Probability: Intuition en Information

In this short course we will study various problems in which chance plays a role. We will investigate how to solve such problems in a systematic way using a subfield of Mathematics called Probability Theory. As we shall see, Probability Theory is based on only a few axioms and definitions. Everything else we do follows from these axioms and definitions by pure logical reasoning and a bit of algebra and calculation.

We will start this course by asking you to solve four puzzles involving chance that people often get wrong because their intuition fails. Do not worry if right now you get the answers wrong: the whole point is to show you that in problems involving chance, you cannot always rely on your intuition.

1 Probability puzzles

The puzzle of the sum of the dice

You have agreed with a friend that you will throw a regular (six-sided) die twice wearing a blindfold, and that he will tell you the sum of the two numbers of dots that you have thrown. You then have to guess whether or not you have thrown a six (at least once).

As agreed, you throw the die twice and your friend tells you that the sum of the two throws is eight. What is the probability that you have thrown a six?

The puzzle of the cards

You have two (playing-)cards. One card has a cross on both sides, the other card has a cross on one side and nothing on the other side. You place the two cards in a box, and shake it well. Then, without looking, you take one card from the box and place it flat on the table. It turns out that on the side of the card facing up, there is a cross. What is the probability that there is also a cross on the other side of the card?

The puzzle of the lucky prisoner

Three prisoners, named A, B and C, are in prison for life. One day, the prison’s governor is in a particularly good mood, and decides that the
following day, one of the prisoners will be released. He decides who will be released on the basis of a fair lottery (meaning that every prisoner has the same chance of being the lucky one).

The guard, who knows who will be released, cannot wait to tell the prisoners the good news, but does not want to give away any information that will change a prisoner’s chances of being the lucky prisoner. So he tells the prisoners, “One of you will be released tomorrow, but I cannot tell you who it is.”

Prisoner A, however, asks the guard in private to toss a fair coin and tell him the name of one of his fellow prisoners (B or C) who will not be released, using the outcome of the coin toss to decide between the names if he has to (and ignore the coin toss otherwise). “After all”, he argues, “I already know that one of them will not be released, so giving me a name will not influence my chances of being the lucky one.” But the guard refuses, claiming that if he gave away a name, this would increase prisoner A’s chances of being released to $1/2$. Who is right?

The puzzle of the positive test result

Approximately 1% of all women between 40 and 50 years old have breast cancer. Suppose that a test for breast cancer is known to have the following characteristics: if a woman does have breast cancer, the result of the test will be positive with a probability of 80%; if a woman does not have breast cancer, the result of the test will be negative with a probability of 90%.

If a woman takes the test and receives a positive test result, then how likely is it that this woman does have breast cancer?

Did you believe the answer to the first puzzle is $1/3$, because there are three combinations of which the sum is eight? Perhaps you thought that the probability that there is also a cross on the other side of the card is $1/2$? Were you convinced that the guard is right? And did you estimate the probability that the woman has breast cancer to be around 80 to 90%?

In all these cases, the spirits of chance fooled you: all these answers are wrong. But not to worry: in this course, we will explain and investigate how we can solve such puzzles in a systematic way using Probability Theory, so that we no longer have to rely on our intuition. Later on in this course, we will return to the four puzzles presented above, and by the end of this course, you will have learned how to solve them.
As you work through this document, please verify that you understand all explanations and all steps in the calculations. Study the examples carefully, and please try to make all the exercises. This will help you to understand the material better.

2 What is a probability?

We assume that you have an intuitive idea about probabilities as a measure for how likely it is that “something” happens. You are probably aware that probabilities lie between 0 and 1 (= 100%), that the probability of “everything” is 1, and that you are allowed to add probabilities when they concern “different things”. We will make these ideas more precise later on. But before we can do that, we have to introduce a few elementary notions used throughout mathematics.

