Conditional probability, Bayes rule and independence

Henry W J Reeve

henry.reeve@bristol.ac.uk

Statistical Computing & Empirical Methods (EMATM0061)

MSc in Data Science, Teaching block 1, 2021.

What will we cover today?

- We will introduce the important concept of conditional probability.
- We introduce 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.

Red & blue spheres

Suppose we have two bags, each containing 50 coloured spheres, so 100 spheres in total.

The first bag contains 40 red and 10 blue spheres. The second contains 10 red and 40 blue.

A fair coin is flipped:

- If the coin lands "heads-up" a sphere is drawn at random from the first bag.
- If the coin lands "tails-up" a sphere is drawn at random from the second bag.

Red & blue spheres

Suppose we have two bags, each containing 50 coloured spheres, so 100 spheres in total.

The first bag contains 40 red and 10 blue spheres. The second contains 10 red and 40 blue.

A fair coin is flipped:

- If the coin lands "heads-up" a sphere is drawn at random from the first bag.
- If the coin lands "tails-up" a sphere is drawn at random from the second bag.

All the spheres have equal probability of being selected, so the probability of a red one is $\frac{1}{2}$.

What if we knew that the coin flip landed "heads-up"?

Red & blue spheres

Suppose we have two bags, each containing 50 coloured spheres, so 100 spheres in total.

The first bag contains 40 red and 10 blue spheres. The second contains 10 red and 40 blue.

A fair coin is flipped:

- If the coin lands "heads-up" a sphere is drawn at random from the first bag.
- If the coin lands "tails-up" a sphere is drawn at random from the second bag.

All the spheres have equal probability of being selected, so the probability of a red one is $\frac{1}{2}$.

What if we knew that the coin flip landed "heads-up"?

Intuitively, given that coin landed "heads-up", the probability that a red sphere is drawn is $\frac{4}{5}$.

Conditional probability gives a precise formulation of this intuition.

Conditional probability

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space with sample space Ω, events $\mathcal{E}$ and probability function $\mathbb{P} : \mathcal{E} \to [0, 1]$.

Let A, $B \in \mathcal{E}$ be events with $\mathbb{P}(B) > 0$.

The **conditional probability** of A given B is given by

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

Conditional probability – blue & red spheres

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$.

Given A, $B \in \mathcal{E}$ with $\mathbb{P}(B) > 0$, the **conditional probability** of A given B is $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Example 1

A fair coin is flipped:

- If the coin lands "heads-up" a sphere is drawn from a bag with 40 red & 10 blue spheres;
- If the coin lands "tails-up" a sphere is drawn from a bag with 10 red & 40 blue spheres.

What's the conditional probability that the sphere is red, given that the coin landed "heads-up"?

Conditional probability – blue & red spheres

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$.

Given A, $B \in \mathcal{E}$ with $\mathbb{P}(B) > 0$, the **conditional probability** of A given B is $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Example 1

A fair coin is flipped:

- If the coin lands "heads-up" a sphere is drawn from a bag with 40 red & 10 blue spheres;
- If the coin lands "tails-up" a sphere is drawn from a bag with 10 red & 40 blue spheres.

What's the conditional probability that the sphere is red, given that the coin landed "heads-up"?

We model this with sample space $\Omega := \{1, \dots, 100\}$ corresponding to each of the spheres.

- $\{1, \dots, 40\}$ correspond to the red spheres in the $\mathbf{1}$st bag;
- $\{41, \dots, 50\}$ correspond to the blue spheres in the $\mathbf{1}$st bag;
- $\{51, \dots, 60\}$ correspond to the red spheres in the $\mathbf{2}$nd bag;
- $\{61, \dots, 100\}$ correspond to the blue spheres in the $\mathbf{2}$nd bag.

Conditional probability – blue & red spheres

Given $A, B \in \mathcal{E}$ with $\mathbb{P}(B) > 0$, the **conditional probability** of A given B is $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Example 1

What's the conditional probability that the sphere is red, given that the coin landed "heads-up"?

We model this with sample space $\Omega := \{1, \ldots, 100\}$ corresponding to each of the spheres.

- $\{1, \ldots, 40\}$ correspond to the red spheres in the **1**st bag;
- $\{41, \ldots, 50\}$ correspond to the blue spheres in the **1**st bag;
- $\{51, \ldots, 60\}$ correspond to the red spheres in the **2**nd bag;
- $\{61, \ldots, 100\}$ correspond to the blue spheres in the **2**nd bag.

