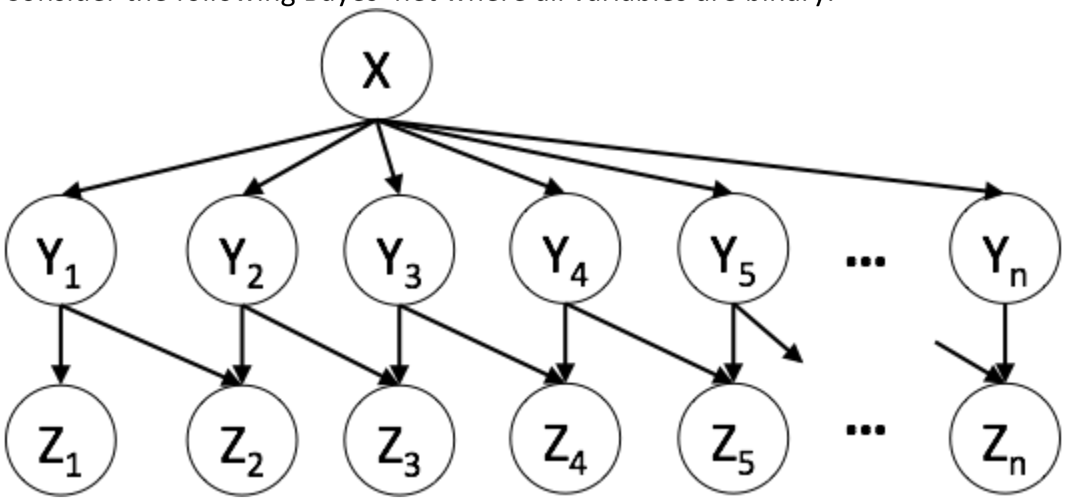
UC Berkeley – Computer Science
CS188: Introduction to Artificial Intelligence

Practice Midterm Solution, Fall 2017

1. Bayes Net Inference

Consider the following Bayes' net where all variables are binary.



I. Assume that we would like to perform inference to obtain
$P(Y_n|Z_1 = z_1, Z_2 = z_2, \cdots, Z_n = z_n)$.

(a) What is the number of rows in the largest factor generated by *inference by enumeration*, for this query $P(Y_n|Z_1 = z_1, Z_2 = z_2, \cdots, Z_n = z_n)$?

____

Solution: 2^{n+1}

(b) Suppose we decide to perform variable elimination to calculate the query $P(Y_n|Z_1 = z_1, Z_2 = z_2, \cdots, Z_n = z_n)$, and we eliminate the variables in the ordering $Y_1, Y_2, \cdots, Y_{n\text{-}1}, X$. Write out the form and the size of factor generated by the elimination of Y_i and X, where $i \in \{2, \cdots, n\text{-}1\}$.

	The new factor generated	The size of the new factor
After Y_i is eliminated	____________________	____________________
After X is eliminated	____________________	____________________

Solution: After Yi is eliminated, the new factor produced is $f_i(X, Y_{i+1}, z_1, z_2, \cdots, z_{i+1}) = \sum_{y_i} p(y_i|X)p(z_i|y_i)p(z_{i+1}|y_i, Y_{i+1})f_{i\text{-}1}(X, Y_i, z_1, \cdots, z_i)$, the size of which is 4. After X is eliminated, the new factor generated is $\sum_x f_{n\text{-}1}(x, Y_n, z_1, \cdots, z_n)$ whose size is 2.

(c) Find the best and the worst variable elimination orderings for this query. The goodness is measured by the sum of the sizes of the factors that are generated.

Best ordering: ____________________

Worst ordering: ___________________

Solution: The best ordering is to eliminate all Y's before X. The worst is to eliminate X first.

Ii. Assume now we want to use variable elimination to calculate *another* query $P(Z_n|Y_1, Y_2, \cdots, Y_{n\text{-}1})$.

(a) Mark all of the following variables that produce a *constant* factor after being eliminated for all possible elimination orderings.

