CS 188 Spring 2014	Introduction to Artificial Intelligence	Written HW8

INSTRUCTIONS

- **Due:** Tuesday, April 8th 2014 11:59 PM
- **Policy:** Can be solved in groups (acknowledge collaborators) but must be written up individually. However, we strongly encourage you to first work alone for about 30 minutes total in order to simulate an exam environment. Late homework will not be accepted.
- **Format:** You must solve the questions on this handout (either through a pdf annotator, or by printing, then scanning; we recommend the latter to match exam setting). Alternatively, you can typeset a pdf on your own that has answers appearing in the same space (check edx/piazza for latex templating files and instructions). **Make sure that your answers (typed or handwritten) are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.**
- **How to submit:** Go to www.pandagrader.com. Log in and click on the class CS188 Spring 2014. Click on the submission titled Written HW 8 and upload your pdf containing your answers. If this is your first time using pandagrader, you will have to set your password before logging in the first time. To do so, click on "Forgot your password" on the login page, and enter your email address on file with the registrar's office (usually your @berkeley.edu email address). You will then receive an email with a link to reset your password.

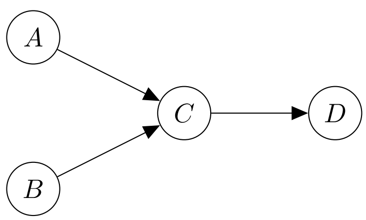
Last Name	Jianzhong
First Name	Chen
SID	23478230
Email	chenjianzhong@berkeley.edu
Collaborators	None

For staff use only

Q. 1	Q. 2	Total
/15	/15	/30

Q1. [12 pts] Bayes' Net Sampling

Assume you are given the following Bayes' net and the corresponding distributions over the variables in the Bayes' net.



$P(A)$	
+a	0.1
-a	0.9

$P(B)$	
+b	.7
-b	.3

$P(C\|A,B)$			
+c	+a	+b	.25
-c	+a	+b	.75
+c	-a	+b	.6
-c	-a	+b	.4
+c	+a	-b	.5
-c	+a	-b	.5
+c	-a	-b	.2
-c	-a	-b	.8

$P(D\|C)$		
+d	+c	.5
-d	+c	.5
+d	-c	.8
-d	-c	.2

(a) [3 pts] Assume we receive evidence that $A = +a$. If we were to draw samples using rejection sampling, on expectation what percentage of the samples will be **rejected**?

1a: on expectation, 90 percent of the samples will be rejected, because P(A=+a)=0.1

(b) [3 pts] Next, assume we observed both $A = +a$ and $D = +d$. What are the weights for the following samples under likelihood weighting sampling?

Sample	Weight
$(+a, -b, +c, +d)$	1 Bi: 0.05
$(+a, -b, -c, +d)$	1 Bii: 0.08
$(+a, +b, -c, +d)$	1 Biii: 0.08

(c) [3 pts] Given the samples in the previous question, estimate $P(-b| + a, +d)$.

1c: $\frac{0.05+0.08}{0.05+0.08+0.08} = \frac{13}{21}$

(d) [3 pts] Assume we need to (approximately) answer two different inference queries for this graph: $P(C| + a)$ and $P(C| + d)$. You are required to answer one query using likelihood weighting and one query using Gibbs sampling. In each case you can only collect a relatively small amount of samples, so for maximal accuracy you need to make sure you cleverly assign algorithm to query based on how well the algorithm fits the query. Which query would you answer with each algorithm?

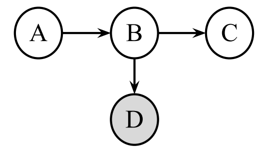
Algorithm	Query
Likelihood Weighting	$P(C\| + a)$

Algorithm	Query
Gibbs Sampling	$P(C\| + d)$

Justification: because likelihood sampling only takes upstream evidence into account, but Gibbs sampling takes both upstream and downstream into account

Q2. [18 pts] Preprocessing Bayes' Net Graphs for Inference

For (a) and (b), consider the Bayes' net shown on the right. You are given the structure but you are not given the conditional probability tables (CPTs). We will consider the query $P(B|+d)$, and reason about which steps in the variable elimination process we might be able to execute based on just knowing the graph structure and not the CPTs.



(a) [2 pts] Assume the first variable we want to eliminate is A and the resulting factor would be $f_1(B)$. Mark which one of the following is true.

○ $f_1(+b) = f_1(-b) = 1$

○ $f_1(+b) = f_1(-b) = 0$

● $f_1(B)$ *cannot* be computed from knowing only the graph structure.

○ $f_1(B)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

2a Explanation: B depends on A, so $f_1(B)$ depends on A

(b) [2 pts] Assume the first variable we eliminate is C and the resulting factor is $g_1(B)$. Mark which one of the following is true.

● $g_1(+b) = g_1(-b) = 1$

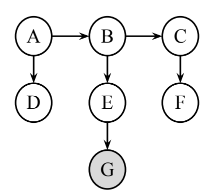
○ $g_1(+b) = g_1(-b) = 0$

○ $g_1(B)$ *cannot* be computed from knowing only the graph structure.