Let $A := \{1, \ldots, 40\} \cup \{51, \ldots, 60\}$ (red) and $B := \{1, \ldots, 50\}$ ("heads-up").
Hence, $A \cap B = \{1, \ldots, 40\}$.

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{1, \ldots, 40\})}{\mathbb{P}(\{1, \ldots, 50\})} = \frac{4}{5}.$$

Conditional probability defines a new probability space

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$.

Given an event $B \in \mathcal{E}$ with $\mathbb{P}(B) > 0$, the conditional probability space $(\Omega, \mathcal{E}, \mathbb{P}(\cdot|B))$ defines a new probability space, where $\mathbb{P}(\cdot|B)$ denotes the map $A \mapsto \mathbb{P}(A|B)$ for events $A \in \mathcal{E}$.

1. For all $A \in \mathcal{E}$, $\mathbb{P}(A|B) \geq 0$;
2. The sample space Ω has probability $\mathbb{P}(\Omega|B) = 1$;
3. Given a sequence of pair-wise disjoint events $A_1, A_2, A_3, \ldots,$ we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \;\middle|\; B\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \mid B\right).$$

Conditional probability defines a new probability space

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$.

Given an event $B \in \mathcal{E}$ with $\mathbb{P}(B) > 0$, the conditional probability space $(\Omega, \mathcal{E}, \mathbb{P}(\cdot|B))$ defines a new probability space, where $\mathbb{P}(\cdot|B)$ denotes the map $A \mapsto \mathbb{P}(A|B)$ for events $A \in \mathcal{E}$.

1. For all $A \in \mathcal{E}$, $\mathbb{P}(A|B) \geq 0$;
2. The sample space Ω has probability $\mathbb{P}(\Omega|B) = 1$;
3. Given a sequence of pair-wise disjoint events $A_1, A_2, A_3, \ldots$, we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \,\middle|\, B\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \,\middle|\, B\right).$$

Proof:

1) Note that $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \geq 0$.

Conditional probability defines a new probability space

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$.

Given an event $B \in \mathcal{E}$ with $\mathbb{P}(B) > 0$, the conditional probability space $(\Omega, \mathcal{E}, \mathbb{P}(\cdot|B))$ defines a new probability space, where $\mathbb{P}(\cdot|B)$ denotes the map $A \mapsto \mathbb{P}(A|B)$ for events $A \in \mathcal{E}$.

1. For all $A \in \mathcal{E}$, $\mathbb{P}(A|B) \geq 0$;
2. The sample space Ω has probability $\mathbb{P}(\Omega|B) = 1$;
3. Given a sequence of pair-wise disjoint events $A_1, A_2, A_3, \ldots$, we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \,\middle|\, B\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \,\middle|\, B\right).$$

Proof:

1) Note that $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \geq 0$.
2) Note that $\mathbb{P}(\Omega|B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.

Conditional probability defines a new probability space

1. For all $A \in \mathcal{E}$, $\mathbb{P}(A|B) \geq 0$;
2. The sample space Ω has probability $\mathbb{P}(\Omega|B) = 1$;
3. Given a sequence of pair-wise disjoint events $A_1, A_2, A_3, \ldots,$ we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \;\middle|\; B\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \;\middle|\; B\right).$$

Proof:

1) Note that $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \geq 0$.
2) Note that $\mathbb{P}(\Omega|B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.
3) If $A_1, A_2, A_3, \ldots,$ are pair-wise disjoint, then so are $A_1 \cap B$,$A_2 \cap B$,$A_3 \cap B$,$\ldots$.
   Moreover, $(\bigcup_{n=1}^{\infty} A_n) \cap B = \bigcup_{n=1}^{\infty} (A_n \cap B)$.

Conditional probability defines a new probability space

1. For all $A \in \mathcal{E}$, $\mathbb{P}(A|B) \geq 0$;
2. The sample space Ω has probability $\mathbb{P}(\Omega|B) = 1$;
3. Given a sequence of pair-wise disjoint events $A_1, A_2, A_3, \ldots,$ we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \;\middle|\; B\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \;\middle|\; B\right).$$

Proof:

