University of BRISTOL

Population, random samples and elementary set theory

Henry W J Reeve

henry.reeve@bristol.ac.uk

Statistical Computing & Empirical Methods (EMATM0061)

MSc in Data Science, Teaching block 1, 2021.

UOB Open

What will we cover today?

- We will introduce the fundamental problem of stochastic variability;

- We introduced the concepts of a random experiment, sample space and event;

- We introduced some fundamental concepts from elementary set theory:

  Intersections, unions, subsets, compliments;

  Cardinality, countable & uncountable infinities;

- We discussed how these set-theoretic concepts can be used to reason about events.

The problem of variability

Adelie

Chinstrap

Are Adelie penguins lighter than Chinstrap penguins?

The problem of variability

We attempt to answer such questions by looking at data.

Our data sets are samples from a much larger population of penguins.

Sample

Population

The problem of variability

											Mean
Adelie	4100	3050	3100	3800	3500	3350	3400	3550	4150	3625	3562
Chinstrap	3600	3650	4800	4400	3800	4400	3500	4500	3500	3300	$\boxed{3945}$

The problem of variability

											Mean
Adelie	4100	3050	3100	3800	3500	3350	3400	3550	4150	3625	3562
Chinstrap	3600	3650	4800	4400	3800	4400	3500	4500	3500	3300	$\boxed{3945}$

											Mean
Adelie	3550	3550	3950	2925	4775	3900	3550	4000	3950	3300	$\boxed{3745}$
Chinstrap	2700	3325	3650	3950	3800	4300	4050	3900	3675	3700	3705

The problem of variability

											Mean
Adelie	4100	3050	3100	3800	3500	3350	3400	3550	4150	3625	3562
Chinstrap	3600	3650	4800	4400	3800	4400	3500	4500	3500	3300	3945

											Mean
Adelie	3550	3550	3950	2925	4775	3900	3550	4000	3950	3300	3745
Chinstrap	2700	3325	3650	3950	3800	4300	4050	3900	3675	3700	3705

Different samples can lead to different conclusions!

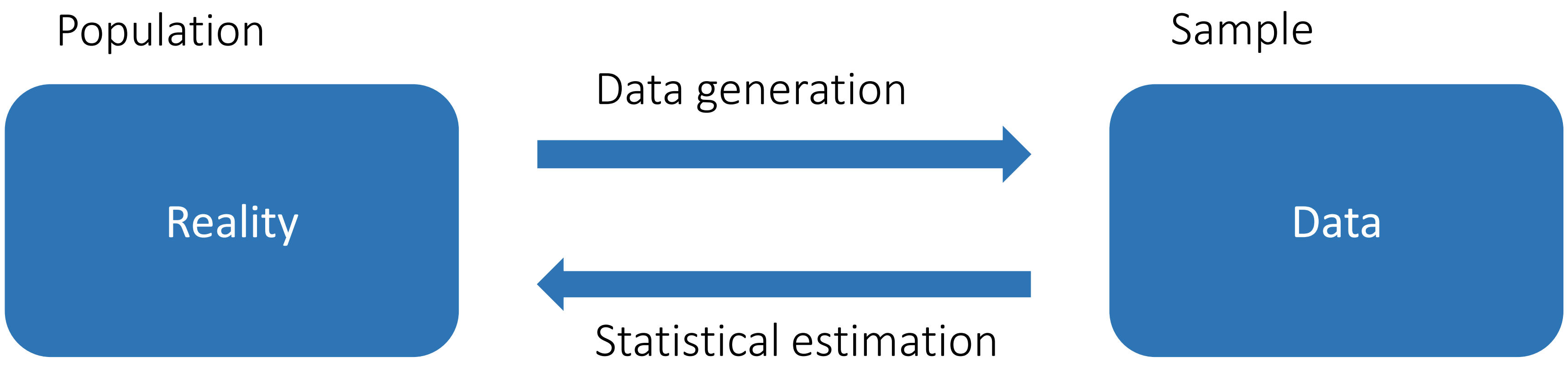
The problem of variability

Even if we use a larger sample the problem of variability persists:

- We can’t weigh every penguin in an entire species..
- We can’t try a new marketing idea on all possible customers…
- We can’t test a new medication on all patients current and future..

