ROB311 Quiz 2

Hanhee Lee

March 14, 2025

Contents

Probabilistic Inference Problems

1 Probabilistic Inference

1.1 Problem Setup

Definition: Given a Bayesian network, $\mathcal{B} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, \dots, X_{|\mathcal{V}|}\}$, we want to find the value of:

$$\text{pr}(\mathbf{Q} \mid \mathbf{E}) := \text{pr}(Q_1, \dots, Q_{|\mathbf{Q}|} \mid E_1, \dots, E_{|\mathbf{E}|}) = \frac{\sum_{\mathcal{V} \setminus (\mathbf{Q} \cup \mathbf{E})} p(X_1, \dots, X_{|\mathcal{V}|})}{\sum_{\mathcal{V} \setminus \mathbf{E}} p(X_1, \dots, X_{|\mathcal{V}|})}$$

$$\text{pr}(\mathbf{Q} \mid \mathbf{E}) \propto \sum_{\mathcal{V} \setminus (\mathbf{Q} \cup \mathbf{E})} \left(p(X_1) \prod_{i \neq 1} p(X_i \mid \text{pts}(X_i)) \right)$$

- $\mathbf{Q} = \{Q_1, \dots, Q_{|\mathbf{Q}|}\}$: Query variables
- $\mathbf{E} = \{E_1, \dots, E_{|\mathbf{E}|}\} \subseteq \mathcal{V}$: Evidence variables
- $\mathbf{Q} \cap \mathbf{E} = \emptyset$.

Warning:
- Denominator: Normalization constant (assuming $\mathbf{E}$ is fixed)
- Therefore, only need to compute numerator (w/o specifying $\mathbf{Q}$), which we can then normalize w.r.t. $\mathbf{Q}$

1.1.1 Joint Distribution in a Bayesian Network

Derivation: For any joint distribution, the following factorization holds:

$$p(X_1, \dots, X_{|p|}) = p(X_1) \prod_{i \neq 1} p(X_i \mid X_1, \dots, X_{i-1})$$

Bayesian Network Conditions: If
- at least 1 variable will be an orphan (i.e. no parents)
- no variable is both ancestor and descendant of another.

then this allows us to order $X_1, \dots, X_{|\mathcal{V}|}$, so that if X_j is a descendent of X_i, then for any $j > i$,

$$\text{pts}(X_i) \subseteq \{X_1, \dots, X_{i-1}\} \text{ and } X_1, \dots, X_{i-1} \notin \text{des}(X_i)$$

Therefore, using the consequence of dependence separation, then

$$p(X_1, \dots, X_{|\mathcal{V}|}) = p(X_1) \prod_{i \neq 1} p(X_i \mid \text{pts}(X_i))$$

1.2 Method 1: Bayesian Network Inference

1.2.1 Markov Blanket

Definition: The **Markov blanket** of a variable X, denoted $\text{mbk}(X)$, consists of the following variables:
- X's children
- X's parents
- The other parents of X's children, excluding X itself.

which is when a variable, X, is "eliminated", the resulting factor's scope is the Markov blanket of X.

1.2.2 Graphical Interpretation

Notes: Pictorially, eliminating X is equivalent to replacing all hyper-edges that include X with their union minus X, and then removing X.

1.2.3 Elimination Ordering

Definition: The order that the variables are eliminated.

- This creates a sequence of hyper-graphs that depend on the elimination ordering.

1.2.4 Elimination Width

Definition: The **elimination width** of a sequence of hyper-graphs is the # of variables in the hyper-edge within the sequence with the most variables.

1.2.5 Heuristics for Elimination Ordering

Definition: Choose the elimination ordering to minimize the elimination width using the following heuristics:

1. Eliminate variable with the fewest parents.
2. Eliminate variable with the smallest domain for its parents, where

$$|\operatorname{dom}(\operatorname{pts}(X))| = \prod_{Z \in \operatorname{pnt}(X)} |\operatorname{dom}(Z)|.$$

3. Eliminate variable with the smallest Markov blanket.
4. Eliminate variable with the smallest domain for its Markov blanket, where

$$|\operatorname{dom}(\operatorname{mbk}(X))| = \prod_{Z \in \operatorname{embk}(X)} |\operatorname{dom}(Z)|.$$

Warning: Choosing the variable with the smallest domain for its Markov blanket is the most effective heuristic.

1.3 Method 2: Inference via Sampling

Definition: Generate a large # of samples and then approximate as:

$$p(\mathbf{Q} \mid \mathbf{E}) \approx \frac{\text{\# of samples w/ } \mathbf{Q} \text{ and } \mathbf{E}}{\text{\# of samples w/ } \mathbf{E}}.$$

- As # of samples $\to \infty$, the approximation becomes exact.

