MAS113 Introduction to Probability and Statistics

Dr Jonathan Jordan

School of Mathematics and Statistics, University of Sheffield

2020–21

Studying probability theory

There are (at least) two ways to think about the study of probability theory.

Pure mathematical approach

Probability theory is a branch of Pure Mathematics.

Pure mathematical approach

Probability theory is a branch of Pure Mathematics.

We start with some **axioms** and investigate the implications of those axioms; what results are we able to prove?

Pure mathematical approach

Probability theory is a branch of Pure Mathematics.

We start with some **axioms** and investigate the implications of those axioms; what results are we able to prove?

Probability theory is a subject in its own right, that we can study (and appreciate!) without concerning ourselves with any possible applications.

Pure mathematical approach

Probability theory is a branch of Pure Mathematics.

We start with some **axioms** and investigate the implications of those axioms; what results are we able to prove?

Probability theory is a subject in its own right, that we can study (and appreciate!) without concerning ourselves with any possible applications.

We can also observe links to other areas of mathematics.

Modelling approach

Probability theory gives us a mathematical model for *explaining* chance phenomena that we observe in the real world.

Modelling approach

Probability theory gives us a mathematical model for *explaining* chance phenomena that we observe in the real world.

As with any mathematical theory, it may not describe reality *perfectly*, but it can still be extremely useful.

Modelling approach

Probability theory gives us a mathematical model for *explaining* chance phenomena that we observe in the real world.

As with any mathematical theory, it may not describe reality *perfectly*, but it can still be extremely useful.

We must think carefully about when and how the theory can be applied in practice.

This course

We will be studying probability from *both* perspectives.

This course

We will be studying probability from *both* perspectives.

Applications of probability theory are important and interesting, but as a mathematician, you should also develop an understanding and an appreciation of the theory in its own right.

Motivation

We are often faced with making decisions in the presence of uncertainty.

- A doctor is treating a patient, and is considering whether to prescribe a particular drug. How likely is it that the drug will cure the patient?

Motivation

We are often faced with making decisions in the presence of uncertainty.

- A doctor is treating a patient, and is considering whether to prescribe a particular drug. How likely is it that the drug will cure the patient?
- A probation board is considering whether to release a prisoner early on parole. How likely is the prisoner to re-offend?

Motivation

We are often faced with making decisions in the presence of uncertainty.

- A doctor is treating a patient, and is considering whether to prescribe a particular drug. How likely is it that the drug will cure the patient?
- A probation board is considering whether to release a prisoner early on parole. How likely is the prisoner to re-offend?
- A bank is considering whether to approve a loan. How likely is it that the loan will be repaid?

Motivation

We are often faced with making decisions in the presence of uncertainty.

- A doctor is treating a patient, and is considering whether to prescribe a particular drug. How likely is it that the drug will cure the patient?
- A probation board is considering whether to release a prisoner early on parole. How likely is the prisoner to re-offend?
- A bank is considering whether to approve a loan. How likely is it that the loan will be repaid?
- A government is considering a CO_2 emissions target. What would the effect of a 20% cut in emissions be on global mean temperatures in 20 years' time?

Describing uncertainty

We often use verbal expressions of uncertainty ("possible", "quite unlikely", "very likely"), but these are often inadequate if we want to communicate with each other about uncertainty.

Describing uncertainty

We often use verbal expressions of uncertainty ("possible", "quite unlikely", "very likely"), but these are often inadequate if we want to communicate with each other about uncertainty.

Consider the following example.

Disease test

You are being screened for a particular disease. You are told that the disease is "quite rare", and that the screening test is accurate but not perfect.

Disease test

You are being screened for a particular disease. You are told that the disease is "quite rare", and that the screening test is accurate but not perfect.

If you have the disease, the test will "almost certainly" detect it, but if you don't have the disease, there is "a small chance" the test will mistakenly report that you have it anyway.

Disease test

You are being screened for a particular disease. You are told that the disease is "quite rare", and that the screening test is accurate but not perfect.

If you have the disease, the test will "almost certainly" detect it, but if you don't have the disease, there is "a small chance" the test will mistakenly report that you have it anyway.

