Exercise Sheet - Reasoning under Uncertainty

Exercise 1. Show from the first principles that $P(a|b \wedge a) = 1$.

Solution: Using conditional probability, i.e. $P(X|Y) = \frac{P(X \wedge Y)}{P(Y)}$, we have:

$$P(a|b \wedge a) = \frac{P(a \wedge (b \wedge a))}{P(b \wedge a)} = \frac{P(b \wedge a)}{P(b \wedge a)} = 1$$

▲

Exercise 2. For each of the following statements, either prove that it is true or give a counterexample.

a) If $P(a|b, c) = P(b|a, c)$, then $P(a|c) = P(b|c)$.

Solution: True. Using conditional probability, i.e. $P(X|Y) = \frac{P(X,Y)}{P(Y)}$, we rewrite $P(a|b, c) = P(b|a, c)$ too:

$$\frac{P(a, b, c,)}{P(b, c)} = \frac{P(a, b, c,)}{P(a, c)}$$

Now, by first dividing the above by $P(a, b, c)$, and then using $P(X, Y) = P(X|Y)P(Y)$, we get:

$$P(b, c) = P(a, c)$$
$$P(b|c)P(c) = P(a|c)P(c)$$
$$P(b|c) = P(a|c)$$

▲

b) If $P(a|b, c) = P(a)$, then $P(b|c) = P(b)$.

Solution: False. $P(a|b, c) = P(a)$ only tells us that a is independent from b and c. Any example where this is true and where b depends on c is a counter example. E.g. if b is the event that the grass is wet and c the event that it rains, and a is some independent event. Then it holds that $P(a|b, c) = P(a)$, but since b depends on c, $P(b|c) \neq P(b)$.

▲

c) If $P(a|b) = P(a|b, c)$, then $P(a|b, c) = P(a|c)$.

Solution: False. $P(a|b) = P(a|b, c)$ means that a is conditionally independent of c given b. E.g. if a depends on b and b depends on c we have that $P(a|b) = P(a|b, c)$ holds, but $P(a|b, c) \neq P(a|c)$, since b would give additional information about a.

▲

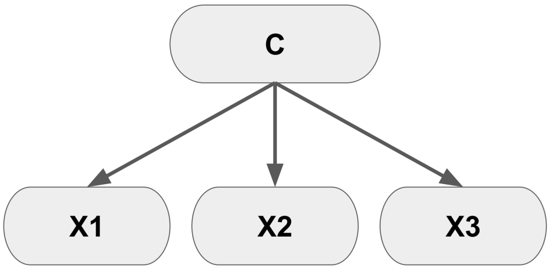

Figure 1: Bayesian network for exercise 4

Exercise 3. After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease and that the test is 99% accurate (i.e., the probability of testing positive when you do have the disease is 0.99, as is the probability of testing negative when you don't have the disease). The good news is that this is a rare disease, striking only 1 in 10,000 people of your age. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

Solution: If 10.000 people take the test, we expect 1 person to have the disease. While the rest does not have the disease, 1% of them will still test positive.

To answer this question we will use Bayes theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

We know $P(positive|disease) = 0.99$, $P(\neg positive|\neg disease) = 0.99$, and $P(disease) = 0.0001$. We get:

$$P(disease|positive) = \frac{P(positive|disease)P(disease)}{P(positive|disease)P(disease) + P(positive|\neg disease)P(\neg disease)}$$

$$= \frac{0.99 * 0.0001}{0.99 * 0.0001 + 0.01 * 0.9999} = 0.0098$$

▲

Exercise 4. We have a bag of three biased coins a, b. and c with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X_1 , X_2, and X_3.

a) Draw the Bayesian network corresponding to this setup and define the necessary CPTs.

Solution: See Figure 1 for the Bayesian network. We need 1 variable for the coins (C), and 1 variable for each coin toss (X_1, X_2, X_3).

We have two conditional probability tables. One for C:

C	$P(C)$
a	$1/3$
b	$1/3$
c	$1/3$

And one for X_i:

X_i	C	$P(X_i|C)$
heads	a	0.2
heads	b	0.6
heads	c	0.8

▲

b) Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

Solution: We want to know for which coin $P(C|2heads, 1tails)$ is the largest. Using Bayes theorem we have:

$$P(C|2h, 1t) = \frac{P(2h, 1t|C)P(C)}{P(2h, 1t)}.$$

Since $P(C)$ is the same for every coin and the term $P(2h, 1t)$ will also be the same, we only have to look at $P(2h, 1t|C)$. We get:

$$P(2h, 1t|C) = P(h|C)P(h|C)P(t|C) * 3.$$

Since there are three orderings for 2 heads and 1 tails, we multiply by 3.