1.3.1 Inference via Sampling with Likelihood Weighting

Motivation: Most of the samples are wasted since they are not consistent with the evidence.

Definition: Generate a large # of samples and then approximate as:

$$p(\mathbf{Q} \mid \mathbf{E}) \approx \frac{\text{weight of samples w/ } \mathbf{Q} \text{ and } \mathbf{E}}{\text{weight of samples w/ } \mathbf{E}}.$$

- Weight for each sample: Probability of forcing the evidence, i.e. probability of the evidence given the sample.

1.4 Canonical Problems:

1.4.1 Bayesian Inference via Variable Elimination

Process:

1. Given Bayesian network w/ variables and their conditional probabilities.
2. Find the probability of the query variable given the evidence variable, $p(\mathbf{Q} \mid \mathbf{E})$.
3. Use $p(\mathbf{Q} \mid \mathbf{E}) = \dfrac{\sum_{\mathcal{V} \setminus (\mathbf{Q} \cup \mathbf{E})} p(X_1, \ldots, X_{|\mathcal{V}|})}{\sum_{\mathcal{V} \setminus \mathbf{E}} p(X_1, \ldots, X_{|\mathcal{V}|})}$.
4. Determine $p(X_1) \prod_{i \neq 1} p(X_i \mid \text{pts}(X_i))$ using the Bayesian network.
5. Write out the summation of the numerator in an order using heuristics to determine elimination ordering.
6. Start with inner summation and work outwards.
7. Calculate the probability of the query variable(s) given the evidence variable(s).

Warning:

- Write the complement probability to make life easier. (HIGHLY RECOMMENDED)
- To determine the conditional probability summation of a variable, look at its parents (inward arrows)
- Inner sum must have all probabilities with that variable in it that you are summing over.
- TO determine $g_?$, look at which variable you aren't summing over and also aren't query/evidence variables, then it will be a fn of the remaining variables.

Example:

1. **Given:**

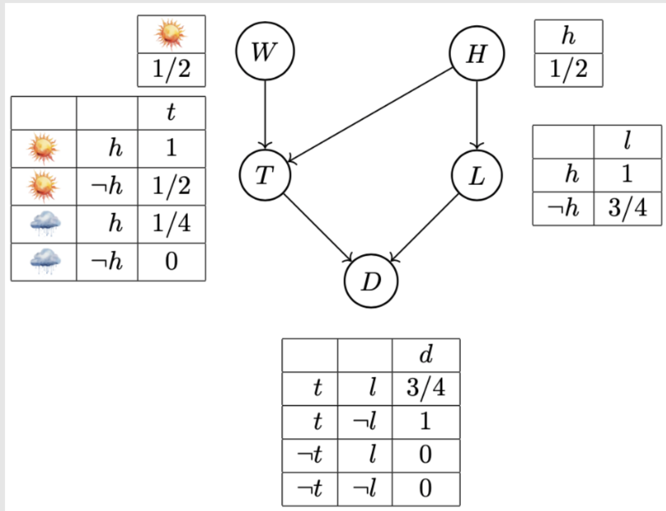


Figure 1

Variables	**Values**
W	$P(\text{Sunny}) = 0.5 \mid P(\text{Rainy}) = 0.5$
H	$P(h) = 0.5 \mid P(\neg h) = 0.5$
T	$P(t \mid \text{Sunny}, h) = 1 \mid P(t \mid \text{Sunny}, \neg h) = 0.5 \mid P(t \mid \text{Rainy}, h) = 0.25 \mid P(t \mid \text{Rainy}, \neg h) = 0$ $P(\neg t \mid \text{Sunny}, h) = 0 \mid P(\neg t \mid \text{Sunny}, \neg h) = 0.5 \mid P(\neg t \mid \text{Rainy}, h) = 0.75 \mid P(\neg t \mid \text{Rainy}, \neg h) = 1$
L	$P(l \mid h) = 1 \mid P(l \mid \neg h) = 0.75$ $P(\neg l \mid h) = 0 \mid P(\neg l \mid \neg h) = 0.25$
D	$P(d \mid t, l) = 0.75 \mid P(d \mid t, \neg l) = 1 \mid P(d \mid \neg t, l) = 0 \mid P(d \mid \neg t, \neg l) = 0$ $P(\neg d \mid t, l) = 0.25 \mid P(\neg d \mid t, \neg l) = 0 \mid P(\neg d \mid \neg t, l) = 1 \mid P(\neg d \mid \neg t, \neg l) = 1$

2. **Problem:** $p(d \mid h)$?
3. **Soln:**
   (a) $p(d \mid h) = \dfrac{p(d,h)}{p(h)} = \dfrac{\sum_{W,T,L} p(W,h,T,L,d)}{\sum_{W,T,L,D} p(W,h,T,L,d)}$ by definition of query and evidence equations.
   (b) $p(W,h,T,L,D) = p(h)p(W)p(L \mid h)p(t \mid W,h)p(D \mid T,L)$ by Bayesian network and $p(X_1,\dots,X_{|\mathcal{V}|}) = p(X_1)\prod_{i\neq 1} p(X_i \mid \text{pts}(X_i))$.