The test result is positive. How certain are you that you really have the disease?

Quantifying uncertainty

Clearly, in this example and in many others, it would be useful if we could *quantify* our uncertainty.

Quantifying uncertainty

Clearly, in this example and in many others, it would be useful if we could *quantify* our uncertainty.

In other words, we would like to *measure* how likely it is that something will happen, or how likely it is that some statement about the world turns out to be true.

Quantifying uncertainty continued

We will introduce a theory for measuring uncertainty: probability theory.

Quantifying uncertainty continued

We will introduce a theory for measuring uncertainty: probability theory.

We have two ingredients: **set theory**, which we will use to describe what sort of things we want to measure our uncertainty about, and, appropriately, **measure theory**, which sets out some basic rules for how to measure things in a sensible way.

Motivation - the need for set theory and measures

If you have studied probability at GCSE or A-level, you may have seen a definition of probability like this:

Motivation - the need for set theory and measures

If you have studied probability at GCSE or A-level, you may have seen a definition of probability like this:

Suppose all the outcomes in an experiment are equally likely.

Motivation - the need for set theory and measures

If you have studied probability at GCSE or A-level, you may have seen a definition of probability like this:

Suppose all the outcomes in an experiment are equally likely.

The probability of an event A is defined to be

$$P(A) := \frac{\text{number of outcomes in which } A \text{ occurs}}{\text{total number of possible outcomes}}. \qquad (1)$$

Example

Example

A deck of 52 playing cards is shuffled thoroughly. What is the probability that the top card is an ace?

Example

Example

A deck of 52 playing cards is shuffled thoroughly. What is the probability that the top card is an ace?

The ‘definition’ can work for examples such as this (with a finite number of equally likely outcomes), but we soon run into difficulties:

Harder example

Example

Imagine generating a random angle between 0 and 2π radians, where any angle is equally likely.

Harder example

Example

Imagine generating a random angle between 0 and 2π radians, where any angle is equally likely.
(For example, consider spinning a roulette wheel, and measuring the angle between the horizontal axis (viewing the wheel from above) and a line from the centre of the wheel that bisects the zero on the wheel).

Harder example

Example

Imagine generating a random angle between 0 and 2π radians, where any angle is equally likely.
(For example, consider spinning a roulette wheel, and measuring the angle between the horizontal axis (viewing the wheel from above) and a line from the centre of the wheel that bisects the zero on the wheel).
What is the probability that the angle is between 0 and $\sqrt{2}$ radians?

Harder example continued

You might guess that the answer should be $\frac{\sqrt{2}}{2\pi}$, but the definition in (1) causes us problems!

Harder example continued

You might guess that the answer should be $\frac{\sqrt{2}}{2\pi}$, but the definition in (1) causes us problems!

There are infinitely many possible angles that could be generated, and there are infinitely many angles between 0 and $\sqrt{2}$ radians. Clearly, we can't divide infinity by infinity and come up with $\frac{\sqrt{2}}{2\pi}$.

Are outcomes equally likely

It may not make much sense to think of outcomes as equally likely:

Are outcomes equally likely

It may not make much sense to think of outcomes as equally likely:

Example

What is the probability that a team from Manchester will win the Premier League this season?

Are outcomes equally likely

It may not make much sense to think of outcomes as equally likely:

Example

What is the probability that a team from Manchester will win the Premier League this season?

We will resolve this by constructing a more general definition of probability, using the tools of set theory and a special type of function called a **measure**, and we'll show how choosing different types of measure will give us sensible answers in both the above examples.

Set theory and probability

Set theory is covered in more detail in MAS110; in this module we consider set theory in the context of probability.

Set theory and probability

Set theory is covered in more detail in MAS110; in this module we consider set theory in the context of probability.

We consider uncertainty in the context of an **experiment**.

Set theory and probability

Set theory is covered in more detail in MAS110; in this module we consider set theory in the context of probability.

We consider uncertainty in the context of an **experiment**.

(Here we use the word experiment in a loose sense to mean observing something in the future, or discovering the true status of something that we are currently uncertain about.)

