Conditional probability, Bayes rule and independence

Statistical Computing and Empirical Methods
Unit EMATM0061, Data Science MSc

Rihuan Ke
rihuan.ke@bristol.ac.uk

Teaching Block 1, 2024

University of BRISTOL

What we will cover today

We will introduce the important concept of conditional probability.

We introduce the Bayes theorem and see how it can be used to "invert" conditional probabilities.

We will also discuss the law of total probability.

Finally, we will discuss the important concept of independence.

Random experiments, sample spaces, probability spaces

A random experiment is a procedure (real or imagined) which:

1. has a well-defined set of possible outcomes;
2. could (at least in principle) be repeated arbitrarily many times.

An event is a set (i.e. a collection) of possible outcomes of an experiment

A sample space is the set of all possible outcomes of interest for a random experiment

A probability space consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$, where Ω is a sample space, $\mathcal{E}$ is a well-behaved collection of events in Ω, and $\mathbb{P} : \mathcal{E} \to \mathbb{R}$ is a function satisfying the three rules of probability.

Rule 1: $\mathbb{P}(A) \geq 0$ for any event $A \in \mathcal{E}$

Rule 2: $\mathbb{P}(\Omega) = 1$ for sample space Ω

Rule 3: For pairwise disjoint events $A_1, A_2, \cdots \in \mathcal{E}$, we have

互斥事件的概率是相加的，因为它们不会同时发生

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

What is conditional probability? An example

Example 1: two bags of spheres

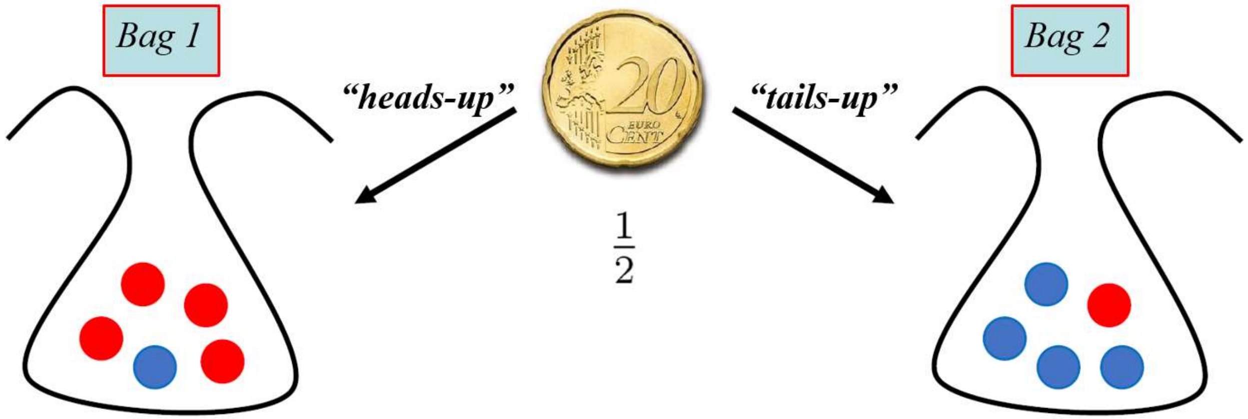
Suppose we have two bags (each with 50 coloured spheres):

- Bag 1: 49 red spheres + 1 blue sphere
- Bag 2: 1 red sphere + 49 blue spheres

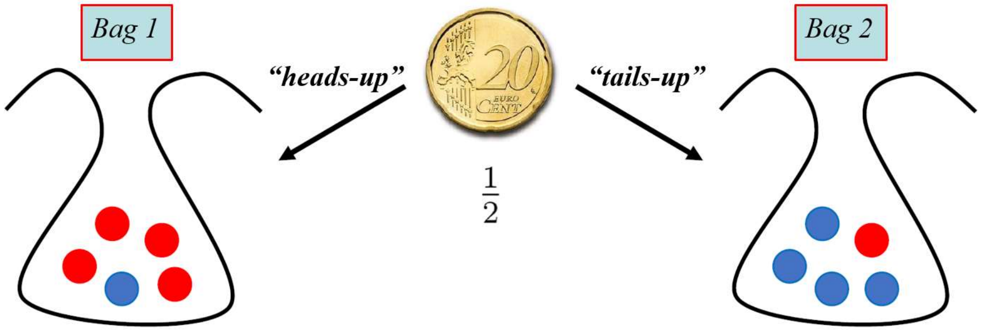
Random experiment:

Step 1: flip a fair coin

Step 2: If the coin lands "heads-up" a sphere is drawn at random from the 1st bag.
If the coin lands "tails-up" a sphere is drawn at random from the 2nd bag.



What is conditional probability? An example



Question 1: What is the probability that a red sphere is drawn?

$$\mathbb{P}(\text{a red sphere is drawn}) = 1/2$$

(because each sphere has an equal probability of being selected and there are 50 red spheres in total, out of the 100 spheres);

Question 2: Given that the coin lands "tails-up", what is the probability that a red sphere is drawn?

Intuitively, given that the coin landed "tails-up", the probability that a red sphere is drawn is $1/50$

$$\mathbb{P}(\text{a red sphere is drawn} \mid \text{the coin landed "tails-up"}) = 1/50.$$

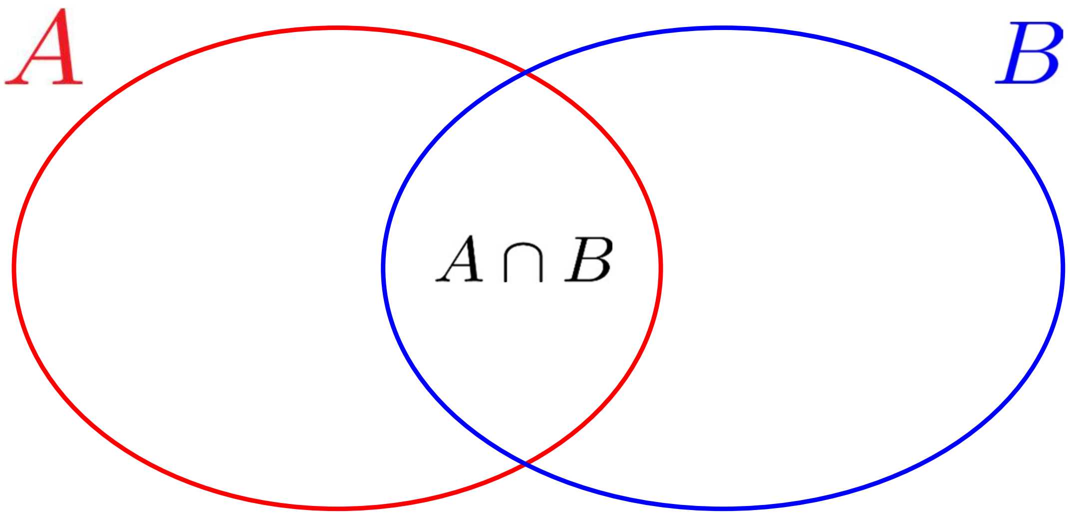
Conditional probability gives a precise formulation of this intuition.

