Worksheet 10: PGMs I

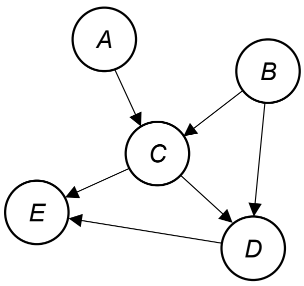
COMP90051 Statistical Machine Learning

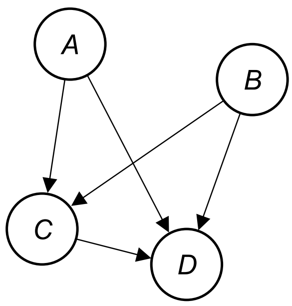
Semester 2, 2020

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.





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

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

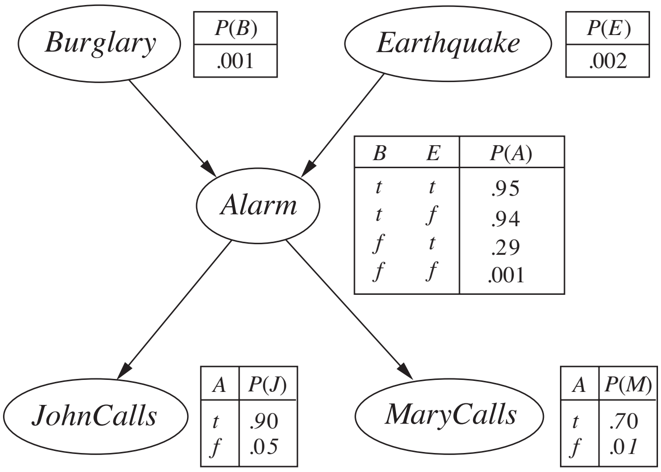
where $\mathsf{Parents}(X_i)$ denotes the parent nodes of node X_i. Consider a If $|\mathsf{Parents}(X_i)| = k$ and all variables are boolean, then 2^k free parameters are required to represent the conditional distribution $P\left(X_i \mid \mathsf{Parents}(X_i)\right)$.

(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.

□

Exercise 2 (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 doesn't 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. Leo would like to determine the likelihood that his home is being burgled if he receives a phone call from both John and Mary. Use Bayes' theorem and marginalisation and the following PGM to help him answer this question.



Solution. For brevity, let B = *Burglary*, E = *Earthquake*, A = *Alarm*, J = *JohnCalls* and M = *MaryCalls*. We begin by factorising the joint distribution according to the topology of the PGM:

$$p(A, B, E, J, M) = p(B)p(E)p(A|B, E)p(J|A)p(M|A)$$

Our query variable is B, and our evidence variables are $J = j$, $M = m$. The hidden or unobserved variables are E and A. Using Bayes' Theorem, we can express the posterior distribution over the query B as proportional to the full joint distribution, then exploit the factorisation and marginalise over the hidden variables to obtain:

$$\begin{aligned} p(B|j, m) &= \frac{1}{Z} \sum_A \sum_E P(A, B, E, J = j, M = m) \\ &= \frac{1}{Z} \sum_A \sum_E p(B)p(E)p(A|B, E)p(j|A)p(m|A) \\ &= \frac{1}{Z} p(B) \sum_E p(E) \sum_A p(A|B, E)p(j|A)p(m|A) \end{aligned}$$

In the last line we push the summations as far right as possible using the distributive property to avoid excessive computation. Performing full enumeration, we would need to proceed left to right, looping over all possible values for E, and for each possible value of E we need to loop over all possible values of A. This may remind you of nested for loops. We proceed in reverse (right to left) order, caching intermediate results to avoid repeated calculations. Start by focusing on the rightmost sum over A, and use the convention that B runs over the columns and E runs over the rows. We note that $P(A|B, E)$ is a $2 \times 2 \times 2$ matrix, and the sum is over 2×2 'slices' of the whole matrix.

$$\begin{aligned} \sum_A p(A|B, E)p(j|A)p(m|A) &= p(a|B, E)p(j|a)p(m|a) + p(\neg a|B, E)p(j|\neg a)p(m|\neg a) \\ &= \begin{pmatrix} p(a|b, e) & p(a|\neg b, e) \\ p(a|b, \neg e) & p(a|\neg b, \neg e) \end{pmatrix} \times 0.7 \times 0.9 + \\ &\quad \begin{pmatrix} p(\neg a|b, e) & p(\neg a|\neg b, e) \\ p(\neg a|b, \neg e) & p(\neg a|\neg b, \neg e) \end{pmatrix} \times 0.05 \times 0.1 \\ &= \begin{pmatrix} 0.95 & 0.29 \\ 0.94 & 0.001 \end{pmatrix} \times 0.7 \times 0.9 + \begin{pmatrix} 0.05 & 0.71 \\ 0.06 & 0.999 \end{pmatrix} \times 0.05 \times 0.1 \\ &= \begin{pmatrix} p(j, m|b, e) & p(j, m|\neg b, e) \\ p(j, m|b, \neg e) & p(j, m|\neg b, \neg e) \end{pmatrix} \end{aligned}$$

