

# U1.1 Probabilistic reasoning: Representation and inference

## Index

1. The qualification problem
2. Probabilistic representation
3. Probabilistic inference
4. Independence
5. Bayes' theorem

# 1 The qualification problem

**Qualification problem:** Practical impossibility of knowing and checking all the **qualifications** (conditions) that would need to be guaranteed in order to ensure the fulfillment of an action

- Example: Leaving the airport 90 minutes before the flight allows me to arrive on time IF there are no traffic jams AND no flat tyres AND ...
- Example: A boat allows us to cross a river IF it is a rowing boat AND it has oars and rowlocks AND they are not broken AND they fit AND ...

**Uncertainty:** Current intelligent systems include **uncertainty** as part of knowledge and represent it by **probabilities** associated with the events (propositions) of interest

## 2 Probabilistic representation

**Joint Probability Distribution:** of the random variables of interest in order to represent probabilistic knowledge

**Example of the dentist:** Knowledge to diagnose caries

*Random variables of interest:*

Toothache:  $T \in \{0, 1\}$   
Caries:  $C \in \{0, 1\}$   
Hole:  $H \in \{0, 1\}$

*Representation:* table below with

$$P(T = t, C = c, H = h) \quad \text{for all } t, c, h \in \{0, 1\}$$

$t$	$c$	$h$	$P$
0	0	0	0.576
0	0	1	0.008
0	1	0	0.144
0	1	1	0.072
1	0	0	0.064
1	0	1	0.012
1	1	0	0.016
1	1	1	0.108
			1.000

# 3 Probabilistic inference

**Sum and product rules:** Basic rules for calculating the probability of any **event (proposition)** of interest from the joint distribution

$$P(x) = \sum_y P(x, y) \quad \text{and} \quad P(x, y) = P(x) P(y | x)$$

**Important note:** In general it is not necessary to know the full table of joint probabilities to calculate the probability of a given event using the sum and product rules

**Dentist example:** Calculation of the probability of observing...

- Caries and observing a hole (at the same time):  $P(c = 1, h = 1) = \sum_{t=0,1} P(t, c = 1, h = 1) = 0.180$
- Hole:  $P(h = 1) = \sum_{t=0,1} \sum_{c=0,1} P(t, c, h = 1) = 0.200$
- Decay after observing a hole:  $P(c = 1 | h = 1) = \frac{P(c = 1, h = 1)}{P(h = 1)} = \frac{0.180}{0.200} = 0.900$

```
In [2]: import numpy as np
T = np.array([[0,0,0,.576], [0,0,1,.008], [0,1,0,.144], [0,1,1,.072],
              [1,0,0,.064], [1,0,1,.012], [1,1,0,.016], [1,1,1,.108]])
Pc1b1 = np.sum(T[(T[:,1]==1) & (T[:,2]==1)], -1)
Pb1 = np.sum(T[T[:,2]==1, -1])
Pc1Db1 = Pc1b1/Pb1
print(f"Pc1b1 = {Pc1b1:.3f}  Pb1 = {Pb1:.3f}  Pc1Db1 = {Pc1Db1:.3f}")

Pc1b1 = 0.180  Pb1 = 0.200  Pc1Db1 = 0.900
```

# 4 Independence

**Independent variables:** Two variables  $x$  and  $y$  are **independent** if

$$P(x, y) = P(x) P(y) \quad \text{or} \quad P(x | y) = P(x) \quad \text{or} \quad P(y | x) = P(y)$$

**Expert knowledge:** independence can be established by expert knowledge and convenience

**Dentist example:**

- We consider a new variable the weather when the patient visits the dentist

$$W \in \{\text{sun, clouds, rain, snow}\}$$

- Let's assume that the three variables we already had are independent of the weather

$$P(t, c, h, w) = P(w) P(t, c, h | w) = P(w) P(t, c, h)$$

- This is how we reduce the number of probabilities to store: 32 vs 4 + 8

# 5 Bayes' theorem

**Bayes' theorem:** It allows us to update our knowledge about a hypothesis  $y$  after observing new evidence  $x$

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

- In other words:  $P(y | x)$  is the probability that the effect  $y$  will occur after observing that the cause  $x$  has occurred