1. Conditional probability

Conditional probability

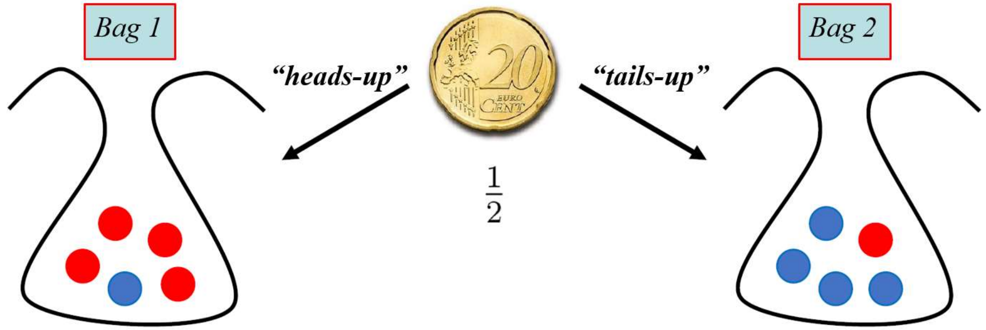
Let $\{\Omega, \mathcal{E}, \mathbb{P}\}$ be a probability space with sample space Ω, a collection of events $\mathcal{E}$ and probability function $\mathbb{P} : \Omega \to [0,1]$. Let $A, B \in \mathcal{E}$ be events and $\mathbb{P}(B) > 0$. The conditional probability of A given B is given by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$



What is conditional probability? An example

Example 1: two bags of spheres



Two bags (each with 50 coloured spheres):
Bag 1: 49 red spheres + 1 blue sphere
Bag 2: 1 red sphere + 49 blue spheres

Question 2: Given that the coin lands "tails-up", what is the probability that a red sphere is drawn?
$\mathbb{P}(\text{a red sphere is drawn} \mid \text{the coin landed "tails-up"}) = 1/50.$

Sample space $\Omega = \{1, 2, \cdots, 100\}$, which corresponds to spheres

$$\underbrace{1, \cdots, 49}_{\text{Red in bag 1}}, \underbrace{50}_{\text{Blue in bag 1}}, \underbrace{51}_{\text{Red in bag 2}}, \underbrace{52, \cdots, 100}_{\text{Blue bag 2}}.$$

A: the event that a red sphere is drawn $= \{1, \cdots, 49, 51\}$
B: the event that the coin landed "tails-up" $= \{51, \cdots, 100\}$
Therefore, $A \cap B$: the coin landed "tails-up" + a red sphere is drawn $= \{51\}$.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{51\})}{\mathbb{P}(\{51, \cdots, 100\})} = \frac{1/100}{50/100} = 1/50.$$

Conditional probability defines a new probability space

Theorem 1 (Conditional probability defines a new probability space)

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space.
Given an event B with $\mathbb{P}(B) > 0$, we can define $\mathbb{Q} := \mathbb{P}(\cdot \mid B)$, that is,

$$\mathbb{Q}(A) = \mathbb{P}(A \mid B) \text{ for any event } A \in \mathcal{E}.$$

Then the conditional probability space $(\Omega, \mathcal{E}, \mathbb{Q})$ defines a new probability space, where $\mathbb{Q}$ is the probability.

To check if $\mathbb{Q}$ is the probability, we need to show that $\mathbb{Q}$ satisfies the three key rules:

Rule 1. For all $A \in \mathcal{E}$, we have $\mathbb{Q}(A) \geq 0$;
Rule 2. The sample space has probability $\mathbb{Q}(\Omega) = 1$.
Rule 3. Given a sequence of pairwise disjoint events $A_1, A_2, \cdots$, we have $\cdots$.
Proof:
1. $\mathbb{Q}(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0$. So Rule 1 is satisfied.
2. $\mathbb{Q}(\Omega) = \mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$. So Rule 2 is satisfied.

Conditional probability defines a new probability space

Recall that $\mathbb{Q}(A) = \mathbb{P}(A \mid B)$ for any event $A \in \mathcal{E}$.