1) Note that $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \geq 0.$
2) Note that $\mathbb{P}(\Omega|B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1.$
3) If $A_1, A_2, A_3, \ldots,$ are pair-wise disjoint, then so are $A_1 \cap B$,$A_2 \cap B$,$A_3 \cap B$,$\ldots$.
   Moreover, $\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B = \bigcup_{n=1}^{\infty} (A_n \cap B)$. Hence,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \;\middle|\; B\right) = \frac{\mathbb{P}\left[\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right]}{\mathbb{P}(B)} = \frac{\mathbb{P}\left\{\bigcup_{n=1}^{\infty} (A_n \cap B)\right\}}{\mathbb{P}(B)}$$

$$= \frac{\sum_{n=1}^{\infty} \mathbb{P}\left(A_n \cap B\right)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n|B).$$

Conditional probability defines a new probability space

Theorem 1. *Suppose that $(\Omega, \mathcal{E}, \mathbb{P})$ is a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$. Let $B \in \mathcal{E}$ be an event with $\mathbb{P}(B) > 0$ and let $\mathbb{P}(\cdot|B)$ denote the map defined by $A \mapsto \mathbb{P}(A|B)$ for events $A \in \mathcal{E}$. Then $(\Omega, \mathcal{E}, \mathbb{P}(\cdot|B))$ is also a probability space.*

1. For all $A \in \mathcal{E}$, $\mathbb{P}(A|B) \geq 0$;
2. The sample space Ω has probability $\mathbb{P}(\Omega|B) = 1$;
3. Given a sequence of pair-wise disjoint events $A_1, A_2, A_3, \ldots$, we have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \,\middle|\, B\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \mid B\right).$$

Conditional probability spaces and red and blue spheres

Theorem 1. *Suppose that $(\Omega, \mathcal{E}, \mathbb{P})$ is a probability space with sample space Ω, events $\mathcal{E}$ and probability $\mathbb{P}$. Let $B \in \mathcal{E}$ be an event with $\mathbb{P}(B) > 0$ and let $\mathbb{P}(\cdot|B)$ denote the map defined by $A \mapsto \mathbb{P}(A|B)$ for events $A \in \mathcal{E}$. Then $(\Omega, \mathcal{E}, \mathbb{P}(\cdot|B))$ is also a probability space.*

Example 1

A fair coin is flipped:

- If the coin lands "heads-up" a sphere is drawn from a bag with 40 red & 10 blue spheres;
- If the coin lands "tails-up" a sphere is drawn from a bag with 10 red & 40 blue spheres.

The conditional probability of a red sphere (A) given a "heads-up" coin (B) is $\mathbb{P}(A|B) = \frac{4}{5}$.

The conditional probability of a blue sphere (A^c) given a "heads-up" coin (B) is

$$\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B) = \frac{1}{5}.$$

Bayes theorem

We often want to "*invert*" probabilities.

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

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

Bayes theorem

We often want to "*invert*" probabilities.

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

We know $\mathbb{P}(A|B)$ . . . but we want to know $\mathbb{P}(B|A)$, 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 the event that the patient has the condition.

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

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

Bayes theorem

Theorem 1 (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*

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

This simple but powerful result allows us to *invert* probabilities!

Bayes theorem

Theorem 1 (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*

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

This simple but powerful result allows us to *invert* probabilities!

Note: Using Bayes theorem does not make you a Bayesian! *All* statistiticians use Bayes theorem some of the time. Bayesian statisticians in some sense use Bayes theorem *all* the time.

Bayes theorem

Theorem 1 (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*

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

This simple but powerful result allows us to *invert* probabilities!

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

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

The result follows by dividing both sides through by $\mathbb{P}(A)$. $\square$

Bayes theorem and our diagnosis example

Bayes theorem says that $\mathbb{P}(B|A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B)/\mathbb{P}(A)$ for events A,B with $\mathbb{P}(A)$,$\mathbb{P}(B) > 0$.

Example 2

A patient tests positive for a medical condition.

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

- Conditional probability of a positive test given the condition is $\mathbb{P}(A|B) = 0.95$;
- Conditional probability of a negative test given the condition's absence is $\mathbb{P}(A^c|B^c) = 0.9$;
- The (unconditional) probability that the patient has the condition is $\mathbb{P}(B) = 0.005$.

We want to know the conditional probability of the the condition given a positive test ($\mathbb{P}(B|A)$).

Bayes theorem and our diagnosis example

Bayes theorem says that $\mathbb{P}(B|A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B)/\mathbb{P}(A)$ for events A,B with $\mathbb{P}(A)$,$\mathbb{P}(B) > 0$.

Example 2

- Conditional probability of a positive test given the condition is $\mathbb{P}(A|B) = 0.95$;
- Conditional probability of a negative test given the condition's absence is $\mathbb{P}(A^c|B^c) = 0.9$;
- The (unconditional) probability that the patient has the condition is $\mathbb{P}(B) = 0.005$.