Summation

$$\text{Numerator}: p(h) \underbrace{\sum_L p(L \mid h) \underbrace{\sum_T p(D \mid T, L) \underbrace{\sum_W p(W) p(T \mid W, h)}_{g_1(T)}}_{g_2(L,D)}}_{g_3(D)}$$

$$g_1(T) = p(\text{Sunny}) p(T \mid \text{Sunny}, h) + p(\text{Rainy}) p(T \mid \text{Rainy}, h)$$

$g_1(t) = p(\text{Sunny}) p(t \mid \text{Sunny}, h) + p(\text{Rainy}) p(t \mid \text{Rainy}, h) = 0.5 \cdot 1 + 0.5 \cdot 0.25 = 0.625$
$g_1(\neg t) = p(\text{Sunny}) p(\neg t \mid \text{Sunny}, h) + p(\text{Rainy}) p(\not t \mid \text{Rainy}, h) = 0.5 \cdot 0 + 0.5 \cdot 0.75 = 0.375$

$$g_2(L, D) = p(D \mid t, L) g_1(t) + p(D \mid \neg t, L) g_1(\neg t)$$

$g_2(l, d) = p(d \mid t, l) g_1(t) + p(d \mid \neg t, l) g_1(\neg t) = 0.75 \cdot 0.625 + 0 \cdot 0.375 = 0.46875$
$g_2(l, \neg d) = p(\neg d \mid t, l) g_1(t) + p(\neg d \mid \neg t, l) g_1(\neg t) = 0.25 \cdot 0.625 + 1 \cdot 0.375 = 0.53125$
$g_2(\neg l, d) = p(d \mid t, \neg l) g_1(t) + p(d \mid \neg t, \neg l) g_1(\neg t) = 1 \cdot 0.625 + 0 \cdot 0.375 = 0.625$
$g_2(\neg l, \neg d) = p(\neg d \mid t, \neg l) g_1(t) + p(\neg d \mid \neg t, \neg l) g_1(\neg t) = 0 \cdot 0.625 + 1 \cdot 0.375 = 0.375$

$$g_3(D) = p(h) p(l \mid h) g_2(l, D) + p(h) p(\neg l \mid h) g_2(\neg l, D)$$

$g_3(d) = p(h) p(l \mid h) g_2(l, d) + p(h) p(\neg l \mid h) g_2(\neg l, d) = (0.5)(1)(0.46875) + (0.5)(0)(0.625) = 0.234375$
$g_3(\neg d) = p(h) p(l \mid h) g_2(l, \neg d) + p(h) p(\neg l \mid h) g_2(\neg l, \neg d) = (0.5)(1)(0.53125) + (0.5)(0)(0.375) = 0.265625$

$$p(d \mid h) = \frac{g_3(d)}{g_3(d) + g_3(\neg d)} = \frac{0.234375}{0.234375 + 0.265625} = \frac{0.234375}{0.5} = 0.46875$$

Example:

Summation

$$\text{Numerator} : \underbrace{p(h) \sum_L p(L \mid h) \underbrace{\sum_W p(W) \underbrace{\sum_T p(T \mid W, h) p(D \mid T, L)}_{g_1(W,D,L)}}_{g_2(D,L)}}_{g_3(D)}$$

$$\text{Numerator} : \underbrace{p(h) \sum_W p(W) \underbrace{\sum_T p(T \mid W, h) \underbrace{\sum_L p(L \mid h) p(D \mid T, L)}_{g_1(D,T)}}_{g_2(D,W)}}_{g_3(D)}$$

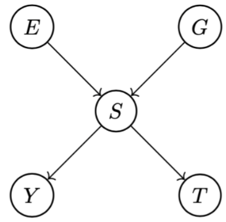
$$\text{Numerator} : \underbrace{p(h) \sum_W p(W) \underbrace{\sum_L p(L \mid h) \underbrace{\sum_T p(T \mid W, h) p(D \mid T, L)}_{g_1(W,D,L)}}_{g_2(W,D)}}_{g_3(D)}$$

$$\text{Numerator} : \underbrace{p(h) \sum_T p(T \mid W, h) \underbrace{\sum_W p(W) \underbrace{\sum_L p(L \mid h) p(D \mid T, L)}_{g_1(D,T)}}_{g_2(D,T)}}_{g_3(D)}$$

