An Introduction to Probability

The study of probability started in the seventeenth century when Antoine Gambaud (who called himself the “Chevalier” de Méré) reached out to the French mathematician Blaise Pascal for an explanation of his gambling loses. De Méré would commonly bet that he could get at least one ace when rolling 4 six-sided dice, and he regularly made money on this bet. When that game started to get old, he started betting on getting at least one double-one in 24 rolls of two dice. Suddenly, he was losing money!

De Méré was dumbfounded. He reasoned that two aces in two rolls are 1/6 as likely as one ace in one roll. To compensate for this lower probability, the two dice should be rolled six times. Finally, to achieve the probability of one ace in four rolls, the number of the rolls should be increased fourfold (to 24). Therefore, you would expect a couple of aces to turn up in 24 double rolls with the same frequency as an ace in four single rolls. As you will see in a minute, although the very first statement is correct, the rest of his argument is not!

1.1 What is Probability?

Let’s start by establishing some common language. For our purposes, an experiment is any action whose outcome cannot necessarily be predicted with certainty; simple examples include the roll of a die and the card drawn from a well-shuffled deck. The outcome space of an experiment is the set of all possible outcomes associated with it; in the case of a die, it is the set \{1, 2, 3, 4, 5, 6\}, while for the card drawn from a deck, the outcome space has 52 elements corresponding to all combinations of 13 numbers (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K) with four suits (hearts, diamonds, clubs, and spades):

\[
\{A\spadesuit, 2\spadesuit, 3\spadesuit, 4\spadesuit, 5\spadesuit, 6\spadesuit, 7\spadesuit, 8\spadesuit, 9\spadesuit, 10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit, \\
A\clubsuit, 2\clubsuit, 3\clubsuit, 4\clubsuit, 5\clubsuit, 6\clubsuit, 7\clubsuit, 8\clubsuit, 9\clubsuit, 10\clubsuit, J\clubsuit, Q\clubsuit, K\clubsuit, \\
A\heartsuit, 2\heartsuit, 3\heartsuit, 4\heartsuit, 5\heartsuit, 6\heartsuit, 7\heartsuit, 8\heartsuit, 9\heartsuit, 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit, \\
A\diamondsuit, 2\diamondsuit, 3\diamondsuit, 4\diamondsuit, 5\diamondsuit, 6\diamondsuit, 7\diamondsuit, 8\diamondsuit, 9\diamondsuit, 10\diamondsuit, J\diamondsuit, Q\diamondsuit, K\diamondsuit\}
\]
A probability is a number between 0 and 1 that we attach to each element of the outcome space. Informally, that number simply describes the chance of that event happening. A probability of 1 means that the event will happen for sure, a probability of 0 means that we are talking about an impossible event, and numbers in between represent various degrees of certainty about the occurrence of the event. In the future, we will denote events using capital letters; for example,

\[ A = \{ \text{It will rain tomorrow} \}, \]
\[ B = \{ \text{The number 6 will come up in the next roll of the dice} \}, \]

while the probability associated with these events is denoted by \( P(A) \) and \( P(B) \). By definition, the probability of at least one event in the outcome space happening is 1, and therefore the sum of the probabilities associated with each of the outcomes also has to be equal to 1. On the other hand, the probability of an event not happening is simply the complement of the probability of the event happening, that is,

\[ P(A) = 1 - P(\bar{A}) \]

where \( \bar{A} \) should be read as “A not happening” or “not A.” For example, if \( A = \{ \text{It will rain tomorrow} \} \), then \( \bar{A} = \{ \text{It will NOT rain tomorrow} \} \).

There are a number of ways in which a probability can be interpreted. Intuitively almost everyone can understand the concept of how likely something is to happen. For instance, everyone will agree on the meaning of statements such as “it is very unlikely to rain tomorrow” or “it is very likely that the LA Lakers will win their next game.” Problems arise when we try to be more precise and quantify (i.e., put into numbers) how likely the event is to occur. Mathematicians usually use two different interpretations of probability, which are often called the frequentist and subjective interpretations.

The frequentist interpretation is used in situations where the experiment in question can be reproduced as many times as desired. Relevant examples for us include rolling a die, drawing cards from a well shuffled deck, or spinning the roulette wheel. In that case, we can think about repeating the experiment a large number of times (call it \( n \)) and recording how many of them result in outcome \( A \) (call it \( z_A \)). The probability of the event \( A \) can be defined by thinking about what happens to the ratio \( \frac{z_A}{n} \) (sometimes called the empirical frequency) as \( n \) grows.

For example, let \( A = \{ \text{A flipped coin comes up heads} \} \). We often assign this event a probability of 1/2, that is, we let \( P(A) = \frac{1}{2} \). This is often argued on the
1.1 What is Probability?

Sidebar 1.1 Random sampling in R

R provides easy-to-use functions to simulate the results of random experiments. When working with discrete outcome spaces such as those that appear with most casino and tabletop games, the function `sample()` is particularly useful. The first argument of `sample()` is a vector whose entries correspond to the elements of the outcome space, the second is the number of samples that we are interested in drawing, and the third indicates whether sampling will be performed with or without replacement (for now we are only drawing with replacement).