**Dentist example:**

- We know that the probability of caries is:  $P(c = 1) = 0.34$
- We know that the probability of toothache is:  $P(t = 1) = 0.20$
- We know that the probability of toothache after observing caries is:  $P(t = 1 | c = 1) = 0.36$
- What is the probability of caries after observing toothache,  $P(c = 1 | t = 1)$ ?

$$P(c = 1 | t = 1) = P(c = 1) \frac{P(t = 1 | c = 1)}{P(t = 1)} = 0.34 \frac{0.36}{0.20} = 0.61$$

```
In [3]: Pc1 = 0.34; Pd1 = 0.20; Pd1c1 = 0.36; Pc1Dd1 = Pc1 * Pd1c1 / Pb1; print(f"Pc1Dd1 = {Pc1Dd1:.2f}")
Pc1Dd1 = 0.61
```

# T1.1 Probabilistic reasoning: representation and inference

**2023\_01\_26\_Question 1:** Given the following table of joint probabilities for three variables of interest:

$A$	0	0	0	0	1	1	1	1
$B$	0	0	1	1	0	0	1	1
$C$	0	1	0	1	0	1	0	1
$P(A, B, C)$	0.093	0.100	0.133	0.163	0.157	0.150	0.117	0.087

What is the value of  $P(A = 1, B = 1 \mid C = 1)$ ?

1.  $P(A = 1, B = 1 \mid C = 1) \leq 0.25$
2.  $0.25 < P(A = 1, B = 1 \mid C = 1) \leq 0.50$
3.  $0.50 < P(A = 1, B = 1 \mid C = 1) \leq 0.75$
4.  $0.75 < P(A = 1, B = 1 \mid C = 1)$

**Solution:** Option 1;  $P(A = 1, B = 1 \mid C = 1) = 0.174$



**2023\_01\_17\_Question 1:** Suppose we have two boxes with 40 oranges in the first and 80 in the second. The first box contains 9 Navelina and 31 Caracara oranges. The second box contains three times more Navelina oranges than Caracara. Now suppose a box is chosen at random, and then an orange is chosen at random from the chosen box. If the orange chosen is Navelina, the probability  $P$  that it comes from the first box is:

1.  $0/4 \leq P < 1/4$ .
2.  $1/4 \leq P < 2/4$ .
3.  $2/4 \leq P < 3/4$ .
4.  $3/4 \leq P \leq 4/4$ .

**Solution:**

$$\begin{aligned} P = P(C = 1 \mid T = N) &= \frac{P(C = 1)P(T = N \mid C = 1)}{P(C = 1)P(T = N \mid C = 1) + P(C = 2)P(T = N \mid C = 2)} \\ &= \frac{40/120 \cdot 9/40}{40/120 \cdot 9/40 + 80/120 \cdot 3/4} = 0.13 \end{aligned}$$

**2022\_01\_27\_Question 4:** Given the following table of joint frequencies of three variables of interest:

$A$	0	0	0	0	1	1	1	1
$B$	0	0	1	1	0	0	1	1
$C$	0	1	0	1	0	1	0	1
$N(A, B, C)$	124	28	227	175	126	222	23	75

What is the value of  $P(A = 1 \mid B = 1, C = 0)$ ?