$\ldots$

Example:

1. **Given:**



G	E	pr$[S=1 \mid G,E]$
1	1	0.7
1	0	0.5
0	1	0.4
0	0	0.1

S	pr$[T=1 \mid S]$
1	0.9
0	0.2

S	pr$[Y=1 \mid S]$
1	0.8
0	0.1

Figure 2

2. **Problem:** Compute $\Pr(s=1 \mid t=1)$ if $\Pr(G=1)=0.3$, $\Pr(E=1)=0.4$, and the conditional probability tables for S, Y, and T are given below.
3. **Solution:**
   (a) Derivation of $p(s=1 \mid t=1)$:

$$\begin{aligned} p(s=1 \mid t=1) &= \frac{p(s=1,t=1)}{p(t=1)} \\ &= \frac{P(s=1,t=1)}{\sum_S P(S,t=1)} \\ &= \frac{\sum_{E,G,Y} P(E,G,Y,s=1,t=1)}{\sum_S \sum_{E,G,Y} P(E,G,Y,S,t=1)} \end{aligned}$$

   (b) Summation Term:

$$\sum_{E,G,Y} p(E)p(G)p(S \mid G,E)p(Y \mid S)p(t=1 \mid S)$$

$$\underbrace{p(t=1 \mid S)\sum_E p(E) \underbrace{\sum_G p(G)p(S \mid G,E) \underbrace{\sum_Y p(Y \mid S)}_{g_1(S)}}_{g_2(E,S)}}_{g_3(S)} \quad \text{one possible ordering}$$

   - Conditional probability and individual probabilities come from Bayesian network, and set $t, s = 1$ due to the query and evidence variables.

   (c) Choose:

$$\underbrace{p(t=1 \mid S)\sum_E p(E) \underbrace{\sum_G p(G)p(S \mid G,E) \underbrace{\sum_Y p(Y \mid S)}_{g_1(S)}}_{g_2(E,S)}}_{g_3(S)}$$

   (d) $g_1(S)$:

$$\begin{aligned} g_1(S) &= p(Y=1 \mid S) + p(Y=0 \mid S) \\ &= \begin{cases} 0.1+0.9=1 & \text{if } S=0 \\ 0.8+0.2=1 & \text{if } S=1 \end{cases} \end{aligned}$$

(e) $g_2(E, S)$:

$$\begin{aligned}
g_2(E, S) &= p(G=1)p(S \mid G=1, E)g_1(S) + p(G=0)p(S \mid G=0, E)g_1(S) \\
g_2(E, S) &= p(G=1)p(S \mid G=1, E) + p(G=0)p(S \mid G=0, E) \\
&= \begin{cases} 0.3(0.5) + 0.7(0.9) & \text{if } E=0, S=0 \\ 0.3(0.5) + 0.7(0.1) & \text{if } E=0, S=1 \\ 0.3(0.3) + 0.7(0.6) & \text{if } E=1, S=0 \\ 0.3(0.7) + 0.7(0.4) & \text{if } E=1, S=1 \end{cases} \\
&= \begin{cases} 0.78 & \text{if } E=0, S=0 \\ 0.22 & \text{if } E=0, S=1 \\ 0.51 & \text{if } E=1, S=0 \\ 0.49 & \text{if } E=1, S=1 \end{cases}
\end{aligned}$$

(f) $g_3(S)$:

$$\begin{aligned}
g_3(S) &= p(t=1 \mid S)p(E=1)g_2(E=1, S) + p(t=1 \mid S)p(E=0)g_2(E=0, S) \\
&= \begin{cases} 0.2(0.4)(0.51) + 0.2(0.6)(0.78) & \text{if } S=0 \\ 0.2(0.4)(0.49) + 0.2(0.6)(0.22) & \text{if } S=1 \end{cases} \\
&= \begin{cases} 0.1344 & \text{if } S=0 \\ 0.2952 & \text{if } S=1 \end{cases}
\end{aligned}$$

(g) $p(s=1 \mid t=1)$:

$$\begin{aligned}
p(s=1 \mid t=1) &= \frac{g_3(1)}{g_3(0) + g_3(1)} \\
&= \frac{0.2952}{0.2952 + 0.1344} \\
&= 0.6875
\end{aligned}$$

Example:

1. **Given:** Consider the following Bayesian network, where A, B, C, D are binary R.V. over $\{0, 1\}$

