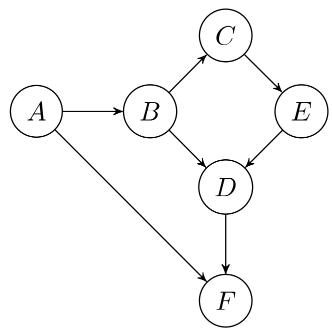
Marc Schlichting • 320-105 (Braun Hall) • email: *mschl@stanford.edu*

PROBLEM SESSION 1: BAYESIAN NETWORKS **January 18, 2023 4:30pm PT**

Question 1. Joint Probabilities of Bayesian Networks.

a) Given is the following Bayesian network.



Find an expression for the joint probability $p(A, B, C, D, E, F)$ using the structure of the Bayesian network.

Solution:
Given the structure of a Bayesian network, the joint probability of the entire Bayesian network can be expressed as product over the probabilities of the individual nodes conditioned on their parents:

$$p(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} p(X_i \mid \text{parents}(X_i)).$$

For the Bayesian network above, we obtain:

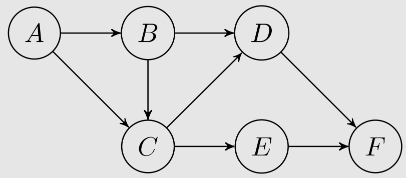
$$p(A, B, C, D, E, F) = p(A)p(B \mid A)p(C \mid B)p(D \mid B, E)p(E \mid C)p(F \mid A, D).$$

b) You are given the following expression for the joint probability of an unknown Bayesian Network:

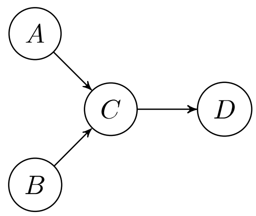
$$p(A, B, C, D, E, F) = p(A)p(B \mid A)p(C \mid A, B)p(D \mid B, C)p(E \mid C)p(F \mid D, E).$$

Draw the Bayesian network.

Solution:
From part 1a, we know that the joint probability of a Bayesian network is the product of the probabilities of the nodes, each conditioned on their parents. For example, $p(X_1 \mid X_2)$ would imply that X_2 is the parent of X_1 and therefore X_1 and X_2 are connected with a directed edge going from X_2 to X_1. Applying this method to the given problem, we obtain the following graphical representation of the Bayesian network:



c) Consider the following Bayesian network



where A is a discrete random variable that can take 3 values, B is a discrete random variable that can take 4 values, C is a discrete random variable that can take 5 values, and D is a continuous random variable that is normally (i.e., Gaussian) distributed. How many independent parameters are necessary to specify the full probability distribution represented by the Bayesian network?

Solution:
For each node discrete X_i with discrete parents, the number of independent parameters can be found using the following formula:

$$\#\text{independent parameters}_i = (k_i - 1) \prod_{j \in \text{parents}(X_i)} k_j,$$

where k_i and k_j are the number of variables of the nodes X_i and X_j, respectively. For a continuously distributed node X_i with discrete parents, the number of independent parameters can be found using the following formula:

$$\#\text{independent parameters}_i = \ell_i \prod_{j \in \text{parents}(X_i)} k_j,$$

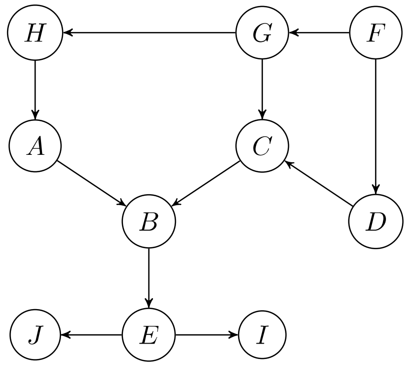
where ℓ_i is the number of parameters required to specify the continuous distribution. For a 1D Gaussian distribution, $\ell = 2$ (i.e., mean and variance). k_j is the number of variables of the node X_j.

For the given Bayesian network we find:

A: $3 - 1 = 2$ independent parameters
B: $4 - 1 = 3$ independent parameters
C: $(5 - 1) \cdot 4 \cdot 3 = 48$ independent parameters
D: $2 \cdot 5 = 10$ independent parameters

The entire Bayesian network has 63 independent parameters.

Question 2. D-Separation for Bayesian Networks.
For this question, consider the following slightly more complex Bayesian network:



a) Check whether D is conditionally independent of H if we observe I and G, i.e., check if $(D \perp H \mid I, G)$.

