CS475 Spring 2006
Homework 2

Due Date: March 8 2006

You may work on this problem set in a group (up to 4 people) and submit one solution per group. Note that some problems refer to the Artificial Intelligence: A Modern Approach (AIMA) book, and you can find the relevant chapters at the course webpage (http://www.cs.cornell.edu/selman/cs475/index.htm).

Problem 1

Solve problems 14.3 and 14.4 in the AIMA book, page 534.

Please provide explanation to you answer in 14.3 a). For problem 14.4, find the answer first symbolically, and then use values $e = 0.1$ and $f = 0.02$ to compute it numerically as well.

Problem 2

Let X be a random variable corresponding to some node in a Bayesian Network. Recall that a Markov Blanket of X is a set of variables consisting of X's parents, children and children's parents in the Network. So:

$$MB(X) = Parents(X) \cup Children(X) \cup [Parents(Children(X)) - \{X\}]$$

Denote by $\mathbf{V}_{-X}$ the set of all variables in the system (i.e. variables that correspond to some node in the Network) *except* for X.

1. Show that $P[X|\mathbf{V}_{-X}] = P[X|MB(X)]$, e.i. a variable is conditionally independent of all other variables given its Markov Blanket. Use the fact that X is conditionally independent of all its non-descendants given its parents (this is true by construction of the Bayesian Network). Be sure to specify what a "non-descendant" node is.

2. Show (e.g. by counter example) that $Parents(Children(X))$ cannot be left out of the definition of Markov Blanket. So you need to show that there is a network in which some X is *not* conditionally independent of all other nodes given $Parents(X) \cup Children(X)$.

Problem 3

This is a problem where you will have a chance to play around with Bayesian Networks and try out the various approaches of inference we discussed in class.

Consider the burglary/earthquake/alarm/call network from the class (see AIMA book, figure 14.2 on page 494). We will be calculating the probability that there is a burglary, given that John calls and there is no earthquake, i.e.:

$$P[burglary|john\_calls, \neg earthquake]$$

The CPTs for this problem are given on page 494 in the AIMA book (figure 14.2).

1. Solve the problem using **exact inference**. You can use Netica program, downloadable from http://www.norsys.com/download.html (a trial version has some size limitations, but it will work fine for our purpose). Construct the network in Netica, and calculate the conditional probability in question (notice that Netica uses "34.5"% instead of "0.345" to quantify probability). Please include a snapshot of your network in the solution.

2. Now solve it using **rejection sampling**. Use your favorite programming language to implement the sampler (Matlab might be a good idea). Submit a "high-level" overview of how your sampler works, do not include the source code. Make a plot of number-of-samples (x-axis) vs. estimate of the desired probability. How long do you need to let is sample?

3. And, finally, use **Markov Chain Monte Carlo** to do the same problem. Construct a Markov Chain corresponding to the state space of the problem that you need to solve. Calculate the desired transition probabilities using Gibbs Sampling method. Show how are the transition probabilities computed in general, and use the formula to find at least one (any) value. You can find the rest of the values using the Netica network you have constructed before. Include the transition probability matrix in your solution (please include mapping between state numbers in your chain and situations (values of variables) in the hypothetical world). Again, plot the number-of-samples vs. estimate of the desired probability plot.

   You can use the applet we used in class (http://www.math.uah.edu/stat/applets/MarkovChainExperiment.xhtml) to do the walk for you (you can click inside the Time-X window, press Ctrl-A to select all values, and copy it out to a text file for analysis of the walk).

4. How would you compare the three methods, what are their strengths and weaknesses?

5. Now do steps 1.–3. for estimating probability of

$$P[john\_calls|marry\_calls, \neg burglary, \neg earthquake]$$

Which of the two probabilities are easier to estimate using the approximate inference methods? If there is any difference, why is one inference easier to do than the other?

Problem 4

In class, we mentioned that Monte Carlo is a very generic approach to solving problems using randomness. Markov Chain Monte Carlo methods are just one family of Monte Carlo algorithms. In this problem, we will use Monte Carlo principle to solve a very different problem: approximating the value of the constant π (i.e. the popular[1] $3.1415926\ldots$ number :-))

Recall that a an area of a unit circle ($radius = 1$)) is π. So if we pick a point uniformly at random from a square that circumvents the circle (with sides of size 2), then the probability that the point will happen to be inside the circle is $\pi/4$ (area of the circle/area of the square). By the Law of Large numbers, if we repeat this experiment for a sufficient number of times, the average number of times we are inside the circle will be close to the true value of $\pi/4$.

1. How do you generate a point uniformly at random from a square of given size? How do you then test if the point is inside the circle that is inscribed into the square?

2. Implement the sampling procedure and run it to estimate the value of π. Notice that both the square and the circle are symmetric along both x and y axis, so you only need to consider one quarter of both. Make a plot that shows number-of-samples vs. closeness-to-π. How long do you have to run it in order to get an approximation that is accurate to the 2nd digit of π?

3. Use the same algorithm to estimate the volume of a 2,3,4,5-dimensional unit sphere (a unit n-dimensional sphere centered at origin is a set of points $(x_1, \ldots, x_n) \in \mathbf{R}^n$ such that $\sum_{i=1}^{n} x_i^2 \leq 1$). The volume can be estimated as the fraction of points chosen uniformly at random from an n-dimensional cube with sides 2, which happen to be in the sphere. Would this simple algorithm be efficient in calculating the volume of a 100 dimensional sphere?

[1] Probably my favorite math expression is $e^{\pi i}+1 = 0$. It puts together some basic constants ($0, 1, e, \pi, i = \sqrt{-1}$), operations (addition, multiplication and exponentiation) and relation (equivalence), all in just one expression. If you also have a "favorite" math expression and don't mind sharing it, please let me know. We can create a "top n" list. :-)