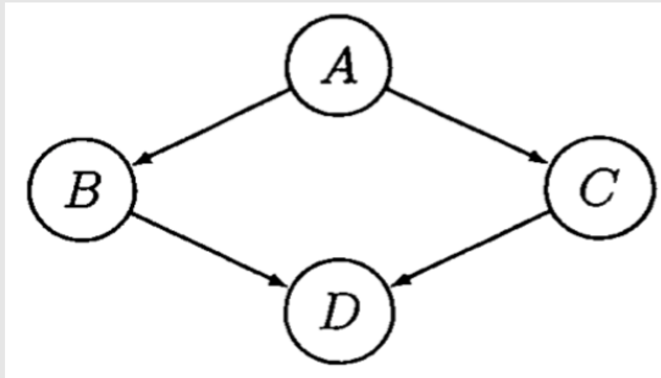


Figure 3

$P(A=1)$	$P(A=0)$
0.1	0.9

A	$P(B=1\|A)$	$P(B=0\|A)$
0	0.4	0.6
1	0.5	0.5

A	$P(C=1\|A)$	$P(C=0\|A)$
0	0.2	0.8
1	0.4	0.6

B	C	$P(D=1\|B,C)$	$P(D=0\|B,C)$
0	0	0.5	0.5
0	1	0.2	0.8
1	0	0.5	0.5
1	1	0.4	0.6

Figure 4

2. **Problem:** Find $P(A = 0 \mid C = 0)$ and $P(D = 1 \mid C = 0)$.
3. **Solution:**
   (a) Derivation of $P(D = 1 \mid C = 0)$:

$$
\begin{aligned}
P(D = 1 \mid C = 0) &= \frac{P(D = 1, C = 0)}{P(C = 0)} && \text{by definition} \\
&= \frac{P(D = 1, C = 0)}{\sum_d P(D = d, C = 0)} && \text{marginalize over } D \\
&= \frac{\sum_{A,B} P(A, B, C = 0, D = 1)}{\sum_d \sum_{A,B} P(A, B, C = 0, D = d)} && \text{equation in problem setup}
\end{aligned}
$$

   - Summing over the variables that are not in the query and evidence variables.

   (b) Summation Term:

$$\sum_{A,B} P(A)P(B \mid A)P(C = 0 \mid A)P(D = d \mid B, C = 0) \quad \text{Bayesian network}$$

$$\sum_{A} P(A)P(C = 0 \mid A) \sum_{B} P(B \mid A)P(D = d \mid B, C = 0) \quad \text{(1st ordering)}$$

$$\sum_{B} P(D = d \mid B, C = 0) \sum_{A} P(A)P(B \mid A)P(C = 0 \mid A) \quad \text{(2nd ordering)}$$

(c) Choose:

$$\underbrace{\sum_B P(D=d \mid B, C=0) \underbrace{\sum_A P(A)P(B \mid A)P(C=0 \mid A)}_{g_1(B)}}_{g_2(d)}$$

(d) $g_1(B)$:

$$\begin{aligned} g_1(B) &= P(A=0)P(B \mid A=0)P(C=0 \mid A=0) + P(A=1)P(B \mid A=1)P(C=0 \mid A=1) \\ &= \begin{cases} 0.9(0.6)(0.8) + 0.1(0.5)(0.6) & \text{if } B=0 \\ 0.9(0.4)(0.8) + 0.1(0.5)(0.6) & \text{if } B=1 \end{cases} \\ &= \begin{cases} 0.462 & \text{if } B=0 \\ 0.318 & \text{if } B=1 \end{cases} \end{aligned}$$

(e) $g_2(d)$:

$$\begin{aligned} g_2(d) &= P(D=d \mid B=0, C=0)g_1(B=0) + P(D=d \mid B=1, C=0)g_1(B=1) \\ &= \begin{cases} 0.5(0.462) + 0.5(0.318) & \text{if } d=0 \\ 0.5(0.462) + 0.5(0.318) & \text{if } d=1 \end{cases} \\ &= \begin{cases} 0.39 & \text{if } d=0 \\ 0.39 & \text{if } d=1 \end{cases} \end{aligned}$$

(f) $P(D=1 \mid C=0) = \dfrac{g_2(1)}{g_2(0)+g_2(1)} = \dfrac{0.39}{0.39+0.39} = 0.5$

4. **Solution 2:**

(a) Derivation of $P(A=0 \mid C=0)$:

$$\begin{aligned} P(A=0 \mid C=0) &= \frac{P(A=0, C=0)}{P(C=0)} \\ &= \frac{P(A=0, C=0)}{\sum_a P(A=a, C=0)} \\ &= \frac{\sum_{B,D} P(A=0, B, C=0, D)}{\sum_a \sum_{B,D} P(A=a, B, C=0, D)} \end{aligned}$$