☐ Z_1 ☐ Z_2 ☐ $Z_{n\text{-}1}$ ☐ X

Solution: everything except X

(b) List all the variables that can be ignored (i.e. the conditional probability tables of the variables can be ruled out from the initial factor set), in performing the query $P(Z_n|Y_1, Y_2, \cdots, Y_{n\text{-}1})$. Briefly present the reason why these variables can be ignored. (Hint: you can use the results in previous part or the conditional independencies encoded in the given BN).

Variables can be ignored when computing the query $P(Z_n|Y_1, Y_2, \cdots, Y_{n\text{-}1})$:

____

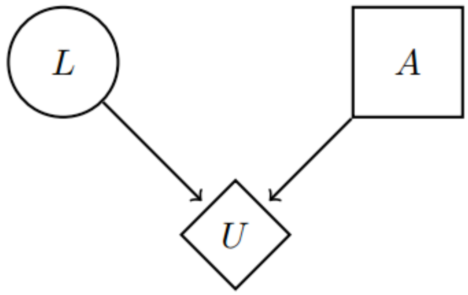
Reason:

____

Solution: The variables can be ignored are $Z_1, Z_2, \cdots, Z_{n\text{-}1}$d because summing out these variables produces constant factors which do not contribute to the query result or these variables are conditional independent from Z_n given $Y_1, \cdots, Y_{n\text{-}1}$

2. Decision Network

In this problem, we will model a lottery as a decision network, where we are deciding whether or not it is worth buying a lottery ticket and playing the lottery.
We can consider the outcome of a lottery to be a node L, with A as the decision to play the lottery or not and U as our utility. The resulting decision network may look something like this:



For all parts of this question, assume that we have an utility function of $U(x) = x^3$, where x is the amount of money we pay and then (hopefully) win in the lottery, and that the price of a lottery ticket is $4.

Q1: For this first part, assume that the lottery is $[0,9/10; 14,1/10]$. Fill in the following table for $U|L, A$.

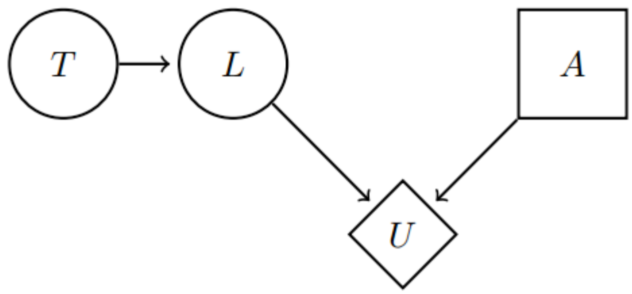
L	A	$U\|L, A$
0	buy	-64
14	buy	1000
0	no buy	0
14	no buy	0

What is $MEU(\{\})$?
Solution: $EU(no\ buy) = 0$ fairly clearly.
In order to calculate $EU(buy)$, we need to incorporate the probabilities of the lottery:
$EU(buy) = EU(buy|L = 14) * P(L = 14) + EU(buy|L = 0) * P(L = 0) = 10^3 * \frac{1}{10}$ - $4^3 * \frac{9}{10} =$ 42.4
Thus, $MEU(\{\}) = max(EU(no\ buy), EU(buy)) = max(0,42.4) = 42.4$.

What action achieves $MEU(\{\})$?
As calculated above, $MEU(\{\}) = EU(buy)$, so it is optimal to buy a lottery ticket.

Now, the organization running the lottery has announced a change in policy: there are two lotteries, and when someone buys a lottery ticket, the ticket is randomly for one of the two lotteries.

We can incorporate this into our model by adding a node T that indicates which ticket gets bought: the resulting decision network looks like this:



In addition, we have the following conditional probability tables, that we will be using for the remaining parts of this problem:

T	$P(T)$
$lottery1$	$1/2$
$lottery2$	$1/2$

T	L	$P(L\|T)$
$lottery1$	0	9/10
$lottery1$	14	1/10
$lottery2$	0	7/10
$lottery2$	5	3/10

Q2: What is the new MEU?