We want to know the conditional probability of the the condition given a positive test ($\mathbb{P}(B|A)$).

First we compute
$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)\\ &= \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c)\\ &= \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \{1 - \mathbb{P}(A^c|B^c)\} \cdot \{1 - \mathbb{P}(B)\}\\ &= 0.95 \cdot 0.05 + (1 - 0.9) \cdot (1 - 0.005) = 0.147.\end{aligned}$$

Bayes theorem and our diagnosis example

Bayes theorem says that $\mathbb{P}(B|A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B)/\mathbb{P}(A)$ for events A,B with $\mathbb{P}(A)$,$\mathbb{P}(B) > 0$.

Example 2

- Conditional probability of a positive test given the condition is $\mathbb{P}(A|B) = 0.95$;
- Conditional probability of a negative test given the condition's absence is $\mathbb{P}(A^c|B^c) = 0.9$;
- The (unconditional) probability that the patient has the condition is $\mathbb{P}(B) = 0.005$.

We want to know the conditional probability of the the condition given a positive test ($\mathbb{P}(B|A)$).

We first compute $\mathbb{P}(A) = 0.147$.

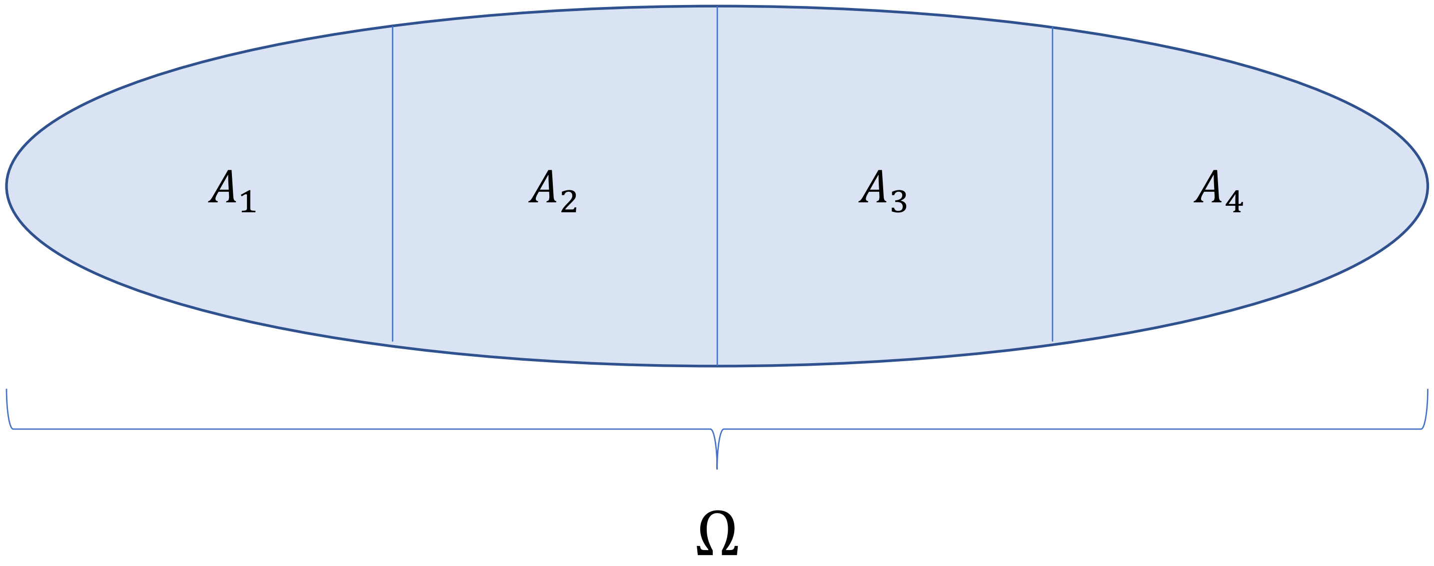
We then apply Bayes theorem

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B) \cdot \mathbb{P}(A|B)}{\mathbb{P}(A)} = \frac{0.005 \cdot 0.95}{0.147} \approx 0.0323.$$

Partitions and the law of total probability

A partition of Ω is a finite or countably infinite sequence of sets $A_1, A_2, A_3, \ldots \subseteq \Omega$ such that:

1. The sequence of $A_1, A_2, \ldots, A_K$ is pair-wise disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$);
2. The sequence covers Ω, so $\Omega = \bigcup_i A_i$.



