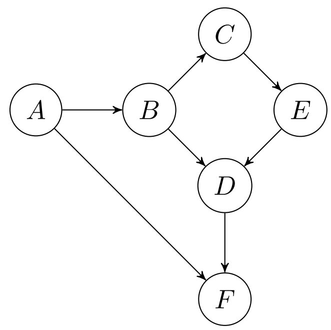
Stanford University
AA228/CS238: Decision Making Under Uncertainty
Fall 2023
Marc Schlichting • Bishop Auditorium • email: *mschl@stanford.edu*

PROBLEM SESSION 1: BAYESIAN NETWORKS **October 4, 2023 4:00pm 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.

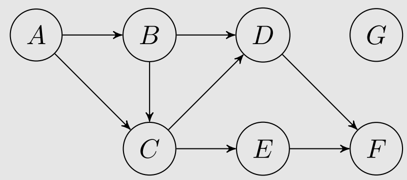
> **Solution:**
> Given the structure of a Bayesian network, the joint probability of the entire Bayesian network can be expressed as product over the probabilities of the individual nodes conditioned on their parents:
>
> $$p(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} p(X_i \mid \text{parents}(X_i)).$$
>
> 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:

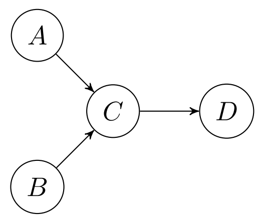
$$p(A, B, C, D, E, F, G) = p(A)p(B \mid A)p(C \mid A, B)p(D \mid B, C)p(E \mid C)p(F \mid D, E)p(G).$$

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:

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

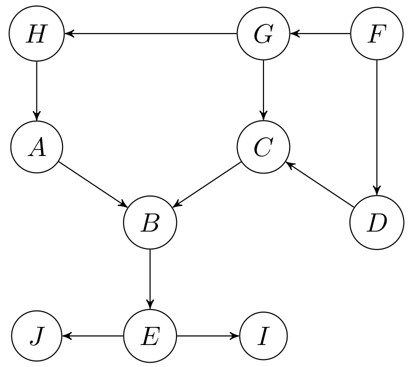
where ℓ_i is the number of parameters required to specify the continuous distribution. For a 1D Gaussian distribution, $\ell = 2$ (i.e., mean and variance). k_j is the number of variables of the node X_j.

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

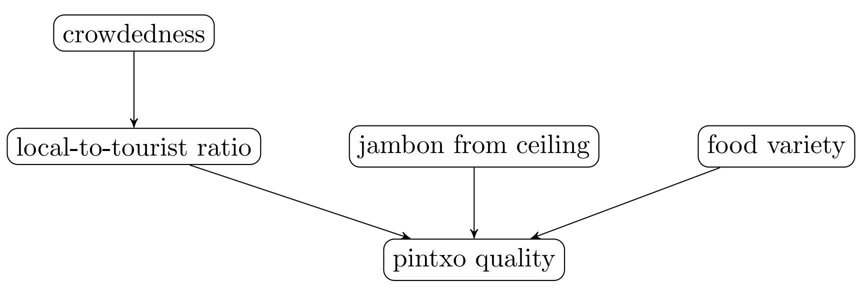
As path 4 now contains the chain $D \rightarrow C \rightarrow B$ and C is in the set of evidence variables, D and H are now d-separated. As we only used C as an additional evidence variable, but not F, we do not need to add F to the final new set of evidence variables $\{C, F, G\}$. F is still part of 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. Sum-Product Variable Elimination.
The Basque country in northern Spain with its charming cities of Bilbao and San Sebastián is known for a culinary specialty called *Pintxos*. Those bread-based delicacies are often topped with regional specialties such as *Jambon de Bayonne*, cheese, or cod.

(a) San Sebastián.

(b) Pintxos.

