Lecture 2

Probability I

In this and the following lecture we will learn about

- why we need to learn about probability
- what probability is
- how to assign probabilities
- how to manipulate probabilities and calculate probabilities of complex events

2.1 Why do we need to learn about probability?

In Lecture 1 we discussed why we need to study statistics, and we saw that statistics plays a crucial role in the scientific process (see figure 2.1). We saw that we use a sample from the population in order to test our hypothesis. There will usually be a very large number of possible samples we could have taken and the conclusions of the statistical test we use will depend on the exact sample we take. It might happen that the sample we take leads us to make the wrong conclusion about the population. Thus we need to know what the chances are of this happening. Probability can be thought of as the study of chance.

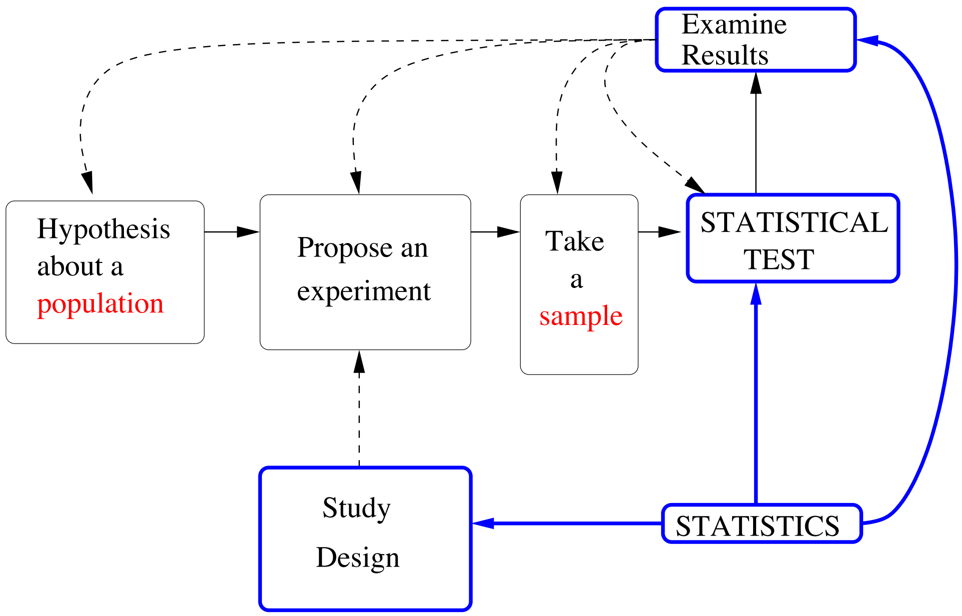


Figure 2.1: The scientific process and role of statistics in this process.

Example 2.1: Random controlled experiment

The Anturane Reinfarction Trial (ART) was a famous study of a drug treatment (anturane) for the aftereffects of myocardial infarction [MDF+81]. Out of 1629 patients, about half (813) were selected to receive anturane; the other 816 patients received an ineffectual ("placebo") pill. The results are summarised in Table 2.1.

Table 2.1: Results of the Anturane Reinfarction Trial.

	Treatment (anturane)	Control (placebo)
# patients	813	816
deaths	74	89
% mortality	9.1%	10.9%

Imagine the following dialogue:

> **Drug Company**: Every hospital needs to use anturane. It saves patients' lives.
> **Skeptical Bureaucrat**: The effect looks pretty small: 15 out of about 800 patients. And the drug is pretty expensive.
> **DC**: Is money all you bean counters can think about? We reduced mortality by 16%.
> **SB**: It was only 2% of the total.
> **DC**: We saved 2% of the patients! What if one of them was your mother?
> **SB**: I'm all in favour of saving 2% more patients. I'm just wondering: You flipped a coin to decide which patients would get the anturane. What if the coins had come up differently? Might we just as well be here talking about how anturane had killed 2% of the patients?

How can we resolve this argument? 163 patients died. Suppose anturane has no effect. Could the apparent benefit of anturane simply reflect the random way the coins fell? Or would such a series of coin flips have been simply too unlikely to countenance? To answer this question, we need to know how to measure the likelihood (or "probability") of sequences of coin flips.

Imagine a box, with cards in it, each one having written on it one way in which the coin flips could have come out, and the patients allocated to treatments. How many of those coinflip cards would have given us the impression that anturane performed well, purely because many of the patients who died happened to end up in the Control (placebo) group? It turns out that it's more than 20% of the cards, so it's really not very unlikely at all.