1. 0.023
2. 0.250
3. 0.092
4. 0.446

**Solution:** Option 3.

**2022\_01\_13\_Question 7:** Follow a probabilistic reasoning problem about road trips, with the random variables of interest:

- Climatology ( $C$ ): {clear (CLE), cloudy (CLO), rain (RAI)}
- Brightness ( $B$ ): {day (DAY), night (NIGHT)}
- Safety ( $S$ ): {safe (SAF), accident (ACC)}

The joint probability of the three variables is given in the table:

$P(S, B, C)$	DAY			NIGHT		
	CLE	CLO	RAI	CLE	CLO	RAI
SAF	0.27	0.23	0.07	0.16	0.07	0.06
ACC	0.02	0.01	0.02	0.02	0.03	0.04

The conditional probability  $P(S = \text{ACC} \mid B = \text{DAY}, C = \text{CLO})$  is:

1. 0.042
2. 0.010
3. 0.240
4. 0.140

**Solution:** Option 1.

# U1.2 Continuous variables and Bayes' rule

## Index

1. Continuous variables
2. Bayes' theorem in the continuous case
3. Bayes' decision rule
4. Generative and discriminative classifiers

# 1 Continuous variables

**Probability density function:** Usual characterization of continuous variables for the representation of probabilistic knowledge

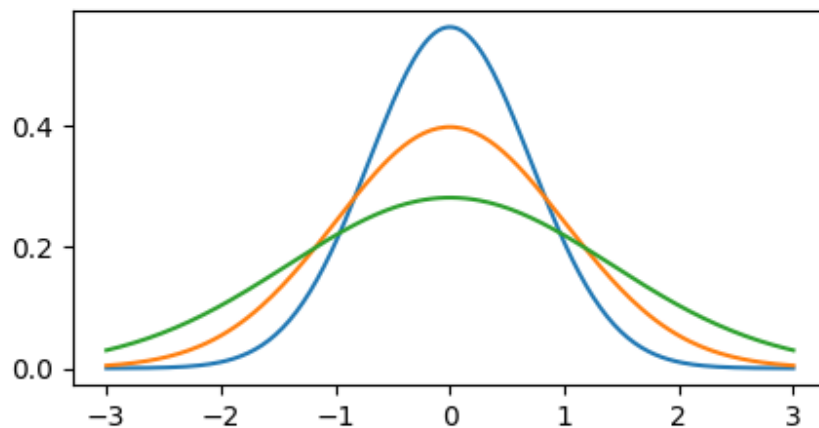
$$p(x) \geq 0 \quad \text{for all } x \quad \text{and} \quad \int p(x) dx = 1$$

**The normal density:**  $p(x) \sim \mathcal{N}(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad P(x \in [\mu \pm 1.96\sigma]) = 0.95$$

**Example:** normal densities with  $\mu = 0$  and  $\sigma^2 = 0.5, 1, 2$

```
In [1]: import numpy as np; from scipy.stats import norm; import matplotlib.pyplot as plt
x = np.linspace(-3, 3, 200)
plt.figure(figsize=(5, 2.5))
plt.plot(x, norm.pdf(x, 0, np.sqrt(0.5))), x, norm.pdf(x, 0, 1), x, norm.pdf(x, 0, np.sqrt(2)));
```





## 2 Bayes theorem in the continuous case

**Bayes theorem in the continuous case:** Probability of a hypothesis  $y$  after observing (new) evidence  $x$

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

**Example:**  $x$  = result of a saliva test for caries diagnosis

- Without caries,  $c = 0$ ,  $p(x | c = 0) \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$
- With caries,  $c = 1$ ,  $p(x | c = 1) \sim \mathcal{N}(\mu = 2, \sigma^2 = 0.5)$
- We know that the probability (a priori) of caries is:  $P(c = 1) = 0.34$
- **If the test gives  $x = 2$ , what is the (posterior) probability of caries?**

$$P(c = 1 | x = 2) = P(c = 1) \frac{p(x = 2 | c = 1)}{p(x = 2)} = 0.340 \frac{0.227}{0.307} = 0.843$$