○ $g_1(B)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

2b Explanation: $g_1(+b) = \sum_c P(c|+b)P(c) = \sum_c P(c|+b)\sum_c P(c) = 1$ and $g_1(-b) = \sum_c P(c|-b)P(c) = \sum_c P(c|-b)\sum_c P(c) = 1$

For (c) through (g), consider the Bayes' net shown on the right. You are given the structure but you are not given the conditional probability tables (CPTs). We will consider the query $P(B|+g)$, and again we will reason about which steps in the variable elimination process we might be able to execute based on just knowing the graph structure and not the CPTs.



(c) [2 pts] Assume the first variable we eliminate is D and the resulting factor is $h_1(A)$. Mark which one of the following is true.

● $h_1(+a) = h_1(-a) = 1$

○ $h_1(+a) = h_1(-a) = 0$

○ $h_1(A)$ *cannot* be computed from knowing only the graph structure.

○ $h_1(A)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

2c Explanation: $h_1(+a) = \sum_d P(d|+a)P(d) = \sum_d P(d|+a)\sum_d P(d) = 1$ and $h_1(-a) = \sum_d P(d|-a)P(d) = \sum_d P(d|-a)\sum_d P(d) = 1$

(d) [2 pts] Assume the first 2 variables we eliminate are A and D and the resulting factor is $i_2(B)$. Mark which one of the following is true.

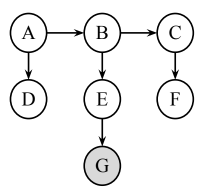
- ○ $i_2(+b) = i_2(-b) = 1$
- ○ $i_2(+b) = i_2(-b) = 0$
- ● $i_2(B)$ *cannot* be computed from knowing only the graph structure.
- ○ $i_2(B)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

2d Explanation: because B depends on A

(e) [2 pts] Assume the first variable we eliminate is F and the resulting factor is $j_1(C)$. Mark which one of the following is true.

- ● $j_1(+c) = j_1(-c) = 1$
- ○ $j_1(+c) = j_1(-c) = 0$
- ○ $j_1(C)$ *cannot* be computed from knowing only the graph structure.
- ○ $j_1(C)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

2e Explanation: $j_1(+c) = \sum_f P(f|+c)P(d) = \sum_f P(f|+c)\sum_f P(f) = 1$ and $j_1(-c) = \sum_f P(f|-c)P(f) = \sum_f P(f|-c)\sum_f P(f) = 1$



For your convenience we included the Bayes' Net structure again on this page.

(f) [2 pts] Assume the first 2 variables we eliminate are $F$and C and the resulting factor is $k_2(B)$. Mark which one of the following is true.

- ● $k_2(+b) = k_2(-b) = 1$
- ○ $k_2(+b) = k_2(-b) = 0$
- ○ $k_2(B)$ *cannot* be computed from knowing only the graph structure.
- ○ $k_2(B)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

2f Explanation: eliminating F and C involves summing up a conditional distribution and a marginal distribution, since both of them are distributions, the sum has to equal to 1

(g) [2 pts] Assume the first variable we eliminate is E and the resulting factor is $l_1(B, +g)$. Mark which one of the following is true.

- ○ $l_1(+b, +g) = l_1(-b, +g) = 1$
- ○ $l_1(+b, +g) = l_1(-b, +g) = 0$
- ● $l_1(B, +g)$ *cannot* be computed from knowing only the graph structure.
- ○ $l_1(B, +g)$ *can* be computed from knowing only the graph structure but is not equal to any of the provided options.

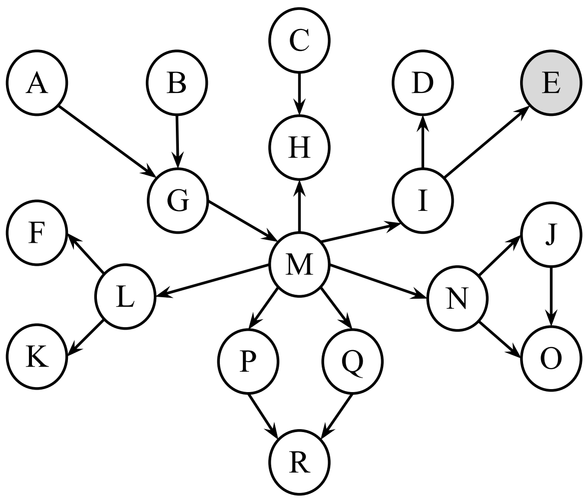
2g Explanation: because G depends on E

(h) [4 pts] In the smaller examples in (a) through (g) you will have observed that sometimes a variable can be eliminated without knowing any of the CPTs. This means that variable's CPT does not affect the answer to the probabilistic inference query and that variable can be removed from the graph for the purposes of that probabilistic inference query. Now consider the following, larger Bayes' net with the query $P(Q|+e)$.

Mark all of the variables whose CPTs do *not* affect the value of $P(Q|+e)$:

○ A	○ B	● C	● D	○ E
● F	○ G	● H	○ I	● J
● K	● L	○ M	● N	● O
● P	○ Q	● R		

2h Explanation: because if a variable is not in the union of the ancestors of the query and the ancestors of the evidence can be ignored.


A
B
C
D
E
H
G
I
F
M
J
L
N
K
P
Q
O
R