To check if $\mathbb{Q}$ is a probability, we need to show that $\mathbb{Q}$ satisfies the three key rules:
Rule 1. For all $A \in \mathcal{E}$, we have $\mathbb{Q}(A) \geq 0$;
Rule 2. The sample space has probability $\mathbb{Q}(\Omega) = 1$.
Rule 3. Given a sequence of pairwise disjoint events $A_1, A_2, \cdots$, we have $\mathbb{Q}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{Q}(A_i)$

Proof: 3. If $A_1, A_2, \cdots$ are pairwise disjoint, so are $A_1 \cap B, A_2 \cap B, \cdots$. Moreover, $(\cup_{i=1}^{\infty} A_i) \cap B = \cup_{i=1}^{\infty}(A_i \cap B)$. Hence

$$\mathbb{Q}(\cup_{i=1}^{\infty} A_i) = \mathbb{P}(\cup_{i=1}^{\infty} A_i \mid B) = \frac{\mathbb{P}((\cup_{i=1}^{\infty} A_i) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{i=1}^{\infty}(A_i \cap B))}{\mathbb{P}(B)}$$

$$= \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B) = \sum_{i=1}^{\infty} \mathbb{Q}(A_i).$$

So Rule 3 is satisfied.

Properties of conditional probability

Rule 1. For all $A \in \mathcal{E}$, we have $\mathbb{Q}(A) \geq 0$;
Rule 2. The sample space has probability $\mathbb{Q}(\Omega) = 1$.
Rule 3. For pairwise disjoint events $A_1, A_2, \cdots$, $\mathbb{Q}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{Q}(A_i)$

Since $\mathbb{Q} := \mathbb{P}(\cdot \mid B)$ defines a probability, the following properties hold as consequences of the three key rules:

1. $\mathbb{P}(\emptyset \mid B) = 0$

2. If $A, C \in \mathcal{E}$ are events and $A \subseteq C$, then $\mathbb{P}(A \mid B) \leq \mathbb{P}(C \mid B)$.

3. For any event $A \in \mathcal{E}$, we have $0 \leq \mathbb{P}(A \mid B) \leq 1$.

4. For events $S_1, S_2, \cdots$, we have $\mathbb{P}(\cup_{i=1}^{\infty} S_i \mid B) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i \mid B)$.

5. For any $A \in \mathcal{E}$, we have $\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B)$

6. For any $A, C \in \mathcal{E}$, we have $\mathbb{P}(A \cup C \mid B) = \mathbb{P}(A \mid B) + \mathbb{P}(C \mid B) - \mathbb{P}(A \cap C \mid B)$.

2. Bayes theorem

We often want to "invert" probabilities.

More precisely, suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

We have the value of $\mathbb{P}(A \mid B)$... but we want to know $\mathbb{P}(B \mid A)$, for some events $A, B \in \mathcal{E}$.

Example 2: A patient tests positive for a medical condition.

Let A be the event that the test is positive, and B be the event that the patient has the medical condition.

Suppose we know the conditional probability of a positive test given the medical condition $\mathbb{P}(A \mid B)$.

However, we want to know the conditional probability that the patient has the medical condition given a positive test result $\mathbb{P}(B \mid A)$.

Bayes theorem

Theorem 2 (Bayes theorem, Bayes, circa. 1760)

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Given events $A, B \in \mathcal{E}$ with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, we have

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B) \cdot \mathbb{P}(A \mid B)}{\mathbb{P}(A)}.$$

This simple but powerful result allows us to invert probabilities!

Proof: By definition we have $\mathbb{P}(B \mid A) := \mathbb{P}(A \cap B)/\mathbb{P}(A)$ and $\mathbb{P}(A \mid B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$. Therefore

$$\mathbb{P}(B \mid A) \cdot \mathbb{P}(A) = \mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B)$$

The result follows by dividing both sides by $\mathbb{P}(A)$. The proof is completed.

Remark: To "invert" $\mathbb{P}(A \mid B)$, we need to know $\mathbb{P}(B)$ and $\mathbb{P}(A)$ in order to apply the Bayes Theorem.