Sets

A set is an unordered collection of elements. Examples are the set of all integer numbers larger than 0, or the set of all fractions between 0 and 1. But the elements of sets do not have to be numbers. The set of all mathematics students at the Vrije Universiteit is also a correct example of a set. Sets play an important role in Probability Theory (although in this context we often call them events, as we shall see).

We usually denote sets by listing their elements between braces (curly brackets). For example, the set consisting of the integer numbers from 1 through 6 is denoted as \{1, 2, 3, 4, 5, 6\}. To simplify the notation when a set contains many elements, we can use dots to denote intermediate elements. By default, the dots stand for all intermediate integer numbers, but if the listing before the dots follows a specific pattern, the dots mean that the listing is to be continued in the same pattern. For example, this allows us to write \{1, 2, 3, 4, 5, 6\} more simply as \{1, \ldots, 6\}, and to denote the set of all even numbers from 0 through 20 as \{0, 2, 4, \ldots, 20\}.

Two sets $A$ and $B$ are equal if they contain exactly the same elements. We then write $A = B$. Since sets are considered to be unordered collections of elements, we have that \{1, 2, 3, 4\} = \{4, 3, 2, 1\} = \{1, 3, 2, 4\}, and so on. It is also immaterial how often an element of a set is listed between the braces, so that for instance \{1, 1, 1, 2\} = \{1, 2, 2\} = \{1, 2\}. 
A set $A$ is called a *subset* of a set $B$ if every element of $A$ is also an element of $B$. We then write $A \subset B$. For example, $\{1, 2\} \subset \{1, 2, 3, 4\}$ and $\{10\} \subset \{0, 2, 4, \ldots, 20\}$. A very special set is the set that has no elements at all. This set is called the *empty set*, and is denoted by the symbol $\emptyset$.

**Sets in Probability Theory**

In Probability Theory we study *experiments* (activities or procedures) of which the outcome is determined by chance. A classical example is throwing a (six-sided) die. There are in this case six possible outcomes, that together form the set $\{1, \ldots, 6\}$, and you cannot predict beforehand which of these outcomes will be realised. The set of all possible outcomes of an experiment is called the *sample space*. We will denote this sample space by the letter $S$.

So for a single toss of a die we have $S = \{1, \ldots, 6\}$.

Often we are not interested in the precise outcome of an experiment, but rather in the probability that a specific statement about the outcome will be true. We call the set of those outcomes that satisfy such a statement an *event*. For example, if we toss a die, we could be interested in whether or not we will throw an even number of dots. We are then interested in the event \{Even number of dots\}, which mathematically is a subset of the sample space $S$, namely $\{\text{Even number of dots}\} = \{2, 4, 6\}$.

**Example 2.1.** If we toss a coin once, the sample space is $S = \{H, T\}$, where $H$ stands for Heads and $T$ for Tails. If we toss the coin more than once, we write each possible outcome as a sequence of $H$’s and $T$’s, where the $i$-th $H$ or $T$ in the sequence indicates whether the $i$-th toss of the coin was Heads or Tails. For example, when we toss a coin two times, the outcome that the first toss was Heads and the second Tails is denoted by $(H, T)$. The sample space for tossing a coin two times is therefore

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

and the event \{First toss is Heads\}, for example, is in this case equal to $\{(H, H), (H, T)\}$.

**Combining events**

When we throw a die, sometimes we are interested in events such as the event that an even number of dots *and* more than four dots were thrown, or the
event that an even number of dots or more than four dots were thrown. We call these events respectively the intersection and union of the two events $A = \{\text{Even number of dots}\}$ and $B = \{\text{More than four dots}\}$. The intersection is denoted as $A \cap B$ and consists of those outcomes that belong to both $A$ and $B$. The union is denoted as $A \cup B$ and consists of those outcomes that belong to at least one of the two events $A$ and $B$ (so the union includes outcomes that belong to both $A$ and $B$).

