**Dr Isabella Gollini**

**Probability and Statistics**

# 2 The Mathematical Theory of Probability

## 2.1 Events

Define an *experiment* (or *trial*) to be any process of observation or measurement, which produces an *outcome*. Often we envisage that the experiment may be repeated again and again, each time resulting in a fresh outcome. The set of all possible outcomes of the experiment is called the *sample space*, $S$.

Historically, many of the basic ideas of probability theory were worked out in the context of games of chance, in which the "experiment" often involved the throwing of dice. Consider, for example, the simple experiment of throwing a single die and observing the outcome, which may be any of the numbers 1, 2, 3, 4, 5, 6. In this case, we may write

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

In mathematical terms, an *event* is a collection of possible outcomes of the experiment, a subset of the sample space, denoted by a capital letter such as $A$ or $B$ .

In our simple example of the throw of a die, the event that "an odd number occurs", $A$, is given by

$$A = \{1, 3, 5\}.$$

For any given realization of the experiment, we can say whether a particular event $A$ has occurred, i.e., whether any of the outcomes contained in $A$ has occurred. If $a$ denotes the outcome of the experiment then the event $A$ occurs if and only if $a \in A$.

As particular extreme cases, we have the *null event*, i.e., the empty set $\emptyset$, and the *certain event*, i.e. the sample space $S$.

Any possible individual outcome $a$ may also be referred to as a *simple event*, $A = \{a\}$, an event that comprises a single outcome.

**Complement** Given $S$ and $A$, the *complement* of $A$, denoted by $A^c$, is the event that $A$ does not occur. It consists of the outcomes that are not contained in $A$.

In our example,

$$A^c = \{2, 4, 6\},$$

the event that an even number occurs.

**Inclusion** Let $A$ and $B$ be two events. We say that $A$ is *included* in $B$, or that the occurrence of $A$ *implies* the occurrence of $B$, if all the outcomes of $A$ are also outcomes of $B$, i.e., if $A$ is a subset of $B$, $A \subseteq B$.

Note that if $A \subseteq B$ then $a \in A \Rightarrow a \in B$.

**Union** The *union* of the events $A$ and $B$, $A \cup B$, is the event that either $A$ or $B$ or both occur, the set of all outcomes that belong to $A$ or $B$ or both.

**Intersection** The *intersection* of the events $A$ and $B$, $A \cap B$, is the event that both $A$ and $B$ occur, the set of all outcomes which belong to both $A$ and $B$.

**Mutual exclusion** Two events $A$ and $B$ are said to be *mutually exclusive* (or *disjoint*) if they have no outcomes in common, i.e., if $A \cap B = \emptyset$, i.e., if the occurrence of one excludes the occurrence of the other.

In our example let $B$ be the event that a score of 4 or less occurs,

$$B = \{1, 2, 3, 4\}.$$

In this case,

$$A \cup B = \{1, 2, 3, 4, 5\},$$

the event that a score of 5 or less occurs, and

$$A \cap B = \{1, 3\},$$

the event that 1 or 3 occurs.

We shall use some of the basic results of set theory — in particular, the *distributive laws*:

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

## 2.2 Probability

Probability is a number assigned to an event, a measure of how likely the event is to occur when the experiment is carried out. Write $\Pr(A)$ for the probability of the event $A$. If in an experiment the event $A$ cannot occur then $\Pr(A) = 0$. If $A$ is certain to occur then $\Pr(A) = 1$. In general

$$0 \leq \Pr(A) \leq 1.$$

The larger the value of $\Pr(A)$ the more likely is the event $A$ to occur.

### 2.2.1 Relative frequency

If we want to explore the interpretation of the concept of probability in more depth, we may consider a number of approaches. One of these is through the idea of probability as *relative frequency*. Consider an experiment that can be repeated indefinitely often, e.g., the toss of a coin, with $S = \{h, t\}$ in an obvious notation, so that we get a sequence of outcomes such as *hhthhtttt*....

For an event $A$, define $f_N(A)$ to be the frequency of the occurrence of the event $A$ in the first $N$ trials, i.e., the number of times that an outcome that belongs to $A$ occurs in the first $N$ trials. In our example, if we take $A = \{h\}$, the event that a head occurs, then $f_N(A) = f_N(h)$ is the number of heads in the first $N$ tosses.

