Introduction to Artificial Intelligence (INFO8006)

Exercises 3 – Reasoning under uncertainty

October 20, 2021

Learning outcomes

At the end of this session you should be able to

- calculate marginal and conditional probabilities given joint distributions.
- apply and understand the Bayes rule.
- formulate a random process as a Bayesian network and conditional probability tables.
- determine whether random variables are independent given a Bayesian network.
- compute probabilities in the context of a (simple) Bayesian network.

Exercise 1   Beliefs

Let A and B be two *events* over the probability space Ω. An agent holds the beliefs $P(A) = 0.4$ and $P(B) = 0.3$. What ranges of probabilities would it be rational for the agent to hold for the events $A \cup B$ and $A \cap B$? What if the agent also believes that $P(A \mid B) = 0.5$?

1. The agent's beliefs should satisfy

$$0.4 = \max\{P(A), P(B)\} \leq P(A \cup B) \leq \min\{1, P(A) + P(B)\} = 0.7$$
$$0 = \max\{0, P(A) + P(B) - 1\} \leq P(A \cap B) \leq \min\{P(A), P(B)\} = 0.3.$$

   This can be illustrated with Venn diagrams.

2. If the agent believes that $P(A \mid B) = 0.5$, it should also believe that

$$P(A \cap B) = P(A \mid B)P(B) = 0.5 \times 0.3 = 0.15,$$

   thanks to Bayes rule. Then, the agent should rationally believe that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.3 - 0.15 = 0.55.$$

Exercise 2   (AIMA, Ex 13.8)

	toothache		no toothache	
	catch	no catch	catch	no catch
cavity	0.108	0.012	0.072	0.008
no cavity	0.016	0.064	0.144	0.576

Given the hereabove probability table, compute the following probabilities:

In the following, we represent the events *toothache*, *catch* and *cavity* by the random variables $T, CT, CV : \Omega \mapsto \{0, 1\}$, respectively. Then, the hereabove table represents the joint probability distribution $P(T, CT, CV)$. When clear from context, we abbreviate the probability of a realization (t, ct, cv) by $P(t, ct, cv)$ instead of $P(T = t, CT = ct, CV = cv)$.

1. $P(\textit{toothache})$

$$P(T = 1) = \sum_{ct} \sum_{cv} P(T = 1, ct, cv) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

2. $P(\textit{cavity})$

$$P(CV = 1) = \sum_{t} \sum_{ct} P(t, ct, CV = 1) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

3. $P(\textit{toothache} \mid \textit{cavity})$

$$P(T = 1 \mid CV = 1) = \frac{P(T = 1, CV = 1)}{P(CV = 1)} = \frac{\sum_{ct} P(T = 1, ct, CV = 1)}{P(CV = 1)}$$
$$= \frac{0.108 + 0.012}{0.2} = 0.6$$

4. $P(\textit{cavity} \mid \textit{toothache} \cup \textit{catch})$

$$P(CV = 1 \mid T = 1 \cup CT = 1) = \frac{P(T = 1 \cup CT = 1, CV = 1)}{P(T = 1 \cup CT = 1)}$$
$$= \frac{P(CV = 1) - P(T = 0, CT = 0, CV = 1)}{1 - P(T = 0, CT = 0)}$$
$$= \frac{0.2 - 0.008}{1 - 0.584} = 0.4615$$

Exercise 3   (AIMA, Ex 13.15)

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 and the probability of testing positive when you don't is 0.01. 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?

It is a good news because the conditional probability will drop due to the weak prior value. Indeed,

$$
\begin{aligned}
P(d \mid t) &= \frac{p(t \mid d)p(d)}{p(t)} \\
&= \frac{p(t \mid d)p(d)}{p(t \mid d)p(d) + p(t \mid \neg d)p(\neg d)} \\
&= \frac{0.99 \times 10^{-4}}{0.99 \times 10^{-4} + 0.01 \times (1 - 10^{-4})} = 0.0098.
\end{aligned}
$$

Exercise 4 (AIMA, Ex 13.13)

For each of the following statements, either prove it is valid or give a counterexample.

1. If $P(a \mid b, c) = P(b \mid a, c)$, then $P(a \mid c) = P(b \mid c)$.

   True. By the conditional probability rule we have $P(a \mid b, c)P(b, c) = P(b \mid a, c)P(a, c)$, which by assumption reduces to $P(b, c) = P(a, c)$ and thereby $P(b \mid c) = P(a \mid c)$.