Partitions and the law of total probability

A partition of Ω is a finite or countably infinite sequence of sets $A_1, A_2, A_3, \ldots \subseteq \Omega$ such that:

1. The sequence of $A_1, A_2, \ldots, A_K$ is pair-wise disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$);
2. The sequence covers Ω, so $\Omega = \bigcup_i A_i$.

Theorem 3 (The law of total probability)**.** *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$*. Suppose that we have a partition* $A_1, A_2, A_3, \ldots \subseteq \Omega$ *consisting of events* $A_i \in \mathcal{E}$*. Given an event* $B \in \mathcal{E}$ *we have,*

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

Partitions and the law of total probability

Theorem 3 (The law of total probability). *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$*. Suppose that we have a partition* $A_1, A_2, A_3, \ldots \subseteq \Omega$ *consisting of events* $A_i \in \mathcal{E}$*. Given an event* $B \in \mathcal{E}$ *we have,*

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

Proof. We consider the sequence S_1, S_2, $\ldots$ with each $S_i = A_i \cap B$. Note that since $A_1, A_2, A_3, \ldots$ are pair-wise disjoint, so are $S_1, S_2, S_3, \ldots$. Hence, the third rule of probability implies $\mathbb{P}(B) = \sum_i \mathbb{P}(A_i \cap B)$, which gives the first equality.

Partitions and the law of total probability

Theorem 3 (The law of total probability)**.** *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$*. Suppose that we have a partition* $A_1, A_2, A_3, \ldots \subseteq \Omega$ *consisting of events* $A_i \in \mathcal{E}$*. Given an event* $B \in \mathcal{E}$ *we have,*

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

Proof. We consider the sequence $S_1,\ S_2,\ \ldots$ with each $S_i = A_i \cap B$. Note that since $A_1, A_2, A_3, \ldots$ are pair-wise disjoint, so are $S_1, S_2, S_3, \ldots$. Hence, the third rule of probability implies $\mathbb{P}(B) = \sum_i \mathbb{P}(A_i \cap B)$, which gives the first equality.

Now if $\mathbb{P}(A_i) > 0$ we have $\mathbb{P}(B|A_i) = \mathbb{P}(A_i \cap B)/\mathbb{P}(A_i)$, by definition. Hence, $\mathbb{P}(A_i \cap B) = \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)$ for $\mathbb{P}(A_i) > 0$. On the other hand, for $\mathbb{P}(A_i) = 0$ then since $A_i \cap B \subseteq A_i$ so $\mathbb{P}(A_i \cap B) \leq \mathbb{P}(A_i) = 0$. Putting this together gives our second equality. $\square$

Partitions and the law of total probability

Theorem 3 (The law of total probability)**.** *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$*. Suppose that we have a partition* $A_1, A_2, A_3, \ldots \subseteq \Omega$ *consisting of events* $A_i \in \mathcal{E}$*. Given an event* $B \in \mathcal{E}$ *we have,*

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

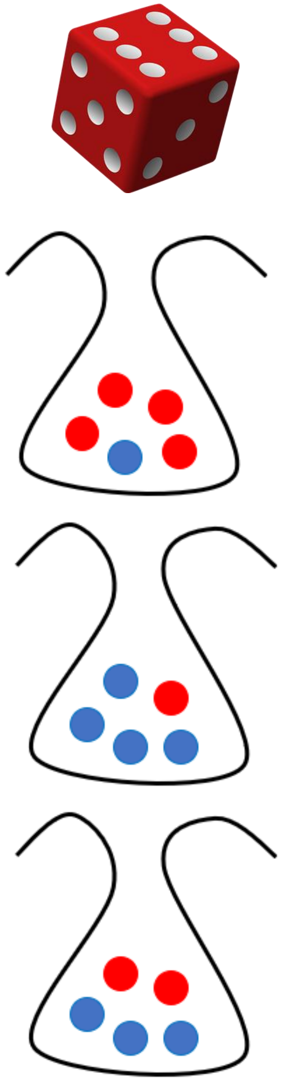
Remark: We have seen one example of this when computing the probability of a positive test:

$$\mathbb{P}\left(\text{Positive test}\right) = \mathbb{P}\left(\text{ Positive test} \mid \text{Positive }\right)\cdot\mathbb{P}\left(\text{ Positive }\right)+\mathbb{P}\left(\text{ Positive test} \mid \text{Negative }\right)\cdot\mathbb{P}\left(\text{ Negative }\right).$$

