To model phenomena which appear to be unpredictable and random, probability theory views them as *outcomes* of an *experiment* (in the large sense of the word).

The outcomes are *elements of a sample space* Ω. *Subsets* of Ω are called *events*. A *probability* is assigned to any event. It is a real number between 0 and 1, which expresses the *likelihood* of the event.

The purpose of this lecture is to define all these concepts in more detail.

1 Sample spaces

1.1 Definition

Probability starts with a **sample space**, which is a set whose elements describe possible **outcomes** in an experiment.

1.2 Examples

- Coin tossing experiment – 1 toss

  The possible outcomes are H and T (for Heads and Tails), so the sample space is

$$\Omega = \{\mathrm{H}, \mathrm{T}\}$$

- Coin tossing experiment – 2 tosses

  The sample space is

$$\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$$

- Coin tossing experiment – N tosses

  The sample space is

$$\Omega = \{\underbrace{\mathrm{HH}\ldots\mathrm{H}}_{\text{N heads}}, \mathrm{HTH}\ldots\mathrm{H}, \mathrm{HTTH}\ldots\mathrm{H}, \ldots, \underbrace{\mathrm{T}\ldots\mathrm{T}}_{\text{N tails}}\}$$

Question: How many elements are there in the sample space above?

- Rolling a dice experiment

  The sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Number of emails received in a week

  The sample space is

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, \ldots\}$$

  In other words, in this case, $\Omega = \mathbb{N}$.

- Day of the week a person randomly chosen on the street was born

  The sample space is

$$\Omega = \{\text{Monday}, \text{Tuesday}, \text{Wednesday}, \text{Thursday}, \text{Friday}, \text{Saturday}, \text{Sunday}\}$$

- Time until the 6 train arrives at Bleecker St station

  The sample space is

$$\Omega = [0, \infty)$$

2 Events

2.1 Definition

Let Ω be a sample space. **A subset** A **of** Ω, written mathematically $A \subseteq \Omega$ **is called an event**.

In probability theory we say that an event A occurs if the outcome of an experiment is an element of the set A.

2.2 Examples

- Consider again the experiment in which one tosses a coin twice. Let A be the event: "One Heads and one Tails are seen".

$$A = \{\text{HT,TH}\}$$

- Consider the experiment with the day of birth of a person randomly chosen on the street. The event "The day of the week starts with an S" is

$$A = \{\text{Saturday}, \text{Sunday}\}$$

2.3 Events and set operations

We will sometimes be interested in the likelihood of combined events. For example, for the experiment with the day of birth just discussed, the event "The day of the week has 6 letters" is

$$A = \{\text{Monday}, \text{Friday}, \text{Sunday}\}$$

and the event "The day of the week includes an 'r'" is

$$B = \{\text{Thursday}, \text{Friday}, \text{Saturday}\}$$

Then, the event "The day of the week is a six-letter day which include an 'r'" is written mathematically

$$C = A \cap B$$

where $\cap$ is the mathematical symbol for intersection. We have

$$C = A \cap B = \{\text{Friday}\}$$

Events are combined according to the usual set operations.

- The union of two events is written

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$$

  $\cup$ is the mathematical symbol for union.

  The event "The day of the week has six letters or includes an 'r'" is

$$D = A \cup B = \{\text{Monday}, \text{Thursday}, \text{Friday}, \text{Saturday}, \text{Sunday}\}$$

- The formal mathematical definition of the intersection of two sets is

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$$

- The complement of an event A is written A^{C}. Its definition is

$$A^{\mathrm{C}} = \{\omega \in \Omega : \omega \notin A\}$$

  In other words, the event A^{C} occurs if and only if A does not occur.

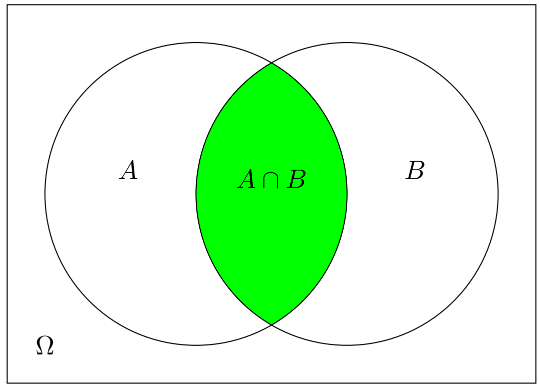
  We write $\emptyset = \Omega^{\mathrm{C}}$. $\emptyset$ is the impossible event.