2. If $P(a \mid b, c) = P(a)$, then $P(b \mid c) = P(b)$.

   False. The statement $P(a \mid b, c) = P(a)$ merely states that a is independent of b and c, it makes no claim regarding the dependence of b and c. Counterexample: a and b record the results of two independent coin flips, and $c = b$.

3. If $P(a \mid b) = P(a)$, then $P(a \mid b, c) = P(a \mid c)$.

   False. While the statement $P(a \mid b) = P(a)$ implies that a is independent of b, it does not imply that a is conditionally independent of b given c. Counterexample: a and b record the results of two independent coin flips, and $c = a \vee b$.

Exercise 5   Bag of coins

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 probability 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 (heads or tail).

1. Draw the Bayesian network corresponding to this setup and define the corresponding conditional probability tables (CPTs).

   The process is described with the following random variables:

   - $Y \in \{a, b, c\}$, the coin drawn from the bag.

   - $X_i \in \{h, t\}$, the realization of the i-th coin toss.

   The variable Y is the parent of X_i, it should be put before in the network ordering.
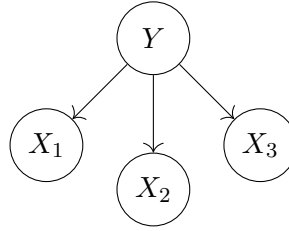
Y	$P(Y)$
a	$\frac{1}{3}$
b	$\frac{1}{3}$
c	$\frac{1}{3}$

Y	$P(X_i = h \mid Y)$
a	0.2
b	0.6
c	0.8

2. Determine which coin (a, b or c) is most likely to have been drawn from the bag if the observed tosses came out heads twice and tail once.

   In order to find the most likely coin, we have to calculate the conditional probability distribution $P(Y \mid X_1 = h, X_2 = h, X_3 = t)$[1]. Thanks to Bayes rule, we have

   $$P(Y \mid X_1 = h, X_2 = h, X_3 = t) \propto P(X_1 = h, X_2 = h, X_3 = t \mid Y)P(Y).$$

   But we know that X_1, X_2 and X_3 are independent with respect to Y, meaning that

   $$P(X_1, X_2, X_3 \mid Y) = P(X_1 \mid Y)P(X_2 \mid Y)P(X_3 \mid Y).$$

   Then, we have

   $$P(Y \mid X_1 = h, X_2 = h, X_3 = t) \propto \frac{1}{3} \times \begin{cases} 0.2 \times 0.2 \times 0.8 = 0.032 & \text{if } Y = a \\ 0.6 \times 0.6 \times 0.4 = 0.144 & \text{if } Y = b \\ 0.8 \times 0.8 \times 0.2 = 0.128 & \text{if } Y = c \end{cases}$$

   meaning that b is the most likely coin to have been drawn. Interestingly, we don't need to normalize the distribution as we only care about the maximum.

---

[1]It should be noted that the order of the tosses is irrelevant as they are performed independently.

Exercise 6   Handedness (AIMA, Ex: 14.6)

Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common *hypothesis* is that left or right-handedness is inherited by a simple mechanism; that is, there is a gene G_x, also with values l or r, and H_x turns out the same as G_x with some probability s.

Furthermore, the gene itself is equally likely to be inherited from either of the individual's parents, with a small nonzero probability m of a random mutation flipping the gene.

1. Which of the following networks claim that $(G_{child} \mid G_{mother}, G_{father}) = P(G_{child})$?
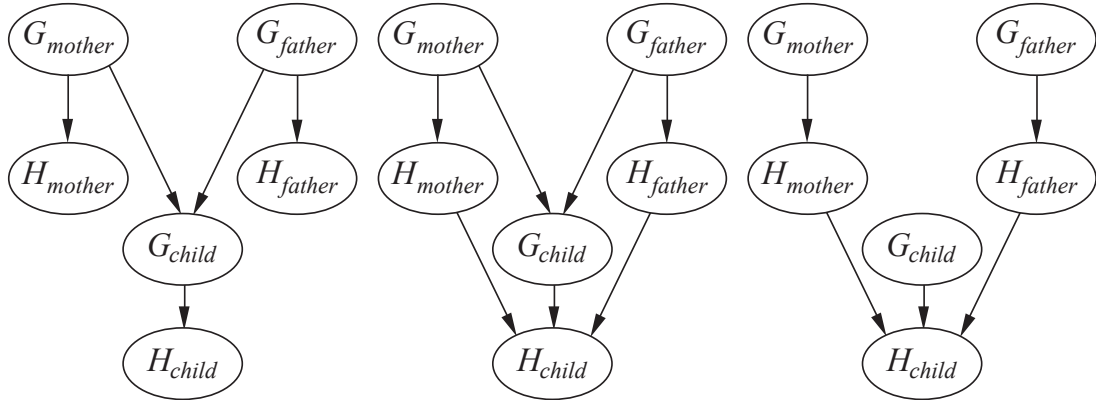
   The third one, as the node G_{child} does not have any parents.