Numerous pintxo bars can be found in the narrow alleys of the cities' old towns, each offering their own unique creations. Thus, the tradition of a *pintxo crawl* was born where much like in a regular bar crawl, one samples the pintxos from different establishments throughout the evening. However, the pintxo quality at each of the bars is not equal and one would like to maximize the quality of the food consumed throughout the evening. There are certain factors like the crowdedness of the place, the variety of food offerings, whether or not they have jambon hanging from the ceiling (several sources mention this as an indicator for good food quality), and the ratio of local people to tourists that can help us to make an informed decision. For this question, we consider the following Bayesian nework:



We can assume that we have knowledge of the following factors: ϕ_1(crowdedness), ϕ_2(jambon from ceiling), ϕ_3(food variety), ϕ_4(crowdedness, local-to-tourist ratio), and ϕ_5(local-to-tourist ratio, jambon from ceiling, food variety, pintxo quality). Furthermore, the food variety and the local-to-tourist ratio is not accessible by an outside inspection only, so we treat those as hidden variables while we directly observe the crowdedness and whether or not there is jambon hanging from the ceiling.

For this problem, use the sum-product variable elimination algorithm to infer:

$$p(\text{pintxo quality}^1 \mid \text{crowdedness}^1, \text{jambon from ceiling}^1),$$

that is the probability for a good pintxo experience, given that we observe a crowded bar that has jambon hanging from the ceiling. Choose the elimination order *food variety* $\rightarrow$ *local-to-tourist ratio*.

Solution:
The sum-product variable elimination iterativeley replaces all the factors that contain hidden variables. The order in which the variables are eliminated needs to be determined *a priori* and determines the efficiency of the elimination. More information can be found in the book in section 3.3.

The first step of the sum-product variable elimination always consists of replacing the factors that contain are observed. In our case ϕ_1 and ϕ_2 can be discarded entirely, as they are only dependent on observed variables. ϕ_4 is replaced by:

$$\phi_6(\text{local-to-tourist ratio}),$$

and ϕ_5 is replaced by:

$$\phi_7(\text{local-to-tourist ratio, food variety, pintxo quality}).$$

The scheme for each variable to be replaced is the same: The product over the factors is taken and then, the variable-to-eliminate is summed out. As *food variety* is the first elimination variable, we take the product of all non-discarded factors (i.e., ϕ_3, ϕ_6, ϕ_7) that contain *food variety* as a variable and sum over *food variety*:

$$\phi_8(\text{local-to-tourist ratio, pintxo quality}) = \sum_{\text{food variety}} \phi_3(\text{food variety})\phi_7(\text{local-to-tourist ratio, food variety, pintxo quality}).$$

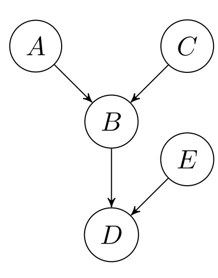
At this point, we discard ϕ_3 and ϕ_7. To keep track, the currently non-discarded factors are ϕ_6 and ϕ_8. The second elimination variable is *local-to-tourist ratio* and both ϕ_6 and ϕ_8 contain the variable. Hence, the new replacement factor is:

$$\phi_9(\text{pintxo quality}) = \sum_{\text{local-to-tourist ratio}} \phi_6(\text{local-to-tourist ratio})\phi_8(\text{local-to-tourist ratio, pintxo quality}).$$

At this point the only remaining factor is ϕ_9. We can expand the final solution and express it with the original factors $\phi_1, \phi_2, \phi_3, \phi_4$, and ϕ_5:

$$p(\text{pintxo quality}^1 \mid \text{crowdedness}^1, \text{jambon from ceiling}^1) \propto$$
$$\phi_1(\text{crowdedness}^1)\phi_2(\text{jambon from ceiling}^1) \sum_{\text{local-to-tourist ratio}} \Bigg(\phi_4(\text{local-to-tourist ratio} \mid \text{crowdedness}^1)$$
$$\sum_{\text{food variety}} \Big(\phi_3(\text{food variety})\phi_5(\text{pintxo quality}^1 \mid \text{local-to-tourist ratio, food variety, jambon from ceiling}^1)\Big)\Bigg).$$

Question 4. Likelihood Weighted Sampling.
For this question, we consider the following Bayesian network:



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:

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