We must think about how are finite sample reflects a larger distribution

Statistical estimation and probability



To model the data generation process we will require some probability theory!

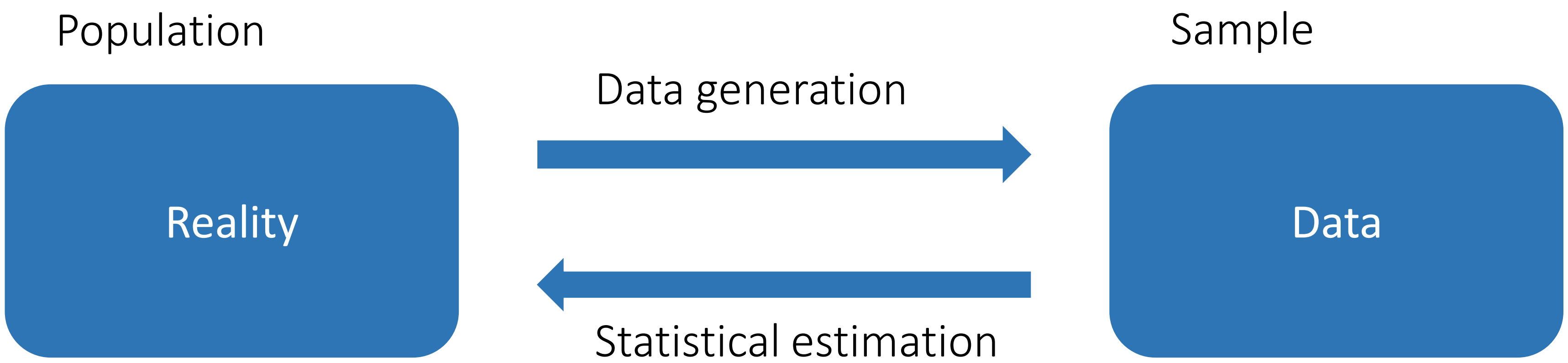
Key motivating questions

We will be exploring how to use **statistics** $\hat{\theta}$ to learn about population quantities θ.

Key questions

1. How can we design an effective statistic $\hat{\theta}$ for a parameter of interest θ?
2. How can we quantify our uncertainty about the quantity θ?
3. How can we use statistics $\hat{\theta}$ to test hypotheses about θ?
4. How can we use our understanding of the population to make predictions about new data?

Statistical estimation and probability



To model the data generation process we will require some probability theory!

Random experiments

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

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

Examples

1. A coin flip for a coin;
2. The roll of a ten dice;
3. A cricket ball is thrown by an athlete;
4. A customer goes into a shop and decides whether to buy coffee or tea.

Note: This is a very broad definition of experiment. There is no suggestion that the experiment is designed or controlled, although this connotation is typical in the natural sciences.

Random experiments and sample spaces

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

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

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

	Random experiment	**Example event**
1.	A coin flip for a coin:	The coin lands head up;
2.	Rolling ten dice:	We roll the sequence $(1,2,3,4,5,1,2,3,4,5)$;
3.	A cricket ball is thrown:	The cricket ball reaches a speed of 10mph;
4.	A customer goes into a shop:	The customer buys coffee.

Random experiments and sample spaces

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

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

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

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

	Random experiment	Sample space
1.	A coin flip for a coin:	Two ways of landing - “heads” & “tails” (we could also include on its side);
2.	Rolling ten dice:	All length ten sequences of numbers between one & six;
3.	A cricket ball is thrown:	All possible trajectories for the cricket ball;
4.	A customer goes into a shop:	All possible purchases (and no purchase at all).