Sample space

In an experiment, suppose we want to consider how likely some particular outcome is.

Sample space

In an experiment, suppose we want to consider how likely some particular outcome is.

We might start by considering what all the possible outcomes of the experiment are. We can use a set to list all the possible outcomes.

Sample space

In an experiment, suppose we want to consider how likely some particular outcome is.

We might start by considering what all the possible outcomes of the experiment are. We can use a set to list all the possible outcomes.

Definition

A **sample space** is a set which lists all the possible outcomes of an 'experiment'.

Universal set

In set theory, we sometimes work with a **universal set** S, which lists all the elements we wish to consider for the situation at hand.

Universal set

In set theory, we sometimes work with a **universal set** S, which lists all the elements we wish to consider for the situation at hand.

(Note that, in spite of the word "universal", the choice of S depends on context.)

Universal set

In set theory, we sometimes work with a **universal set** S, which lists all the elements we wish to consider for the situation at hand.

(Note that, in spite of the word "universal", the choice of S depends on context.)

In the context of probability, the sample space will play the role of the universal set S.

Examples

Example

Examples of sample spaces

Events

Definition

An **event** is a subset of a sample space.

Events

Definition

An **event** is a subset of a sample space.
If the outcome of the experiment is a member of the event, we say that the event has **occurred**.

Events

Definition

An **event** is a subset of a sample space.
If the outcome of the experiment is a member of the event, we say that the event has **occurred**.

Example

Examples of events

Set operations and events

Given a sample space S and two subsets/events A and B, the set operations union, intersection, complement and difference all define further events.

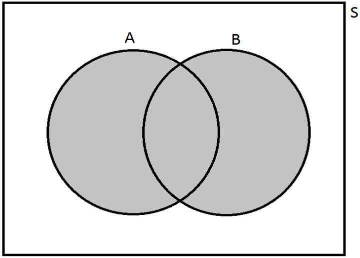
Union

The **union** $A \cup B$ corresponds to either A occurring or to B occuring (*or* to both occurring).

Union

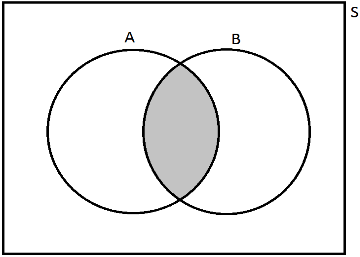
The **union** $A \cup B$ corresponds to either A occurring or to B occuring (*or* to both occurring).

We can visualise this using a *Venn diagram*. The rectangle represents the universal set, the two circles represent the subsets A and B, and the shaded area represents the set $A \cup B$.



Intersection

The **intersection** $A \cap B$ corresponds to both A and B occurring.



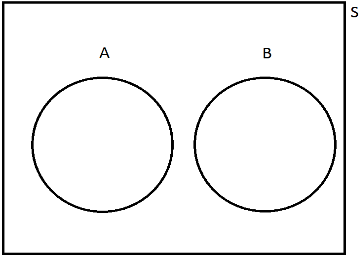
Empty set

If there are no elements in S that are both in A and in B, then $A \cap B = \varnothing$.

Empty set

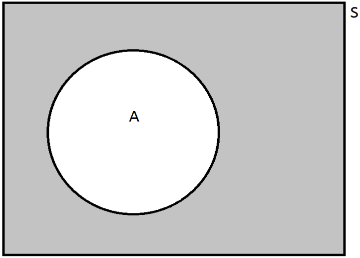
If there are no elements in S that are both in A and in B, then $A \cap B = \varnothing$.

In this case we say that A and B are **mutually exclusive**.



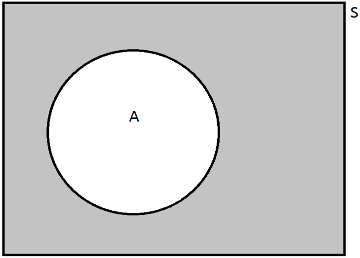
Complement

The **complement** of A, $\overline{A}$, corresponds to A not occurring.