Partitions and the law of total probability

Theorem 3 (The law of total probability)**.** *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$. *Suppose that we have a partition* $A_1, A_2, A_3, \ldots \subseteq \Omega$ *consisting of events* $A_i \in \mathcal{E}$. *Given an event* $B \in \mathcal{E}$ *we have,*

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



Example 3

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

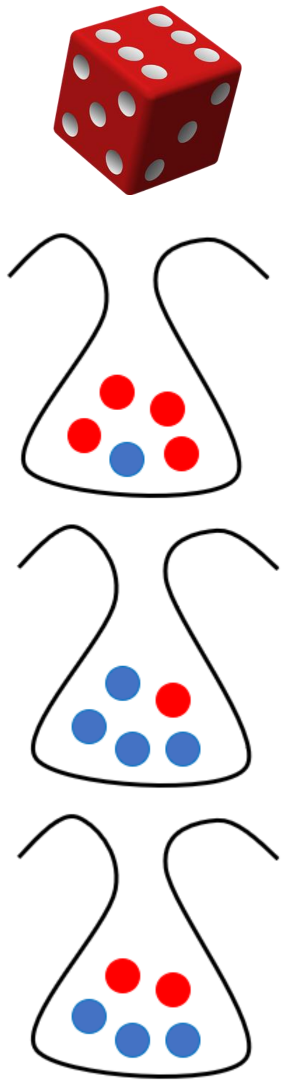
We roll a fair dice. If our dice lands with the i-th face up then we pick a ball at random from the i-th bag.

What is the probability of picking a red ball?

Partitions and the law of total probability

Example 3

Suppose that we have six bags, each containing 10 spheres. The i-th bag contains i red balls, and $10 - i$ blue balls. We roll a fair dice. If our dice lands with the i-th face up then we pick a ball at random from the i-th bag.



What is the probability of picking a red ball?

Let $\Omega := \{1, \ldots, 6\} \times \{\text{red}, \text{blue}\}$. The first coordinate corresponds to the roll of the dice and the second to the colour of the ball.

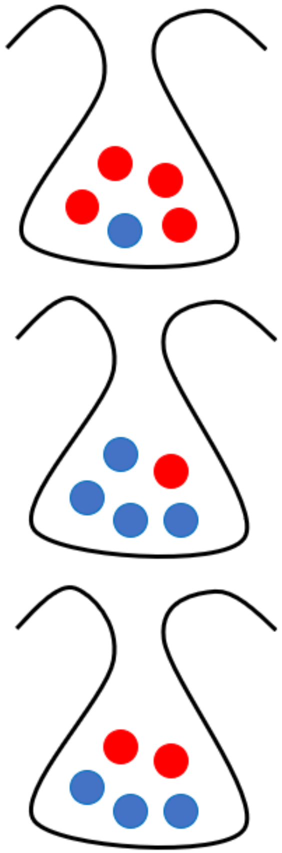
We consider the partition into events $A_1, \ldots, A_6$ where A_i is the event that the dice lands with the i-th face up.

For each i, we have $\mathbb{P}(\text{ red } | A_i) = \frac{i}{10}$.

Partitions and the law of total probability

Example 3

Suppose that we have six bags, each containing 10 spheres. The i-th bag contains i red balls, and $10 - i$ blue balls. We roll a fair dice. If our dice lands with the i-th face up then we pick a ball at random from the i-th bag.



What is the probability of picking a red ball?

Let $\Omega := \{1, \ldots, 6\} \times \{\text{red}, \text{blue}\}$. The first coordinate corresponds to the roll of the dice and the second to the colour of the ball.

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

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

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

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 GBPUSD-exchange rate is above is above 1.37 tomorrow is closely tied to whether or not it is today.

The event that customer in a music shop is a Jimi Hendrix fan is closely tied to whether or not they are an Eric Clapton fan.

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

Independence

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

Independence

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

Example 4

Suppose we roll a dice and flip a fair coin.

We model the scenario via a simple probability space with $\Omega = \{1, \ldots, 6\} \times \{\mathsf{H}, \mathsf{T}\}$, so $|\Omega| = 12$.

Let A be the event that we roll a 6, so $A = \{(6, \mathsf{H}), (6, \mathsf{T})\}$.

Let B be the event that the coin lands heads up, so $B = \{(1, \mathsf{H}), \ldots, (6, \mathsf{H})\}$.

