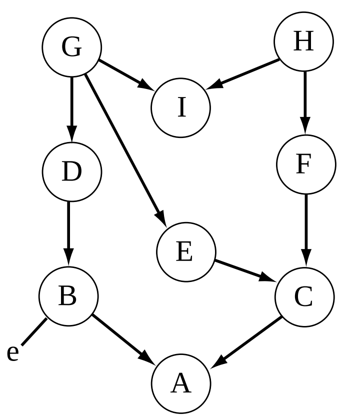
For the following graphs, determine which variables are *d-separated* from A. For each variable that is not d-separated, specify an unblocked path by listing the nodes in the order they are visited.



(i)



(ii)



(iii)

4. [1% of grade]
Let $\mathcal{G}(V, E)$ be a directed acyclic graph with n vertices. Each node $j \in V$ represents a variable $x_j \in X$, establishing a one-to-one correspondence between V and X. Let $\pi_j = \{i \in V | (i, j) \in E\}$ denote the set of parents of node j.

A distribution $p(X)$ is said to factorize according to $\mathcal{G}$ if we can write

$$p(X) = \prod_{i \in I} f(x_i; x_{\pi_i}),$$

where I is a topological ordering of the nodes in V and $f(x_i; x_{\pi_i})$ are the potential functions from Theorem PS2-1. We refer to this property of $p(X)$ and $\mathcal{G}$ as *directed factorization* ($\mathcal{DF}$). Further, we describe the d-separation properties of $\mathcal{G}$ by a set of relations $\{A_i \text{ d-sep } C_i \mid B_i\}_i$, where A_i, B_i, C_i are disjoint sets of nodes in V. They could be computed using the Bayes Ball algorithm, for example. Similarly, the conditional independencies of $p(X)$ consist of the relations $\{A_i \perp\!\!\!\perp C_i \mid B_i\}_i$.

With respect to the above definitions, Theorem PS2-1 states that

$$\mathcal{DF} \Rightarrow f(x_i; x_{\pi_i}) = p(x_i | x_{\pi_i}) \; \forall i,$$

and Theorem 4.6 on p. 67 of the text states that

$$\mathcal{DF} \Rightarrow \text{ if } A_i \text{ d-sep } C_i \mid B_i \text{ holds in } \mathcal{G}, \text{ then } A_i \perp\!\!\!\perp C_i \mid B_i \text{ holds in } p(X) \; \forall i.$$

Prove the converse of Theorem 4.6, that the equivalence of d-separation and conditional independencies implies $\mathcal{DF}$. Note that your text proves this result using undirected graphs. Here you are asked to show the result using directed graphs, by following the induction approach used in Theorem 4.6: Assume $\mathcal{DF}$ holds for $n - 1$ nodes obtained by removing a node from $p(X)$. Show using the properties of d-separation that $\mathcal{DF}$ holds for $p(X)$ as well. Specifically, you are asked to prove:

Theorem PS2-2 (converse of Th 4.6) Given $\mathcal{G}$ and $p(X)$ with the property that if A_i d-sep $C_i \mid B_i$ holds for $\mathcal{G}$, then $A_i \perp\!\!\!\perp C_i \mid B_i$ holds for $p(X)$. Then we have $p(X) = \prod_{i \in I} p(x_i | x_{\pi_i})$.

5. [2% of grade]

This problem is under development, an updated problem set will be posted shortly. In the meantime, familiarize yourself with BNT or some equivalent Bayes net software.