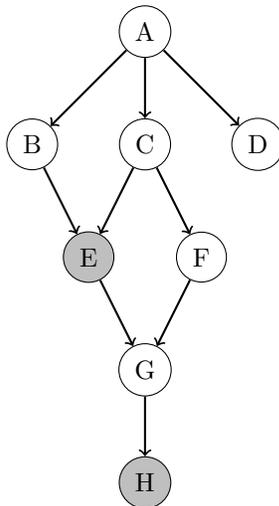


Q1. Bayes' Nets: Inference

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.

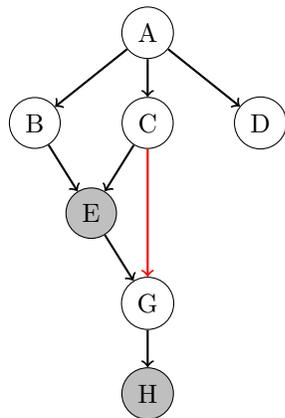
The sum of the sizes of factors that are generated for the variable elimination ordering G, F, C, A is $2^1 + 2^1 + 2^2 + 2^2$ rows, which is smaller than for any of the other variable elimination orderings. The ordering F, G, C, A is close but the sum of the sizes of factors is slightly bigger, with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

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

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.

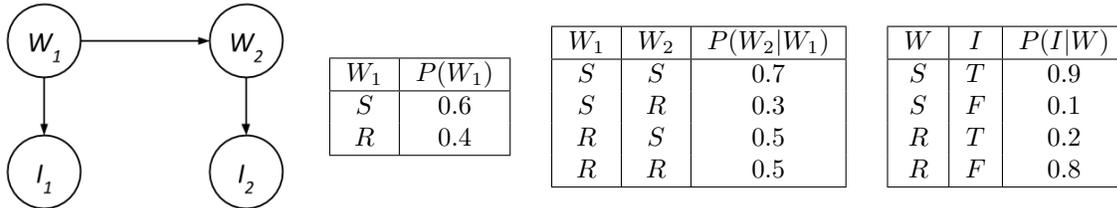


No additional directed edges needed

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.



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~~

- What is $\hat{P}(W_2 = R)$, the probability that sampling assigns to the event $W_2 = R$?
 Number of samples in which $W_2 = R$: 5. Total number of samples: 10. Answer $5/10 = 0.5$.
- Cross off samples above which are rejected by rejection sampling if we're computing $P(W_2|I_1 = T, I_2 = F)$.

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 $I_1 = T$ and $I_2 = F$:

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

- 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 $I_1 = T, I_2 = F$. The weight of the first sample is therefore

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

- Use likelihood weighting to estimate $P(W_2|I_1 = T, I_2 = F)$.

The sample weights are given by

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

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

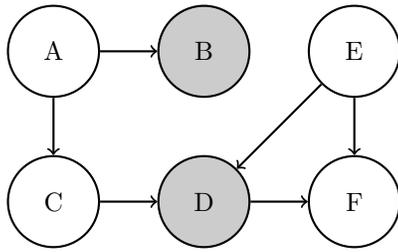
To compute the probabilities, we thus normalize the weights and find

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

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

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 $[A=1, B=0, C=0, D=0, E=1, F=0]$?

- $0.25 \cdot 0.1 \cdot 0.3 \cdot 0.9 \cdot 0.8 \cdot 0.7$
- $0.25 \cdot 0.5 \cdot 0.7 \cdot 0.5 \cdot 0.9 \cdot 0.2$
- $0.75 \cdot 0.1 \cdot 0.3 \cdot 0.9 \cdot 0.5 \cdot 0.8$
- $0.25 \cdot 0.9 \cdot 0.7 \cdot 0.1 \cdot 0.5 \cdot 0.6$
- $0.25 \cdot 0.5 \cdot 0.3 \cdot 0.2 \cdot 0.9 \cdot 0.2$
- $0.75 \cdot 0.1 \cdot 0.3 \cdot 0.9 \cdot 0.5 \cdot 0.2 + 0.25 \cdot 0.5 \cdot 0.7 \cdot 0.5 \cdot 0.9 \cdot 0.2$

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?