It is an experimentally observable fact that in general, as $N$ increases, the *relative frequency* $f_N(A)/N$, the proportion of times that the event $A$ occurs, appears to converge to some limiting value $p$, say, which then necessarily satisfies $0 \le p \le 1$. This value is taken to be the probability $\Pr(A)$. Thus, as $N \to \infty$,

$$\frac{f_N(A)}{N} \to \Pr(A).$$

In other words, the probability of an event is the long-term proportion of times that the event occurs, i.e., the relative frequency with which the event occurs, as the experiment is repeated arbitrarily often. Such a probability could in principle be determined, at least approximately, through repeated experimentation. The relative frequency interpretation of probability is sometimes referred to as *statistical probability* or *objective probability.*

For instance, consider the simulated coin tossing data presented in Figure 1. Here, after 1000 tosses, the relative frequency of "heads" appears to be converging to a limiting value of about 0.5, which then, to the extent that the data allows us to determine it, is approximately the probability that "heads" occurs for any given toss of the coin.

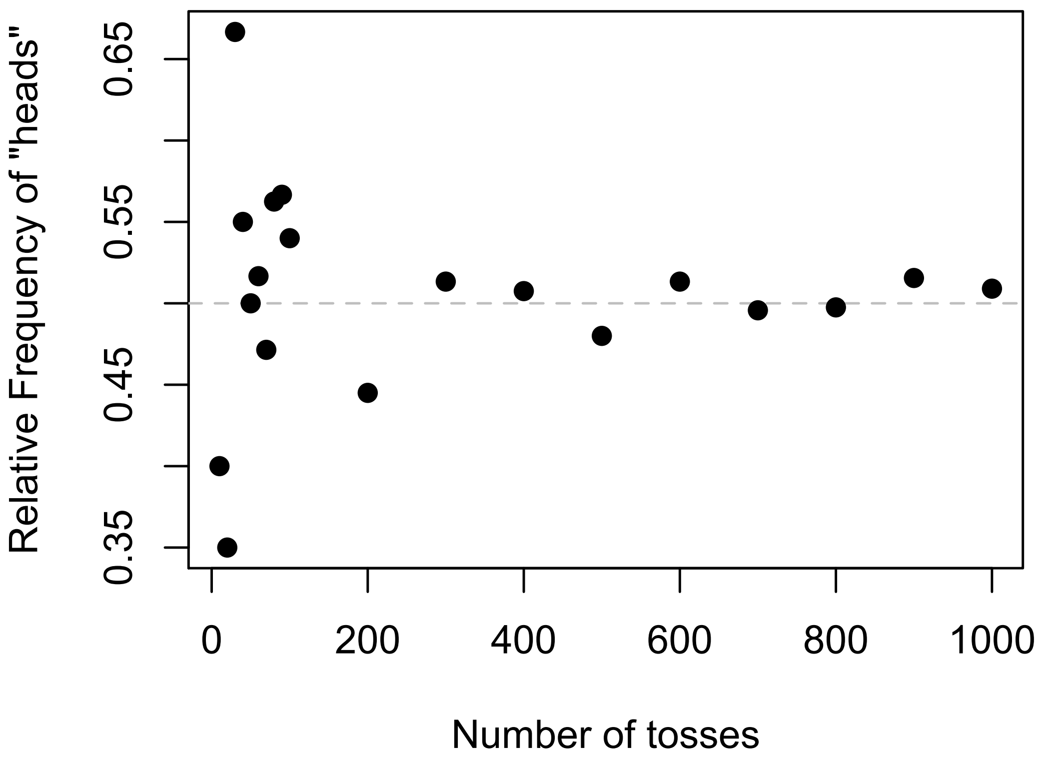


Figure 1: Simulated coin tossing data

From this relative frequency approach, we may deduce a property that probabilities should satisfy. Let $A$ and $B$ be a pair of mutually exclusive events. Then

$$f_N(A \cup B) = f_N(A) + f_N(B).$$

Dividing through by $N$ and letting $N \to \infty$, we obtain the result that, for any pair of mutually exclusive events $A$ and $B$,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$

### 2.2.2 Subjective probability

An alternative to the relative frequency approach to probability is through the idea of what is known as *personal probability* or *subjective probability* — probability which is a measure of an individual's strength of belief in the likelihood of the occurrence of some event. So if we say that the probability of "heads" occurring at the next toss of a coin is $\frac{1}{2}$ then this reflects our belief that the coin is fair, i.e., that "heads" is as likely to occur as the alternative of "tails".