2. Which of the networks make independence claims that are consistent with the hypothesis about the inheritance of handedness ?

   The first and second ones. It should be noted that the second one is also valid as it makes *less* assumptions about the variables dependencies, *i.e.* it is more general than the first one.

3. Write down the CPT for the G_{child} nodes in first network, in terms of s and m.

G_{mother}	G_{father}	$P(G_{child} = l \mid G_{mother}, G_{father})$
l	l	1 - m
l	r	0.5
r	l	0.5
r	r	m

4. Suppose that $P(G_{mother} = l) = P(G_{father} = l) = q$. In the first network, derive an expression for $P(G_{child} = l)$ in terms of s, m and q.

$$P(G_{child} = l) = \sum_{g_m} \sum_{g_f} P(G_{child} = l, g_m, g_f)$$
$$= \sum_{g_m} \sum_{g_f} P(G_{child} = l \mid g_m, g_f) P(g_m) P(g_f)$$
$$= (1 - m)q^2 + 0.5(1 - q)q + 0.5q(1 - q) + m(1 - q)^2$$
$$= q + m - 2qm$$

5. Under conditions of genetic equilibrium, we expect the distribution of genes to be the same across generations. Use this to calculate the value of q, and, given what you know

about handedness in humans, explain why the hypothesis described at the beginning of this question must be wrong.

Under genetic equilibrium, we have $P(G_{child} = l) = P(G_{parent} = l) = q$. Therefore, $q = q + m - 2qm$ meaning that $q = 0.5$. In practice, being left-handed is much more rare than being right-handed which invalidates the initial hypothesis.

Exercise 7 Independence
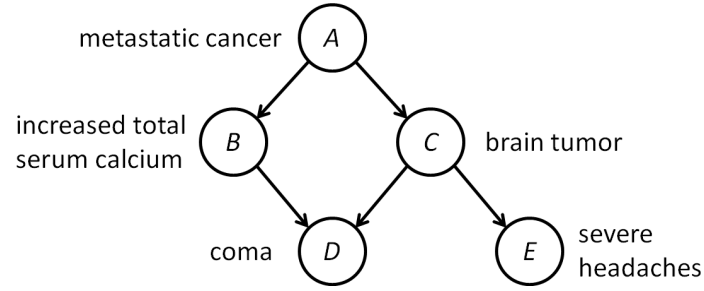
Considering the hereabove Bayesian network, which of the following statements are asserted, *i.e.* enforced, by the network structure ?

We apply the d-separation[2] algorithm. To show that two variables X and Y could be dependent, it is sufficient to find a single *active* (undirected) path from X to Y. A path is active if all of its consecutive triples are active.

1. $P(B,C) = P(B)P(C)$

   This is true iff (if and only if) $B \perp C$ (B is independent from C). The two paths that link B and C are (B,A,C) and (B,D,C). (B,A,C) is active because A is unknown ($\wedge$-structure). We cannot guarantee that $B \perp C$.

2. $P(B,C \mid A) = P(B \mid A)P(B \mid A)$

   This is true iff $B \perp C \mid A$. This time, (B,A,C) is not active because A is known. (B,D,C) is not active either because D is unknown ($\vee$-structure). We can guarantee that $B \perp C \mid A$.

3. $P(B,C \mid A,D) = P(B \mid A,D)P(B \mid A,D)$

   This is true iff $B \perp C \mid A,D$. This time, (B,D,C) is active because D is known. We cannot assert that $B \perp C \mid A,D$.

4. $P(B,E \mid A) = P(B \mid A)P(E \mid A)$

   This is true iff $B \perp E \mid A$. We know that, knowing A, no path between B and C are active. As C is the only parent of E, all paths between B and E go trough C, meaning that they are not active. We can guarantee that $B \perp E \mid A$.