To figure this out, we are going to need to understand

1. How to enumerate all the ways the coins could come up. How many ways are there? The number depends on the exact procedure, but if we flip one coin for each patient, the number of cards in the box would be 2^{1629}, which is vastly

more than the number of atoms in the universe. Clearly, we don't want to have to count up the cards individually.

2. How coin flips get associated with a result, as measured in apparent success or failure of anturane. Since the number of "cards" is so large, we need to do this without having to go through the results one by one.

■

Example 2.2: Baby-boom

Consider the Baby-boom dataset we saw in Lecture 1. Suppose we have a hypothesis that in the population boys weigh more than girls at birth. We can use our sample of boys and girls to examine this hypothesis. One intuitive way of doing this would be to calculate the mean weights of the boys and girls in the sample and compare the difference between these two means

$$\text{Sample mean of boys weights} = \bar{x}_{\text{boys}} = 3375$$
$$\text{Sample mean of girls weights} = \bar{x}_{\text{girls}} = 3132$$
$$\Rightarrow D = \bar{x}_{\text{boys}} - \bar{x}_{\text{girls}} = 3375 - 3132 = 243$$

Does this allow us to conclude that in the population boys are born heavier than girls? On what scale to we assess the size of D? Maybe boys and girls weigh the same at birth and we obtained a sample with heavier boys just by chance. To be able to conclude confidently that boys in the population are heavier than girls we need to know what the chances are of obtaining a difference between the means that is 243 or greater, i.e. we need to know the probability of getting such a large value of D. If the chances are small then we can be confident that in the population boys are heavier than girls on average at birth. ■

2.2 What is probability?

The examples we have discussed so far look very complicated. They aren't really, but in order to see the simple underlying structure, we need to introduce a few new concepts. To do so, we want to work with a much simpler example:

Example 2.3: Rolling a die

Consider a simple **experiment** of rolling a fair six-sided die.

When we toss the die there are six possible outcomes i.e. 1, 2, 3, 4, 5 and 6. We say that the **sample space** of our experiment is the set $S = \{1, 2, 3, 4, 5, 6\}$.

The outcome "the top face shows a three" is the **sample point** 3.

The **event** A_1, that the die shows an even number is the subset $A_1 = \{2, 4, 6\}$ of the sample space.

The **event** A_2 that the die shows a number larger than 4 is the subset $A_2 = \{5, 6\}$ of S_2. ■

2.2.1 Definitions

The example above introduced some terminology that we will use repeatedly when we talk about probabilities.

An **experiment** is some activity with an observable outcome.

The set of all possible outcomes of the experiment is called the **sample space**.

A particular outcome is called a **sample point**.

A collection of possible outcomes is called an **event**.

2.2.2 Calculating simple probabilities

Simply speaking, the probability of an event is a number between 0 and 1, inclusive, that indicates how likely the event is to occur.

In some settings (like the example of the fair die considered above) it is natural to assume that all the sample points are equally likely.
In this case, we can calculate the probability of an event A as

$$P(A) = \frac{|A|}{|S|},$$

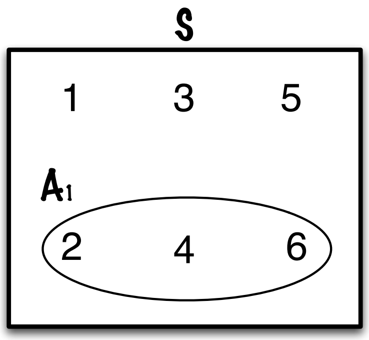
where $|A|$ denotes the number of sample points in the event A.

2.2.3 Example 2.3 continued

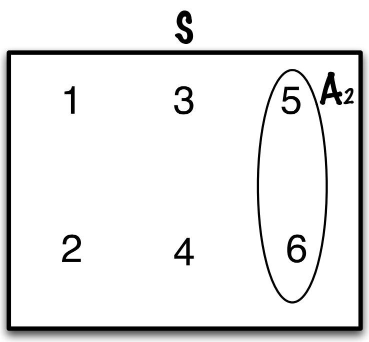
It is often useful in simple examples like this to draw a diagram (known as a "Venn diagram") to represent the sample space, and then label specific events in the diagram by grouping together individual sample points.