Complement

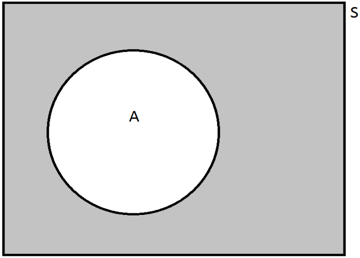
The **complement** of A, $\overline{A}$, corresponds to A not occurring.



(**Alternative notation**: $\bar{A}$ is also written as A^C and A'.)

Complement

The **complement** of A, $\overline{A}$, corresponds to A not occurring.

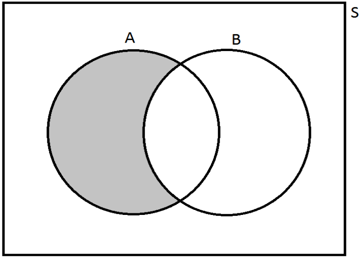


(**Alternative notation**: $\bar{A}$ is also written as A^C and A'.)

Note that $\varnothing = \bar{S}$, and that the event $\varnothing$ is one which cannot happen, as it contains no elements.

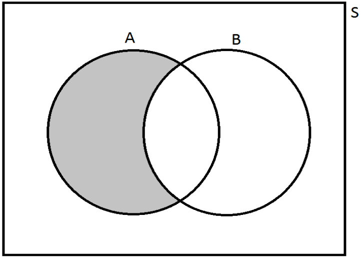
Set difference

The **set difference** $A \setminus B$ corresponds to A occurring but B not.



Set difference

The **set difference** $A \setminus B$ corresponds to A occurring but B not.



Note that

$$A \setminus B = A \cap \bar{B}.$$

De Morgan's Laws

Theorem

(De Morgan's Laws) For any two sets A and B,

$$\begin{aligned}\overline{(A \cup B)} &= \bar{A} \cap \bar{B},\\ \overline{(A \cap B)} &= \bar{A} \cup \bar{B}.\end{aligned}$$

Proof: an exercise in MAS110.

Examples

Example

Combinations of events

The difference between an outcome and an event

Let's clarify the difference between an "outcome" and an "event".

The difference between an outcome and an event

Let's clarify the difference between an "outcome" and an "event".

An outcome is an *element* of the sample space S, whereas an event is a *subset* of S.

The difference between an outcome and an event

Let's clarify the difference between an "outcome" and an "event".

An outcome is an *element* of the sample space S, whereas an event is a *subset* of S.

Consider the following example.

Spain v. Germany

Suppose Spain play Germany at football this evening.

Spain v. Germany

Suppose Spain play Germany at football this evening.

The possible outcomes here can be thought of as ordered pairs of non-negative integers, e.g. $(2, 1)$ representing the outcome that Spain win by 2 goals to 1.

Spain v. Germany

Suppose Spain play Germany at football this evening.

The possible outcomes here can be thought of as ordered pairs of non-negative integers, e.g. $(2, 1)$ representing the outcome that Spain win by 2 goals to 1.

So the sample space is the set of such pairs, $\mathbb{N}_0 \times \mathbb{N}_0$.

Spain v. Germany continued

- Before the match, you can bet on an *event*, eg “Either Spain win or draw”.

Spain v. Germany continued

- Before the match, you can bet on an *event*, eg “Either Spain win or draw”.
- Tomorrow, a news reporter will tell you the *outcome*, eg “Spain won 2-1”. The news reporter would not report the event “either Spain won or the match was drawn last night”.

Spain v. Germany continued

- Before the match, you can bet on an *event*, eg "Either Spain win or draw".
- Tomorrow, a news reporter will tell you the *outcome*, eg "Spain won 2-1". The news reporter would not report the event "either Spain won or the match was drawn last night".
- If the observed outcome belongs to the event "either Spain win or draw", the event has occurred, and you have won your bet.

Spain v. Germany continued

- Before the match, you can bet on an *event*, eg "Either Spain win or draw".
- Tomorrow, a news reporter will tell you the *outcome*, eg "Spain won 2-1". The news reporter would not report the event "either Spain won or the match was drawn last night".
- If the observed outcome belongs to the event "either Spain win or draw", the event has occurred, and you have won your bet.
- An event could be a single outcome (you could bet on "Spain win 2-1").