(b) Summation Term:

$$\sum_{B,D} P(A=a)P(B \mid A=a)P(C=0 \mid A=a)P(D \mid B, C=0) \quad \text{Bayesian network}$$

$$P(C=0 \mid A=a) \sum_B P(B \mid A=a)P(A=a \mid B, C=0) \sum_D P(D \mid B, C=0) \quad \text{(1st ordering)}$$

$$P(C=0 \mid A=a) \sum_D P(D \mid B, C=0) \sum_B P(B \mid A=a)P(A=a \mid B, C=0) \quad \text{(2nd ordering)}$$

(c) Choose:

$$P(C=0 \mid A=a) \underbrace{\sum_B P(B \mid A=a)P(A=a \mid B, C=0) \underbrace{\sum_D P(D \mid B, C=0)}_{g_1(B)}}_{g_2(A)}$$

(d) Same as before.

1.4.2 Hypergraph

Process: Process of eliminating a variable.

1. Create a Hyper-graph by creating a node for each variable.
2. Create hyper-edges (factors) by circling the nodes based on of its parents (i.e. arrows pointing into a variable). If no parents, circle itself.
3. Select a variable v that we are summing over.
   (a) Circle all the variables that have v in their hyperedge into one big hyperedge (i.e. union of hyper-edges).
   (b) Eliminate v by removing the node.
   (c) Calculate the factor by multiplying the support of the variables in the union of hyperedges.
4. Repeat the process for all other v.
5. Select the smallest factor to eliminate first.
6. Repeat until all variables are eliminated to determine the best ordering of elimination.
   - The first eliminated variable will be the inner sum.

Example:

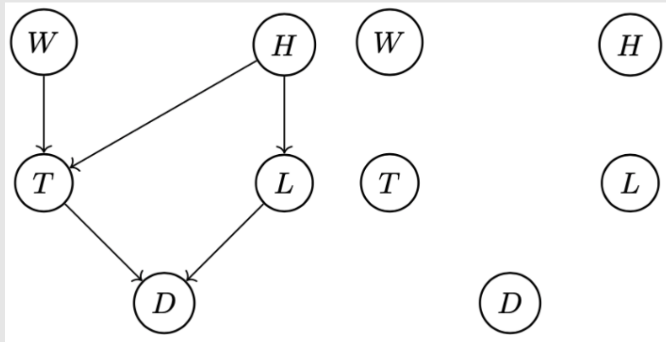


Figure 5

- Since these are all binary variables, we are selecting the factor with the least number of variables to eliminate first.

