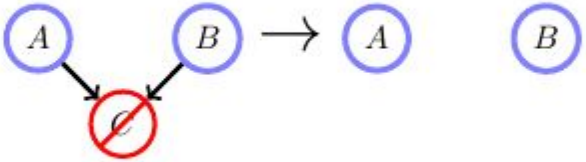
Bayesian Networks

Tutorial - 29th February 2020

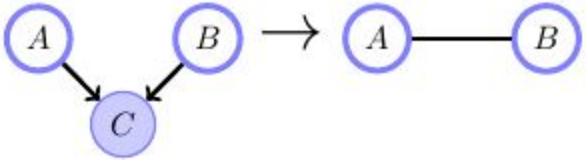
Is x independent of y given z?

“given” = “conditioned on”

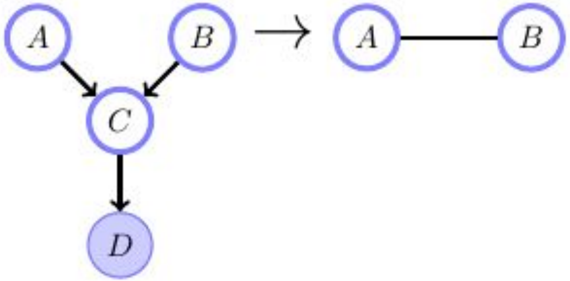
Marginalizing and Conditionalizing



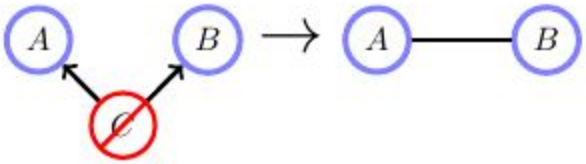
Marginalising over C makes A and B independent. A and B are (unconditionally) independent : $p(A, B) = p(A)p(B)$. In the absence of any information about the effect C, we retain this belief.



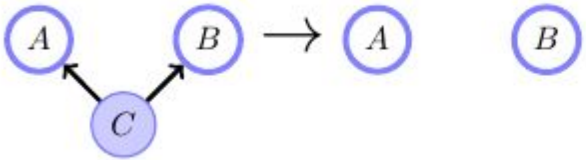
Conditioning on C makes A and B (graphically) dependent — in general $p(A, B|C) \neq p(A|C)p(B|C)$. Although the causes are a priori independent, knowing the effect C in general tells us something about how the causes colluded to bring about the effect observed.



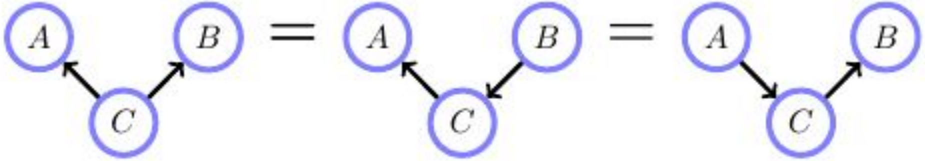
Conditioning on D, a descendent of a collider C, makes A and B (graphically) dependent — in general $p(A, B|D) \neq p(A|D)p(B|D)$.



Marginalising over C makes A and B (graphically) dependent. In general, $p(A, B) \neq p(A)p(B)$. Although we don't know the 'cause', the 'effects' will nevertheless be dependent.



Conditioning on C makes A and B independent: $p(A, B|C) = p(A|C)p(B|C)$. If you know the 'cause' C, you know everything about how each effect occurs, independent of the other effect. This is also true for reversing the arrow from A to C – in this case A would 'cause' C and then C 'cause' B. Conditioning on C blocks the ability of A to influence B.



These graphs all express the same conditional independence assumptions.

Collider

Definition 3.2. Given a path $\mathcal{P}$, a *collider* is a node c on $\mathcal{P}$ with neighbours a and b on $\mathcal{P}$ such that $a \rightarrow c \leftarrow b$. Note that a collider is path specific, see fig(3.8).