We can also be interested in the probability that a specific event $A$ does not occur. To help up talk about this, we introduce the complement $A^c$ of the event $A$. By definition, $A^c$ consists of all those outcomes in the sample space that are not elements of the event $A$. With this definition, the event $A$ does not occur precisely when the event $A^c$ does occur, and vice versa.

**Example 2.2.** If we toss a die, we have the two events

$$A = \{\text{Even number of dots}\} = \{2, 4, 6\}$$
$$B = \{\text{More than four dots}\} = \{5, 6\}$$

We see that the outcome 6 is an element of both $A$ and $B$. In the union of $A$ and $B$, however, we only have to write down the outcome 6 once. So the union and intersection are equal to $A \cup B = \{2, 4, 5, 6\}$ and $A \cap B = \{6\}$. The complements of the events $A$ and $B$ are, respectively, $A^c = \{1, 3, 5\}$ and $B^c = \{1, 2, 3, 4\}$.

**Example 2.3.** We toss a coin twice. In mathematical notation, the two events $A = \{\text{First toss Heads}\}$ and $B = \{\text{Second toss Tails}\}$ are

$$A = \{(H, H), (H, T)\}$$
$$B = \{(H, T), (T, T)\}$$

So the union and intersection of $A$ and $B$ are

$$A \cup B = \{(H, H), (H, T), (T, T)\}$$
$$A \cap B = \{(H, T)\}$$

and the complements of $A$ and $B$ are $A^c = \{(T, H), (T, T)\}$ and $B^c = \{(H, H), (T, H)\}$.

We say that two events $A$ and $B$ are disjoint if they have no elements in common, that is, if $A \cap B = \emptyset$. For example, when we toss a die, the events $\{\text{Even number of dots}\}$ and $\{\text{Odd number of dots}\}$ are disjoint. Note that, by definition, any event and its complement are always disjoint.
Definition of Probability

We are now in a position to give a precise definition of probability. Suppose we consider an experiment with a finite sample space $S$. Then we define a probability function $P$ on $S$ as a function that assigns to every event $A$ a number $P(A)$ with the following three properties:

1. For all events $A$, $0 \leq P(A) \leq 1$.
2. $P(S) = 1$.
3. If the events $A$ and $B$ are disjoint, then $P(A \cup B) = P(A) + P(B)$.

Observe that these three properties correspond to the intuitive ideas about probabilities mentioned at the start of this section. From now on, we will refer to the three properties above as, respectively, the first, second and third probability axiom.

Often, all possible outcomes of an experiment are believed to have the same probability (usually because of a symmetry). We claim that if this is the case, then the three probability axioms imply that the probability of an event $A$ can be computed using the formula

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S}$$

To see this, suppose that the sample space $S$ consists of $n$ outcomes that are equally likely, and the event $A$ contains $k$ outcomes ($k \leq n$). Let us denote the outcomes in $A$ by $a_1, a_2, \ldots, a_k$. Then, by the third probability axiom,

$$P(\{a_1, a_2\}) = P(a_1) + P(a_2),$$
$$P(\{a_1, a_2, a_3\}) = P(\{a_1, a_2\}) + P(a_3) = P(a_1) + P(a_2) + P(a_3),$$

and so on, so that in particular,

$$P(A) = P(\{a_1, \ldots, a_k\}) = P(a_1) + P(a_2) + \cdots + P(a_k) = kP(a_1),$$

where the last step follows because all outcomes have the same probability. In the special case $A = S$, this equation implies that $P(S) = nP(a_1) = 1$ by the second probability axiom. Hence $P(a_1) = 1/n$, so that the probability of an event $A$ consisting of $k$ outcomes is given by $P(A) = kP(a_1) = k/n$.

