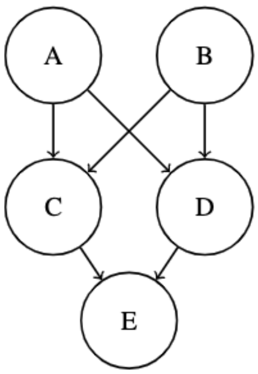
CS 188 Spring 2021 | Introduction to Artificial Intelligence | Exam Prep 6 Solutions

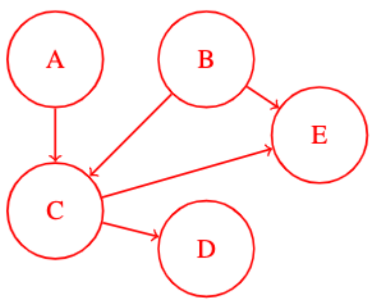
Q1. Bayes Nets and Joint Distributions

(a) Write down the joint probability distribution associated with the following Bayes Net. Express the answer as a product of terms representing individual conditional probabilities tables associated with this Bayes Net:



$P(A)P(B)P(C|A,B)P(D|A,B)P(E|C,D)$

(b) Draw the Bayes net associated with the following joint distribution:
$P(A) \cdot P(B) \cdot P(C|A,B) \cdot P(D|C) \cdot P(E|B,C)$



(c) Do the following products of factors correspond to a valid joint distribution over the variables A, B, C, D? (Circle FALSE or TRUE.)

(i) **[FALSE]** TRUE $P(A) \cdot P(B) \cdot P(C|A) \cdot P(C|B) \cdot P(D|C)$

(ii) FALSE **[TRUE]** $P(A) \cdot P(B|A) \cdot P(C) \cdot P(D|B,C)$

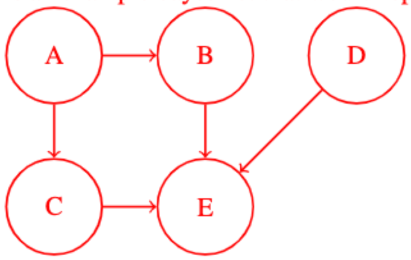
(iii) **[FALSE]** TRUE $P(A) \cdot P(B|A) \cdot P(C) \cdot P(C|A) \cdot P(D)$

(iv) **[FALSE]** TRUE $P(A|B) \cdot P(B|C) \cdot P(C|D) \cdot P(D|A)$

(d) What factor can be multiplied with the following factors to form a valid joint distribution? (Write "none" if the given set of factors can't be turned into a joint by the inclusion of exactly one more factor.)

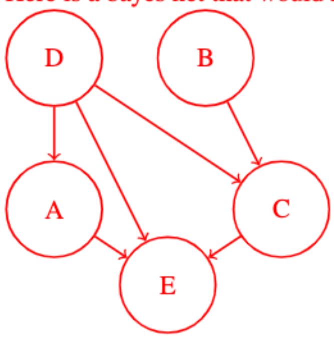
(i) $P(A) \cdot P(B|A) \cdot P(C|A) \cdot P(E|B,C,D)$

P(D) is missing. D could also be conditioned on A,B, and/or C without creating a cycle (e.g. $P(D|A,B,C)$). Here is an example bayes net that would represent the distribution after adding in $P(D)$:

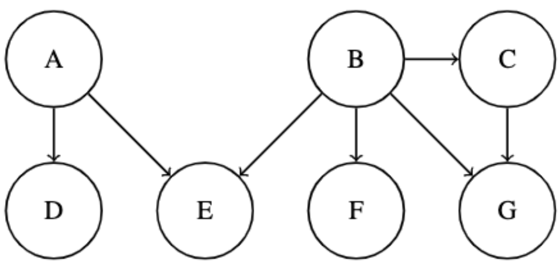


(ii) $P(D) \cdot P(B) \cdot P(C|D,B) \cdot P(E|C,D,A)$

P(A) is missing to form a valid joint distributions. A could also be conditioned on B, C, and/or D (e.g. $P(A|B,C,D)$. Here is a bayes net that would represent the distribution is $P(A|D)$ was added in.



(e) Answer the next questions based off of the Bayes Net below:
All variables have domains of {-1, 0, 1}



(i) Before eliminating any variables or including any evidence, how many entries does the factor at G have?

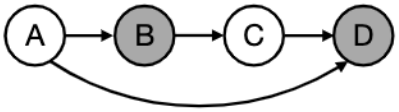
The factor is $P(G|B,C)$, so that gives $3^3 = 27$ entries.

(ii) Now we observe $e = 1$ and want to query $P(D|e = 1)$, and you get to pick the first variable to be eliminated.