E.g. for coin a we get:

$$P(h|a)P(h|a)P(t|a) * 3 = 0.2 * 0.2 * 0.8 * 3 = 0.096.$$

Doing this for all three coins shows that coin b is the most likely.

▲

function ELIMINATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  factors ← [ ]
  for each var **in** ORDER(bn.VARS) **do**
    factors ← [MAKE-FACTOR(var, **e**)|*factors*]
    if var is a hidden variable **then** *factors* ← SUM-OUT(var, *factors*)
  return NORMALIZE(POINTWISE-PRODUCT(*factors*))

Exercise 5. From Russell & Norvig, ex. 14.15. This question is about section 14.4
Consider the variable elimination algorithm above.

(a) Perform variable elimination to the query:
$P(Burglary|JohnCalls = true, MaryCalls = true)$

(b) Count the number of arithmetic operations performed, and compare it with the number performed by the enumeration algorithm.

(c) Suppose a network has the form of a chain: a sequence of Boolean variables $X_1, \ldots, X_n$, where $Parents(X_i) = \{X_{i-1}\}$ for $i = 2, \ldots, n$. What is the complexity of computing $P(X_1|X_{n=true})$ using enumeration? Using variable elimination?

Solution: 5 a.

$$
\begin{aligned}
P(B|j,m) &= \alpha P(B) \sum_e P(e) \sum_a P(a|b,e)P(j|a)P(m|a) \\
&= \alpha P(B) \sum_e P(e) \left[.9 \times .7 \times \begin{pmatrix} .95 & .29 \\ .94 & .001 \end{pmatrix} + .05 \times .01 \times \begin{pmatrix} .05 & .71 \\ .06 & .999 \end{pmatrix} \right] \\
&= \alpha P(B) \sum_e P(e) \begin{pmatrix} .598525 & .183055 \\ .59223 & .0011295 \end{pmatrix} \\
&= \alpha P(B) \left[.002 \times \begin{pmatrix} .598525 \\ .183055 \end{pmatrix} + .998 \times \begin{pmatrix} .59223 \\ .0011295 \end{pmatrix} \right] \\
&= \alpha \begin{pmatrix} .001 \\ .999 \end{pmatrix} \times \begin{pmatrix} .59224259 \\ .001493351 \end{pmatrix} \\
&= \alpha \begin{pmatrix} .00059224259 \\ .0014918576 \end{pmatrix} \\
&\approx \langle .284, .716 \rangle
\end{aligned}
$$

5 b. Including the normalization there are 7 additions, 16 multiplications and 2 divisions.

For the enumeration algorithm, we have for $P(b|j,m)$:

$$
\begin{aligned}
P(b|j,m) &= \alpha P(b) \sum_e P(e) \sum_a P(a|b,e)P(j|a)P(m|a) \\
&= \alpha P(b) \Big[P(e) \Big[P(a|b,e)P(j|a)P(m|a) + P(\neg a|b,e)P(j|a)P(m|a) \Big] \\
&+ P(\neg e) \Big[P(a|b,\neg e)P(j|a)P(m|a) + P(\neg a|b,\neg e)P(j|a)P(m|a) \Big] \Big]
\end{aligned}
$$

Where we have 11 multiplications and 3 summations. For $P(\neg b|j,m)$ we have the same number of operations, bringing us to 22 multiplications and 6 summations.

Then, for the normalization step, we have 1 more addition and 2 divisions. So 7 additions, 22 multiplications and 2 divisions in total. The enumeration has 6 extra multiplications compared to variable elimination.

5 c. To compute $P(X_1|X_n = true)$ using enumeration, we have to evaluate two complete binary trees(one for each value of X_1), each of depth $n-2$, so the total work is $O(2^n)$. Using variable eliminaion, the factors never grow beyond two variables. For example, the first step is

$$
\begin{aligned}
P(X_1|X_n = true) &= \alpha P(X_1) \ldots \sum_{x_{n-2}} P(x_{n-2}|x_{n-3}) \sum_{x_{n-1}} P(x_{n-1}|x_{n-2}) P(X_n = true|x_{n-1}) \\
&= \alpha P(X_1) \ldots \sum_{x_{n-2}} P(x_{n-2}|x_{n-3}) \sum_{x_{n-1}} f_{X_{n-1}}(x_{n-1}, x_{n-2}) f_{X_n}(x_{n-1}) \\
&= \alpha P(X_1) \ldots \sum_{x_{n-2}} P(x_{n-2}|x_{n-3}) f_{\overline{X_{n-1}} X_n}(x_{n-2})
\end{aligned}
$$

The last line is isomorphic to the problem with $n-1$ variables instead of n; the work done on the first step is a constant independent of n, hence (by in duction on n, if you want to be formal) the total work is $O(n)$.

▲