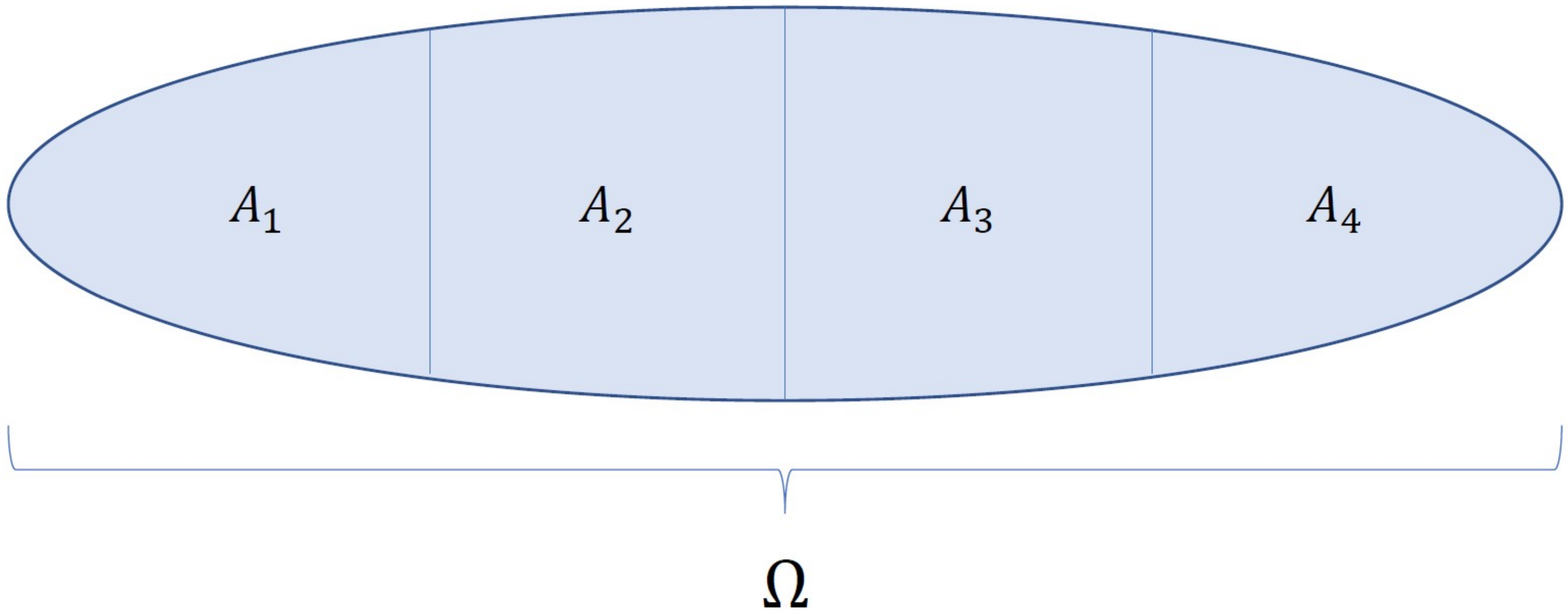
3. The law of total probability

Partition of a set

Recall that a partition of a set B is a sequence of disjoint sets whose union is B.

Formally, a partition of Ω is a finite or countably infinite sequence of sets $A_1, A_2, \cdots \subseteq \Omega$ such that:

1. The sequence of $A_1, A_2, \cdots$ is pairwise disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$).
2. The sequence covers Ω, so, $\Omega = \cup_i A_i$.



The law of total probability

Theorem 3 (The law of total probability)

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, and $A_1, A_2, \cdots \in \mathcal{E}$ forms a partition of Ω. For any event $B \in \mathcal{E}$, we have

$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i \cap B) = \sum_{\{i:\mathbb{P}(A_i)>0\}} \mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i).$$

Proof. We consider the sequence $S_1, S_2, \cdots$ with each $S_i := A_i \cap B$.

(1). Note that since $A_1, A_2, \cdots$ are pairwise disjoint, so are $S_1, S_2, \cdots$. Also, $B = B \cap \Omega = B \cap (\cup_i A_i) = \cup_i S_i$. Hence, the third rule of probability implies $\mathbb{P}(B) = \sum_i \mathbb{P}(S_i)$, which gives the first equality.

