Artificial Intelligence

SCALAB
Grupo de Inteligencia Artificial

Universidad Carlos III de Madrid

2017-2018

Probabilistic Reasoning – Exercises

Exercise 1: full joint distribution

A dentist registered the following information for a total of 1000 patients. These data relates toothache, cavity and catch (the dentist's nasty steel probe catches in the tooth)

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	108	12	72	8
¬ cavity	16	64	144	576

- What is the full joint distribution?
- What is the probability that a patient has a cavity, $P(\text{cavity})$?
- What is the probability of a catch and the patient has not toothache, $P(\neg\text{toothache}, \text{catch})$?
- We know a patient has toothache. What is the probability of cavity?

Exercise 1 – solution

A dentist registered the following information for a total of 1000 patients. These data relates toothache, cavity and catch (the dentist's nasty steel probe catches in the tooth)

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- What is the full joint distribution? $P(\mathrm{Cavity}, \mathrm{Toothache}, \mathrm{Catch})$
- What is the probability that a patient has a cavity, $P(\mathrm{cavity})$?

$$P(\mathrm{cavity}) = \sum_{\substack{Toothache,\\ Catch}} P(\mathrm{Cavity}, \mathrm{Toothache}, \mathrm{Catch}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

- What is the probability of a catch and the patient has not toothache, $P(\neg \mathrm{toothache}, \mathrm{catch})$?

$$P(\neg \mathrm{toothache}, \mathrm{catch}) = \sum_{Cavity} P(\mathrm{Cavity}, \neg \mathrm{Toothache}, \mathrm{Catch}) = 0.072 + 0.144 = 0.216$$

Exercise 1 – solution

A dentist registered the following information for a total of 1000 patients. These data relates toothache, cavity and catch (the dentist's nasty steel probe catches in the tooth)

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- We know a patient has toothache. What is the probability of cavity?

Exercise 1 – solution

A dentist registered the following information for a total of 1000 patients. These data relates toothache, cavity and catch (the dentist's nasty steel probe catches in the tooth)

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- We know a patient has toothache. What is the probability of cavity?

$$P(\text{cavity}/\text{toothache}) = \frac{P(\text{cavity}, \text{toothache})}{P(\text{toothache})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Exercise 1 – solution

A dentist registered the following information for a total of 1000 patients. These data relates toothache, cavity and catch (the dentist's nasty steel probe catches in the tooth)

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- We know a patient has toothache. What is the probability of cavity?

$$P(\text{cavity}/\text{toothache}) = \frac{P(\text{cavity}, \text{toothache})}{P(\text{toothache})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Normalization: $P(\text{toothache})$ is not needed

Exercise 1 – solution

A dentist registered the following information for a total of 1000 patients. These data relates toothache, cavity and catch (the dentist's nasty steel probe catches in the tooth)

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- We know a patient has toothache. What is the probability of cavity?

$$P(\text{cavity}/\text{toothache}) = \frac{P(\text{cavity}, \text{toothache})}{P(\text{toothache})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Normalization: $P(\text{toothache})$ is not needed

$$\begin{aligned} P(\text{Cavity}/\text{Toothache}) &= \alpha P(\text{Cavity}, \text{Toothache}) \\ &= \alpha[P(\text{Cavity}, \text{toothache}, \text{catch}) + P(\text{Cavity}, \text{toothache}, \neg\text{catch})] \\ &= (0.6, 0.4) \end{aligned}$$

Exercise 2:

- 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?

Exercise 2: Solution

- Step 1: State notation (random variables):

Exercise 2: Solution

- Step 1: State notation (random variables):
  - Test: Boolean variable
    - test: tested positive
      ¬test: tested negative
  - Disease: Boolean variable
    - disease: you do have the disease
    - ¬disease: you do not have the disease
- Step 2: State the probabilities known

Exercise 2: Solution

- Step 1: State notation (random variables):
  - Test: Boolean variable
    - test: tested positive
      ¬test: tested negative
  - Disease: Boolean variable
    - disease: you do have the disease
    - ¬disease: you do not have the disease
- Step 2: State the probabilities known
  - Test=test
  - $P(test|disease)=0.99$
  - $P(\neg test|\neg disease)=0.99$
  - $P(disease)=0.0001$