What action achieves $MEU(\{\})$?

Solution: Conceptually, it is perhaps simplest if we simply condense this nested lottery into a single lottery: $[0, 16/20; 5, 3/10; 13, 1/20]$. Now, we can use the same approach as in the first part:

EU(buy) = 4/5 * -4^3 + 3/20 * 1^3 + 1/20 * 10^3 = -1.05

Since $EU(buy) < EU(no\ buy)$, the optimal action is to not buy a lottery ticket.

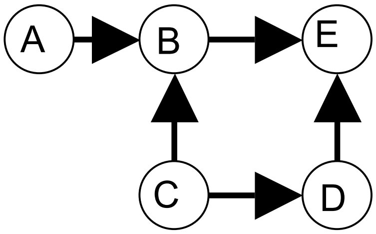
Q3: What is VPI(T)?

MEU(T = lottery1) = max(0, (9/10 – 4^3 + 1/10 * 10^3)) = max(0, 42.4) = 42.4

MEU(T = lottery2) = max(0, (7/10 – 4^3 + 3/10 * 1^3)) = max(0, -44.5) = 0

VPI(T) = ΣP(T = t) MEU(T = t) – MEU({}) = ½ * 42/4 + ½ * 0 - 0 = 21.2

3. Sampling



A	P(A)
0	0.6
1	0.4

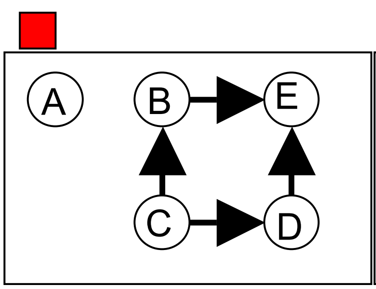
C	D	P(D|C)
0	0	0.2
0	1	0.8
1	0	0.3
1	1	0.7

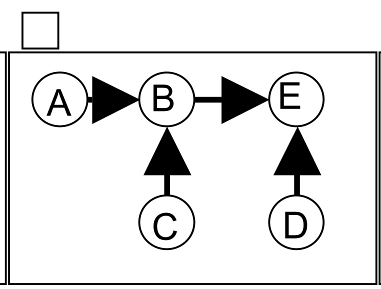
A	C	B	P(B|A,C)
0	0	0	0.2
0	0	1	0.8
0	1	0	0.4
0	1	1	0.6
1	0	0	0.2
1	0	1	0.8
1	1	0	0.4
1	1	1	0.6

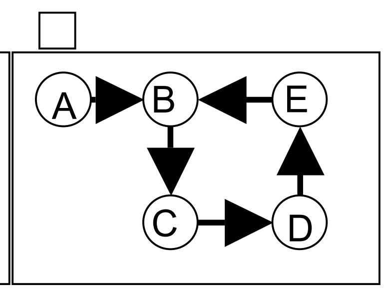
B	D	E	P(E|B,D)
0	0	0	0.75
0	0	1	0.25
0	1	0	0.5
0	1	1	0.5
1	0	0	0.85
1	0	1	0.15
1	1	0	0.4
1	1	1	0.6

C	P(C)
0	0.4
1	0.6

i. (3 pts) Check the boxes above the Bayes' nets below that <u>could also be valid</u> for the above probability tables.







Justification: The first Bayes' net implies B is independent of A given C, which is true: looking at P(B|A,C), we see that $P(b|a,c) \ = \ P(b|c)$for all a.
The second Bayes' net implies C and D are independent, which isn't true: for instance,
$P(D=1) \ = \ P(D=1|C=0)P(C=0) \ + \ P(D=1|C=1)P(C=1) \ = \ (0.8)(0.4) \ + \ (0.7)(0.6) \ = \ 0.74$
but$P(D=1|C=0) \ = \ 0.8$.
The third Bayes' net is cyclic, and is therefore not a valid Bayes' net.