Then $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{2}{12} = \frac{1}{6}$ and $\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{6}{12} = \frac{1}{2}$.

Moreover, $\mathbb{P}(A \cap B) = \mathbb{P}\left(\{(6, \mathsf{H})\}\right) = \frac{1}{12} = \left(\frac{1}{6}\right) \cdot \left(\frac{1}{2}\right) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

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 $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

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 5

We have two bags of five spheres. The first has one red (R) and four blue (B). The second has four red (R) and one blue (B).

We flip a fair coin. If the coin lands "heads-up" (H) we choose a ball at random from the first bag. If the coin lands "tails-up" (T) we choose a ball at random from the second bag.

Independence and dependence

Example 5

We have two bags of five spheres. The first has one red (R) and four blue (B). The second has four red (R) and one blue (B).

We flip a fair coin. If the coin lands "heads-up" (H) we choose a ball at random from the first bag. If the coin lands "tails-up" (T) we choose a ball at random from the second bag.

Hence, we have sample space $\Omega := \{(\mathrm{H}, \mathrm{R}), (\mathrm{H}, \mathrm{B}), (\mathrm{T}, \mathrm{R}), (\mathrm{T}, \mathrm{B})\}$.

Independence and dependence

Example 5

We have two bags of five spheres. The first has one red (R) and four blue (B). The second has four red (R) and one blue (B).

We flip a fair coin. If the coin lands "heads-up" (H) we choose a ball at random from the first bag. If the coin lands "tails-up" (T) we choose a ball at random from the second bag.

Hence, we have sample space $\Omega := \{(\mathsf{H}, \mathsf{R}), (\mathsf{H}, \mathsf{B}), (\mathsf{T}, \mathsf{R}), (\mathsf{T}, \mathsf{B})\}$.

- Let A_H denote the event that the coin lands "heads-up", so $A_H = \{(\mathsf{H}, \mathsf{R}), (\mathsf{H}, \mathsf{B})\}$.
- Let A_R denote the event that the sphere is red so $A_R = \{(\mathsf{H}, \mathsf{R}), (\mathsf{T}, \mathsf{R})\}$.

Independence and dependence

Example 5

We have two bags of five spheres. The first has one red (R) and four blue (B). The second has four red (R) and one blue (B).

We flip a fair coin. If the coin lands "heads-up" (H) we choose a ball at random from the first bag. If the coin lands "tails-up" (T) we choose a ball at random from the second bag.

Hence, we have sample space $\Omega := \{(\mathsf{H},\mathsf{R}),(\mathsf{H},\mathsf{B}),(\mathsf{T},\mathsf{R}),(\mathsf{T},\mathsf{B})\}$.

- Let A_H denote the event that the coin lands "heads-up", so $A_H = \{(\mathsf{H},\mathsf{R}),(\mathsf{H},\mathsf{B})\}$.
- Let A_R denote the event that the sphere is red so $A_R = \{(\mathsf{H},\mathsf{R}),(\mathsf{T},\mathsf{R})\}$.
- We have a fair coin, so $\mathbb{P}(A_H) = \mathbb{P}(A_H^c) = \frac{1}{2}$.

Independence and dependence

Example 5

We have two bags of five spheres. The first has one red (R) and four blue (B). The second has four red (R) and one blue (B).

We flip a fair coin. If the coin lands "heads-up" (H) we choose a ball at random from the first bag. If the coin lands "tails-up" (T) we choose a ball at random from the second bag.

Hence, we have sample space $\Omega := \{(\mathrm{H},\mathrm{R}),(\mathrm{H},\mathrm{B}),(\mathrm{T},\mathrm{R}),(\mathrm{T},\mathrm{B})\}$.

- Let A_H denote the event that the coin lands "heads-up", so $A_H = \{(\mathrm{H},\mathrm{R}),(\mathrm{H},\mathrm{B})\}$.
- Let A_R denote the event that the sphere is red so $A_R = \{(\mathrm{H},\mathrm{R}),(\mathrm{T},\mathrm{R})\}$.
- We have a fair coin, so $\mathbb{P}(A_H) = \mathbb{P}(A_H^c) = \frac{1}{2}$.
- There is one red in the first bag, out of a total of five, so $\mathbb{P}(A_R|A_H) = \frac{1}{5}$.
- There is four reds in the second bag, out of a total of five, so $\mathbb{P}(A_R|A_H^c) = \frac{4}{5}$.