More generally, subjective probability is a number assigned by an individual to an event or a proposition, a measure of how likely he/she thinks that the event is to occur or that the proposition is true, given the information currently available to the individual. Probability in this sense reflects an individual's personal relationship to an event or proposition. Different individuals may well have different probabilities for the same event or proposition, and an individual's probability may change over time as he/she acquires more information. For example, what is the probability that the Conservative Party will win an overall majority at the next general election? Note that in this example the idea of probability as relative frequency is not applicable, as we are not in a situation where the experiment of holding the election can be repeated again and again.

### 2.2.3 Equally likely outcomes

In practice for many probability calculations we may be able to assume that all the outcomes in the sample space $S$ are equally likely to occur. In the experiment of tossing a coin with $S = \{h, t\}$, assuming that the coin is "fair" or "unbiased", the two outcomes $h$ and $t$ may be taken to be equally likely, so that $\Pr(h) = \Pr(t) = \frac{1}{2}$.

In many classical probability problems that involve, for example, the throwing of dice, there is usually an element of symmetry in that the underlying experiment is assumed to have a number of equally likely outcomes. The probability of any event $A$ is then calculated to be the ratio of the number of outcomes in $A$ to the total number of outcomes in the sample space $S$,

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

where $|A|$ denotes the number of outcomes/elements in the event/set $A$.

For example, in the simple experiment of throwing a single die, discussed in Section 2.1, $|S| = 6$. If $B$ is the event that a score of 4 or less occurs, $B = \{1, 2, 3, 4\}$, then $|B| = 4$, so that $\Pr(B) = 4/6 = 2/3$.

In general, if $A$ and $B$ are mutually exclusive events then

$$|A \cup B| = |A| + |B|.$$

It follows that

$$\Pr(A \cup B) = \frac{|A \cup B|}{|S|} = \frac{|A| + |B|}{|S|} = \frac{|A|}{|S|} + \frac{|B|}{|S|} = \Pr(A) + \Pr(B).$$

So again we see the result that for any pair of mutually exclusive events $A$ and $B$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$

## 2.3 The probability axioms

Irrespective of what approach we take to interpreting the concept of probability, mathematically probability is defined as below. Although in the simplest cases the sample space $S$ is finite, it can be any set, finite or infinite.

For a given sample space $S$, probability is a function Pr defined on events in $S$ that satisfies the following axioms:

1. For any event $A$,
$$\Pr(A) \geq 0.$$
2. $$\Pr(S) = 1.$$
3. For any finite or infinite sequence of pairwise mutually exclusive events, $A_1, A_2, \ldots, A_i, \ldots$,
$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i).$$

- Axiom 3 is an extension of what we demonstrated in Section 2.2 that, for any pair of mutually exclusive events $A$ and $B$, $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- By a sequence of pairwise mutually exclusive events, $A_1, A_2, \ldots, A_i, \ldots$ we mean that $A_i \cap A_j = \emptyset$ for all pairs $i \neq j$.

**Lemma 1** *For any event $A$,*
$$\Pr(A^c) = 1 - \Pr(A).$$

*Proof.* $A \cap A^c = \emptyset$ and $A \cup A^c = S$. Hence applying Axiom 3 and Axiom 2 successively,
$$\Pr(A) + \Pr(A^c) = \Pr(A \cup A^c) = \Pr(S) = 1,$$
from which the result of the lemma follows.

**Corollary 2** *For any event $A$,* $\Pr(A) \leq 1$.

*Proof.* Applying Axiom 1, $\Pr(A^c) \geq 0$. Hence, from Lemma 1, $1 - \Pr(A) \geq 0$.