Now take a break!

UOB Open

Elementary set theory

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

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

To reason about **events** and **sample spaces** we use some elementary set-theory.

A **set** is just a collection of objects of interest (our interest is in sets of possible outcomes).

Examples

1. The set $\mathbb{N}$ consisting of all positive whole numbers;
2. The set $\mathbb{R}$ consisting of all real numbers;
3. The set $[0, 1]$ consisting of all real numbers between zero and one;
4. The empty-set $\emptyset$ which doesn't contain any objects.

UOB Open

Elementary set theory

We often use curly braces $\{\ldots\}$ (containing a list) to denote finite sets of objects.

Example: $\{1,2,3,4,5\}$ denotes the set of whole numbers less than or equal to five.

We write $x \in A$ to denote that x is an element of the set A ,

Example: We have $1 \in \{1,2,3,4,5\}$.

We write $x \notin A$ to denote that x is not an element of the set A .

Example: We have $6 \notin \{1,2,3,4,5\}$.

Given a set A and a property F we write $\{x \in A \; : \; F(x)\}$ for the set of all elements x in the set A which satisfy the property F .

Examples: If $A = \{1,2,3,4,5\}$ then $\{x \in A : x \text{ is odd}\} = \{1,3,5\}$ and $\{x \in A : x \text{ is even}\} = \{2,4\}$.

Elementary set theory: Finite & infinite sets

The cardinality of a set is just the number of elements

We say a set A is finite if the cardinality of A is a non-negative integer i.e. $n \in \{0, 1, 2, 3, \ldots\}$.

We say that a set is infinite whenever it is not finite.

Examples

1. The empty set $\emptyset$ is finite with cardinality zero;
2. The sets $\mathrm{A} = \{1,2,3,4,5\}$ and $\mathrm{B} = \{2,4,6,8,10\}$ both have cardinality 5;
3. The set $\mathbb{N}$ consisting of all natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$;
4. The set $\mathbb{Q} = \{\pm m/n \ : \ m, n \in \mathbb{N} \cup \{0\}\}$ of all rational numbers;
5. The set $\mathbb{R}$ consisting of all real numbers;

UOB Open

Elementary set theory: Finite & infinite sets

The cardinality of a set is just the number of elements

We say a set A is finite if the cardinality of A is a non-negative integer i.e. $n \in \{0, 1, 2, 3, \dots\}$.

We say that a set is infinite whenever it is not finite.

Examples

1. The empty set $\emptyset$ is finite with cardinality zero;
2. The sets $A = \{1,2,3,4,5\}$ and $B = \{2,4,6,8,10\}$ both have cardinality 5;
3. The set $\mathbb{N}$ consisting of all natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$;
4. The set $\mathbb{Q} = \{\pm m/n \ : \ m, n \in \mathbb{N} \cup \{0\}\}$ of all rational numbers;
5. The set $\mathbb{R}$ consisting of all real numbers;

Finite

Infinite

Elementary set theory: Countable & uncountably infinities

We say an infinite set A is countably infinite if there exists an enumeration $a_1, a_2, \ldots, a_n, a_{n+1}, \ldots$

such that $A = \{a_1, a_2, \ldots, a_n, a_{n+1}, \ldots\} = \{a_n : n \in \mathbb{N}\}$

We say that a set A is uncountably infinite whenever A is infinite but not countably infinite.