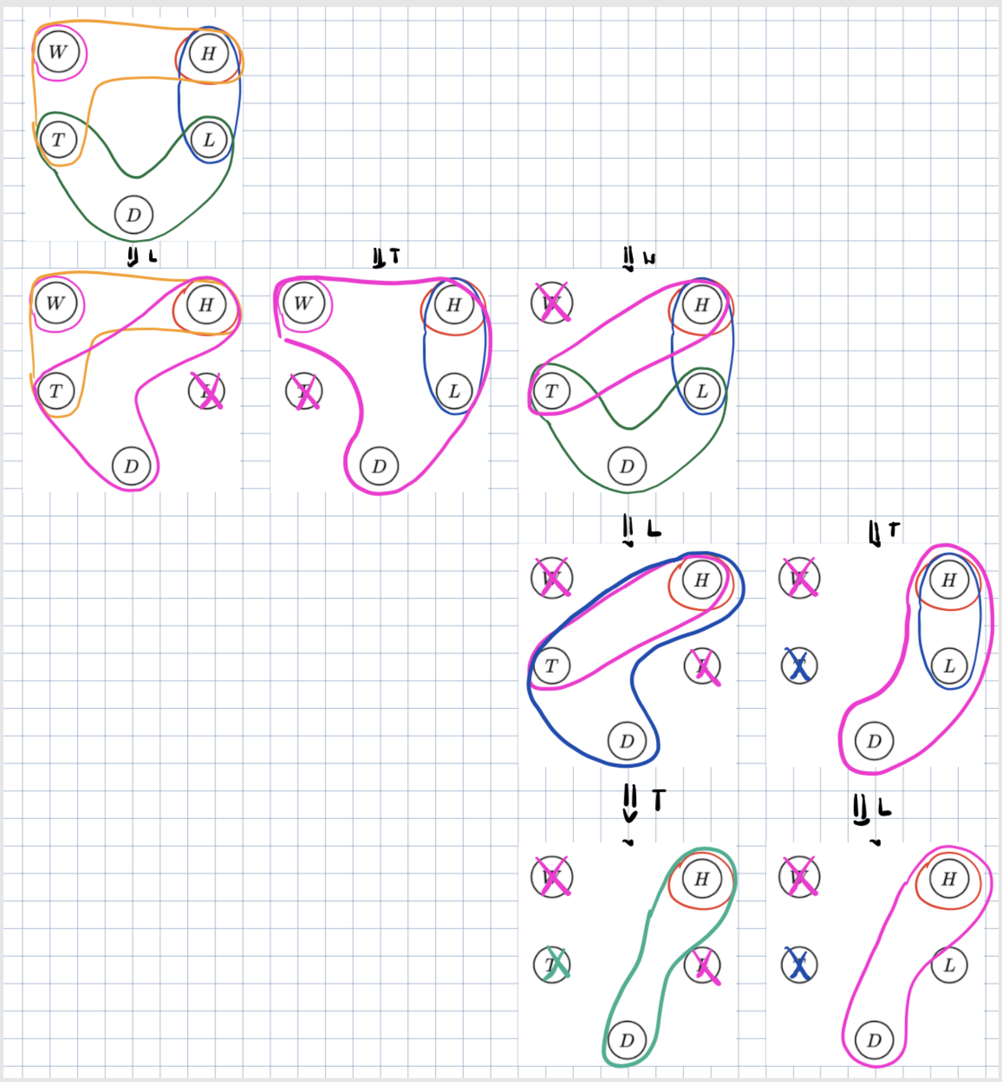


Figure 6

Example:

1. **Given:** Bayesian network

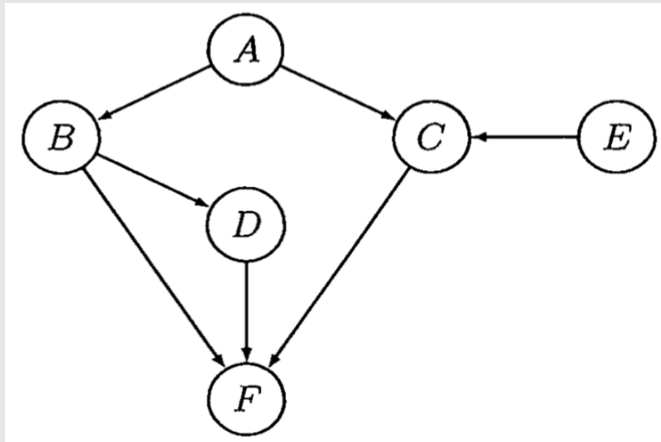


Figure 7

with cardinality of the support of each variable (i.e. number of values each variable can take on) as follows:
   - A: 2^4
   - B: 2^2
   - C: 2^{12}
   - D: 2^2
   - E: 2^3
   - F: 2^6

   Suppose elimination ordering is chosen so that the next variable eliminated is the one that results in the smallest factor (breaking ties alphabetically).
2. **Problem 1:** How many variables must be eliminated to compute $P(A, F \mid C)$?
3. **Solution 1:**
   (a) Since A, F are query, and C is evidence, we must eliminate B, D, and E, so 3 variables must be eliminated.
4. **Problem 2:** What is the first variable to be eliminated to compute $P(F \mid A)$?
5. **Solution 2:**
   (a) Try eliminating all variables that aren't query or evidence and count # of variables in union of hyperedges.

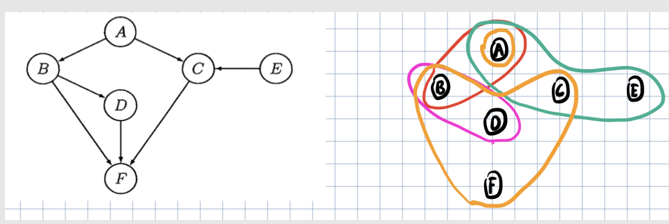


Figure 8

   i. Eliminate B: Hyperunion is ACDF $\to 2^4 \cdot 2^{12} \cdot 2^2 \cdot 2^6 = 2^{24}$
   ii. Eliminate C: Hyperunion is ABDEF $\to 2^4 \cdot 2^2 \cdot 2^2 \cdot 2^3 \cdot 2^6 = 2^{17}$
   iii. Eliminate D: Hyperunion is BCF $\to 2^2 \cdot 2^{12} \cdot 2^6 = 2^{20}$
   iv. Eliminate E: Hyperunion is AC $\to \boxed{2^4 \cdot 2^{12} = 2^{16}}$

   (b) Choose E as the first variable to be eliminated because it has the lowest support in its hyperunion.
6. **Problem 3:** What is the second variable to be eliminated to compute $P(F \mid A)$?
7. **Solution 3:**
   (a) Try eliminating all variable except F, A, E.