So we have marginalized, or 'eliminated' the hidden variable A from our intermediate results. Now to evaluate the sum over E. Note that the top row corresponds to $E = e$, and the bottom corresponds to $E = \neg e$, so we have to perform vector matrix multiplication:

$$\begin{aligned}\sum_E p(E)\begin{pmatrix} p(j,m|b,e) & p(j,m|\neg b,e) \\ p(j,m|b,\neg e) & p(j,m|\neg b,\neg e)\end{pmatrix} &= \begin{pmatrix} p(e) & p(\neg e)\end{pmatrix}\begin{pmatrix} p(j,m|b,e) & p(j,m|\neg b,e) \\ p(j,m|b,\neg e) & p(j,m|\neg b,\neg e)\end{pmatrix} \\ &= \begin{pmatrix} 0.002 & 0.998\end{pmatrix} \\ &= \begin{pmatrix} p(j,m|b) & p(j,m|\neg b)\end{pmatrix}\end{aligned}$$

Now we have marginalized/eliminated the hidden variable E from our intermediate result to leave a factor only depending on B. Now take the Schur (elementwise) product for the last part (because the left column corresponds to $B = b$ and the right to $B = \neg b$) to find:

$$\begin{aligned} p(B|j,m) &= \frac{1}{Z} p(B) \odot \begin{pmatrix} p(j,m|b) & p(j,m|\neg b)\end{pmatrix} \\ &= \frac{1}{Z}\begin{pmatrix} p(j,m|b) & p(j,m|\neg b)\end{pmatrix} \\ &= \frac{1}{Z}\begin{pmatrix} 0.00059224 & 0.00149186\end{pmatrix}\end{aligned}$$

Normalization implies that $Z = \sum_B p(B,j,m) = 0.002084100239$, giving the posterior distribution:

$$p(B|j,m) = \begin{pmatrix} 0.28417184, & 0.71582816\end{pmatrix}$$

You may have found the above sequence of matrix products quite chaotic. It is sensible if you keep track of which variable is being eliminated at each step. This allows you to establish a clear correspondence between the axes of your matrix/tensor and the variables to be eliminated. For a more efficient approach, you may use the 'pointwise' product method outlined in RN Ch. 14.4.2 or the elimination algorithm presented in lecture. □

Exercise 3 (Based on RN 14.4)**.** The following questions relate to the PGM from Exercise 2.

(a) If no evidence is observed, are the *Burglary* and *Earthquake* nodes independent?

(b) Assume we observe $Alarm = t$, now are *Burglary* and *Earthquake* independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence.

(c) **[Discussion]** Explain intuitively why knowledge of the *Alarm* random variable renders the random variables *Earthquake* and *Burglary* non-conditionally independent.

Solution.

1. Marginalizing over unobserved variables in the joint distribution and simplifying the summation, we have

$$\begin{aligned} p(B,E) &= \sum_A\sum_J\sum_M p(A,B,E,J,M) \\ &= \sum_A\sum_J\sum_M p(B)p(E)p(A|B,E)p(J|A)p(M|A) \\ &= p(B)p(E)\sum_A p(A|B,E)\sum_J p(J|A)\sum_M p(M|A)\end{aligned}$$

Note that every sum above is equal to 1 by the law of total probability—i.e. the variables A, J, M are irrelevant to the query. More generally, every variable that is not an ancestor of a query/evidence variable is irrelevant. So we have $p(B, E) = p(B)p(E)$ in this case.

2. By Bayes' Theorem

$$\begin{aligned} p(B|E, a) &= \frac{1}{Z} p(B, E, a) \\ &= \frac{1}{Z} p(a|B, E)p(B)p(E) \\ &= \frac{p(a|B, E)p(B)}{\sum_B p(a|B, E)p(B)} \end{aligned}$$

Where $Z = \sum_B p(a|B, E)p(B)p(E)$ is a normalizing constant (confusingly, also called the 'evidence'). For B and E to be conditionally independent given $A = a$, $p(B|E, a)$ should not depend on E. This is not possible as $p(a|B, E)$ depends on E (validate this numerically).

3. Under our assumptions, two possible explanations for the alarm going off ($\mathit{Alarm} = t$) are an earthquake occurring ($\mathit{Earthquake} = t$) and a burglary occurring ($\mathit{Burglary} = t$). If we know that $\mathit{Alarm} = t$ and $\mathit{Earthquake} = t$ then intuitively we think that the likelihood a burglary has occurred is low ($\mathit{Burglary} = f$) (Leo and the burglar would have to be very unlucky to have an earthquake and burglary occur at the same time):

$$p(b|a, e) < p(\neg b|a, e) \tag{1}$$

This is reasonably intuitive: for observed data with two possible explanations, knowing one of the explanations has occurred decreases the probability that the other possible explanation has also occurred—there is now a conditional dependence between causes, given observation of the children.

Hence for any PGM with the structure $X \to Z \leftarrow Y$, knowledge of Z couples X and Y.

□