- Two events A and B are **disjoint** if $A \cap B = \emptyset$. They do not have any outcome in common. We may equivalently say that they are **mutually exclusive**.

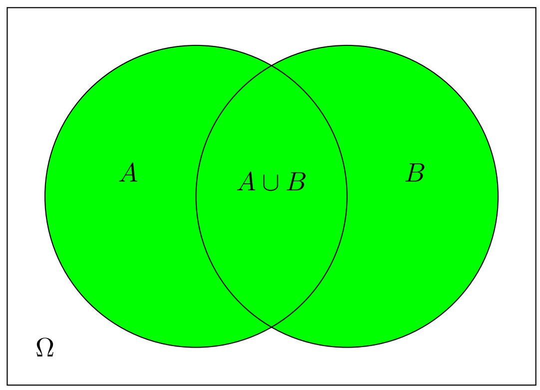
- An event A **implies** an event B if $A \subset B$.

  All the elements of A belong to B.

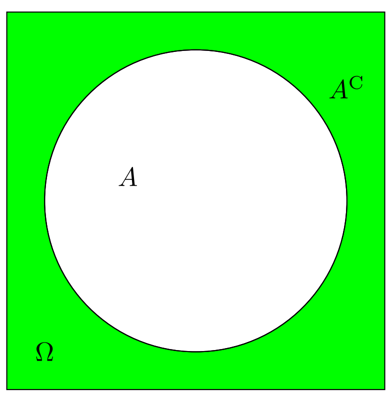
It is helpful to visualize these concepts using simple two-dimensional situations, as shown in the figures below..



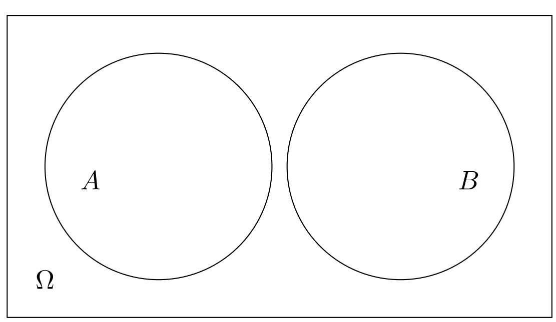
The event $A \cap B$ is shaded in green.



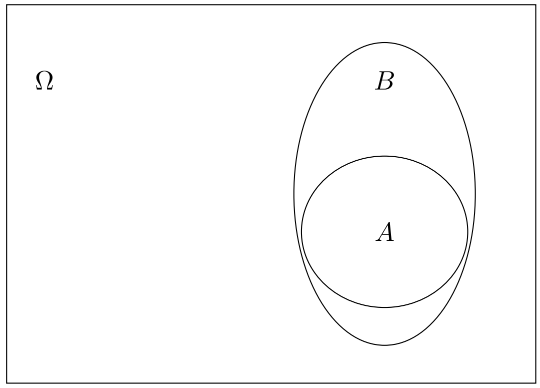
The event $A \cup B$ is shaded in green.



The complement of A, A^{C} is shaded in green.



The events A and B are disjoint: $A \cap B = \emptyset$. A and B are mutually exclusive events.



$A \subset B$:the event A implies the event B.

De Morgan's law: It can be readily shown (I recommend the exercise for you to practice on these concepts) that for any two events A and B, the following equalities hold:

$$(A \cup B)^{\mathrm{C}} = A^{\mathrm{C}} \cap B^{\mathrm{C}} \qquad\qquad (A \cap B)^{\mathrm{C}} = A^{\mathrm{C}} \cup B^{\mathrm{C}}$$

By looking at the figures in the previous page, it is relatively easy to convince yourself that these equalities indeed hold.

3 Probability

3.1 Definition

A probability P is an assignment of a **real value** to each event A in a sample space Ω, such that:

1. $P(A) \in [0, 1]$
2. $P(\Omega) = 1$
3. If A and B are disjoint, $P(A \cup B) = P(A) + P(B)$

3.2 Immediate properties

- By definition, A, and A^{C} are disjoint. Hence, if P is a probability function,

$$\begin{aligned} & P(A \cup A^{\mathrm{C}}) = P(A) + P(A^{\mathrm{C}}) \\ \Leftrightarrow\ & P(\Omega) = P(A) + P(A^{\mathrm{C}}) \\ \Leftrightarrow\ & P(A^{\mathrm{C}}) = 1 - P(A) \end{aligned}$$

  In particular, since $\Omega^{\mathrm{C}} = \emptyset$,

$$P(\emptyset) = 1 - P(\Omega) = 0$$

  The probability of the impossible event is 0.