5. $P(C \mid A,D,E) = P(C \mid A,B,D,E)$

   This is true iff $B \perp C \mid A,D,E$. Once again, (B,D,C) is active because D is known. We can cannot guarantee that $B \perp C \mid A,D,E$.

6. $P(B,E) = \sum_{a,c,d} P(E \mid c)P(d \mid B,c)P(c \mid a)P(B \mid a)P(a)$

   This comes down to evaluate if

   $$P(A,B,C,D,E) = P(E \mid C)P(D \mid B,C)P(C \mid A)P(B \mid A)P(A),$$

   is guaranteed. This exactly corresponds to

   $$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i)),$$

   for the considered network, which we know is guaranteed.

---

[2]Slides 64 and 65, Lecture 4.

For the same network, use inference by variable elimination to compute $P(E \mid A = 1, B = 1)$.

We have

$$P(E \mid A = 1, B = 1) \propto \sum_{c,d} P(A = 1, B = 1, c, d, E)$$

$$\propto \sum_{c,d} P(A = 1)P(B = 1|A = 1)P(c|A = 1)P(d|B = 1, c)P(E|c)$$

$$\propto P(A = 1)P(B = 1|A = 1)\sum_c P(c|A = 1)P(E|c)\sum_d P(d|B = 1, c).$$

We define the initial factors as $f_1 = P(A = 1)$, $f_2 = P(B = 1|A = 1)$, $f_3(C) = P(C|A = 1)$, $f_4(C, D) = P(D|B = 1, C)$, $f_5(E, C) = P(E|C)$.

Then we compute $f_6(C) = \sum_d f_4(C, d)$, then $f_7(C, E) = f_3(C) \times f_5(E, C) \times f_6(C)$ and $f_8(E) = \sum_c f_7(c, E)$, then $f_9(E) = f_1 \times f_2 \times f_7(E)$. Finally

$$P(E \mid A = 1, B = 1) = \frac{f_9(E)}{\sum_e f_9(e)}.$$

Exercise 8 Car Diagnosis (AIMA, Ex: 14.8)

Let be the following Bayesian network describing some features of a car's electrical system and engine. Each variable is Boolean, and the true value indicates that the corresponding aspect of the vehicle is in working order.
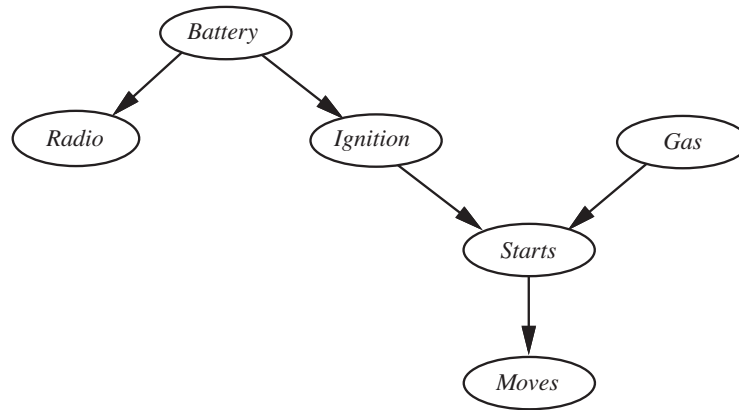
1. Extend the network with the Boolean variables IcyWeather and StarterMotor.

   IcyWeather is not caused by any of the car-related variables, so needs no parents. It directly affects the battery and the starter motor. StarterMotor is an additional precondition for Starts.
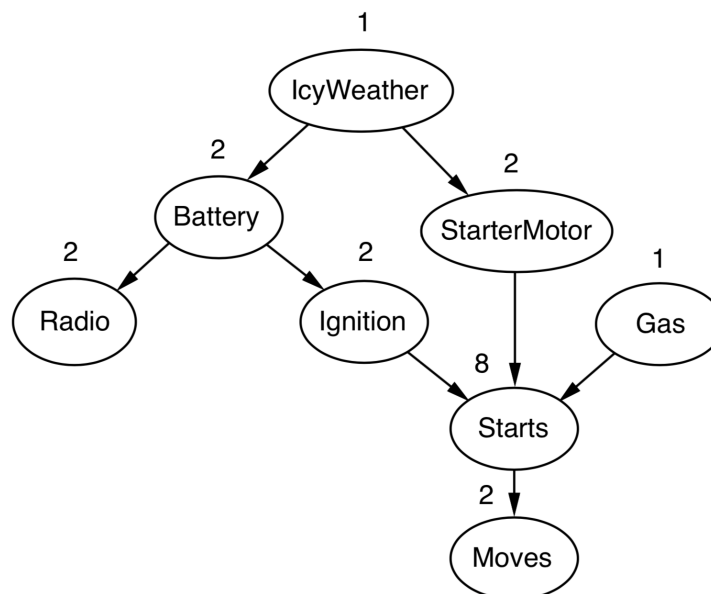
2. Give reasonable conditional probability tables for all the nodes.