Notation

For a sample space S, outcome a and event A, we would use the notation $a \in S$ and $A \subset S$. Make sure you understand when to use the symbol $\in$ and when to use the symbol $\subset$ or $\subseteq$.

Set functions

We can define set functions that take sets as their inputs and produce real numbers as outputs.

Set functions

We can define set functions that take sets as their inputs and produce real numbers as outputs.

Not all set functions make sense for all possible sets, so when defining a set function, we'll first say what the sample space is, and then define set functions that operate on subsets of a sample space.

Set functions continued

Example

Examples of set functions

Set functions continued

Example

Examples of set functions

Formally, we can think of a set function as a function whose domain is a set whose elements are subsets of S.

Measures

We're close to being able to give a formal definition of probability: we're going to define probability as a special type of set function that we apply to a set.

Measures

We're close to being able to give a formal definition of probability: we're going to define probability as a special type of set function that we apply to a set.

We first define a **measure** as a special type of set function.

Definition

Definition

Given a sample space (or universal set) S, a (finite) **measure** is a set function that assigns a non-negative real number $m(A)$ to a set $A \subseteq S$, which satisfies the following condition:

Definition

Definition

Given a sample space (or universal set) S, a (finite) **measure** is a set function that assigns a non-negative real number $m(A)$ to a set $A \subseteq S$, which satisfies the following condition:
For any two disjoint subsets A and B of S (ie $A \cap B = \varnothing$),

$$m(A \cup B) = m(A) + m(B). \tag{2}$$

Intuition

Informally, the above definition can be put into words as follows.

Intuition

Informally, the above definition can be put into words as follows.

A measure is something which assigns a number to a set, produces a non-negative number as its output, and 'behaves additively':

Intuition

Informally, the above definition can be put into words as follows.

A measure is something which assigns a number to a set, produces a non-negative number as its output, and 'behaves additively':

If we take two non-overlapping (disjoint) sets, we can either combine the sets and then apply the measure, or apply the measure to each set and then sum the results, and we will get the same value either way.

Finite unions

We can extend the condition in (2) to finite unions of disjoint sets.

Finite unions

We can extend the condition in (2) to finite unions of disjoint sets.

If three sets A, B and C are all mutually disjoint ($A \cap B = A \cap C = B \cap C = \varnothing$), if we define $D = B \cup C$, then

$$m(A \cup B \cup C) \quad = \quad m(A \cup D)$$

Finite unions

We can extend the condition in (2) to finite unions of disjoint sets.

If three sets A, B and C are all mutually disjoint ($A \cap B = A \cap C = B \cap C = \varnothing$), if we define $D = B \cup C$, then

$$\begin{aligned} m(A \cup B \cup C) &= m(A \cup D) \\ &= m(A) + m(D) \end{aligned}$$

Finite unions

We can extend the condition in (2) to finite unions of disjoint sets.

If three sets A, B and C are all mutually disjoint ($A \cap B = A \cap C = B \cap C = \varnothing$), if we define $D = B \cup C$, then

$$\begin{aligned} m(A \cup B \cup C) &= m(A \cup D) \\ &= m(A) + m(D) \\ &= m(A) + m(B \cup C) \end{aligned}$$

Finite unions

We can extend the condition in (2) to finite unions of disjoint sets.

If three sets A, B and C are all mutually disjoint ($A \cap B = A \cap C = B \cap C = \varnothing$), if we define $D = B \cup C$, then

$$\begin{aligned} m(A \cup B \cup C) &= m(A \cup D) \\ &= m(A) + m(D) \\ &= m(A) + m(B \cup C) \\ &= m(A) + m(B) + m(C). \end{aligned}$$

Examples

Example

Examples of measures

Properties of measures

Theorem

(Properties of measures)
Let m be a measure on some set S, with $A, B \subseteq S$.

Properties of measures

Theorem

(Properties of measures)
Let m be a measure on some set S, with $A, B \subseteq S$.