- From the additive property of a probability function for disjoint events, we can derive an equality for the probability of the union of events which are not disjoint. Let A and B be those events.

$$A = (A \cap B) \cup (A \cap B^{\mathrm{C}})$$

  Since B and B^{C} are disjoint, so are $A \cap B$ and $A \cap B^{\mathrm{C}}$. Hence

$$P(A) = P(A \cap B) + P(A \cap B^{\mathrm{C}}) \tag{1}$$

  Likewise,

$$A \cup B = ((A \cup B) \cap B) \cup ((A \cup B) \cap B^{\mathrm{C}})$$

The first event on the right-hand side is just B. The second event is $A \cap B^{\mathrm{C}}$. Hence,

$$P(A \cup B) = P(B) + P(A \cap B^{\mathrm{C}})$$

. This can be rewritten as

$$P(A \cap B^{\mathrm{C}}) = P(A \cup B) - P(B)$$

Using this result in Eq.(1), we get

$$P(A) = P(A \cap B) + P(A \cup B) - P(B)$$

We thus proved the following important relationship:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)} \tag{2}$$

This relationship holds for any two events A and B.

- Let us now assume that A and B are two events in a sample space Ω such that $A \subset B$. Then A and $B \setminus A$ are disjoint events. Hence,

  $$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$$

  from which we conclude that $P(A) \leq P(B)$ since P is a probability function, which must therefore satisfy $P(B \setminus A) \geq 0$. We conclude that

  $$\boxed{A \subseteq B \qquad \Rightarrow \qquad P(A) \leq P(B)} \tag{3}$$

3.3 Examples

- <u>Coin tossing experiment – 1 toss</u>

  If the coin is fair (and we ignore the VERY unlikely situation in which the coin lands on its side), we have

  $$P(\{\mathrm{H}\}) = P(\{\mathrm{T}\}) = \frac{1}{2}$$

- <u>Coin tossing experiment – 2 successive tosses</u>

  If the coin is fair, the probability of the event "One heads and one tails are seen" is

  $$P(\{\mathrm{HT}, \mathrm{TH}\}) = \frac{1}{2}$$

  For that example, one can also verify the additive property of probabilities:

  $$P(\{\mathrm{HT}, \mathrm{TH}\}) = P(\{\mathrm{HT}\}) + P(\{\mathrm{TH}\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

- <u>Rolling a dice experiment</u> The probability of obtaining a 5 by rolling the dice once is, if the dice is fair

  $$P(\{5\}) = \frac{1}{6}$$

- <u>Day of birth experiment</u> For the day of the week for the birth of a random person chosen in the street, the probability that the day of birth was a Sunday is

  $$P(\{\mathrm{Sunday}\}) = \frac{1}{7}$$

  The probability that the day of the week has six letters is

  $$P(\{\mathrm{Monday}, \mathrm{Friday}, \mathrm{Sunday}\}) = \frac{3}{7}$$

  As expected, $P(\{\mathrm{Sunday}\}) \leq P(\{\mathrm{Monday}, \mathrm{Friday}, \mathrm{Sunday}\})$, since $\{\mathrm{Sunday}\} \subset \{\mathrm{Monday}, \mathrm{Friday}, \mathrm{Sunday}\}$.

3.4 Product of sample spaces

As we have seen with coin tossing, it will often be the case that an experiment consists in a succession of experiments. The two successive coin tosses experiment is made of two individual coin tossing experiments. Its sample space Ω can be written as

$$\Omega = \{\mathrm{H}, \mathrm{T}\} \times \{\mathrm{H}, \mathrm{T}\} = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$$

Imagine now that the experiment consists in first tossing a coin, and then rolling a dice. The sample space is

$$\Omega = \{\mathrm{H}, \mathrm{T}\} \times \{1, 2, 3, 4, 5, 6\} = \{\mathrm{H1}, \mathrm{H2}, \mathrm{H3}, \mathrm{H4}, \mathrm{H5}, \mathrm{H6}, \mathrm{T1}, \mathrm{T2}, \mathrm{T3}, \mathrm{T4}, \mathrm{T5}, \mathrm{T6}\}$$

In general, if $\Omega = \Omega_1 \times \Omega_2$, and Ω_1 has m elements and Ω_2 has n elements, then Ω has $m \times n$ elements, as can be verified for the examples above.

Consider now the particular, but nevertheless important, situation in which all outcomes in Ω_1 are equally likely, all outcomes in Ω_2 are equally likely, and the outcomes in Ω_1 and Ω_2 do not influence each other. Then all outcomes in $\Omega_1 \times \Omega_2$ are equally likely.