Exercise 2: Solution

- Step 1: State notation (random variables):
  - Test: Boolean variable
    - test: tested positive
      ¬test: tested negative
  - Disease: Boolean variable
    - disease: you do have the disease
    - ¬disease: you do not have the disease
- Step 2: State the probabilities known
  - Test=test
  - $P(test|disease)=0.99$
  - $P(\neg test|\neg disease)=0.99$
  - $P(disease)=0.0001$
- Note that we also know:
  - $P(\neg disease)=0.9999$, $P(\neg test|disease)=0.01$, $P(test|\neg disease)=0.01$

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive
- We want P(disease|test)

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive
- We want P(disease|test)

$$P(\text{disease}|\text{test}) = \frac{P(test|disease)P(disease)}{P(test)}$$

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive
- We want P(disease|test)

  $$\text{P(disease|test)} = \frac{P(test|disease)P(disease)}{P(test)}$$

- We need P(test)?

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive
- We want P(disease|test)

  $$\text{P(disease}|\text{test)} = \frac{P(test|disease)P(disease)}{P(test)}$$

- We need P(test)?

  $$\begin{aligned}\text{P(test)} &= \text{P(test}|\text{disease)P(disease)} + \text{P(test}|\neg\text{disease)P(}\neg\text{disease)}\\ &= (0.99)(0.0001) + (0.01)(0.9999)\\ &= 0.010098\end{aligned}$$

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive
- We want P(disease|test)

$$P(\text{disease}|\text{test}) = \frac{P(test|disease)P(disease)}{P(test)}$$

- We need P(test)?

$$\begin{aligned} P(\text{test}) &= P(\text{test}|\text{disease})P(\text{disease}) + P(\text{test}|\neg\text{disease})P(\neg\text{disease}) \\ &= (0.99)(0.0001) + (0.01)(0.9999) \\ &= 0.010098 \end{aligned}$$

- We go back to computing P(disease|test)

Exercise 2: Solution

- What we want to know is the probability that you actually have the disease given that you tested positive
- We want P(disease|test)

$$P(\text{disease}|\text{test}) = \frac{P(test|disease)P(disease)}{P(test)}$$

- We need P(test)?

$$\begin{aligned} P(\text{test}) &= P(\text{test}|\text{disease})P(\text{disease}) + P(\text{test}|\neg\text{disease})P(\neg\text{disease}) \\ &= (0.99)(0.0001) + (0.01)(0.9999) \\ &= 0.010098 \end{aligned}$$

- We go back to computing P(disease|test)

$$\begin{aligned} P(\text{disease}|\text{test}) &= \frac{P(test|disease)P(disease)}{P(test)} \\ &= \frac{(0.99)(0.0001)}{0.010098} \\ &= 0.009804 \end{aligned}$$

We do not really need P(test) since it can be seen as a normalization factor

Exercise 3:

- In Barcelona, 51 % of the adults are males. (It doesn't take too much advanced mathematics to deduce that the other 49 % are females.) One adult is randomly selected for a survey involving credit card usage.
  1. Find the prior probability that the selected person is a male.
  2. It is later learned that the selected survey subject was smoking a cigar. Also, 9.5 % of males smoke cigars, whereas 1.7 % of females smoke cigars (based on data from the Substance Abuse and Mental Health Services Administration). Use this additional information to find the probability that the selected subject is a male.

Exercise 3: Solution

- Let's use the following notation:
  - *M* propositional:
    - m = male
    - ¬m = female (or not male)
  - *C* propositional:
    - c = cigar smoker
    - ¬c = not a cigar smoker.

1. Before using the information given in part b, we know only that 51 % of the adults in barcelona are males, so the probability of randomly selecting an adult and getting a male is given by P(m) = 0.51.

Exercise 3: Solution

- Based on the additional given information, we have the following:
  - $P(m) = 0.51$ because 51 % of the adults are males
  - $P(\neg m) = 0.49$ because 49 % of the adults are females (not males)
  - $P(c|m) = 0.095$ because 9.5 % of the males smoke cigars (That is, the probability of getting someone who smokes cigars, given that the person is a male, is 0.095.)
  - $P(c|\neg m) = 0.017$. because 1.7 % of the females smoke cigars (That is, the probability of getting someone who smokes cigars, given that the person is a female, is 0.017.)