**Corollary 3** $\Pr(\emptyset) = 0$.

*Proof.* $\emptyset = S^c$. Hence, from Lemma 1 together with Axiom 2,
$$\Pr(\emptyset) = 1 - \Pr(S) = 1 - 1 = 0.$$

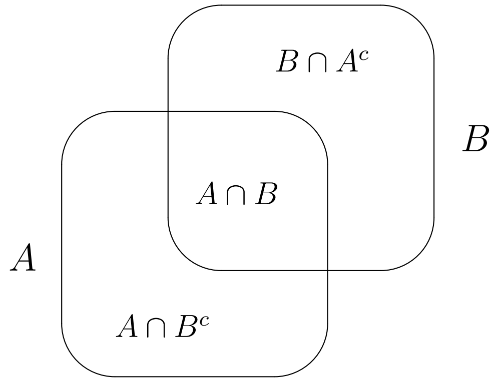


Figure 2: The union and intersection of two events $A$ and $B$

**Lemma 4** *For any pair of events $A$ and $B$,*

$$\Pr(A \cap B^c) = \Pr(A) - \Pr(A \cap B).$$

*Proof.* [See Figure 1 for illustration.] Consider the pair of events $A \cap B$ and $A \cap B^c$. They are mutually exclusive since

$$(A \cap B) \cap (A \cap B^c) \subseteq B \cap B^c = \emptyset.$$

Using the distributive law for sets,

$$(A \cap B) \cup (A \cap B^c) = A \cap (B \cup B^c) = A \cap S = A.$$

Hence by Axiom 3

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c),$$

from which the result of the lemma follows.

**Theorem 5** *For any pair of events $A$ and $B$,*

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

*Proof.* [See Figure 1 for illustration.] Consider the pair of events $A \cap B^c$ and $B$. They are mutually exclusive since

$$(A \cap B^c) \cap B \subseteq B^c \cap B = \emptyset.$$

Using the distributive law for sets,

$$(A \cap B^c) \cup B = (A \cup B) \cap (B^c \cup B) = (A \cup B) \cap S = A \cup B.$$

Hence by Axiom 3

$$\Pr(A \cup B) = \Pr(A \cap B^c) + \Pr(B).$$

Using the result of Lemma 4,

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

**Example**

Consider drawing a card at random from a standard pack of 52 cards. What is the probability that a spade or a king is drawn?

$|S| = 52$. Let $A$ be the event that a spade is drawn. Let $B$ be the event that a king is drawn. Thus $|A| = 13$ and $|B| = 4$, from which it follows that $\Pr(A) = 13/52 = 1/4$ and $\Pr(B) = 4/52 = 1/13$.

Now $A \cap B = \{\spadesuit K\}$, so that $|A \cap B| = 1$ and hence $\Pr(A \cap B) = 1/52$.

The event that a spade or a king is drawn is $A \cup B$. Using the result of Theorem 4,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \frac{1}{4} + \frac{1}{13} - \frac{1}{52} = \frac{4}{13}.$$

## 2.4 The birthday problem

In a group of $r$ people, what is the probability $p(r)$ that no pair of them have birthdays on the same day of the year?

Ignoring the leap day issue, and making the simplifying approximation that birthdays are evenly distributed throughout the year, we may use the equally likely outcomes approach. Using some simple combinatorial mathematics,

$$p(r) = \frac{365 \times 364 \times \ldots \times (365 - r + 1)}{365^r}.$$

Table 2 gives the values of $p(r)$ for some selected values of $r$. Note that for $r$ as small as 23 it is more likely that there are some common birthdays than not. In reality, since birthdays are not evenly distributed throughout the year, it is even more likely that there are some common birthdays than the tabulated values imply.

| r | p(r) |
|---|---|
| 10 | 0.883 |
| 20 | 0.589 |
| 21 | 0.556 |
| 22 | 0.524 |
| 23 | 0.493 |
| 24 | 0.462 |
| 25 | 0.431 |
| 26 | 0.402 |
| 27 | 0.373 |
| 28 | 0.346 |
| 29 | 0.319 |
| 30 | 0.294 |
| 40 | 0.109 |
| 50 | 0.030 |

Table 1: Probabilities of no common birthdays for groups of size $r$