Examples

1. The set $\mathbb{N}$ consisting of all natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$;
2. The set of all even numbers $\{2, 4, 6, 8, \ldots\}$; — Countable
3. The set $\mathbb{Q} = \{\pm m/n \ : \ m, n \in \mathbb{N} \cup \{0\}\}$ of all rational numbers;
4. The set $\mathbb{R}$ consisting of all real numbers;
5. Open intervals $(a, b) := \{x \in \mathbb{R} : a < x < b\}$ for real numbers $a, b \in \mathbb{R}$; — Uncountable
6. Closed intervals $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ for real numbers $a, b \in \mathbb{R}$;

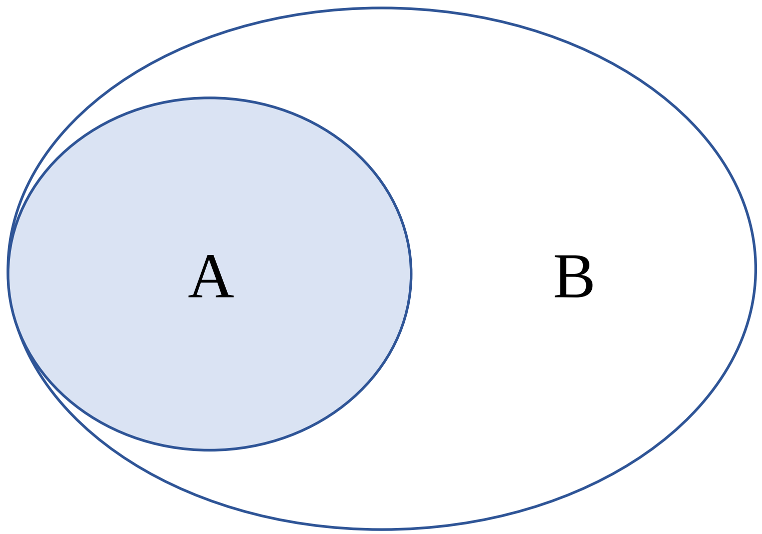
Elementary set theory: Subsets

Suppose that A and B are sets.

We say that A is a **subset** of B if every element of A is also an element of B.

We write $A \subseteq B$.

The event A implies the event B.



Example

$\{1,2,3\} \subseteq \{1, 2, 3,4,5\}$

Given any event A and a sample space Ω we have $A \subseteq \Omega$.

Elementary set theory: Complements

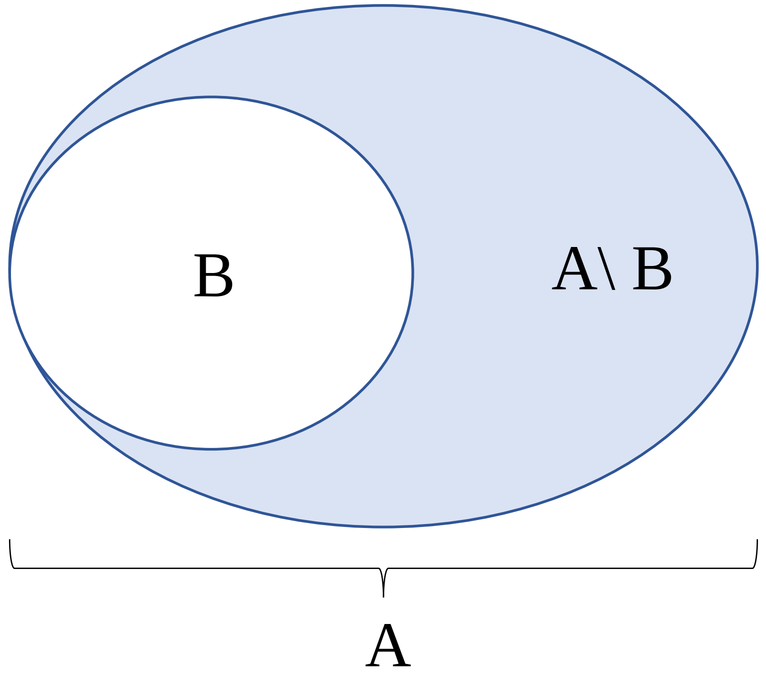
Suppose that A and B are sets.

We write $A \setminus B$ for the **complement** of B in A.

More precisely, $A \setminus B := \{x \in A : x \notin B\}$,

consisting of all elements in A and not B.