Solution:
We need to check whether D and H are d-separated given I and G. If this is the case, they are conditionally independent. To check for d-separation, the first step consists of finding all paths between D and H:

Path 1: $D \leftarrow F \rightarrow G \rightarrow H$
Path 2: $D \rightarrow C \leftarrow G \rightarrow H$
Path 3: $D \leftarrow F \rightarrow G \rightarrow C \rightarrow B \leftarrow A \leftarrow H$
Path 4: $D \rightarrow C \rightarrow B \leftarrow A \leftarrow H$

For each of the paths any of the following rules must be true that $(D \perp H \mid I, G)$:

1. The path contains a chain $X \rightarrow Y \rightarrow Z$ such that Y is in the set of evidence variables (i.e., I and G).
2. The path contains a fork $X \leftarrow Y \rightarrow Z$ such that Y is in the set of evidence variables.
3. The path contains a v-structure $X \rightarrow Y \leftarrow Z$ such that neither Y nor any of its descendants is in the set of evidence variables.

Examining each of the paths, we find:

Path 1: $D \leftarrow \underbrace{F \rightarrow G \rightarrow H}_{chain}$ ✓
Path 2: $D \rightarrow \underbrace{C \leftarrow G \rightarrow H}_{fork}$ ✓
Path 3: $D \leftarrow \underbrace{F \rightarrow G \rightarrow C}_{chain} \rightarrow B \leftarrow A \leftarrow H$ ✓
Path 4: $D \rightarrow \underbrace{C \rightarrow B \leftarrow A}_{v-structure} \leftarrow H$ ×

Path 4 contains a v-structure with a descendant (I) that is in the set of evidence variables. Therefore, D is not conditionally independant of H.

b) What is the Markov blanket of D? What variable(s) need to be added to the evidence variables (from the Markov blanket) to solve the problem you encountered in part a)? Verify that the added evidence variable(s) solve the problem!

Solution:
The Markov blanket of a node consists of the node's parents, children and the other parents of the

children. For node D, the Markov blanket consists of F, C, and G. The new preliminary set of evidence variables consists of C, F, G, and I. As adding more evidence variables never leads to d-connection if we haven't used the v-structure argument, we only need to verify that the fourth path from part a) is no longer d-connected:

Path 4: $\underbrace{D \to C \to B}_{chain} \leftarrow A \leftarrow H \quad \checkmark$

As path 4 now contains the chain $D \to C \to B$ and C is in the set of evidence variables, D and H are now d-separated. As we only used C as an additional evidence variable, but not F, we do not need to add F to the final new set of evidence variables $\{C, F, G\}$. F is still part of the Markov blanket of D is its purpose is to identify the set of evidence variables necessary to make D d-separated from all other remaining nodes.

Question 3. Exact Inference.
For the last part of the problem session, we want to consider the following Bayesian network:

$$
\begin{tikzcd}
A \arrow[dr] & & C \arrow[dl] \\
& B \arrow[d] & E \arrow[dl] \\
& D &
\end{tikzcd}
$$

All variables in the Bayesian network are continuous random variables. We are interested in $p(D \mid A, E)$. The naive expression for obtaining $p(D \mid A, E)$ is:

$$p(D \mid A, E) = \frac{\int_B \int_C p(A)p(b \mid A, c)p(c)p(D \mid b, E)p(E)\ dc\, db}{\int_B \int_C \int_D p(A)p(b \mid A, c)p(c)p(\tilde{d} \mid b, E)p(E)\ d\tilde{d}\, dc\, db}.$$

Show that this expression is equal to the following, much simpler expression:

$$p(D \mid A, E) = \int_B \int_C p(D \mid b, E)p(b \mid A, c)p(c)\ dc\, db.$$

Solution:
We start by using the definition of conditional probability and the chain rule:

$$
\begin{aligned}
p(D \mid A, E) &= \frac{p(A, D, E)}{p(A, E)} \\
&= \frac{1}{p(A, E)} \int_B \int_C p(A, b, c, D, E)\ dc\, db \\
&= \int_B \int_C \frac{p(D \mid A, b, c, E)p(b, c \mid A, E)p(A, E)}{p(A, E)}\ dc\, db \\
&= \int_B \int_C p(D \mid A, b, c, E)p(b, c \mid A, E)\ dc\, db
\end{aligned}
$$