$S = \{1, 2, 3, 4, 5, 6\}$ $A_1 = \{2, 4, 6\}$ $A_2 = \{5, 6\}$

$$P(A_1) = \frac{|A_1|}{|S|} = \frac{3}{6} = \frac{1}{2}$$



$$P(A_2) = \frac{|A_2|}{|S|} = \frac{2}{6} = \frac{1}{3}$$



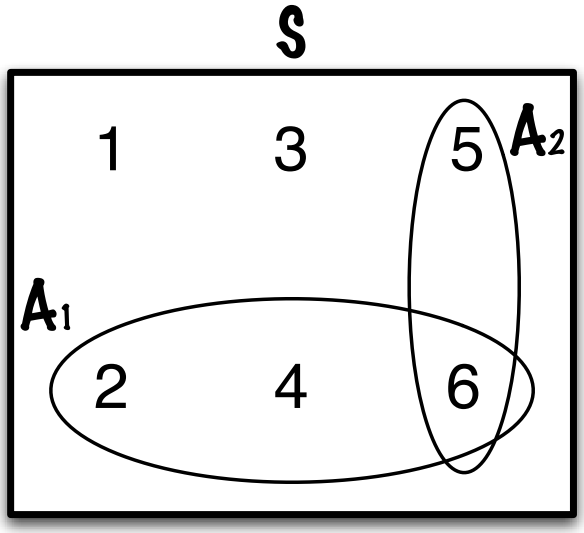
2.2.4 Intersection

What about P(face is even, **and** larger than 4) ?

We can write this event in set notation as $A_1 \cap A_2$.

This is the **intersection** of the two events, A_1 and A_2
i.e the set of elements which belong to both A_1 and A_2.

$$A_1 \cap A_2 = \{6\} \quad \Rightarrow \quad P(A_1 \cap A_2) = \frac{|A_1 \cap A_2|}{|S|} = \frac{1}{6}$$



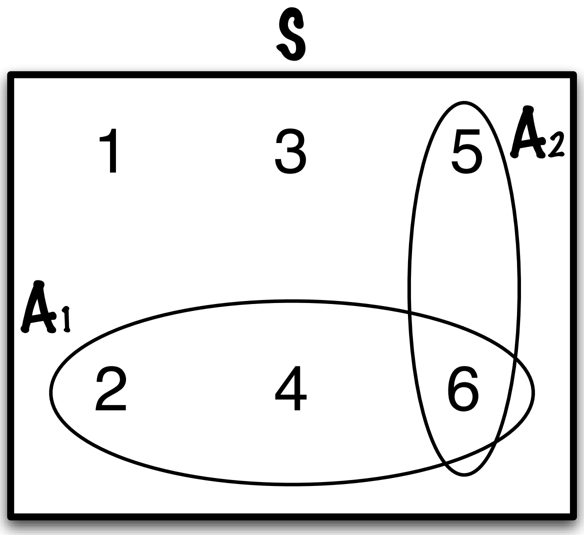
2.2.5 Union

What about P(face is even, **or** larger than 4) ?

We can write this event in set notation as $A_1 \cup A_2$.

This is the **union** of the two events, A_1 and A_2
i.e the set of elements which belong either A_1 and A_2 or both.

$$A_1 \cup A_2 = \{2, 4, 5, 6\} \quad \Rightarrow \quad P(A_1 \cup A_2) = \frac{|A_1 \cup A_2|}{|S|} = \frac{4}{6} = \frac{2}{3}$$



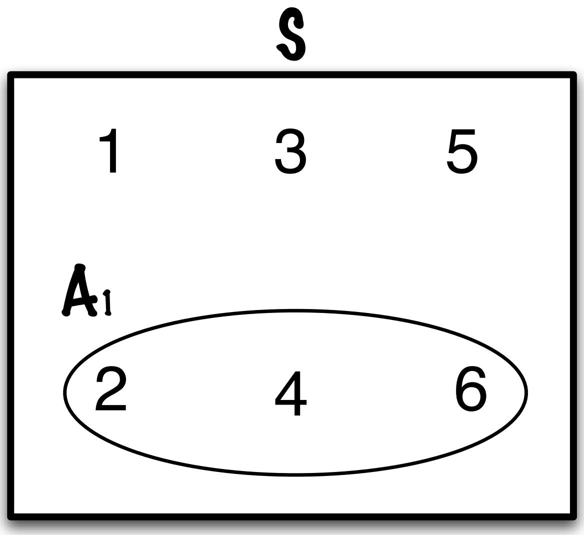
2.2.6 Complement

What about P(face is **not** even) ?

We can write this event in set notation as A_1^c.