**Example 2.4.** We toss a die once. As we know, the sample space is then $S = \{1, \ldots, 6\}$. Since a die is symmetric, the probability of every outcome is equally likely. This means that

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S} = \frac{\text{number of outcomes in } A}{6}$$
A few examples will make this clear. The event \( A = \{ \text{Even number of dots} \} \) contains three outcomes: \( A = \{2, 4, 6\} \). Hence, \( P(A) = 3/6 = 1/2 \). The event \( B = \{ \text{More than four dots} \} = \{5, 6\} \) consists of two outcomes, so \( P(B) = 2/6 = 1/3 \). Furthermore (as we saw before), \( A \cup B = \{2, 4, 5, 6\} \) and \( A \cap B = \{6\} \), from which it follows that \( P(A \cup B) = 2/3 \) and \( P(A \cap B) = 1/6 \).

In all the examples considered so far, every outcome of the experiment was equally likely. This is not always the case.

**Example 2.5.** In a football match, Robin has to shoot a penalty. He makes a choice between six directions for his aim:

\[
S = \{ \text{low left, high left, low centre, high centre, low right, high right} \}
\]

Suppose that the probabilities with which Robin chooses his aim are

\[
\begin{align*}
P(\text{low left}) &= 0.2 & P(\text{low centre}) &= 0.05 & P(\text{low right}) &= 0.2 \\
P(\text{high left}) &= 0.15 & P(\text{high centre}) &= 0.1 & P(\text{high right}) &= 0.3
\end{align*}
\]

Using the third probability axiom, it now follows that

\[
\begin{align*}
P(\text{left}) &= P(\text{low left}) + P(\text{high left}) = 0.35 \\
P(\text{low}) &= P(\text{low left}) + P(\text{low centre}) + P(\text{low right}) = 0.45
\end{align*}
\]

**Exercises**

**Question 2.1.** Of each statement below, decide whether it is true or false:
1. \( 0 \in \{1, \ldots, 100\} \)
2. \( -2 \in \{-100, -98, -96, \ldots, 100\} \)
3. \( \{3, 3, 3, 2, 2, 1\} = \{1, 2\} \)
4. \( \{15, 13, 11, \ldots, 1\} = \{1, 3, 5, \ldots, 15\} \)
5. \( \{1, 4, 9\} \subset \{1, 4\} \)
6. \( \{1, 4\} \subset \{1, 4, 9\} \)
7. \( \emptyset \subset \{1, 4, 9\} \)

**Question 2.2.** Consider an experiment in which we toss a fair coin twice. Is it true or false that the two events \( \{ \text{First toss Heads} \} \) and \( \{ \text{First toss Tails} \} \) are disjoint? And is it true or false that the two events \( \{ \text{First toss Heads} \} \) and \( \{ \text{Second toss Heads} \} \) are disjoint?

**Question 2.3.** Suppose we toss a fair coin three times. What is the sample space \( S \) for this experiment?
Question 2.4. Consider again an experiment in which we toss a fair coin three times. Explicitly write out the four events $A = \{\text{First toss Heads}\}$, $B = \{\text{Even number of Heads}\}$, $A \cap B$ and $A \cup B$ as a set of outcomes from the sample space $S$ determined in the previous exercise. What are the respective probabilities of these four events?

Question 2.5. Suppose we throw a regular six-sided die twice. If, for example, we first throw 5 dots and then 2 dots, we will denote the outcome as $(5, 2)$. What is the sample space $S$ for this experiment, and how many outcomes does this sample space consist of?

Question 2.6. Again suppose we throw a six-sided die twice. Consider the events

$$A = \{\text{Sum of the throws is 7}\}$$
$$B = \{\text{Sum of the throws is 8}\}$$
$$C = \{\text{Second throw is 5}\}$$

and also the events $A \cap C$ and $A \cup C$. Write each of these events explicitly as a set of outcomes from the sample space $S$ determined in the previous exercise. Use this to calculate $P(A)$, $P(B)$, $P(C)$, $P(A \cap C)$ and $P(A \cup C)$.

Question 2.7. In Example 2.5, what is the probability that Robin aims to the right? And what is the probability that he shoots low or right?