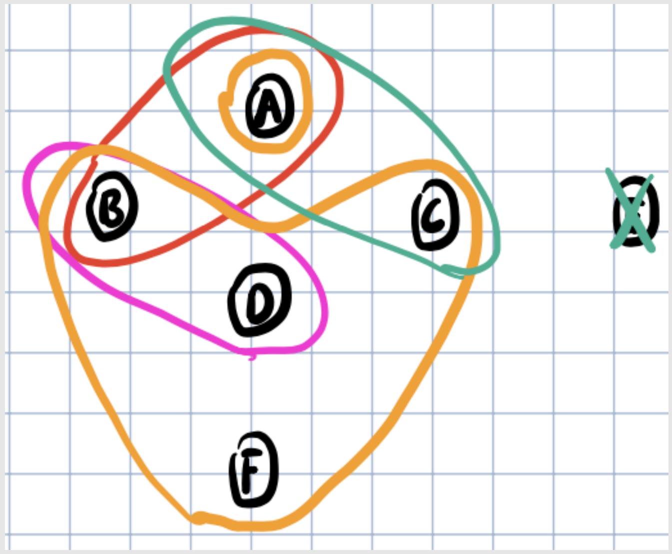


Figure 9

i. Eliminate B: Hyperunion is ACDF $\rightarrow 2^4 \cdot 2^{12} \cdot 2^2 \cdot 2^6 = 2^{24}$
ii. Eliminate C: Hyperunion is ABDF $\rightarrow \boxed{2^4 \cdot 2^2 \cdot 2^2 \cdot 2^6 = 2^{14}}$
iii. Eliminate D: Hyperunion is BCF $\rightarrow 2^2 \cdot 2^{12} \cdot 2^6 = 2^{20}$

(b) Choose C as the second variable to be eliminated because it has the lowest support in its hyperunion.

1.4.3 Inference via Sampling

Process:
1. Given samples
2. Calculate number of samples w/ the query and evidence variables.
3. Calculate number of samples w/ the evidence variables.
4. Approximate the probability of the query variable given the evidence variable by dividing the # of samples w/ the query and evidence variables by the # of samples w/ the evidence variables.

Example:
1. **Given:** Samples

W	H	T	L	D
	h	t	l	d
	h	t	l	d
	$\neg h$	$\neg t$	l	$\neg d$
	$\neg h$	t	l	d
	h	t	l	$\neg d$
	h	$\neg t$	l	d
	$\neg h$	$\neg t$	l	d
	$\neg h$	$\neg t$	$\neg l$	$\neg d$
	h	$\neg t$	$\neg l$	$\neg d$
	$\neg h$	$\neg t$	$\neg l$	d

Figure 10

2. **Problem:** Find the probability of $p(d \mid h)$.
3. **Soln:** $p(d \mid h) \approx \frac{\#\text{ of samples w/ } d \text{ and } h}{\#\text{ of samples w/ } h} = \frac{3}{5} = 0.6.$

1.4.4 Probability Review

Example:

(a) [1 pts] Assume A, B, C are random variables where $A \perp B$, $B \perp C$, and $C \perp A$. Which of the following expressions are equivalent to $P(A, B)$?

$\boxed{\sum_c P(A, B, C = c)}$: Marginalizing over C

$\sum_c P(A|C = c)P(B|C = c)$

$\boxed{\sum_c P(A)P(B|A)P(C = c|A, B)}$: Writing out in terms of $\sum_c P(A, B, C = c)$

$\boxed{P(A)P(B)}$: A and B are independent.

None of the above

(b) [1 pts] Let A, B, C be random variables with $A \not\perp B$. Is it possible for $A \perp B|C$?

$\boxed{\text{Yes}}$: Not being indepedent doesn't imply conditional independence.

No

(c) [1 pts] Let $\mathcal{V}$ denote the set of variables in a Bayesian network. Suppose $X \in \mathcal{V}$, and let $\texttt{pts}(X)$, $\texttt{chl}(X)$, $\texttt{ans}(X)$, and $\texttt{des}(X)$ represent the parents, children, ancestors, and descendants of X. Provide a general independence rule of the form:

$$X \perp \mathcal{V} \setminus \mathrm{des}(X) \mid \texttt{pts}(X)$$

You may use the set operations, $\cap, \cup$ and/or $\setminus$ if necessary.

(d) [1 pts] Assume A, B, C, and D are binary random variables, and E is a trinary random variable over $\{1, 2, 3\}$. What is $\sum_{A,B} P(A|B, E = 1)$?

$P(A = 0 \mid B = 0, E = 1) + P(A = 1 \mid B = 0, E = 1) = 1$
$P(A = 0 \mid B = 1, E = 1) + P(A = 1 \mid B = 1, E = 1) = 1$
Therefore, 2.