$A \setminus B$ is also referred to as the set difference between A & B



Example

$\{4,5\} = \{1, 2, 3,4,5\} \setminus \{1,2,3\}.$

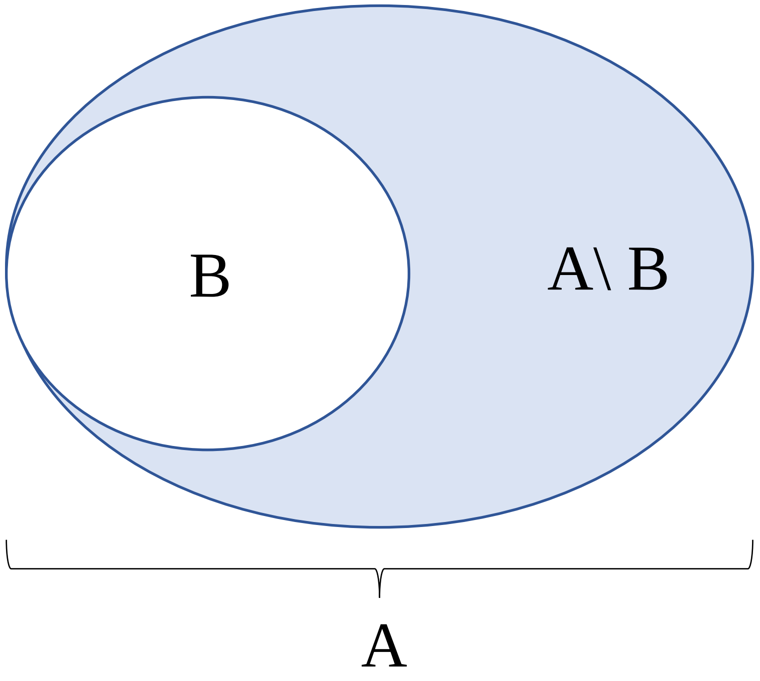
Elementary set theory: Complements

Suppose that A and B are sets.

We write $A \setminus B$ for the **complement** of B in A.

More precisely, $A \setminus B := \{x \in A : x \notin B\}$,

consisting of all elements in A and not B.



Important example

Suppose A is an event, Ω is a sample space and $A \subseteq \Omega$.

Then $\Omega \setminus A$ is the compliment event in which A does not occur.

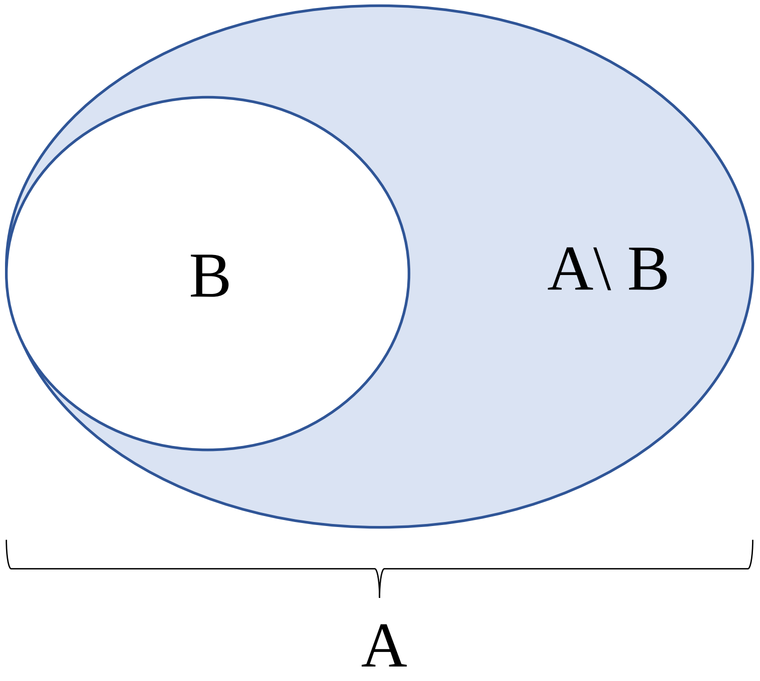
Elementary set theory: Complements

Suppose that A and B are sets.