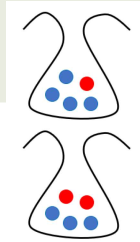
(2). If $\mathbb{P}(A_i) > 0$, we have $\mathbb{P}(B \mid A_i) = \mathbb{P}(A_i \cap B)/\mathbb{P}(A_i)$. Hence

$$\mathbb{P}(A_i \cap B) = \mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i).$$

If $\mathbb{P}(A_i) = 0$, since $A_i \cap B \subseteq A_i$, so $\mathbb{P}(A_i \cap B) \leq \mathbb{P}(A_i) = 0$.

Therefore, the second equality holds. The proof is completed.

The law of total probability: Example



Example 3. Six bags of spheres

Suppose that we have six bags, each containing 10 spheres. The i-th bag contains i red spheres, and $10 - i$ blue spheres.

Random experiment: We roll a fair dice. If our dice lands with the i-th face up then we pick a sphere at random from the i-th bag.

Question: What is the probability of picking a red sphere?

Sample space. Let $\Omega = \{1, \cdots, 6\} \times \{\mathrm{red}, \mathrm{blue}\}$. The first coordinate corresponds to the roll of the dice and the second to the colour of the sphere.

We consider the partition of Ω into events $A_1, A_2, \cdots, A_6$ where A_i is the event that the dice lands with the i-th face up.

For each i, we have $\mathbb{P}(\mathrm{red} \mid A_i) = \frac{i}{10}$. Hence by the *law of total probability*

$$\begin{aligned}\mathbb{P}(\mathrm{red}) &= \mathbb{P}(\mathrm{red} \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(\mathrm{red} \mid A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(\mathrm{red} \mid A_6)\mathbb{P}(A_6) \\ &= \frac{1}{10} \cdot \frac{1}{6} + \frac{2}{10} \cdot \frac{1}{6} + \cdots + \frac{6}{10} \cdot \frac{1}{6} = \frac{7}{20}.\end{aligned}$$

Bayes theorem and our diagnosis example

Example 2: A patient tests positive for a medical condition.

Let A be the event that the test is positive, and B be the event that the patient has the medical condition.

Question. We want to know the conditional probability that the patient has the medical condition given a positive test result, $\mathbb{P}(B \mid A)$.

Assume that we have the data:

Conditional probability of a positive test given the condition is $\mathbb{P}(A \mid B) = 0.95$.

Conditional probability of a negative test given the condition's absence is $\mathbb{P}(A^c \mid B^c) = 0.9$.

The (unconditional) probability that the patient has the condition is $\mathbb{P}(B) = 0.005$.

Bayes theorem and our diagnosis example

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B) \cdot \mathbb{P}(A \mid B)}{\mathbb{P}(A)}$$

$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i \cap B) = \sum_{\{i:\mathbb{P}(A_i)>0\}} \mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i).$$

Example 2: A patient tests positive for a medical condition.

Question. We want to know the conditional probability that the patient has the medical condition given a positive test result, $\mathbb{P}(B \mid A)$.

Assume that $\mathbb{P}(A \mid B) = 0.95$, $\mathbb{P}(A^c \mid B^c) = 0.9$, and $\mathbb{P}(B) = 0.005$.

First, we compute the probability of A (the event of a positive test)

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid B) \cdot \mathbb{P}(B) + \mathbb{P}(A \mid B^c) \cdot \mathbb{P}(B^c)\\ &= \mathbb{P}(A \mid B) \cdot \mathbb{P}(B) + [1 - \mathbb{P}(A^c \mid B^c)] \cdot [1 - \mathbb{P}(B)]\\ &= 0.95 \times 0.005 + (1 - 0.9) \times (1 - 0.005) = 0.10425.\end{aligned}$$

Second, we apply the Bayes Theorem

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B) \cdot \mathbb{P}(A \mid B)}{\mathbb{P}(A)} = \frac{0.005 \times 0.95}{0.10425} \approx 0.0456$$

So here $\mathbb{P}(B \mid A)$ is much smaller than $\mathbb{P}(A \mid B)$ (because $\mathbb{P}(B)$ is much smaller than $\mathbb{P}(A)$)!