The probability of any outcome of Ω_1 is $\frac{1}{m}$.

The probability of any outcome of Ω_2 is $\frac{1}{n}$.

The probability of any outcome of $\Omega_1 \times \Omega_2$ is $\frac{1}{mn}$.

We observe that

$$\frac{1}{mn} = \frac{1}{m} \times \frac{1}{n}$$

This is a general result. Let us write it formally. Let

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

If outcome $\omega_{1i} \in \Omega_1$ has probability p_i and outcome $\omega_{2j} \in \Omega_2$ jas probability p_j, and **if the outcomes in Ω_1 and Ω_2 are independent of each other**, then the probability of the outcome $(\omega_{1i}, \omega_{2j}) \in \Omega$ is

$$\underline{P(\{(\omega_{1i}, \omega_{2j})\}) = p_i p_j}$$

This can be generalized to the sample space Ω for N experiments: $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_N$. If an outcome $\omega_{1i} \in \Omega_1$ has probability p_i, if $\omega_{2j} \in \Omega_2$ has probability p_j, etc., if $\omega_{Nk} \in \Omega_N$ has probability p_k, then the probability of the outcome $(\omega_{1i}, \omega_{2j}, \ldots, \omega_{Nk}) \in \Omega$ is

$$\underline{P(\{(\omega_{1i}, \omega_{2j}, \ldots, \omega_{Nk})\}) = p_i p_j \ldots p_k}$$

if the outcomes of each individual experiment are independent of each other.

3.5 Infinite sample spaces

In the example sample spaces we considered in Section 1.2, we saw two cases of *infinite* sample spaces: a *countably infinite* sample space (the email example), and an infinite sample space which is *not countable* (the subway train example). Uncountably infinite sample spaces lead to unique mathematical challenges, which we will treat later in this course.

On the other hand, we already possess all the tools necessary to treat countably infinite sample spaces, as we will now see.

Let us first prove the following interesting property:

Property: the outcomes of a countably infinite sample space cannot all have the same probability.

Proof: Assume it were possible, i.e. there is a countably infinite sample space $\Omega = \{\omega_1, \omega_2, \ldots\}$ such that

$$\forall (i, j) \in \mathbb{N}^2\,,\;\; P(\{\omega_i\}) = P(\{\omega_j\})$$

Now, if $i \neq j$, $\{\omega_i\}$ and $\{\omega_j\}$ are disjoint events. Therefore,

$$\forall N \in \mathbb{N}\ ,\ P(\bigcup_{i=1}^{N}\{\omega_i\}) = \sum_{i=1}^{N} P(\{\omega_i\}) = NP(\{\omega_1\})$$
$$\Leftrightarrow P(\{\omega_1\}) = \frac{1}{N} P(\bigcup_{i=1}^{N}\{\omega_i\}) \tag{4}$$

In the limit $N \to +\infty$, $\bigcup\limits_{i=1}^{N}\{\omega_i\} \to \Omega$, so

$$P(\bigcup_{i=1}^{N}\{\omega_i\}) \xrightarrow[N\to+\infty]{} P(\Omega) = 1$$

Eq.(4) then tells us that

$$P(\{\omega_1\}) \xrightarrow[N\to+\infty]{} 0$$

from which we conclude that all the outcomes have probability zero, which is impossible $\square$

To finish this section, let us consider a specific example. Imagine the experiment in which one counts the number of tosses of a fair coin it takes to have the first occurence of a heads. The sample space is:

$$\Omega = \{1, 2, 3, \ldots\} = \mathbb{N}^*$$

We have

$$P(\{1\}) = \frac{1}{2}\ \ ,\ \ P(\{2\}) = \left(1 - \frac{1}{2}\right)\frac{1}{2} = \frac{1}{4}\ \ ,\ \ P(\{3\}) = \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{2}\right)\frac{1}{2} = \frac{1}{8}$$

The general result is

$$P(\{N\}) = \left(1 - \frac{1}{2}\right)^{N-1}\frac{1}{2} = \frac{1}{2^N}\quad ,\quad \forall\, N \in \mathbb{N}$$

Let us numerically verify that $P(\Omega) = 1$:

$$P(\Omega) = P(\bigcup_{i=1}^{+\infty}\{i\}) = \sum_{i=1}^{+\infty} P(\{i\}) = \sum_{i=1}^{+\infty}\frac{1}{2^i}$$

We recognize a geometric series, so

$$\sum_{i=1}^{+\infty}\frac{1}{2^i} = \frac{1}{2}\frac{1}{1-\frac{1}{2}} = 1$$

We thus have $P(\Omega) = 1$ as expected.