Independence and dependence

Example 5

Hence, we have sample space $\Omega := \{(\mathsf{H},\mathsf{R}),(\mathsf{H},\mathsf{B}),(\mathsf{T},\mathsf{R}),(\mathsf{T},\mathsf{B})\}$.

- Let A_H denote the event that the coin lands "heads-up", so $A_H = \{(\mathsf{H},\mathsf{R}),(\mathsf{H},\mathsf{B})\}$.
- Let A_R denote the event that the sphere is red so $A_R = \{(\mathsf{H},\mathsf{R}),(\mathsf{T},\mathsf{R})\}$.
- We have a fair coin, so $\mathbb{P}(A_H) = \mathbb{P}(A_H^c) = \frac{1}{2}$.
- There is one red in the first bag, out of a total of five, so $\mathbb{P}(A_R|A_H) = \frac{1}{5}$.
- There is four reds in the second bag, out of a total of five, so $\mathbb{P}(A_R|A_H^c) = \frac{4}{5}$.

By the **law of total probability** we have

$$\mathbb{P}(A_R) = \mathbb{P}(A_R|A_H)\cdot\mathbb{P}(A_H) + \mathbb{P}(A_R|A_H^c)\cdot\mathbb{P}(A_H^c) = \frac{1}{5}\cdot\frac{1}{2} + \frac{4}{5}\cdot\frac{1}{2} = \frac{1}{2}.$$

Independence and dependence

Example 5

Hence, we have sample space $\Omega := \{(\mathsf{H},\mathsf{R}),(\mathsf{H},\mathsf{B}),(\mathsf{T},\mathsf{R}),(\mathsf{T},\mathsf{B})\}$.

- Let A_H denote the event that the coin lands "heads-up", so $A_H = \{(\mathsf{H},\mathsf{R}),(\mathsf{H},\mathsf{B})\}$.
- Let A_R denote the event that the sphere is red so $A_R = \{(\mathsf{H},\mathsf{R}),(\mathsf{T},\mathsf{R})\}$.
- We have a fair coin, so $\mathbb{P}(A_H) = \mathbb{P}(A_H^c) = \frac{1}{2}$.
- There is one red in the first bag, out of a total of five, so $\mathbb{P}(A_R|A_H) = \frac{1}{5}$.
- There is four reds in the second bag, out of a total of five, so $\mathbb{P}(A_R|A_H^c) = \frac{4}{5}$.

By the **law of total probability** we have

$$\mathbb{P}(A_R) = \mathbb{P}(A_R|A_H)\cdot\mathbb{P}(A_H) + \mathbb{P}(A_R|A_H^c)\cdot\mathbb{P}(A_H^c) = \frac{1}{5}\cdot\frac{1}{2}+\frac{4}{5}\cdot\frac{1}{2} = \frac{1}{2}.$$

We also note that $\mathbb{P}(A_H \cap A_R) = \mathbb{P}(A_R|A_H)\cdot\mathbb{P}(A_H) = \frac{1}{5}\cdot\frac{1}{2} = \frac{1}{10}$.

The events A_H and A_R are **dependent** since $\mathbb{P}(A_R \cap A_H) = \frac{1}{10} \neq \left(\frac{1}{2}\right)\cdot\left(\frac{1}{2}\right) = \mathbb{P}(A_R)\cdot\mathbb{P}(A_H)$.

Independence and conditional probabilities

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 $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

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

Independence and conditional probabilities

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 $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

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

Proof. By definition $\mathbb{P}(A|B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$, so

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad \Leftrightarrow \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

$\square$

Independence and conditional probabilities

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 $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

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

Proof. By definition $\mathbb{P}(A|B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$, so

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad \Leftrightarrow \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

$\sqcap$

Remark: If $A,B \in \mathcal{E}$ and $\mathbb{P}(B) = 0$ then $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = 0$.

Independence for a sequence of events

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

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

$$\mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}).$$

Independence for a sequence of events

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

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

$$\mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}).$$

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

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

Independence for a sequence of events

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

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

$$\mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \ldots \cdot \mathbb{P}(A_{i_k}).$$

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

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

Remarks:

1. Pairwise-independence does **not** imply mutual-independence.
2. For sequences $A_1, \ldots, A_n$ **independence** typically refers to **mutual-independence**.

What have we covered?

- 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 connections with conditional probability.

University of BRISTOL

Thanks for listening!

Henry W J Reeve

henry.reeve@bristol.ac.uk

Include EMATM0061 in the subject of your email.