- Which choice would create the **largest** factor f_1?
  Eliminating B first would give the largest f_1:, $f_1(A,F,G,C,e) = \sum_{B=b} P(b)P(e|A,b)P(F|b)P(G|b,C)P(C|b)$. This factor has 3^4 entries.
- Which choice would create the **smallest** factor f_1?
  eliminating F first would give smallest factors of 3 entries: $f_1(B) = \sum_f P(f|B)$. Eliminating D is not correct because D is the query variable.

Q2. Bayes Nets: Sampling

Consider the following Bayes Net, where we have observed that $B = +b$ and $D = +d$.



$P(A)$	
$+a$	0.5
$-a$	0.5

$P(B\|A)$		
$+a$	$+b$	0.8
$+a$	$-b$	0.2
$-a$	$+b$	0.4
$-a$	$-b$	0.6

$P(C\|B)$		
$+b$	$+c$	0.1
$+b$	$-c$	0.9
$-b$	$+c$	0.7
$-b$	$-c$	0.3

$P(D\|A,C)$			
$+a$	$+c$	$+d$	0.6
$+a$	$+c$	$-d$	0.4
$+a$	$-c$	$+d$	0.1
$+a$	$-c$	$-d$	0.9
$-a$	$+c$	$+d$	0.2
$-a$	$+c$	$-d$	0.8
$-a$	$-c$	$+d$	0.5
$-a$	$-c$	$-d$	0.5

(a) Consider doing Gibbs sampling for this example. Assume that we have initialized all variables to the values $+a, +b, +c, +d$. We then unassign the variable C, such that we have $A = +a$, $B = +b$, $C = ?$, $D = +d$. Calculate the probabilities for new values of C at this stage of the Gibbs sampling procedure.

$P(C = +c$ at the next step of Gibbs sampling$) =$ $\frac{0.1 \cdot 0.6}{0.1 \cdot 0.6 + 0.9 \cdot 0.1} = \boxed{\frac{2}{5}}$

$P(C = -c$ at the next step of Gibbs sampling$) =$ $\frac{0.9 \cdot 0.1}{0.1 \cdot 0.6 + 0.9 \cdot 0.1} = \boxed{\frac{3}{5}}$

(b) Consider a sampling scheme that is a hybrid of rejection sampling and likelihood-weighted sampling. Under this scheme, we first perform rejection sampling for the variables A and B. We then take the sampled values for A and B and extend the sample to include values for variables C and D, using likelihood-weighted sampling.

(i) Below is a list of candidate samples. Mark the samples that would be rejected by the rejection sampling portion of the hybrid scheme.

- ■ $-a$ $-b$
- ☐ $+a$ $+b$
- ■ $+a$ $-b$
- ☐ $-a$ $+b$

(ii) To decouple from part (i), you now receive a *new* set of samples shown below. Fill in the weights for these samples under our hybrid scheme.

				Weight
$-a$	$+b$	$-c$	$+d$	0.5
$+a$	$+b$	$-c$	$+d$	0.1
$+a$	$+b$	$-c$	$+d$	0.1
$-a$	$+b$	$+c$	$+d$	0.2
$+a$	$+b$	$+c$	$+d$	0.6

(iii) Use the weighted samples from part (ii) to calculate an estimate for $P(+a \mid +b, +d)$.

The estimate of $P(+a \mid +b, +d)$ is $\frac{0.1 + 0.1 + 0.6}{0.5 + 0.1 + 0.1 + 0.2 + 0.6} = \boxed{\frac{8}{15}}$

(c) We now attempt to design an alternative hybrid sampling scheme that combines elements of likelihood-weighted and rejection sampling. For each proposed scheme, indicate whether it is valid, i.e. whether the weighted samples it produces correctly approximate the distribution $P(A, C \mid +b, +d)$.

(i) *First collect a likelihood-weighted sample for the variables A and B. Then switch to rejection sampling for the variables C and D. In case of rejection, the values of A and B and the sample weight are* ***thrown away****. Sampling then restarts from node* ***A****.*

● Valid ○ Invalid

(ii) *First collect a likelihood-weighted sample for the variables A and B. Then switch to rejection sampling for the variables C and D. In case of rejection, the values of A and B and the sample weight are* ***retained****. Sampling then restarts from node* ***C****.*

○ Valid ● Invalid

The sampling procedure in part (i) is the correct way of combining likelihood-weighted and rejection sampling: any time a node gets rejected, the sample must be thrown out in its entirety. In part (ii), however, the evidence that $D = +d$ has no effect on which values of A are sampled or on the sample weights. This means that values for A would be sampled according to $P(A \mid +b)$, not $P(A \mid +b, +d)$.

As an extreme case, suppose node D had a different probability table where $P(+d \mid +a) = 0$. Following the procedure from part (ii), we might start by sampling $(+a, +b)$ and assigning a weight according to $P(+b \mid +a)$. However, when we move on to rejection sampling we will be forced to continuously reject all possible values because our evidence $+d$ is inconsistent with our the assignment of $A = +a$. This means that the procedure from part (ii) is flawed to the extent that it might fail to generate a sample altogether!