Exercise 3: Solution

- We want to compute P(m|c)

Exercise 3: Solution

- We want to compute P(m|c)

  P(m|c) = $\frac{P(c|m)P(m)}{P(c)}$

Exercise 3: Solution

- We want to compute P(m|c)

$$P(m|c) = \frac{P(c|m)P(m)}{P(c)}$$
$$\frac{(0.095)(0.51)}{P(c)}$$
$$\frac{(0.095)(0.51)}{P(c|m)P(m)+P(c|\neg m)P(\neg m))}$$
$$\frac{(0.095)(0.51)}{(0.095)(0.51)+(0.017)(0.49)}$$
$$\frac{0.04845}{0.05678}$$
$$= 0.85329341$$

- Before we knew that the survey subject smoked a cigar, there is a 0.51 probability that the survey subject is male (because 51 % of the adults in Barcelona are males). However, after learning that the subject smoked a cigar, we revised the probability to 0.853. There is a 0.853 probability that the cigar -smoking respondent is a male. This makes sense, because the likelihood of a male increases dramatically with the additional information that the subject smokes cigars (because so many more males smoke cigars than females).

Exercise 4

- Company A makes 80 % of a product, the Company B makes 15 % of them, and Company C makes the other 5 %. The products made by Company A have a 4 % rate of defects, the Company B products have a 6 % rate of defects, and the Company C products have a 9 % rate of defects (which helps to explain why Company C has the lowest market share).
    1. If a product is randomly selected from the general population of all products, find the probability that it was made by the Company A.
    2. If a randomly selected product is then tested and is found to be defective, find the probability that it was made by the Company A.

Exercise 4: Solution

- We use the following notation:
  - Company: discrete variable
    - a = product manufactured by Company A
    - b = product manufactured by Company B
    - c = product manufactured by Company C
  - D: defective: propositional
    - d = product is defective
    - ¬d = product is not defective (or it is good)

Exercise 4: Solution

1. P(a)=?

Exercise 4: Solution

1. P(a)=?
   If a product is randomly selected from the general population of all products, the probability that it was made by Company A is 0.8 (because Company A manufactures 80 % of them).
   P(a)=0.8

Exercise 4: Solution

- State all probabilities known:
  - P(a)=0.80
  - Pb)= 0.15
  - P(c)=0.05
  - P(d|a)=0.04
  - P(d|b)=0.06
  - P(d|c)=0.09
- What we want is:

Exercise 4: Solution

- State all probabilities known:
  - P(a)=0.80
  - Pb)= 0.15
  - P(c)=0.05
  - $P(d|a)$=0.04
  - $P(d|b)$=0.06
  - $P(d|c)$=0.09
- What we want is:$P(a|d)$

Exercise 4: Solution

- State all probabilities known:
  - P(a)=0.80
  - Pb)= 0.15
  - P(c)=0.05
  - P(d|a)=0.04
  - P(d|b)=0.06
  - P(d|c)=0.09
- What we want is:P(a|d)

$$P(a|d) = \frac{P(d|a)P(a)}{P(d)}$$

Exercise 4: Solution

- State all probabilities known:
  - P(a)=0.80
  - Pb)= 0.15
  - P(c)=0.05
  - P(d|a)=0.04
  - P(d|b)=0.06
  - P(d|c)=0.09
- What we want is:P(a|d)

$$P(a|d) = \frac{P(d|a)P(a)}{P(d)}$$
$$= \frac{P(d|a)P(a)}{P(d|a)P(a)+P(d|b)(b)+P(d|c)P(c)}$$
$$= \frac{(0.04)(0.80)}{(0.04)(0.80)+(0.06)(0.15))+(0.09))(0.05)}$$
$$= 0.703$$

Exercise 5

- A couple has four children. The general probability of having a male child is 50 %.
  1. What is the probability that the four children are the same sex?
  2. What is the probability that only two children are men?
  3. What is the probability that at least two children are male?

Exercise 5: Solution

- State all possibilities

Exercise 5: Solution

- State all possibilities

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. What is the probability that the four children are the same sex?

Exercise 5: Solution

- State all possibilities

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. What is the probability that the four children are the same sex? 2/16
2. What is the probability that only two children are men?

Exercise 5: Solution

- State all possibilities

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. What is the probability that the four children are the same sex? 2/16
2. What is the probability that only two children are men? 6/16
3. What is the probability that at least two children are male?

Exercise 5: Solution

- State all possibilities

MMMM	MMMF	MMFM	MMFF
MFMM	MFMF	MFFM	MFFF
FMMM	FMMF	FMFM	FMFF
FFMM	FFMF	FFFM	FFFF

1. What is the probability that the four children are the same sex? 2/16
2. What is the probability that only two children are men? 6/16
3. What is the probability that at least two children are male? 11/16