We write $A \setminus B$ for the **complement** of B in A .

More precisely, $A \setminus B := \{x \in A : x \notin B\}$,

consisting of all elements in A and not B .



Important example

Suppose A is an event, Ω is a sample space and $A \subseteq \Omega$.

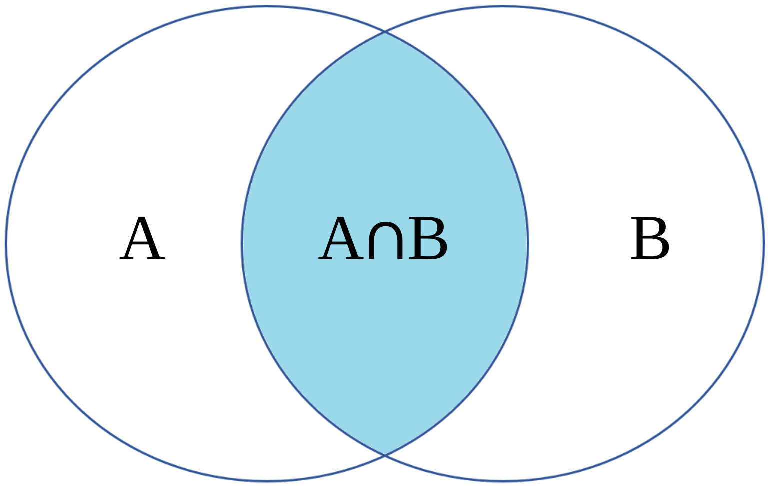
Then $\Omega \setminus A$ is the compliment event in which A does not occur. We often just write A^c.

Elementary set theory: Intersections

Suppose that A and B are sets.

We write $A \cap B$ for the set of all elements in both and of A and B .

We refer to $A \cap B$ as the **intersection** of A and B .



Example

1) $\{3\} = \{1, 2, 3\} \cap \{3,4,5\}$.

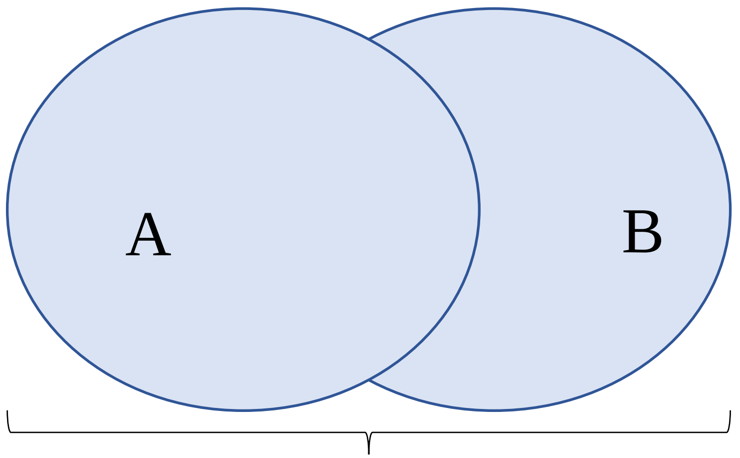
2) Given events A and B, the intersection $A \cap B$ denotes the event where both A & B occur.

Elementary set theory: Unions

Suppose that A and B are sets.

We write $A \cup B$ for the set of all elements in either of A or B (or both) .

We refer to $A \cup B$ as the **union** of A and B .



$A \cup B$

Example

1) $\{1,2,3,4,5\} = \{1,2,3\} \cup \{3,4,5\}$.

2) Given events A and B, the union $A \cup B$ denotes the event where at least one of A, B occur.