- $0.25 \cdot 0.5 \cdot 0.7 \cdot 0.5 \cdot 0.9 \cdot 0.8$
- $0.25 \cdot 0.7 \cdot 0.9 \cdot 0.8 + 0.75 \cdot 0.3 \cdot 0.9 \cdot 0.8$
- $0.25 \cdot 0.5 \cdot 0.7 \cdot 0.5 \cdot 0.8$
- 0
- $0.5 \cdot 0.5$
- $0.9 \cdot 0.5 + 0.1 \cdot 0.5$

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

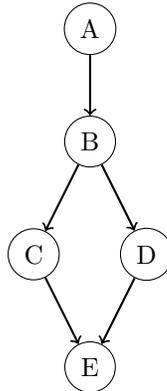
(c) You decide to use Gibb's sampling instead. Starting with the initialization $[A = 1, B=1, C=0, D=0, E=0, F=0]$, suppose you resample F first, what is the probability that the next sample drawn is $[A = 1, B=1, C=0, D=0, E=0, F=1]$?

- 0.4
- $0.6 \cdot 0.1 \cdot 0.5$
- $0.25 \cdot 0.5 \cdot 0.7 \cdot 0.5 \cdot 0.1 \cdot 0.3$
- 0.6
- 0
- $0.9 \cdot 0.5 + 0.1 \cdot 0.5$

Other _____ In Gibb's sampling, you resample individual variables 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 $F=0$.

Q4. Bayes' Nets: Sampling

Assume we are given the following Bayes' net, with the associated conditional probability tables (CPTs).



A	P(A)
+a	0.5
-a	0.5

A	B	P(B A)
+a	+b	0.2
+a	-b	0.8
-a	+b	0.5
-a	-b	0.5

B	C	P(C B)
+b	+c	0.4
+b	-c	0.6
-b	+c	0.8
-b	-c	0.2

B	D	P(D B)
+b	+d	0.2
+b	-d	0.8
-b	+d	0.2
-b	-d	0.8

C	D	E	P(E C, D)
+c	+d	+e	0.6
+c	+d	-e	0.4
+c	-d	+e	0.2
+c	-d	-e	0.8
-c	+d	+e	0.4
-c	+d	-e	0.6
-c	-d	+e	0.8
-c	-d	-e	0.2

You are given a set of the following samples, but are not told whether they were collected with rejection sampling or likelihood weighting.

-a -b +c +d +e
 -a +b +c -d +e
 -a -b -c -d +e
 -a -b +c -d +e
 -a +b +c +d +e

Throughout this problem, you may answer as either numeric expressions (e.g. $0.1 \cdot 0.5$) or numeric values (e.g. 0.05).

- (a) Assuming these samples were generated from *rejection sampling*, what is the sample based estimate of $P(+b \mid -a, +e)$?

Answer: 0.4

The answer is the number of samples satisfying the query variable's assignment (in this case, $B = +b$) divided by the total number of samples, so the answer is $2 / 5 = 0.4$.

- (b) Assuming these samples were generated from *likelihood weighting*, what is the sample-based estimate of $P(+b \mid -a, +e)$?

Answer: $\frac{1}{3}$

Based on likelihood weighting, we know the weight of each sample is $P(A = a) * P(E = e | C = c, D = d)$. The weights are: 0.3 (= 0.5 * 0.6), 0.1 (= 0.5 * 0.2), 0.4 (= 0.5 * 0.8), 0.1 (same assignments to C and D as second sample), 0.3 (same assignments to C and D as first sample). The estimate is then $(0.1 + 0.3) / (0.3 + 0.1 + 0.4 + 0.1 + 0.3) = 0.4 / 1.20 = 1/3 = 0.333$.

- (c) Again, assume these samples were generated from *likelihood weighting*. However, you are not sure about the original CPT for $P(E | C, D)$ given above being the CPT associated with the Bayes' Net: With 50% chance, the CPT associated with the Bayes' Net is the original one. With the other 50% chance, the CPT is actually the CPT below.