ii. (2 pts) Caryn wants to compute the distribution P(A,C|E=1) using ***prior sampling*** on Model-Q1 (given at the top of this page). She draws a bunch of samples. The first of these is (0, 0, 1, 1, 0), given in (A, B, C, D, E) format. What's the probability of drawing this sample?
Solution: P(A=0, B=0, C=1, D=1, E=0)
= P(A=0) * P(C=1) * P(B=0 | A=0, C=1) * P(D=1 | C=1) * P(E=0 | B=0, D=1)
= 0.6 * 0.6 * 0.4 * 0.7 * 0.5
= 0.0504

iii. (2 pts) Give an example of an inference query for Model-Q1 with **one query variable and one evidence variable** that could be estimated more efficiently (in terms of runtime) using ***rejection sampling*** than by using ***prior sampling***. If none exist, state "not possible".
Example Solution: P(C|A=0)
Explanation: Rejection sampling provides an efficiency advantage when it allows us to realize that a sample is going to be unusable before it is fully generated. Thus, the query needs to be able to reject a sample (by comparing against the evidence variable) before all variables are sampled. The last variable sampled must be E (since it is the only sink (no outgoing edges) in the Bayes's net), so any solution where the evidence is not E (and the query is not the same as the evidence), then rejection sampling will be more efficient than prior sampling.

Common Mistakes:

- It was important to actually give evidence, e.g. P(A|B=1), and not P(A|B). The latter doesn't actually incorporate evidence for rejection sampling to use.
- While it is true that A and B are independent, Model-Q1 doesn't incorporate this information, and sampling methods applied to this model won't take advantage of this information either. (Key point: sampling methods are applied to a model, and not directly on the joint distribution.)
- Note that the domains of the variables (as visible in the CPTs) are {0, 1}. Specifically, they are not {+a, -a} or anything similar.

iv. (2 pts) Give an example of an inference query for Model-Q1 with **one query variable and one evidence variable** for which ***rejection sampling*** provides no efficiency advantage (in terms of runtime) over using ***prior sampling***. If none exist, state "not possible".
Example Solution: P(C|E=0)
Explanation: For the same logic as the previous part, if E is the evidence, then rejection and prior sampling will have the same runtime.

v. (2 pts) Now Caryn wants to determine P(A,C|E=1) for Model-Q1 using likelihood weighting. She draws the five samples shown below, which are given in (A, B, C, D, E) format, where the leftmost sample is "Sample 1" and the rightmost is "Sample 5". What are the weights of the samples S1 and S3?

weight(S1): $P(E=1 \mid B=0, D=1) = 0.5$ weight(S3):
$P(E=1 \mid B=0, D=1) = 0.5$

S1: (0, 0, 1, 1, 1)   S2: (0, 0, 1, 1, 1)   S3: (1, 0, 1, 1, 1)   S4: (0, 1, 0, 0, 1)   S5: (0, 1, 0, 0, 1)

vi. (1 pt) For the same samples as in part v, compute P(A=1,C=1|E=1) for Model-Q1. Express your answer as a simplified fraction (e.g. 2 / 3 instead of 4 / 6).

The weights are (left to right): 0.5, 0.5, 0.5, 0.15, 0.15. This can be found using just two table lookups: the weights of S1, S2, and S3 are all the same as calculated in the previous part, and the weights of S4 and S5 are $P(E = 1 \mid B = 1, D = 0) = 0.15$.
Only S3 matches the query A=1, C=1, so we infer P(A=1,C=1|E=1) to be the following:

weight(S3) / (sum of all weights) = 0.5 / (0.5 + 0.5 + 0.5 + 0.15 + 0.15) = 5 / 18

vii. (2 pts) Select True or False for each of the following:

True	False	
○	●	When there is no evidence, prior sampling is guaranteed to yield the exact same answer as inference by enumeration.
●	○	When collecting a sample during likelihood weighting, evidence variables are not sampled.
○	●	When collecting a sample during rejection sampling, variables can be sampled in any order.
●	○	Gibbs sampling is a technique for performing approximate inference, not exact inference.