For example, suppose that you want to flip a balanced coin (i.e., a coin that has the same probability of heads and tails) multiple times:

```r
> outspc = c("Heads", "Tails")  # Outcome space
> z = sample(outspc, 20, replace=TRUE)  # Flip 20 times
> z
[1] "Tails" "Heads" "Tails" "Tails" "Tails" "Tails" "Heads"
[8] "Heads" "Tails" "Tails" "Tails" "Heads" "Heads" "Heads"
[15] "Heads" "Tails" "Tails" "Tails" "Heads" "Heads" "Tails"
```

Similarly, if we want to roll a six-sided die 15 times:

```r
> outspc = seq(1,6)  # Outcome space
> z = sample(outspc, 15, replace=TRUE)
> z
[1] 5 2 5 4 2 1 1 2 3 1 6 5 2 6 3
```

basis of symmetry: there is no apparent reason why one side of a regular coin would be more likely to come up than the other. Since you can flip a coin as many times as you want, the frequentist interpretation of probability can be used to interpret the value 1/2.

Because flipping the coin by hand is very time-consuming, we instead use a computer to simulate 5000 flips of a coin and plot the cumulative empirical frequency of heads using the following R code (please see Sidebar 1.1 for details on how to simulate random outcomes in R and Figure 1.1 for the output).

```r
> n = 5000
> outc = sample(c("Head","Tail"), n, replace=T)
> z = cumsum(outc=="Head")/seq(1,n)
> plot(z, xlab="Flips", ylab="Frequency of Heads", type="l")
> abline(h=0.5, col="grey")
```
An Introduction to Probability

Note that the empirical frequency fluctuates, particularly when you have flipped the coin just a few times. However, as the number of flips \( n \) in our formula becomes larger and larger, the empirical frequency gets closer and closer to the “true” probability \( \frac{1}{2} \) and fluctuates less and less around it.

The convergence of the empirical frequency to the true probability of an event is captured by the so-called law of large numbers.

**Law of Large Numbers for Probabilities**

Let \( z_n \) represent the number of times that event \( A \) happens in a total of \( n \) identical repetitions of an experiment, and let \( P(A) \) denote the probability of event \( A \). Then \( \frac{z_n}{n} \) approaches \( P(A) \) as \( n \) grows.

This version of the law of large numbers implies that, no matter how rare a non-zero probability event is, if you try enough times, you will eventually observe it. Besides providing a justification for the concept of probability, the law of large numbers also provides a way to compute the probability of complex events by repeating an experiment multiple times and computing the empirical frequency associated with it. In the future, we will do this by using a computer (as we did in our simple coin flipping example before) rather than by physically rolling dice or drawing cards from a deck.
Even though the frequency interpretation of probability we just described is appealing, it cannot be applied to situations where the experiment cannot be repeated. For example, consider the event

\[ A = \{ \text{It will rain tomorrow} \}. \]

There will be only one tomorrow, so we will only get to observe the “experiment” (whether it rains or not) once. In spite of that, we can still assign a probability to \( A \) based on our knowledge of the season, today’s weather, and our prior experience of what that implies for the weather tomorrow. In this case, \( P(A) \) corresponds to our “degree of belief” on tomorrow’s rain. This is a subjective probability, in the sense that two reasonable people might not necessarily agree on the number.

To summarize, although it is easy for us to qualitatively say how likely some event is to happen, it is very challenging if we try to put a number to it. There are a couple of ways in which we can think about this number:

- The frequentist interpretation of probability that is useful when we can repeat and observe an experiment as many times as we want.
- The subjective interpretation of probability, which is useful in almost any probability experiment where we can make a judgment of how likely an event is to happen, even if the experiment cannot be repeated.

### 1.2 Odds and Probabilities

In casinos and gambling dens, it is very common to express the probability of events in the form of odds (either in favor or against). The \textit{odds in favor} of an event \( A \) is simply the ratio of the probability of that event happening divided by the probability of the event not happening, that is,

\[
\text{Odds in favor of } A = \frac{P(A)}{1 - P(A)}
\]

Similarly, the \textit{odds against} \( A \) are simply the reciprocal of the odds in favor, that is,

\[
\text{Odds against } A = \frac{1}{\text{Odds in favor of } A} = \frac{1 - P(A)}{P(A)}
\]

The odds are typically represented as a ratio of integer numbers. For example, you will often hear that the odds in favor of any given number in American roulette are 1 to 37, or 1:37. Note that you can recover \( P(A) \) from the odds in
favor of $A$ through the formula,

\[
P(A) = \frac{\text{Odds in favor of } A}{1 + \text{Odds in favor of } A}
\]

In the context of casino games, the odds we have just discussed are sometimes called the \textit{winning odds} (or the \textit{losing odds}). In that context, you will also hear sometimes about \textit{payoff odds}. This is a bit of a misnomer, as these represent the ratio of payoffs, rather than the ratio of probabilities.

\[
\text{Payoff odds in favor of } A = \frac{\text{Payoff to player if } A \text{ happens}}{\text{Payoff to player if } A \text{ does not happen}}
\]

For example, the winning odds in favor of any given number in American roulette are 1 to 37, but the payoff odds for the same number are just 1 to 35 (which means that, if you win, every dollar you bet will bring back $35 in profit). This distinction is important, as many of the odds on display in casinos refer to these payoff odds rather than the winning odds. Keep this in mind!

