STUDENT NAME: _____

STUDENT ID: _____

MIDTERM EXAMINATION

Probabilistic reasoning in AI - Winter 2005

March 1, 2005

You are allowed one double sided cheat sheet.

There are 10 questions, for a total of 100 points. Please read the questions first, as they cover a whole range of topics.

Answer all questions on the exam booklet

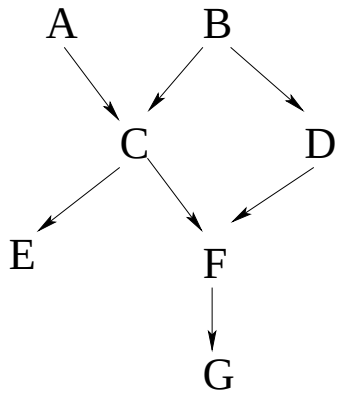
Good luck!

1. [5 points] **Probabilities**

   Let A and B be two binary random variables. Show that if $p(B = 1|A) < p(B = 1)$ then $p(B = 0|A) > p(B = 0)$.

2. [25 points] **Independence, exact inference**

Consider the following Bayes network:



(a) [3 points] For each of the following assertions state whether they are true or false:

i. $A \perp\!\!\!\perp B$

ii. $A \perp\!\!\!\perp B | C$

iii. $A \perp\!\!\!\perp B | D$

(b) [5 points] Write the parameters on the network picture above. Assuming a tabular representation of the CPDs, and binary variables, how many parameters will there be?

(c) [2 points] Draw the moral graph of the network

(d) [2 points] What is the Markov blanket of node C?

(e) [3 points] Using Bayes rule and the definition of conditional probabilities, express $p(C|E)$ based on the parameters in the network.

(f) [5 points] Draw the junction tree for the the network

(g) [5 points] Write the initial parameters of the junction tree

3. [5 points] **Short questions**

   (a) [2 points] True or false: $p(A) \geq p(A, B), \forall A, B$

   (b) [3 points] You have a Bayes net with 100 nodes. Imagine the lexicographic ordering of the nodes, such that a node comes after all its parents in the ordering. You have a query conditioned on a variable that is almost at the end of this ordering. You want to use approximate inference to compute conditional probability queries for this variable. Should you use likelihood weighting or Gibbs sampling? Justify your answer in one sentence

4. [12 points] Consider the Bayes net below:

$$X \rightarrow Y \rightarrow Z$$

in which X, Y and Z are Boolean variables and $p(X = 1) = 0.5$, $p(Y = 1|X = 1) = 0.7$, $p(Y = 1|X = 0) = 0.2$, $p(Z = 1|Y = 1) = 0.9$ and $p(Z = 1|Y = 0) = 0.6$

(a) [2 points] Suppose that you have evidence $Y = 1$. Give a sample generated by likelihood weighting with this evidence, and its weight.

(b) [5 points] Show the Markov chain corresponding to Gibbs sampling on this network, with evidence $Y = 1$, assuming that variables to be flipped are picked at random. You do not need to compute the numerical result for the transitions, just show the formulas and plug in the numbers.

(c) [5 points] Give an equation satisfied by the steady-state distribution on this chain. Explain how it relates to conditional probabilities in the original network.

5. [10 points] **Maximum likelihood estimation**

   Let X be a random variable drawn uniformly from and interval $[a, b]$, where a and b are unknown. More precisely,

$$p(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{if } x > b \end{cases}$$

   You observe a sequence of samples $x_1, \ldots x_n$, with $n \geq 2$. Assume that at least two of these samples have different values.

   Compute the maximum likelihood values for a and b.

6. [5 points] **Undirected models**

   Let X_i, $i = 1 \ldots n$ and Y be random variables. Draw a Markov network such that, for all $i \neq j$, $X_i \perp\!\!\!\perp X_j | Y$.

7. [15 points] **Parameter estimation**

   You are given the following table of data, corresponding to a naive Bayes model:

   $$F \leftarrow D \rightarrow S$$

   Each row corresponds to a patient.

Disease (D)	Fever (F)	Sore throat (S)
1	1	1
0	0	1
0	0	0
0	0	0
1	1	1
1	0	1
1	1	0
1	1	0
0	1	0
0	0	0

   (a) [5 points] What are the maximum likelihood parameters based on this data?

   (b) [5 points] Suppose a patient comes in with a fever and no sore throat. What is your diagnosis?

(c) [5 points] Suppose you have 5 more examples of patients which have the disease and had a fever, but no one asked them about sore throat. Suppose that you initialize the parameters of the network with the estimates from part a. Show one iteration of the EM algorithm, and compute the new parameter values.

8. [8 points] **Using Bayes nets**

   A car insurance company wants to build a probabilistic model of customer behavior. In the future, they plan to use this in order to determine premiums for new customers, based on the estimated probability of accidents. They hire you as a consultant and present you with a data set of 50 customers, for which they recorded: age (5 brackets), gender (2 values), occupation (5 categories), car make (100 values), number of accidents within a one-year period (integer). They want you to build a model. What do you do?

   (a) Tell them to go get a serious data set and then come find you. There is too little information to do anything.

   (b) Convince a company expert to sit down with you and make up a reasonable model based on their experience. There is too little data to use it.

   (c) Convince a company expert to sit down with you and draw a model structure, then you use maximum likelihood to fit parameters.

   (d) Convince a company expert to sit down with you and draw a model structure, then use Bayesian learning to fit parameters.

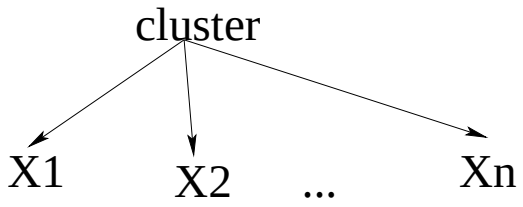
   (e) Use the data to do structure learning with simulated annealing and MDL scoring

   (f) None of the above. In this case, state your choice.

   Justify your choice in maximum 3 sentences.

9. [10 points] **Latent variables**

One of the interesting problems in machine learning and data mining is that of clustering: given some data points, described by attributes $X_1, \ldots X_n$, find a number of clusters that capture the data well. This problem can be formulated as learning a graphical model of the following form:



In this case, the Cluster variable is never observed, and the number of values is unknown.

(a) Suppose Cluster can take only 1 possible value. When you fit parameters using MLE, how many parameters will you have?

(b) Do you expect this model to give a high likelihood to the data?

(c) Suppose you have N instances, and you are assuming N possible values for the Cluster variable. How many parameters does your model have?

(d) Is N a good choice for the number of values for the cluster? Justify your answer.

(e) Describe in at most 3 sentences an algorithm for choosing an appropriate number of clusters.

10. [10 points] **Entropy and mutual information**

    Let X and Y be two discrete random variables, with a finite number of values. Let $H(X)$ and $H(Y)$ be the entropy of X and Y respectively. Let $H(X,Y)$ be the entropy of their joint distribution. Prove that $H(X,Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent.