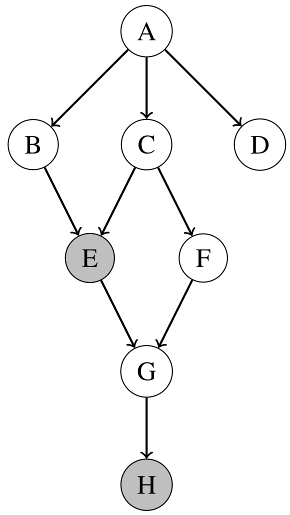
Q1. Bayes' Nets: Inference

Assume we are given the following Bayes' net, and would like to perform inference to obtain $P(B, D \mid E = e, H = h)$.



(a) What is the number of rows in the largest factor generated by *inference by enumeration*, for this query $P(B, D \mid E = e, H = h)$? Assume all the variables are binary.

○ 2^2 ○ 2^3 ● 2^6 ○ 2^8

○ None of the above.

Since the inference by enumeration first joins all the factors in the Bayes' net, that factor will contain six (unobserved) variables. The question assumes all variables are binary, so the answer is 2^6.

(b) Mark all of the following variable elimination orderings that are optimal for calculating the answer for the query $P(B, D \mid E = e, H = h)$. Optimality is measured by the sum of the sizes of the factors that are generated. Assume all the variables are binary.

☐ C, A, F, G ☐ F, G, C, A ☐ A, C, F, G ■ G, F, C, A

☐ None of the above.

The sum of the sizes of factors that are generated for the variable elimination ordering G, F, C, A is $2^1 + 2^1 + 2^2 + 2^2$ rows, which is smaller than for any of the other variable elimination orderings. The ordering F, G, C, A is close but the sum of the sizes of factors is slightly bigger, with $2^2 + 2^1 + 2^2 + 2^2$ rows.

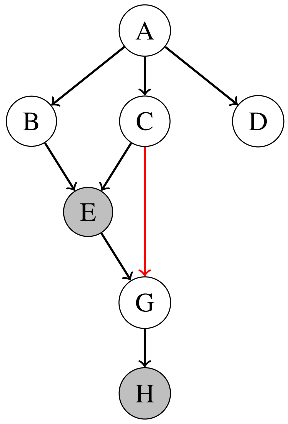
(c) Suppose we decide to perform variable elimination to calculate the query $P(B, D \mid E = e, H = h)$, and choose to eliminate F first.

(i) When F is eliminated, what intermediate factor is generated and how is it calculated? Make sure it is clear which variable(s) come before the conditioning bar and which variable(s) come after.

$$f_1(\underline{\quad G \mid C, e \quad}) = \sum_f \underline{\quad P(f \mid C)P(G|f, e) \quad}$$

This follows from the first step of variable elimination, which is to join all factors containing F, and then marginalize over F to obtain the intermediate factor f_1.

(ii) Now consider the set of distributions that can be represented by the remaining factors *after F is eliminated*. Draw the minimal number of directed edges on the following Bayes' Net structure, so that it can represent any distribution in this set. If no additional directed edges are needed, please fill in that option below.

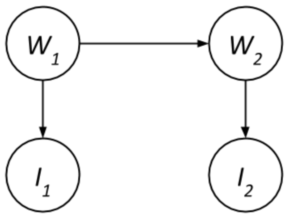


○ No additional directed edges needed

An additional edge from C to G is necessary, because the intermediate factor is of the form $f_1(G|C)$. Without this edge from C to G, the Bayes' net would not be able to express the dependence of G on C. (Note that adding an edge from G to C is not allowed, since that would introduce a cycle.)

2 Sampling and Dynamic Bayes Nets

We would like to analyze people's ice cream eating habits on sunny and rainy days Suppose we consider the weather, along with a person's ice-cream eating, over the span of two days. We'll have four random variables: W_1 and W_2 stand for the weather on days 1 and 2, which can either be rainy R or sunny S, and the variables I_1 and I_2 represent whether or not the person ate ice cream on days 1 and 2, and take values T (for truly eating ice cream) or F. We can model this as the following Bayes Net with these probabilities.



W_1	$P(W_1)$
S	0.6
R	0.4

W_1	W_2	$P(W_2\|W_1)$
S	S	0.7
S	R	0.3
R	S	0.5
R	R	0.5

W	I	$P(I\|W)$
S	T	0.9
S	F	0.1
R	T	0.2
R	F	0.8

Suppose we produce the following samples of (W_1, I_1, W_2, I_2) from the ice-cream model:

~~R, F, R, F~~ ~~R, F, R, F~~ ~~S, F, S, T~~ ~~S, T, S, T~~ S, T, R, F
~~R, F, R, T~~ ~~S, T, S, T~~ ~~S, T, S, T~~ S, T, R, F ~~R, F, S, T~~

1. What is $\widehat{P}(W_2 = \mathtt{R})$, the probability that sampling assigns to the event $W_2 = \mathtt{R}$?
   Number of samples in which $W_2 = \mathtt{R}$: 5. Total number of samples: 10. Answer $5/10 = 0.5$.

2. Cross off samples above which are rejected by rejection sampling if we're computing $P(W_2 | I_1 = \mathtt{T}, I_2 = \mathtt{F})$.