C	D	E	P(E C, D)
+c	+d	+e	0.8
+c	+d	-e	0.2
+c	-d	+e	0.4
+c	-d	-e	0.6
-c	+d	+e	0.2
-c	+d	-e	0.8
-c	-d	+e	0.6
-c	-d	-e	0.4

Samples from previous page copied below for convenience:

-a -b +c +d +e
 -a +b +c -d +e
 -a -b -c -d +e
 -a -b +c -d +e
 -a +b +c +d +e

Given this uncertainty, what is the sample-based estimate of $P(+b | -a, +e)$?

Answer: $\frac{10}{27}$

The weight of each sample is $P(A = a) * (0.5 * P_1(E = e | C = c, D = d) + 0.5 * P_2(E = e | C = c, D = d))$. The new weights are 0.35 (= 0.5 * (0.5 * 0.6 + 0.5 * 0.8)), 0.15 (= 0.5 * (0.5 * 0.2 + 0.5 * 0.4)), 0.35 (= 0.5 * (0.5 * 0.8 + 0.5 * 0.6)), 0.15, and 0.35. The estimate is then $(0.15 + 0.35) / (0.35 * 3 + 0.15 * 2) = 0.5 / 1.35 = 10 / 27$

- (d) Now assume you can only sample a *small, limited number of samples*, and you want to estimate $P(+b, +d | -a)$ and $P(+b, +d | +e)$. You are allowed to estimate the answer to one query with likelihood weighting, and the other answer with rejection sampling. In order to obtain the best estimates for both queries, *which query should you estimate with likelihood weighting?* (The other query will have to be estimated with rejection sampling.)

- $P(+b, +d | -a)$
 $P(+b, +d | +e)$
 Either – both choices allow you to obtain the best estimates for both queries.

The evidence +e is at the leaves of the Bayes' net, which means it's possible to sample all the other variables, but have to reject the last node E. We can avoid this problem by using likelihood weighting for sampling, since it fixes the values of observed random variables to that of the fixed evidence.

- (e) Suppose you choose to use Gibbs sampling to estimate $P(B, E | +c, -d)$. Assume the CPTs are the same as the ones for parts (a) and (b). Currently your assignments are the following:

-a -b +c -d +e

- (i) Suppose the next step is to resample E.
 What is the probability that the new assignment to E will be +e?

Answer: 0.2 In order to sample E , we need to calculate $P(E | -a, -b, +c, -d)$, which is equal to $P(E | +c, -d)$ since E is conditionally independent of A and B , given C and D . The value

for $P(+e \mid +c, -d)$ is given directly in the CPT for $P(E|C, D)$, and it is 0.2.

(ii) Instead, suppose the next step is to resample A .

What is the probability that the new assignment to A will be $+a$?

Answer: $\frac{8}{13}$ In order to sample A , we need to calculate $P(A \mid -b, +c, -d, +e)$, which is equal to $P(A \mid B)$ since A is conditionally independent of C, D , and E , given B . We can calculate this using Bayes' rule: $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$. Thus, $P(+a \mid -b) = \frac{P(-b|+a)P(+a)}{P(-b)} = \frac{P(-b|+a)P(+a)}{\sum_{a \in \{+a, -a\}} P(-b|a)P(a)} = \frac{0.8*0.5}{0.8*0.5+0.5*0.5} = \frac{0.4}{0.65} = \frac{8}{13}$

(iii) Instead, suppose the next step is to resample B .

What is the probability that the new assignment to B will be $+b$?

Answer: $\frac{1}{3}$ In order to sample B , we need to calculate $P(B \mid -a, +c, -d, +e)$, which is equal to $P(B \mid -a, +c, -d)$ since B is conditionally independent of E , given C and D . The CPT tables that are involved in calculating $P(B \mid -a, +c, -d, +e)$ are $P(A)$, $P(B|A)$, $P(C|B)$, $P(D|B)$. First, we remove rows of the CPTs that do not agree with the evidence $-a, +c, -d, +e$. We then join the resulting CPTs to obtain $P(-a, B, +c, -d)$. We select $P(-a, +b, +c, -d)$ and $P(-a, -b, +c, -d)$ from this table, and normalize so that the two probabilities sum to one (i.e., to transform them to $P(+b \mid -a, +c, -d)$ and $P(-b \mid -a, +c, -d)$).