UOB Open

Elementary set theory: Intersections and unions

We can also have intersections and unions of many set $A_1, \ldots, A_N \subseteq \Omega$

$$\bigcap_{n=1}^{N} A_n := A_1 \cap \ldots \cap A_N := \{x \in \Omega : x \in A_n \text{ for all } n = 1, \ldots, N\}$$

$$\bigcup_{n=1}^{N} A_n := A_1 \cup \ldots \cup A_N := \{x \in \Omega : x \in A_n \text{ for at least one } n = 1, \ldots, N\}$$

Elementary set theory: Intersections and unions

We can also have intersections and unions of many set $A_1, \ldots, A_N \subseteq \Omega$

$$\bigcap_{n=1}^{N} A_n := A_1 \cap \ldots \cap A_N := \{x \in \Omega : x \in A_n \text{ for all } n = 1, \ldots, N\}$$

$$\bigcup_{n=1}^{N} A_n := A_1 \cup \ldots \cup A_N := \{x \in \Omega : x \in A_n \text{ for at least one } n = 1, \ldots, N\}$$

Given infinitely many sets $A_1, A_2, A_3, \ldots$

$$\bigcap_{n \in \mathbb{N}} A_n := \bigcap_{n=1}^{\infty} A_n := \{x \in \Omega : x \in A_n \text{ for all } n \in \mathbb{N}\}$$

$$\bigcup_{n \in \mathbb{N}} A_n := \bigcup_{n=1}^{\infty} A_n := \{x \in \Omega : x \in A_n \text{ for at least one } n \in \mathbb{N}$$

Elementary set theory and sample spaces

An **event** is a set (i.e. a collection) of possible outcomes for a random experiment.

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

We can use ideas from set theory to reason about events and sample spaces:

Given a pair of events A and B we have:

1. $A \subseteq B$ means that event A implies event B;

2. $A \cap B$ (the intersection) denotes the event in which both A and B occur;
3. $A \cup B$ (the union) denotes the event in which at least one of A or B occur;
4. $A \setminus B$ (the compliment) denotes the event in which A occurs but B does not occur.

Indicator functions

Let $A \subseteq \Omega$ be a set (or event).

We can associate A with binary function $\mathbb{1}_A : \Omega \to \{0, 1\}$ for $\omega \in \Omega$, by

$$\mathbb{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The function $\mathbb{1}_A$ is referred to as the indicator function of A.

Indicator functions

We can associate A with binary function $\mathbb{1}_A : \Omega \to \{0,1\}$ for $\omega \in \Omega$, by

$$\mathbb{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Given a pair of sets A,$B \subseteq \Omega$ we have:

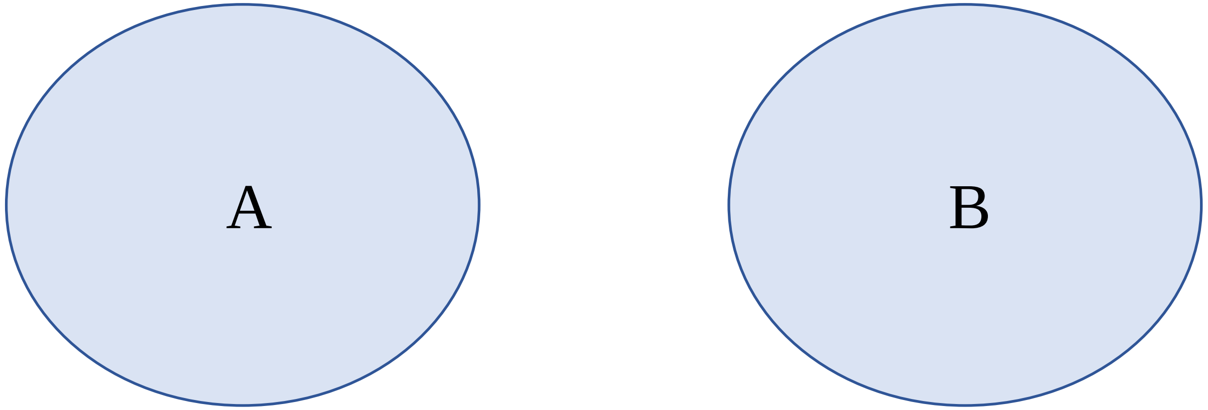
1. If $A \subseteq B$ then $\mathbb{1}_A(\omega) \leq \mathbb{1}_B(\omega)$ for all $\omega \in \Omega$;
2. We have $\mathbb{1}_{A \cap B}(\omega) = \mathbb{1}_A \cdot \mathbb{1}_B$ for all $\omega \in \Omega$;
3. We have $\mathbb{1}_{A \cup B}(\omega) = \max\{\mathbb{1}_A(\omega), \mathbb{1}_B(\omega)\}$ for all $\omega \in \Omega$;
4. We have $\mathbb{1}_{A \setminus B}(\omega) = \mathbb{1}_A(\omega) \cdot (1 - \mathbb{1}_B(\omega))$ for all $\omega \in \Omega$.