   Rejection sampling seems to be wasting a lot of effort, so we decide to switch to likelihood weighting. Assume we generate the following six samples given the evidence $I_1 = \mathtt{T}$ and $I_2 = \mathtt{F}$:

$$(W_1, I_1, W_2, I_2) = \left\{(\mathtt{S, T, R, F}), (\mathtt{R, T, R, F}), (\mathtt{S, T, R, F}), (\mathtt{S, T, S, F}), (\mathtt{S, T, S, F}), (\mathtt{R, T, S, F})\right\}$$

3. What is the weight of the first sample (S, T, R, F) above?

   The weight given to a sample in likelihood weighting is

$$\prod_{\text{Evidence variables } e} \Pr(e|\text{Parents}(e)).$$

   In this case, the evidence is $I_1 = \mathtt{T}, I_2 = \mathtt{F}$. The weight of the first sample is therefore

$$w = \Pr(I_1 = \mathtt{T} | W_1 = \mathtt{S}) \cdot \Pr(I_2 = \mathtt{F} | W_2 = \mathtt{R}) = 0.9 \cdot 0.8 = 0.72$$

4. Use likelihood weighting to estimate $P(W_2 | I_1 = \mathtt{T}, I_2 = \mathtt{F})$.
   The sample weights are given by

(W_1, I_1, W_2, I_2)	w	(W_1, I_1, W_2, I_2)	w
S, T, R, F	0.72	S, T, S, F	0.09
R, T, R, F	0.16	S, T, S, F	0.09
S, T, R, F	0.72	R, T, S, F	0.02

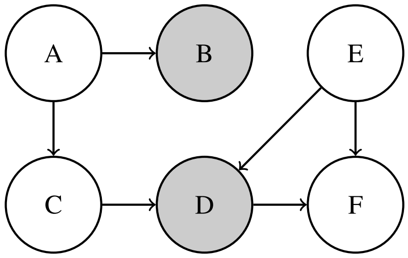
To compute the probabilities, we thus normalize the weights and find

$$\widehat{P}(W_2 = \texttt{R} | I_1 = \texttt{T}, I_2 = \texttt{F}) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02} = 0.889$$

$$\widehat{P}(W_2 = \texttt{S} | I_1 = \texttt{T}, I_2 = \texttt{F}) = 1 - 0.889 = 0.111.$$

Q3. Bayes Nets and Sampling

You are given a bayes net with the following probability tables:



E	D	F	$P(F\|E,D)$
0	0	0	0.6
0	0	1	0.4
0	1	0	0.7
0	1	1	0.3
1	0	0	0.2
1	0	1	0.8
1	1	0	0.7
1	1	1	0.3

A	$P(A)$
0	0.75
1	0.25

A	B	$P(B\|A)$
0	0	0.1
0	1	0.9
1	0	0.5
1	1	0.5

A	C	$P(C\|A)$
0	0	0.3
0	1	0.7
1	0	0.7
1	1	0.3

E	$P(E)$
0	0.1
1	0.9

E	C	D	$P(D\|E,C)$
0	0	0	0.5
0	0	1	0.5
0	1	0	0.2
0	1	1	0.8
1	0	0	0.5
1	0	1	0.5
1	1	0	0.2
1	1	1	0.8

You want to know $P(C = 0|B = 1, D = 0)$ and decide to use sampling to approximate it.

(a) With prior sampling, what would be the likelihood of obtaining the sample [A=1, B=0, C=0, D=0, E=1, F=0]?

- ○ 0.25*0.1*0.3*0.9*0.8*0.7
- ● 0.25*0.5*0.7*0.5*0.9*0.2
- ○ 0.75*0.1*0.3*0.9*0.5*0.8
- ○ 0.25*0.5*0.3*0.2*0.9*0.2
- ○ 0.25*0.9*0.7*0.1*0.5*0.6
- ○ 0.75*0.1*0.3*0.9*0.5*0.2 + 0.25*0.5*0.7*0.5*0.9*0.2

○ Other _______________ Prior sampling samples without taking the evidence into account, so the probability of the sample is P(A)P(B|A)P(C|A)P(D|C,E)P(E)P(F|E,D)

(b) Assume you obtained the sample [A = 1, B=1, C=0, D=0, E=1, F=1] through likelihood weighting. What is its weight?

- ○ 0.25*0.5*0.7*0.5*0.9*0.8
- ○ 0
- ○ 0.25*0.7*0.9*0.8 + 0.75*0.3*0.9*0.8
- ● 0.5*0.5
- ○ 0.25*0.5*0.7*0.5*0.8
- ○ 0.9*0.5 + 0.1*0.5

○ Other _______________ The weight of a sample in gibbs sampling is the probability of the evidence given their parents: P(D=0|E=1,C=0)*P(B=1|A=1)

(c) You decide to use Gibb's sampling instead. Starting with the initialization [A = 1, B=1, C=0, D=0, E=0, F=0], suppose you resample F first, what is the probability that the next sample drawn is [A = 1, B=1, C=0, D=0, E=0, F=1]?

- ● 0.4
- ○ 0.6
- ○ 0.6*0.1*0.5
- ○ 0
- ○ 0.25*0.5*0.7*0.5*0.1*0.3
- ○ 0.9*0.5 + 0.1*0.5

○ Other _______________ In Gibb's sampling, you resample individual vairables conditioned on the rest of the sample. The distribution of F given the rest of the sample is 0.4 for F=1 and 0.6 for F=0.