4. Independence and dependence

Events in the real world often exhibit interesting dependencies upon one another.

Examples

Whether or not a patient catches a virus is closely tied to whether or not their friends do.

Whether or not the temperature at Bristol is higher than 25°C is closely tied to whether or not the temperature at Bath is.

Whether or not the EUR/USD exchange rate is above 1 tomorrow is closely tied to whether or not it is today

Conditional probability plays a fundamental role in our understanding of independence.

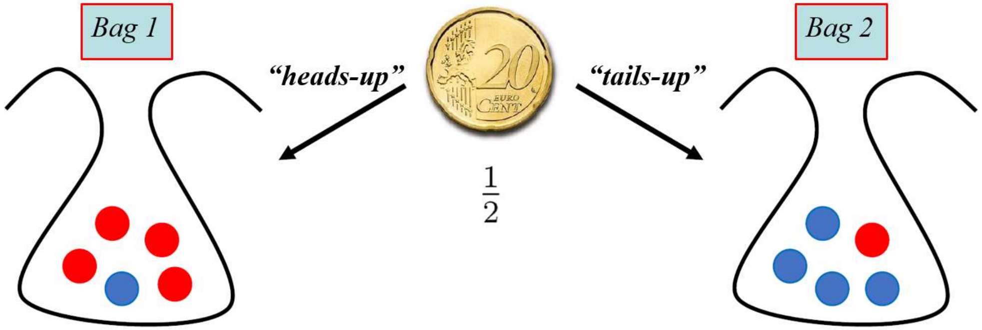
Independence and dependence

Independence and dependence

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space.

A pair of events $A, B \in \mathcal{E}$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
A pair of events $A, B \in \mathcal{E}$ are said to be dependent if $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example 1: two bags of spheres



Two bags (each with 50 coloured spheres):
Bag 1: 49 red spheres + 1 blue sphere
Bag 2: 1 red sphere + 49 blue spheres

A: the event that a red sphere is drawn $= \{1, \cdots, 49, 51\}$

B: the event that the coin landed "tails-up" $= \{51, \cdots, 100\}$

Therefore, $A \cap B$: the coin landed "tails-up" + a red sphere is drawn $= \{51\}$.

$$\mathbb{P}(A \cap B) = 1/100 \neq \mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{50}{100} \cdot \frac{50}{100} = 1/4.$$

So the events A and B are dependent.

Independence and dependence

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space.

A pair of events $A, B \in \mathcal{E}$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
A pair of events $A, B \in \mathcal{E}$ are said to be dependent if $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example 4. Rolling a dice and flipping a fair coin

Suppose we roll a dice and flip a fair coin.

Sample space: We model the scenario via a simple probability space with $\Omega = \{1, 2, \cdots, 6\} \times \{\mathrm{H}, \mathrm{T}\}$, so $|\Omega| = 12$.

Let A be the event that we roll a 6, so $A = \{(6, \mathrm{H}), (6, \mathrm{T})\}$.
Let B be the event that the coin lands "heads" up, so $B = \{(1, \mathrm{H}), ..., (6, \mathrm{H})\}$.

Then A and B are independent, because

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{(6, \mathrm{H})\}) = \frac{1}{12},$$
$$\mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

and hence $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Equivalent condition for independence

Lemma 1

Let $A, B \in \mathcal{E}$ be events with $P(B) > 0$. Then A and B are independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Proof. By definition $\mathbb{P}(A \mid B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$. So

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \Longleftrightarrow \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A) \Longleftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

The proof is completed.