- Note that first it was necessary to find the (density of) probability (a priori) of test  $x = 2$ :

$$\begin{aligned} p(x = 2) &= P(c = 0)p(x = 2 | c = 0) + P(c = 1)p(x = 2 | c = 1) \\ &= (1 - 0.34) \cdot 0.054 + 0.34 \cdot 0.564 = 0.227 \end{aligned}$$

```
In [2]: Pc1 = 0.34; px2Dc0 = norm.pdf(2, 0, 1); px2Dc1 = norm.pdf(2, 2, np.sqrt(0.5))
px2 = (1-Pc1) * px2Dc0 + Pc1 * px2Dc1; Pc1Dx2 = Pc1 * px2Dc1 / px2
print(f"px2Dc0 = {px2Dc0:.3f}  px2Dc1 = {px2Dc1:.3f}  px2 = {px2:.3f}  Pc1Dx2 = {Pc1Dx2:.3f}")

px2Dc0 = 0.054  px2Dc1 = 0.564  px2 = 0.227  Pc1Dx2 = 0.843
```

### 3 Bayes' decision rule

**Bayes' decision rule:** Predicts a hypothesis after observing some evidence  $x$  by choosing, from a set of possible hypotheses  $\mathcal{C}$ , a hypothesis of maximum **posterior probability** (from the observation of the evidence)

$$c^*(x) = \operatorname{argmax}_{c \in \mathcal{C}} P(c \mid x)$$

**Probability of error:** That is, the probability that the predicted hypothesis is different from the correct one

$$P(\text{error} \mid x) = 1 - P(c^*(x) \mid x)$$

**Bayes' rule optimality:** No other choice would improve this error probability!

**Dentist example:**

$$c^*(x = 2) = \operatorname{argmax}_c \begin{pmatrix} P(c = 0 \mid x = 2) = 0.116 \\ P(c = 1 \mid x = 2) = 0.884 \end{pmatrix} = 1$$

**Bayes' rule based on prior probabilities and conditional (densities) of classes:** Instead of (arg-)maximize  $P(c \mid x)$  in  $c$ , we do it as a function of  $P(c) p(x \mid c)$  since the result is the same

$$c^*(x) = \operatorname{argmax}_{c \in \mathcal{C}} P(c \mid x) = \operatorname{argmax}_{c \in \mathcal{C}} P(c) \frac{p(x \mid c)}{p(x)} = \operatorname{argmax}_{c \in \mathcal{C}} P(c) p(x \mid c)$$

**Dentist example:**

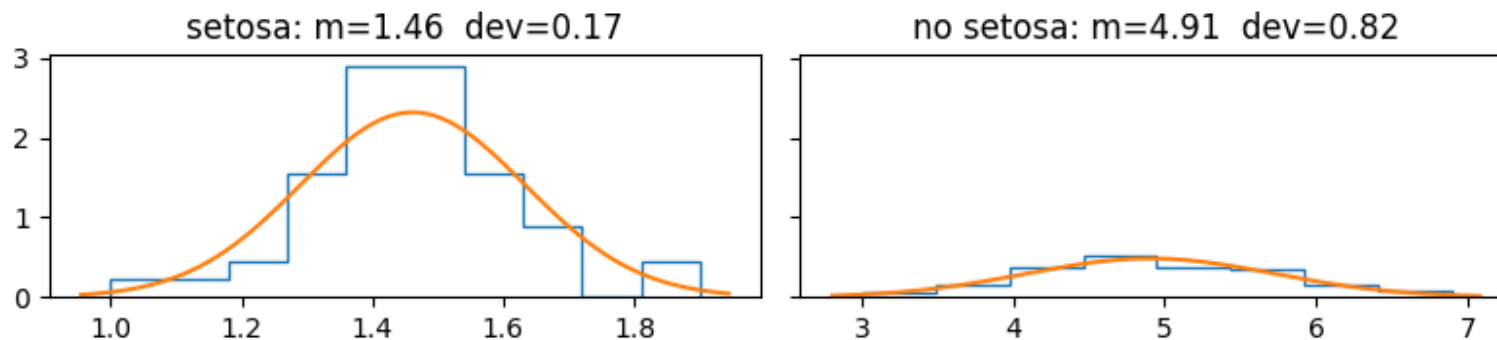
$$c^*(x = 2) = \operatorname{argmax}_c \begin{pmatrix} P(c = 0) p(x = 2 \mid c = 0) = 0.036 \\ P(c = 1) p(x = 2 \mid c = 1) = 0.271 \end{pmatrix} = 1$$

