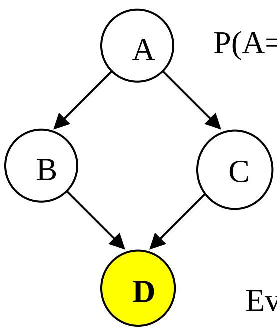
COMPSCI 276
Homework Assignment 7
Spring 2017

Instructor: Rina Dechter   Due: Monday, June 5th

For sampling consult class slides, the “Cutset sampling in Bayesian networks” http://www.ics.uci.edu/~csp/r137.pdf and Darwiche chapter 15 .

Question 1 (20).

Consider a Bayesian network with variables A, B, C, D, and evidence D=0:



P(A=0)=0.9, P(A=1)=0.1

Evidence: D=0

A	P(B=0)	P(B=1)
0	0.9	0.1
1	0.1	0.9

A	P(C=0)	P(C=1)
0	0.9	0.1
1	0.1	0.9

BC	P(D=0)	P(D=1)
00	0.99	0.01
01	0.1	0.9
10	0.1	0.9
11	0.05	0.95

a. Compute the exact posterior probabilities P(A|D=0), P(B|D=0), P(C|D=0).

b. Show how to generate a sample using forward sampling. Generate 10 samples using forward sampling. How many samples do you get where D=0 and how many did you have to reject? Estimate posterior probabilities for B, C, A based on those 10 samples?

c. Show how you would generate a sample using ordered Gibbs sampler. Generate 10 samples and compute the posterior probabilities (beliefs).

d. Show how you would generate a sample using Gibbs cutset sampling by sampling only cutset {B,C}. Estimate the posterior probabilities for B, C, A by 1) counting the fraction of samples with desired variable value; 2) using mixture estimator.
Which estimate was better? How do Gibbs estimates compare to forward sampling?

e. Write a few sentences to outline the main trade-offs (the good and the bad) between using forward sampling and Gibbs sampling.
Note: You can hard-code the sampling process in the language of your choice or you may perform calculations by hand. Mainly, you need a random number generator that generates the numbers in the range [0, 1] to draw a random value. You can generate a sequence of random numbers using C/C++/Java random number generator or even Excel (just type =rand() in the Excel cell). You can build the network and compute exact posterior marginals as well as necessary sampling probabilities for Gibbs sampler using a Bayesian Network applet:
http://www.aispace.org/bayes/index.shtml (click the button to start the applet at the link above)

Question 2 (20).

Consider a simple network in Figure 1.
Assume all nodes are bi-valued with domains D(X) = {0,1}.
Assume we know following CPT:

P(I|G,H)

G	H	P(I=0|G,H)	P(I=1|G,H)
0	0	1	0
0	1	0.91	0.09
1	0	0.91	0.09
1	1	0.95	0.05

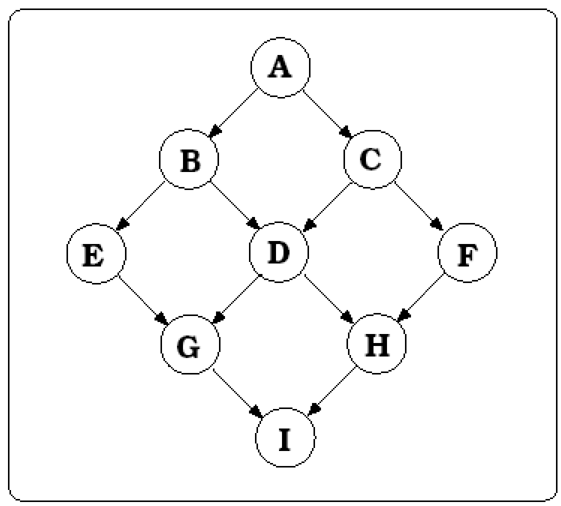


Figure 1: A Bayesian network

a) Is this network ergodic? Explain.

b) Assume node I is observed: {I = 1}. We say that assignment of values to the variables is *consistent* with evidence if they do not conflict with the evidence values.
Assume we want to apply Gibbs sampling to this network and we assigned initial values to the network nodes as follows:
S° = {A = 0,…,G = 0, H = 0}
Is this instantiation *consistent* with evidence? Explain.

c) Assume node I is observed: {I=i}. Find a cycle-cutset of the network C and trace **cutset sampling** algorithm (Gibbs sampling on cutset C):
Show step-by-step how the first two samples would be computed.
Discuss what kind of algorithm you can use to compute $P(C_i \mid C \setminus C_i)$ and what the complexity would be.

d) Given I=1 show how to generate a sample using LW (likelihood weighting)

Question 3 (Extra credit, 10).

In this question, we consider the application of importance sampling to Markov networks. Assume you are given a Markov network that expresses a probability distribution.
1. Explain intuitively why we cannot simply apply likelihood weighting to Markov networks.
2. Show how likelihood weighting can be applied to chordal Markov networks (namely to a decomposable distribution). Is this approach interesting? Explain.
3. Provide a general method that allows the application of importance sampling to Markov networks. Be sure to define both a reasonable proposal distribution and an algorithmic technique for computing the weights.

a	$p(a)$
0	0.3
1	0.7

b	$p(b)$
0	0.6
1	0.4

e	$p(e)$
0	0.7
1	0.3

y	x	$p(x\|y)$
0	0	0.10
0	1	0.90
1	0	0.30
1	1	0.70

z	y	x	$p(x\|y,z)$
0	0	0	0.25
0	0	1	0.75
0	1	0	0.60
0	1	1	0.40
1	0	0	0.10
1	0	1	0.90
1	1	0	0.20
1	1	1	0.80

Figure 3: Conditional probability tables

Relevant reading: Class notes (Dechter chapter 9) and Darwiche chapter 15.

4. (20) Assume you are given a 4x4 directed grid (like in Figure 1 but having dimension 4 rather than 3).
(a) Provide an arc-labeled minimal dual graph for the 4x4 grid network.
(b) Generate a join-graph whose maximal cluster size is 4. Show the functions in each cluster and the variables in each cluster. Label the arcs with the appropriate separators.
(c) Show the schematic messages that would pass in one iteration of IJGP on your join graph.