Remark: If $A, B \in \mathcal{E}$ and $\mathbb{P}(B) = 0$, then $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Therefore, if $\mathbb{P}(B) = 0$, then B is independent of any other events in $\mathcal{E}$.

Remark: For any $A \in \mathcal{E}$, A and Ω are independent.

$$\mathbb{P}(A \cap \Omega) = \mathbb{P}(A) = \mathbb{P}(A) \cdot 1 = \mathbb{P}(A)\mathbb{P}(\Omega).$$

Independence for a sequence of events

Independence for a sequence of events

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space.

A sequence of events $A_1, A_2, \cdots A_n$ is said to be mutually-independent, if for any subset $\{i_1, i_2, \cdots, i_n\} \subseteq \{1, \cdots, n\}$ with $i_1 < i_2 < \cdots < i_k$, we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k})$$

A sequence of events $A_1, A_2, \cdots A_n$ is said to be pairwise-independent, if for any pair $\{i_1, i_2\} \subseteq \{1, \cdots, n\}$ with $i_1 \neq i_2$, we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2})$$

Remarks.

1. Mutual-independency implies pairwise-independency, but pairwise-independency does not imply mutual-independency.

2. For sequence $A_1, A_2, \cdots A_n$, independency typically refers to mutual-independency.

Independence for a sequence of events: Example

Example 5. Rolling a dice for three times

Suppose we roll a fair dice and record which faces land up.

Sample space: We model the scenario via a simple probability space with $\Omega = \{1, 2, \cdots, 6\} \times \{1, 2, \cdots, 6\} \times \{1, 2, \cdots, 6\}$, so $|\Omega| = 216$.

A_1: we get a 6 in the first roll, so $A_1 = \{(6, i, j) : i, j \in \{1, \cdots, 6\}\}$. So $|A_1| = 36$.

A_2: we get a 6 in the 2nd roll, so $A_2 = \{(i, 6, j) : i, j \in \{1, \cdots, 6\}\}$. So $|A_2| = 36$.

A_3: we get a 6 in the 3rd roll, so $A_3 = \{(i, j, 6) : i, j \in \{1, \cdots, 6\}\}$. So $|A_3| = 36$.

$A_1 \cap A_2 \cap A_3 = \{(6, 6, 6)\}$. So $|A_1 \cap A_2 \cap A_3| = 1$.

Then A_1, A_2, A_3 are independent (i.e., mutually-independent), because

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{216} = \frac{36}{216} \cdot \frac{36}{216} \cdot \frac{36}{216} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Summary

Conditional probability

Let $\{\Omega, \mathcal{E}, \mathbb{P}\}$ be a probability space with sample space Ω, a collection of events $\mathcal{E}$ and probability function $\mathbb{P} : \Omega \to [0, 1]$. Let $A, B \in \mathcal{E}$ be events and $\mathbb{P}(B) > 0$. The conditional probability of A given B is given by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Theorem 2 (Bayes theorem, Bayes, circa. 1760)

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Given events $A, B \in \mathcal{E}$ with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, we have

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B) \cdot \mathbb{P}(A \mid B)}{\mathbb{P}(A)}.$$

Theorem 3 (The law of total probability)

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, and $A_1, A_2, \dots \in \mathcal{E}$ forms a partition of Ω. For any event $B \in \mathcal{E}$, we have

$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i \cap B) = \sum_{\{i : \mathbb{P}(A_i) > 0\}} \mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i).$$

Independence and dependence

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space.

A pair of events $A, B \in \mathcal{E}$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

A pair of events $A, B \in \mathcal{E}$ are said to be dependent if $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

What we have covered today

We introduced the important concept of conditional probability.

- We verified that conditional probabilities are indeed a type of probability

We introduced Bayes theorem and saw how it can be used to "invert" conditional probabilities.

We also discussed the law of total probability.

We introduced the concept of independence and discussed its connections with conditional probability.

University of BRISTOL

Thanks for listening!

Dr. Rihuan Ke

rihuan.ke@bristol.ac.uk

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science