

# Introduction to Artificial Intelligence (INFO8006)

## Exercise session 3

### Kolmogorov's axioms

A probability space is a sample space equipped with a probability function, i.e. an assignment  $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  such that

1.  $P(\omega) \in \mathbb{R}, 0 \leq P(\omega)$  for all  $\omega \in \Omega$ .
2.  $P(\Omega) = 1$ .
3.  $P(\{\omega_1, \dots, \omega_n\}) = \sum_{i=1}^n P(\omega_i)$ .

with  $\Omega$  a **sample space**,  $\omega \in \Omega$  a **sample point** and  $\mathcal{P}(\Omega)$  the power set of  $\Omega$ .

### Probability rules

The **product rule** states that

$$P(a, b) = P(b)P(a \mid b) = P(a)P(b \mid a)$$

from which **Bayes' formula** is derived as

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)}$$

with  $P(a, b)$  being the joint probability of events  $a$  and  $b$ ,  $P(b)$  -  $P(a)$  the marginals and  $P(a \mid b)$  -  $P(b \mid a)$  the conditionals.

The **chain rule** states that

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid x_1, \dots, x_{i-1}).$$

**In session exercises:** Ex. 1, Ex. 2, Ex. 3, Ex. 5

## Exercise 1 Beliefs

Let  $A$  and  $B$  be two *events* over the probability space  $\Omega$ . 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  $P(A|B) = 0.5$ ?

1. The agent's beliefs should satisfy

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

This can be illustrated with Venn diagrams.

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

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

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.  $P(\text{toothache})$

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

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

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

3.  $P(\text{toothache}|\text{cavity})$

$$\begin{aligned} P(T = 1|CV = 1) &= \frac{P(T = 1, CV = 1)}{P(CV = 1)} = \frac{\sum_{ct} P(T = 1, ct, CV = 1)}{P(CV = 1)} \\ &= \frac{0.108 + 0.012}{0.2} = 0.6 \end{aligned}$$

4.  $P(\text{cavity}|\text{toothache} \cup \text{catch})$

$$\begin{aligned} P(CV = 1|T = 1 \cup CT = 1) &= \frac{P(T = 1 \cup CT = 1, CV = 1)}{P(T = 1 \cup CT = 1)} \\ &= \frac{P(CV = 1) - P(T = 0, CT = 0, CV = 1)}{1 - P(T = 0, CT = 0)} \\ &= \frac{0.2 - 0.008}{1 - 0.584} = 0.4615 \end{aligned}$$

### 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,

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

## Exercise 4 (AIMA, Ex 13.13)

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

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

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

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

False. The statement  $P(a|b, c) = P(a)$  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  $P(a|b) = P(a)$ , then  $P(a|b, c) = P(a|c)$ .

False. While the statement  $P(a|b) = P(a)$  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

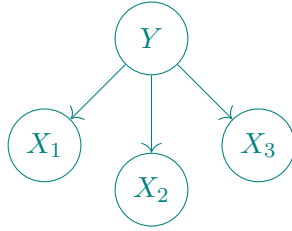
We have a bag of three biased coins  $a$ ,  $b$ , and  $c$  with probabilities of coming up heads of 20 %, 60 %, and 80 %, respectively. One coin is drawn randomly from the bag (with equal probability of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes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



$Y$	$P(Y)$	$P(X_i = h Y)$
$a$	$\frac{1}{3}$	0.2
$b$	$\frac{1}{3}$	0.6
$c$	$\frac{1}{3}$	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  $P(Y|X_1 = h, X_2 = h, X_3 = t)$ <sup>1</sup>. Thanks to Bayes rule, we have

$$P(Y|X_1 = h, X_2 = h, X_3 = t) = \alpha P(X_1 = h, X_2 = h, X_3 = t|Y)P(Y).$$

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

$$P(X_1, X_2, X_3|Y) = P(X_1|Y)P(X_2|Y)P(X_3|Y).$$

Then, we have

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

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.

---

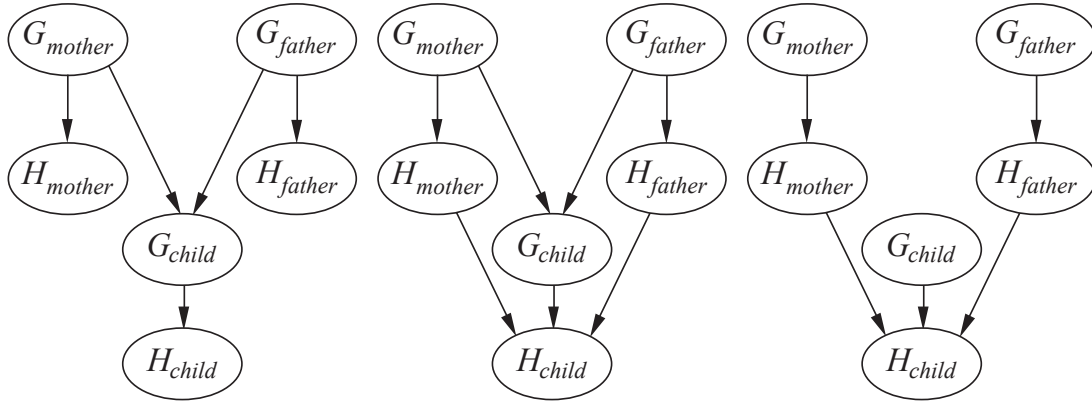
<sup>1</sup>It should be noted that the order of the tosses is irrelevant as they are performed independently.

## 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  $P(G_{child}|G_{mother}, G_{father}) = P(G_{child})$  ?



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  $m$ .

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

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  $q$ .

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

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.



## Quiz

The Bayes' rule states that ...

☐  $P(x | y) = \frac{P(y|x)P(y)}{P(x)}.$

☒  $P(x | y) = \frac{P(y|x)P(x)}{P(y)}.$

☐  $P(y | x) = \frac{P(x|y)P(x)}{P(y)}.$

☐  $P(x) = \frac{P(x|y)P(y)}{P(x)}.$

Which of the following is true?

☐  $P(x, y | z) = \frac{P(x|y,z)}{P(y,z|x)}.$

☐  $P(x, y, z) = P(x)P(y)P(z).$

☒  $P(x, y | z) = P(y | x, z)P(x | z).$

☐  $P(x | y, z) = \frac{P(x,y,z)}{P(y)P(z)}.$

Naive Bayes model assumes ...

☒ pairwise conditional independence between effects given the cause.

☐ pairwise independence between effects.

☐ independence between effects and cause.

☐ conditional independence between each effect and the cause, given another effect.

The Bayes' rule can be read as "The posterior is equal to ...

☐ the product of the likelihood and the prior".

☐ the ratio between the posterior distribution and the evidence, times the prior".

☐ the likelihood to prior ratio multiplied by the evidence".

☒ the likelihood to evidence ratio multiplied by the prior".