Worksheet 10: PGMs I

COMP90051 Statistical Machine Learning

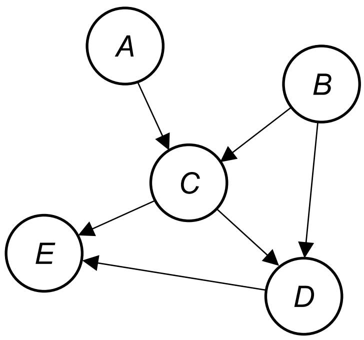
Semester 1, 2021

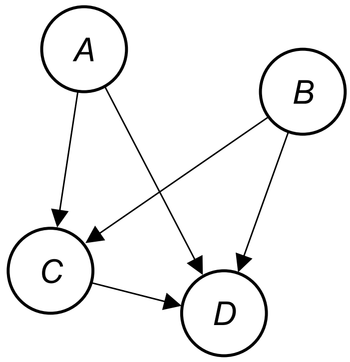
Exercise 1. For the following PGMs:

(a) Find the factorized joint distribution.

(b) Count the number of free parameters in the conditional probability tables, assuming each variable is boolean.

(c) Find the conditional distribution $P(A|B)$, simplifying as much as possible.





Solution. For a directed PGM with nodes $\{X_1, \ldots, X_n\}$ we can factorise the joint distribution as follows:

$$P(X_1, \ldots X_n) = \prod_{i=1}^{n} P\left(X_i \mid \mathsf{Parents}(X_i)\right)$$

where $\mathsf{Parents}(X_i)$ denotes the parent nodes of node X_i. Consider a If $|\mathsf{Parents}(X_i)| = k$ and all variables are boolean, then 2^k free parameters are required to represent the conditional distribution $P\left(X_i \mid \mathsf{Parents}(X_i)\right)$.

(a) $P(A, B, C, D, E) = P(A)P(B)P(C|A, B)P(D|B, C)P(E|C, D)$. Hence the joint CPT has $2^0 + 2^0 + 2^2 + 2^2 + 2^2 = 14$ free parameters.

(b) $P(A, B, C, D) = P(A)P(B)P(C|A, B)P(D|A, B, C)$. Similarly, $2^0 + 2^0 + 2^2 + 2^3 = 14$ free parameters.

(c) To compute $P(A|B)$ we need to expand this using conditional Bayes rule and then express in terms of the joint distribution':

$$\begin{aligned}
P(A|B) &= \frac{P(A,B)}{P(B)} \\
&= \frac{\sum_{C,D,E} P(A)P(B)P(C|A,B)P(D|B,C)P(E|C,D)}{P(B)} \\
&= \frac{P(A)P(B)\sum_C P(C|A,B)\sum_D P(D|B,C)\sum_E P(E|C,D)}{P(B)} \\
&= \frac{P(A)P(B)}{P(B)} = P(A)
\end{aligned}$$

The middle step involves distributing the sums as far right as possible, and recognising that the sum over all events in a distribution is 1. Finally, this tells us that A and B are *marginally independent* RV's. (See also Lecture 19, which includes methods based on the graph topology.)
For the right graph, the process is similar and we reach the same end result:

$$\begin{aligned}
P(A|B) &= \frac{P(A,B)}{P(B)} \\
&= \frac{\sum_{C,D} P(A)P(B)P(C|A,B)P(D|A,B)}{P(B)} \\
&= \frac{P(A)P(B)}{P(B)} = P(A)
\end{aligned}$$

□

Exercise 2. Regarding the rightmost PGM above:

1. Illustrate the form of the conditional probability table for $P(A)$ and $P(D|A,B,C)$. What data structure might be a good choice to implement each of these tables?

2. Considering the pairs of random variables $(A,B);(A,D);(C,D))$, state whether they are *marginally independent* or not. Note that X and Y are marginally independent if $P(X,Y) = P(X)P(Y)$, or equivalently $P(X|Y) = P(X)$.

Solution. 1. $P(A)$ can be simply represented as a single number, denoting $P(A=1)$; which implicitly encodes $P(A=0) = 1 - P(A=1)$.

$P(D|A,B,C)$ is a table of 2^3 values, one for each possible configuration of A,B,C. Each value denotes $P(D=1|A=a,B=b,C=c)$ for combinations of a,b,c. As before $P(D=0|\cdots)$ can be computed from this number. A natural representation for this is a 3-d cube, with each dimension corresponding one of A,B,C. Alternatively it could be flattened into a 2-dimensional table, e.g., with dimensions for A (2 rows, for $A=1$ and $A=0$) and B,C (4 rows for all combinations of the two random variables).

In terms of data structures, this would naturally suit a multi-dimensional array, of shape (1) or (2,2,2). Of course, other data structures could be used, even a python dictionary from sequence of truth values to a float, however a multi-dimensional array lends itself to clean and efficient implementation, including the use of linear algebra.

2. For each pair we can express the distribution in terms of the joint (marginalising out the remaining RVs), and then attempt to simplify. Only in the first case is this possible. Hence only A,B are marginally dependent.

A, B	marginally independent (showed this above, as part of 1c)
A, D	$P(A, D) = \sum_{B,C} P(A)P(B)P(C\|A, B)P(D\|A, B)$ ... no
C, D	$P(D, D) = \sum_{A,B} P(A)P(B)P(C\|A, B)P(D\|A, B) = \sum_{B} P(A)P(B)P(C\|A, B)$ ... no

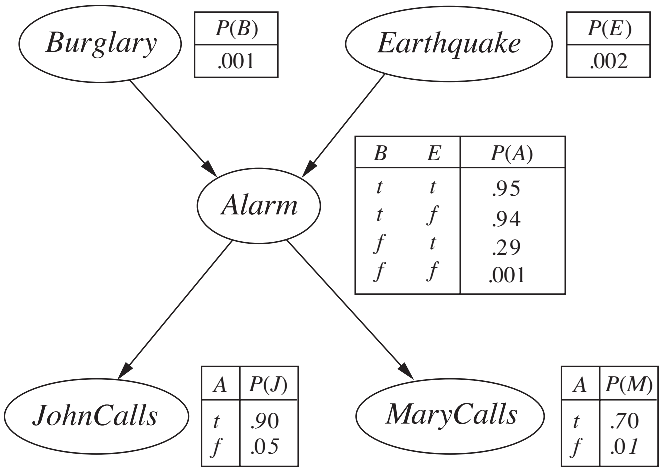
Please see also the lecture this week, showing how we can read independence relations directly from the graph, based on the paths connecting the two nodes.

□

Exercise 3 (Based on RN 14.15)**.** Leo is a botanist who lives in the Bay Area. His neighbourhood is a hotspot for burglars, so his house is fitted with an alarm system. Unfortunately, the alarm is not perfectly reliable: it doesnt always trigger during a home invasion, and it may be erroneously triggered during minor earthquakes, which occur occasionally. Leo has asked his neighbours John and Mary (who don't know each other) to call him if they hear the alarm.

Construct a PGM to help model this scenario, with a view to allowing Leo to perform queries such as determining how likely his home has been invaded given he has heard from one or another of his neighbours. You will have to choose what random variables to use, what edges to add, and the direction of the edges to form the most natural model.

Solution. This is one way it might be modelled, although bear in mind there are many correct answers. Generally we try to orient edges in the direction of causality, leading to a more intuitive and interpretable model.



Note that the CPTs illustrated are not part of the question. □