4 Counting (a.k.a. Elementary combinatorics)

4.1 Computing probabilities

In general, assigning probabilities to outcomes of an experiment can be extremely challenging, especially if the experiment depends on intricate sociological, psychological, physical, biological, etc. factors. It then becomes the job of experts in the discipline.

However, we will now learn a straightforward method to compute the probabilities of events for the simple case **when the probability of each outcome is the same** (rolling a dice, coin tossing, poker hand, etc.).

Let Ω be a sample space with an associated probability function P such that $P(\{\omega_i\}) = P(\{\omega_j\})$ for all $(i, j) \in \mathbb{N}^2$. Then the probability of an event $A \subseteq \Omega$ is given by:

$$\boxed{P(A) = \frac{|A|}{|\Omega|}}$$

where $|B|$ is the mathematical notation for the number of elements in the set B.
Proof: Observe first that Ω cannot be infinite, because of what we proved in the previous section.
Let N be the number of elements in Ω.

$$\forall i \in \mathbb{N}^* \ , \ P(\{\omega_i\}) = \frac{1}{N}$$

Let $A = \{\omega_1, \omega_2, \ldots, \omega_k\} \subseteq \Omega$ be an event, where we have relabeled the ω_i if necessary.

$$A = \bigcup_{i=1}^{k} \{\omega_i\}$$

so

$$P(A) = \sum_{i=1}^{k} P(\{\omega_i\}) = \frac{k}{N} = \frac{|A|}{|\Omega|} \qquad \square$$

We just proved that for the simple case when the probability of each outcome is the same, the probability of an event A is easy to calculate if we know how to count the number of elements in A (and in Ω). This is why we close this lecture with standard concepts and methods for counting.

4.2 Counting

- We have already encountered a principle for counting when we learned about the product of sample spaces: if experiment 1 has m possible outcomes and experiment 2 has n possible outcomes, than the experiment combining 1 and 2 has mn possible outcomes.

- Selecting an ordered sample from a set

  A) Without replacement

  Consider a horse race with 15 horses. A trifecta consists in picking the first three finishers in exact order. There are

  $$15 \times 14 \times 13$$

  different ways of choosing a trifecta.

  In general, if there are N horses and one wants to choose the first P in the proper order, there are

  $$\boxed{\frac{N!}{(N-P)!}}$$

  different ways of choosing, where ! is the symbol for factorial.

  B) With replacement

  What if instead of picking the order for a horse race, one rolls a dice 3 times and wants to predict the 3 digit number one will get?

  In that situation, there are

  $$6 \times 6 \times 6 = 6^3$$

  different numbers possible.

  In general, if the sample space for one experiment has N elements, and there are P successive experiments, the number of possible outcomes is

  $$\boxed{N^P}$$

- Selecting an unordered sample from a set

  In horse racing, it is also possible to bet on the first three horses to finish the race without specifying their finishing order.

The number $15 \times 14 \times 13$ does not correspond to the number of choices for that bet, because it counts different finishing orders as different choices. To address this, we have to *divide* $15 \times 14 \times 13$ by the number of ways there exists to reorder the finishing order of the 3 horses. Mathematically, that number is **the number of permutations of the finishing order**.

There are

$$3 \times 2 \times 1 = 3!$$

permutations of the finishing order of the first 3 horses.

Hence, the number of choices for that type of bet is

$$\frac{15 \times 14 \times 13}{3!} = \frac{15!}{(15-3)!3!}$$

In mathematics, and combinatorics in particular, that particular combination appears so often that mathematicians have chosen to give it a dedicated notation:

$$\frac{15!}{(15-3)!3!} = \binom{15}{3}$$

The number $\binom{15}{3}$ is read "15 choose 3".

The general formula for N horses in the race and choosing the first P in *any* order is

$$\boxed{\binom{N}{P} = \frac{N!}{(N-P)!P!}}$$

If all 15 horses in the race have equal strength and speed (which is never the case in reality), the probability of betting on the right trifecta is

$$P_{\text{Tri}} = \frac{1}{15 \times 14 \times 13} = \frac{1}{2730} \approx 0.00037$$

The probability of guessing the first three horses to finish the race, in any order, is

$$P_3 = \frac{1}{\binom{15}{3}} = \frac{1}{455} = 6P_{\text{Tri}} \approx 0.0022$$