Disjoint sets

A pair of sets A and B are said to be **disjoint** if $A \cap B = \emptyset$.

As events A and B are disjoint if they cannot both occur.

Example The events "Manchester wins the league" and "Liverpool wins league" are disjoint.



Disjoint sets and partitions

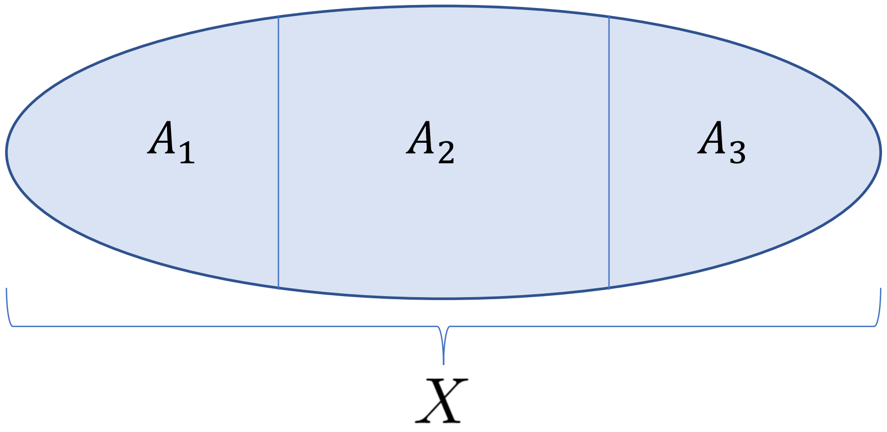
A pair of sets A and B are said to be **disjoint** if $A \cap B = \emptyset$.

As events A and B are disjoint if they cannot both occur.

We say a family of sets $A_1, \ldots, A_K$ is said to be **pair-wise disjoint** if $A_i \cap A_j = \emptyset$ for $i \neq j$.

A **partition** of a set X is a family $A_1, \ldots, A_K$ consisting of pair-wise disjoint sets for which

$$X = A_1 \cup \ldots \cup A_K = \bigcup_{k \in \{1, \ldots, K\}} A_k .$$



What have we covered?

- We introduced the fundamental problem of stochastic variability;
- We introduced the concepts of a random experiment, sample space and event;
- We introduced some fundamental concepts from elementary set theory:

  Intersections, unions, subsets, compliments;

  Cardinality, countable & uncountable infinities;
- We discussed how these set-theoretic concepts can be used to reason about events;
- Next we will look at how these concepts allow us to formalize the fundamental rules of probability.

University of BRISTOL

Thanks for listening!

Henry W J Reeve

henry.reeve@bristol.ac.uk

Include EMATM0061 in the subject of your email.

UOB Open