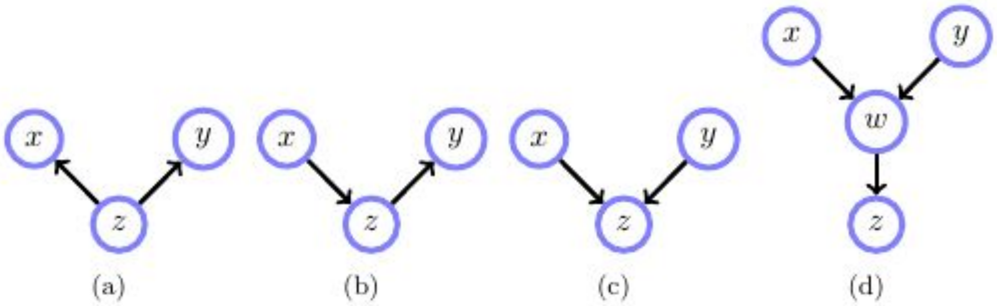


Figure 3.7: In graphs (a) and (b), variable z is not a collider. **(c)**: Variable z is a collider. Graphs (a) and (b) represent conditional independence $x \perp\!\!\!\perp y \mid z$. In graphs (c) and (d), x and y are 'graphically' conditionally dependent given variable z.

D-Separation

One may also phrase this as follows. For every variable $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, check every path U between x and y. A path U is said to be *blocked* if there is a node w on U such that either

1. w is a collider and neither w nor any of its descendants is in $\mathcal{Z}$, or
2. w is not a collider on U and w is in $\mathcal{Z}$.

If all such paths are blocked then $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$. If the variable sets $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$, they are independent conditional on $\mathcal{Z}$ in all probability distributions such a graph can represent.

$$\boxed{\mathcal{X} \text{ and } \mathcal{Y} \text{ d-separated by } \mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}}$$

Markov Blanket (MB)

The Markov blanket of a node is its **parents**, **children** and **parents of its children**.

Significance?

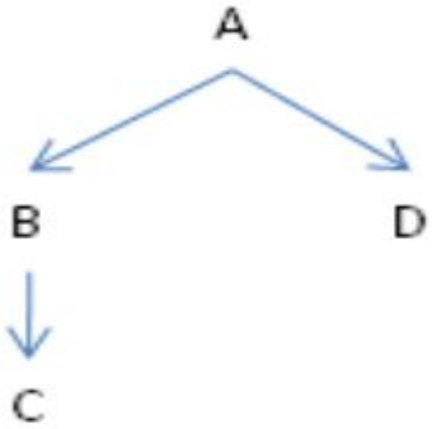
Markov blanket of a node is the only knowledge needed to predict the behavior of that node and its children.

$$\Pr(A \mid \mathrm{MB}(A), B) = \Pr(A \mid \mathrm{MB}(A)).$$

Yudea Pearl came up with this!

Exercises

Q1: A Bayesian network has the following graphical structure:



Assume you have all the conditional probability tables necessary to define the network completely. Derive the formula for computing the conditional probability $P(C|a)$

Ans. $P(C|a) = P(C, a)/P(a)$

Now,

$$P(C,a) = \Sigma_{B,D} P(a, B, C, D).$$

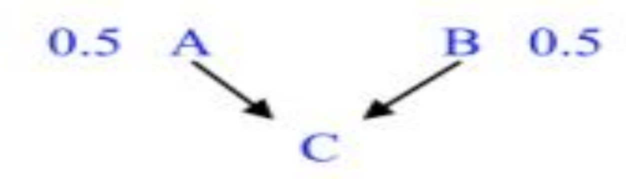
Also,

$$P(a, B, C, D) = P(a) \times P(B|a) \times P(C|B) \times P(D|a)$$

We get,

$$P(C|a) = \Sigma_{B,D} P(B|a) \times P(C|B) \times P(D|a)$$

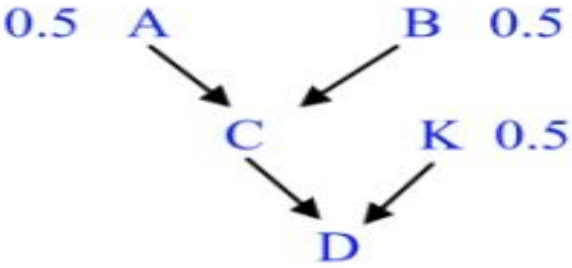
Q2a. Let A, B and C be Boolean variables with possible values 0 and 1. The variables are related such that C = XOR(A,B) (that is C = (A+B) mod 2). Draw a Bayesian network and define the corresponding probabilities for this network which correspond to this relation. You may assume prior probabilities of A and B are 0.5.