This is the **complement** of the event, A_1
i.e the set of elements which do not belong to A_1.

$$A_1^c = \{1, 3, 5\} \quad \Rightarrow \quad P(A_1^c) = \frac{|A_1^c|}{|S|} = \frac{3}{6} = \frac{1}{2}$$



2.3 Probability in more general settings

In many settings, either the sample space is infinite or all possible outcomes of the experiment are not equally likely. We still wish to associate probabilities with events of interest. Luckily, there are some rules/laws that allow us to calculate and manipulate such probabilities with ease.

2.3.1 Probability Axioms (Building Blocks)

Before we consider the probability rules we need to know about the axioms (or mathematical building blocks) upon which these rules are built. There are three axioms which we need in order to develop our laws

(i). $0 \leq P(A) \leq 1$ for any event A.
This axiom says that probabilities must lie between 0 and 1

(ii). $P(S) = 1$.
This axiom says that the probability of everything in the sample space is 1. This says that the sample space is complete and that there are no sample points or events that allow outside the sample space that can occur in our experiment.

(iii). If $A_1, \ldots, A_n$ are **mutually exclusive** events, then

$$P(A_1 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n).$$

A set of events are **mutually exclusive** if at most one of the events can occur in a given experiment.
This axiom says that to calculate the probability of the union of distinct events we can simply add their individual probabilities.

2.3.2 Complement Law

If A is an event, the set of all outcomes that are not in A is called the **complement** of the event A, denoted A^{C}.

This is (pronounced 'A complement').

The rule is

$$\boxed{A^{\mathsf{C}} = 1 - P(A)} \qquad \text{(Law 1)}$$

Example 2.4: Complements

Let S (the sample space) be the set of students at Oxford. We are picking a student at random.

Let A = The event that the randomly selected student suffers from depression

We are told that 8% of students suffer from depression, so $P(A) = 0.08$. What is the probability that a student does not suffer from depression?

The event {student does not suffer from depression} is A^{C}. If $P(A) = 0.08$ then $P(A^{\mathsf{C}}) = 1 - 0.08 = 0.92$. ■

2.3.3 Addition Law (Union)

Suppose,

A	=	The event that a randomly selected student from the class has brown eyes
B	=	The event that a randomly selected student from the class has blue eyes

What is the probability that a student has brown eyes **OR** blue eyes?

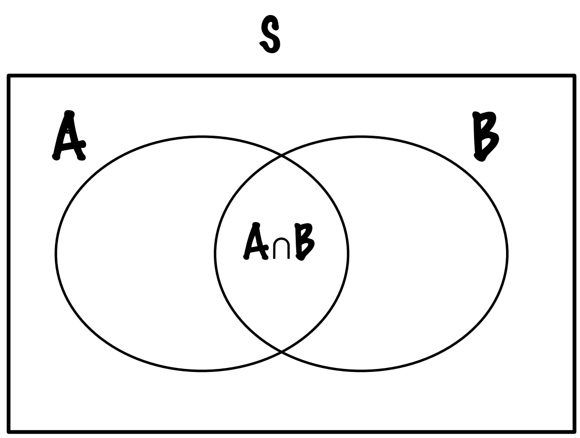
This is the **union** of the two events A and B, denoted A∪B (pronounced 'A or B')

We want to calculate P(A∪B).

In general for two events

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)} \qquad \text{(Addition Law)}$$

To understand this law consider a Venn diagram of the situation (below) in which we have two events A and B. The event $A \cup B$ is represented in such a diagram by the combined sample points enclosed by A or B. If we simply add $P(A)$ and $P(B)$ we will count the sample points in the intersection $A \cap B$ twice and thus we need to subtract $P(A \cap B)$ from $P(A) + P(B)$ to calculate $P(A \cup B)$.



Example 2.5: SNPs

Single nucleotide polymorphisms (SNPs) are nucleotide positions in a genome which exhibit variation amongst individuals in a species. In some studies in humans, SNPs are discovered in European populations. Suppose that of such SNPs, 70% also show variation in an African population, 80% show variation in an

Asian population and 60% exhibit variation in both the African and Asian population.

Suppose one such SNP is chosen at random, what is the probability that it is variable in either the African or the Asian population?

Write A for the event that the SNP is variable in Africa, and B for the event that it is variable in Asia. We are told

$$\begin{aligned} P(A) &= 0.7 \\ P(B) &= 0.8 \\ P(A \cap B) &= 0.6. \end{aligned}$$

We require $P(A \cup B)$. From the addition rule:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.7 + 0.8 - 0.6 \\ &= 0.9. \end{aligned}$$

■