Using the graphical model, the definition of conditional probability, and the law of total probability, we can find an expression for $p(D \mid A, B, C, E)$:

$$\begin{aligned}
p(D \mid A, B, C, E) &= \frac{p(A, B, C, D, E)}{\int_D p(A, B, C, \tilde{d}, E)\ d\tilde{d}} \\
&= \frac{p(A)p(B \mid A, C)p(C)p(D \mid B, E)p(E)}{p(A)p(B \mid A, C)p(C)p(E) \underbrace{\int_D p(\tilde{d} \mid B, E)\ d\tilde{d}}_{=1}} \\
&= p(D \mid B, E)
\end{aligned}$$

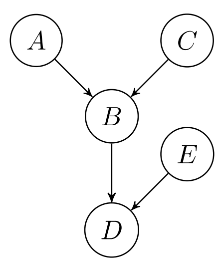
Using the previously found results and the structure of the graphical model, we can find a simplified expression for $p(B, C \mid A, E)$:

$$\begin{aligned}
p(B, C \mid A, E) &= \frac{\int_D p(A, B, C, \tilde{d}, E)\ d\tilde{d}}{\int_B \int_C \int_D p(A, b, c, \tilde{d}, E)\ d\tilde{d}\, dc\, db} \\
&= \frac{p(A)p(B \mid A, C)p(C)p(E)}{p(A)p(E) \underbrace{\int_B \int_C \underbrace{p(b \mid A, c)p(c)p(A)}_{p(b,c|A)}\ dc\, db}_{=1}} \\
&= p(B \mid A, C)p(C)
\end{aligned}$$

Note that $p(A) = 1$ as A is observed. Finally, we can show that the simplification is correct by combining our results:

$$p(D \mid A, E) = \int_B \int_C p(D \mid b, E)p(b \mid A, c)p(c)\ dc\, db.$$

Question 4. Likelihood Weighted Sampling.
For this question we consider the same Bayesian network as in problem 3:



a) Similar to most other sampling methods, likelihood weighted sampling requires a topological sort. Give a topological sort for the given Bayesian network. Is this sort unique?

Solution:
The requirement for a topological sort is that a variable only appears after all its parents have appeared in the sort. With few exceptions are topological sort non-unique. Possible topological sorts for the given Bayesian network are ACBED, CABED, AECBD, CEABD, EACBD, ECABD, ACEBD, and CAEBD.

b) For the subsequent problems, the goal is to find an estimate for $p(d^2 \mid b^0, e^1)$. Assume that A, B, C, and E are binary, while D can take 3 values. Generate 5 samples for likelihood weighted sampling.

Solution:
There are many correct solutions, however, it is important that the samples are consistent with the evidence as well that the values for each of the variables are in the correct range. A possible solution would be:

i	A	B	C	D	E
1	0	0	1	2	1
2	1	0	0	1	1
3	1	0	1	2	1
4	0	0	0	0	1
5	1	0	1	0	1

c) Use the samples from part b) to calculate the weights w_i and finally an estimate for $p(d^2 \mid b^0, e^1)$. Assume that $p(e^0) = 0.3$. The following conditional probability table might be helpful:

A	B	C	$p(B \mid A, C)$
0	0	0	0.2
0	0	1	0.7
1	0	0	0.3
1	0	1	0.9

Solution:
The weights for each sample are calculated as the product of the conditional probabilities of the observed variables. We obtain $p(b^0 \mid A, C)$ from the table above and note that $p(e^1) = 1 - p(e^0) = 0.7$. The weights are:

i	A	B	C	D	E	w_i
1	0	0	1	2	1	$p(b^0 \mid a^0, c^1)p(e^1) = 0.7 \cdot 0.7 = 0.49$
2	1	0	0	1	1	$p(b^0 \mid a^1, c^0)p(e^1) = 0.3 \cdot 0.7 = 0.21$
3	1	0	1	2	1	$p(b^0 \mid a^1, c^1)p(e^1) = 0.9 \cdot 0.7 = 0.63$
4	0	0	0	0	1	$p(b^0 \mid a^0, c^0)p(e^1) = 0.2 \cdot 0.7 = 0.14$
5	1	0	1	0	1	$p(b^0 \mid a^1, c^1)p(e^1) = 0.9 \cdot 0.7 = 0.63$

We note that for $i = \{1, 3\}$ $d^{(i)} = 2$. We then obtain the estimate for $p(d^2 \mid b^0, e^1)$ as:

$$p(d^2 \mid b^0, e^1) \approx \frac{\sum_i w_i(d^{(i)} = 2)}{\sum_i w_i} = \frac{0.49 + 0.63}{0.49 + 0.21 + 0.63 + 0.14 + 0.63} = 0.53.$$