## U1.2 Continuous variables and Bayes' rule

**Problem:** Consider the classification of iris flowers as setosa or non-setosa based on petal length,  $x$ . It can be shown that distributions of  $x$  for setosa and non-setosa can be approximated by normal distributions of means and standard deviations:

$$p(x \mid c = \text{set}) \sim \mathcal{N}(\mu_{\text{set}} = 1.46, \sigma_{\text{set}} = 0.17) \quad \text{and} \quad p(x \mid c = \text{nos}) \sim \mathcal{N}(\mu_{\text{nos}} = 4.91, \sigma_{\text{nos}} = 0.82)$$

```
In [3]: import numpy as np; import matplotlib.pyplot as plt
from sklearn.datasets import load_iris; from scipy.stats import norm
iris = load_iris(); X = iris.data.astype(np.float16); y = iris.target.astype(np.uint)
x_set = np.squeeze(X[np.where(y==0), 2]); x_nos = np.squeeze(X[np.where(y!=0), 2])
fig, axs = plt.subplots(1, 2, figsize=(8, 2), sharey=True, tight_layout=True)
axs[0].hist(x_set, bins='auto', density=True, histtype='step')
x_set_range = np.arange(*axs[0].get_xlim(), .01)
x_set_mean = x_set.mean(); x_set_dev = np.sqrt(x_set.var())
axs[0].set_title(f'setosa: m={x_set_mean:.2f} dev={x_set_dev:.2f}')
axs[0].plot(x_set_range, norm.pdf(x_set_range, x_set_mean, x_set_dev))
axs[1].hist(x_nos, bins='auto', density=True, histtype='step')
x_nos_range = np.arange(*axs[1].get_xlim(), .01)
x_nos_mean = x_nos.mean(); x_nos_dev = np.sqrt(x_nos.var())
axs[1].set_title(f'no setosa: m={x_nos_mean:.2f} dev={x_nos_dev:.2f}')
axs[1].plot(x_nos_range, norm.pdf(x_nos_range, x_nos_mean, x_nos_dev));
```



If the estimated normal densities are true and the prior probability of setosa is  $1/3$ , what is the posterior probability that a flower of petal length 2 is setosa?

**Solution:**

$$\begin{aligned} P(c = \text{set} \mid x = 2) &= \frac{P(c = \text{set}) p(x = 2 \mid c = \text{set})}{p(x = 2)} \\ &= \frac{P(c = \text{set}) p(x = 2 \mid c = \text{set})}{P(c = \text{set}) p(x = 2 \mid c = \text{set}) + P(c = \text{nos}) p(x = 2 \mid c = \text{nos})} \\ &= \frac{1/3 \cdot \mathcal{N}(x = 2 \mid \mu_{\text{set}} = 1.46, \sigma_{\text{set}} = 0.17)}{1/3 \cdot \mathcal{N}(x = 2 \mid \mu_{\text{set}} = 1.46, \sigma_{\text{set}} = 0.17) + 2/3 \cdot \mathcal{N}(x = 2 \mid \mu_{\text{nos}} = 4.91, \sigma_{\text{nos}} = 0.82)} \\ &= \frac{\frac{1}{0.17} \exp\left(-\frac{(2-1.46)^2}{2 \cdot 0.17^2}\right)}{\frac{1}{0.17} \exp\left(-\frac{(2-1.46)^2}{2 \cdot 0.17^2}\right) + \frac{2}{0.82} \exp\left(-\frac{(2-4.91)^2}{2 \cdot 0.82^2}\right)} = \frac{0.0379}{0.0379 + 0.0045} = 0.89 \end{aligned}$$