A	B	p(C)
0	0	0.0
1	0	1.0
0	1	1.0
1	1	0.0

Q2b. Add to the XOR network network two additional nodes, **D and K**, and the corresponding links and probabilities so that this new network represents the following situation. We are testing in a written exam whether a student knows the operation XOR. In the **exam problem**, the **student is given the values of A and B, an is asked to calculate XOR(A,B)**. The **students answer is D**. D should ideally be equal C. But D may be different from C if the student does not know about XOR. Even if the student knows about XOR, the answer may still be incorrect due to a silly mistake. Let the **variable K = 1 if the student knows the XOR operation, otherwise K = 0**. **If the student knows the operation then his or her answer D will be correct in 99% of the cases**. If the **student does not know XOR, then the answer D will be chosen completely randomly with equal probabilities of 0 and 1**. Draw the Bayesian network to represent this situation. You may assume that the the prior probability of K is 0.5.

Ans.



C	K	p(D)
0	0	0.5
1	0	0.5
0	1	0.01
1	1	0.99

Q2c. In the previous question, let A = 0, B = 1 and D = 1. What is the (approximate) probability that the student knows about XOR?

Ans. **a** signifies **A = 1** and, **¬a** signifies **A = 0**

We want to use the network to compute **P(K = 1|¬a, b, d)**

- $P(K|\neg a, b, d) = \alpha P(K, \neg a, b, d)$
- $P(K, \neg a, b, d) = \Sigma_C P(K, \neg a, b, C, d)$
- $P(K, \neg a, b, C, d) = P(\neg a) \times P(b) \times P(C|\neg a, b) \times P(K) \times P(d|C,K)$

- Solving separately for K = 1 and K = 0, we get:

$P(k, \neg a, b, d) = P(\neg a)P(b)P(c|\neg a, b)P(k)P(d|c, k) + P(\neg a)P(b)P(\neg c|\neg a,b)P(k)P(d|\neg c,k)$

- Using the CPTs, this is,

  $P(k, \neg a, b, d) = (0.5)(0.5)(1)(0.5)(0.99) + (0.5)(0.5)(0)(0.5)(0.01) =$ **0.12375**

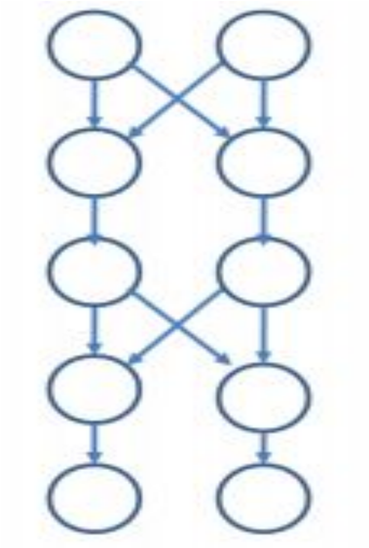
- Similarly for ¬k, we get:

  $P(\neg k, \neg a, b, d) = (0.5)(0.5)(1)(0.5)(0.5) + 0 =$ **0.0625**

- Adding and normalising, we get $P(k|\neg a, b, d) =$

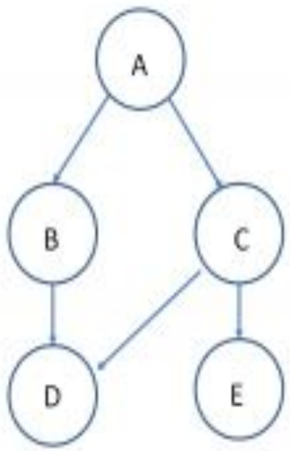
  $0.12375/(0.12375+0.0625) =$ **0.6644295**

Q3. If all the r.v's in the graph shown below are binary, and their joint distribution satisfies the Markov condition, how many entries are needed: **(a)** in the full joint distribution; and **(b)** in the factorized conditional distributions:



Ans. (a) 2^{10}-1 = 1023
(b) 26

Q4. For the following Bayesian network, list out the parents, non-descendents and conditional independencies identified by the Markov condition of each node.



Ans.

X	PA_X	ND_X	Cond. Indep.
A	$\emptyset$	$\emptyset$	–
B	A	C, E	B c.i. C, E, given A
C	A	B	C c.i. B given A
D	B, C	A, E	D c.i. A, E given B, C
E	C	A, B, D	E c.i. A, B, D given C

Resources

- Bayesian Reasoning and Machine Learning - David Barber [PDF] (Simple/Intuitive)
- Probabilistic Graphical Models - Daphne Koller (Advanced/Comprehensive)