(M1) *If $B \subseteq A$, then*

$$m(A \setminus B) = m(A) - m(B).$$

Properties of measures

Theorem

(Properties of measures)
Let m be a measure on some set S, with $A, B \subseteq S$.

(M1) *If $B \subseteq A$, then*

$$m(A \setminus B) = m(A) - m(B).$$

More generally,
$m(A \setminus B) = m(A \setminus (A \cap B)) = m(A) - m(A \cap B)$.

Properties of measures

Theorem

(Properties of measures)
Let m be a measure on some set S, with $A, B \subseteq S$.

(M1) *If $B \subseteq A$, then*

$$m(A \setminus B) = m(A) - m(B).$$

More generally,
$m(A \setminus B) = m(A \setminus (A \cap B)) = m(A) - m(A \cap B).$

(M2) *If $B \subseteq A$, then*

$$m(B) \leq m(A).$$

Properties of measures

Theorem

(Properties of measures)
Let m be a measure on some set S, with $A, B \subseteq S$.

(M1) *If $B \subseteq A$, then*

$$m(A \setminus B) = m(A) - m(B).$$

More generally,
$m(A \setminus B) = m(A \setminus (A \cap B)) = m(A) - m(A \cap B).$

(M2) *If $B \subseteq A$, then*

$$m(B) \leq m(A).$$

(M3) $m(\varnothing) = 0$

Properties of measures continued

Theorem

(Properties of measures)

(M4) $m(A \cup B) = m(A) + m(B) - m(A \cap B)$.

Properties of measures continued

Theorem

(Properties of measures)

(M4) $m(A \cup B) = m(A) + m(B) - m(A \cap B)$.

(M5) *For any constant* $c > 0$, *the function* $g(A) = c \times m(A)$ *is also a measure.*

Properties of measures continued

Theorem

(Properties of measures)

(M4) $m(A \cup B) = m(A) + m(B) - m(A \cap B)$.

(M5) *For any constant* $c > 0$, *the function* $g(A) = c \times m(A)$ *is also a measure.*

(M6) *If* m *and* n *are both measures, the function* $h(A) = m(A) + n(A)$ *is also a measure.*

Measures on partitions of sets

Definition

A **partition** of a set S is a collection of sets
$\mathcal{E} = \{E_1, E_2, \dots, E_n\}$

Measures on partitions of sets

Definition

A **partition** of a set S is a collection of sets $\mathcal{E} = \{E_1, E_2, \ldots, E_n\}$ such that

1. $E_i \cap E_j = \varnothing$, for all $1 \leq i, j \leq n$ with $i \neq j$;

Measures on partitions of sets

Definition

A **partition** of a set S is a collection of sets $\mathcal{E} = \{E_1, E_2, \dots, E_n\}$ such that

1. $E_i \cap E_j = \varnothing$, for all $1 \leq i, j \leq n$ with $i \neq j$;
2. $E_1 \cup E_2 \cup \dots \cup E_n = S$.

Measures on partitions of sets

Definition

A **partition** of a set S is a collection of sets $\mathcal{E} = \{E_1, E_2, \dots, E_n\}$ such that

1. $E_i \cap E_j = \varnothing$, for all $1 \leq i, j \leq n$ with $i \neq j$;
2. $E_1 \cup E_2 \cup \dots \cup E_n = S$.

We say that $\{E_1, E_2, \dots, E_n\}$ are *mutually exclusive* and *exhaustive*.

Measures on partitions of sets

Definition

A **partition** of a set S is a collection of sets $\mathcal{E} = \{E_1, E_2, \dots, E_n\}$ such that

1. $E_i \cap E_j = \varnothing$, for all $1 \leq i, j \leq n$ with $i \neq j$;
2. $E_1 \cup E_2 \cup \dots \cup E_n = S$.

We say that $\{E_1, E_2, \dots, E_n\}$ are *mutually exclusive* and *exhaustive*.

If S is a sample space, one of the events in $\mathcal{E}$ *must* occur, but no two events can occur simultaneously.

Partitions continued

From (2), we have

$$m(S) = \sum_{i=1}^{n} m(E_i).$$