   Reasonable probabilities may vary a lot depending on the kind of car and perhaps the personal experience of the assessor. The following values indicate the general order of magnitude and relative values that would be reasonable:

   - A reasonable prior for IcyWeather might be 0.05 (depending on location and season).
   - $P(Battery|IcyWeather) = 0.95, P(Battery|\neg IcyWeather) = 0.997$.
   - $P(StarterMotor|IcyWeather) = 0.98, \ P(Battery|\neg IcyWeather) = 0.999$.
   - $P(Radio|Battery) = 0.9999, P(Radio|\neg Battery) = 0.05$.
   - $P(Ignition|Battery) = 0.998, P(Ignition|\neg Battery) = 0.01$.

- $P(Gas) = 0.995$.

- $P(Starts|Ignition, StarterMotor, Gas) = 0.9999, other entries 0.0$.

- $P(Moves|Starts) = 0.998$.

3. How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming that no conditional independence relations are known to hold among them?

   With 8 Boolean variables, the joint has $2^8 - 1 = 255$ independent entries.

4. How many independent probability values do your network tables contain?

   Given the topology with IcyWeather and StarterMotor, the total number of independent CPT entries is $1 + 2 + 2 + 2 + 2 + 1 + 8 + 2 = 20$.

Exercise 9   Nuclear Power Plant (AIMA, Ex: 14.11)

In your local nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables A (alarm sounds), F_A (alarm is faulty), and F_G (gauge is faulty) and the multi-valued nodes G (gauge reading) and T (actual core temperature).

1. Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.
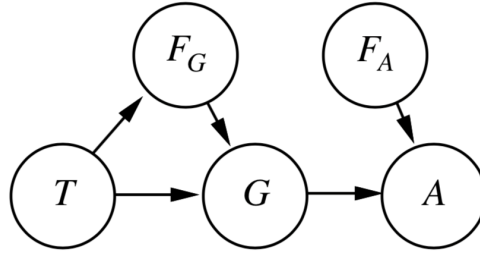
2. Suppose there are just two possible actual and measured temperatures: low (l) and high (h). The probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Give the conditional probability table associated with G.

   The CPT for G is shown below. Students should pay careful attention to the semantics of F_G, which is true when the gauge is faulty, *i.e.* not working properly.

T	F_G	$P(G = l \mid T, F_G)$
l	0	x
l	1	y
h	0	$1 - x$
h	1	$1 - y$

3. Suppose the alarm is always triggered by high measured temperatures, unless it is faulty, in which case it never sounds. Give the conditional probability table associated with A.

G	F_A	$P(A = 1 \mid G, F_A)$
l	0	0
l	1	0
h	0	0
h	1	1

4. Suppose the gauge is not faulty and the alarm is triggered. Calculate an expression for the probability that the temperature of the core is too high, in terms of the various conditional probabilities in the network.

   The probability of interest here is $P(T = h \mid A = 1, F_G = 0)$. Because the alarm's behavior is deterministic, we can deduce that G indicates high (h). Hence, $P(T = h \mid A = 1, F_G = 0) = P(T = h \mid A = 1, G = h, F_G = 0)$. We also see that $A \perp T \mid G$. Therefore, we only need to calculate

   $$P(T = h \mid G = h, F_G = 0) = \frac{P(T = h, G = h, F_G = 0)}{P(G = h, F_G = 0)}$$
   $$= \frac{P(G = h \mid T = h, F_G = 0)P(F_G = 0 \mid T = h)P(T = h)}{\sum_t P(G = h \mid t, F_G = 0)P(F_G = 0 \mid t)P(t)},$$

   which we cannot develop more as we don't know $P(T)$ and $P(F_G \mid T)$.

Supplementary materials

- Hygiène Mentale – La Pensée Bayésienne

- The Book of Why

- Chapters 13 and 14 of the reference textbook.