### 1.3 Equiprobable Outcome Spaces and De Méré’s Problem

In many problems, we can use symmetry arguments to come up with reasonable values for the probability of simple events. For example, consider a very simple experiment consisting of rolling a perfect, six-sided (cubic) die. This type of dice typically has its sides marked with the numbers 1–6. We could ask about the probability that a specific number (say, 3) comes up on top. Since the six sides are the only possible outcomes (we discount the possibility of the die resting on edges or vertexes!) and they are symmetric with respect to each other, there is no reason to think that one is more likely to come up than another. Therefore, it is natural to assign probability 1/6 to each side of the die.

Outcome spaces where all outcomes are assumed to have the same probability (such as the outcome space associated with the roll of a six-sided die) are called equiprobable spaces. In \textit{equiprobable spaces}, the probabilities of different events can be computed using a simple formula:

\[
P(A) = \frac{\text{Number of outcomes consistent with } A}{\text{Total number of possible outcomes}}.
\]
Note the similarities with the law of large numbers and the frequentist interpretation of probability.

Although the concept of equiprobable spaces is very simple, some care needs to be exercised when applying the formula. Let’s go back to Chevalier de Méré’s predicament. Recall that De Méré would commonly bet that he could get at least one ace when rolling 4 (fair) six-sided dice, and he would regularly make money on this bet. To make the game more interesting, he started betting on getting at least one double-one in 24 rolls of two dice, after which he started to lose money.

Before analyzing in detail De Méré’s bets, let’s consider the outcome space associated with rolling two dice. The same symmetry arguments we used in the case of a single die can be used in this case, so it is natural to think of this outcome space as equiprobable. However, there are two ways in which we could construct the outcome space, depending on whether we consider the order of the dice relevant or not (see Table 1.1). The first construction leads to the conclusion that getting a double one has probability \(\frac{1}{21}\approx 0.0476190\), while the second leads to a probability of \(\frac{1}{36}\approx 0.027778\). The question is, which one is the correct one?

In order to gain some intuition, let’s run another simulation in R in which two dice are rolled 100,000 times each.

```r
> n = 100000
> die1 = sample(seq(1,6), n, replace=T)
> die2 = sample(seq(1,6), n, replace=T)
> sum(die1==1 & die2==1)/n
[1] 0.02765
```

The result of the simulation is very close to \(\frac{1}{36}\), which suggests that this is the right answer. A formal argument can be constructed by thinking of the dice as being rolled sequentially rather than simultaneously. Since there

<table>
<thead>
<tr>
<th>Order is irrelevant</th>
<th>Order is relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 outcomes in total</td>
<td>36 outcomes in total</td>
</tr>
<tr>
<td>1–1 2–2 3–3 4–4 5–5 6–6</td>
<td>1–1 2–1 3–1 4–1 5–1 6–1</td>
</tr>
<tr>
<td>1–2 2–3 3–4 4–5 5–6</td>
<td>1–2 2–2 3–2 4–2 5–2 6–2</td>
</tr>
<tr>
<td>1–3 2–4 3–5 4–6</td>
<td>1–3 2–3 3–3 4–3 5–3 6–3</td>
</tr>
<tr>
<td>1–4 2–5 3–6</td>
<td>1–4 2–4 3–4 4–4 5–4 6–4</td>
</tr>
<tr>
<td>1–5 2–6</td>
<td>1–5 2–5 3–5 4–5 5–5 6–5</td>
</tr>
<tr>
<td>1–6</td>
<td>1–6 2–6 3–6 4–6 5–6 6–6</td>
</tr>
</tbody>
</table>
are 6 possible outcomes of the first roll and another 6 possible outcomes for
the second one, there is a total of 36 combined outcomes. Since just 1 of these
36 outcomes corresponds to a pair of ones, our formula for the probability of
events in equiprobable spaces leads to the probability of 2 ones being 1/36. Underlying this result is a simple principle that we will call the multiplication
principle of counting.

**Multiplication Principle for Counting**
If events \( A \), \( B \), \( C \), … can each happen in \( n_a \), \( n_b \), \( n_c \), … ways then they can happen together in \( n_a \times n_b \times n_c \times \cdots \) ways.

Now, let’s go back to De Mérés’s problem and use the multiplication rule to
calculate the probability of winning each of his two bets. In this context, it is
ejasier to first compute the probability of losing the bet and, because no ties are
possible, then obtain the probability of winning the bet as

\[
P(\text{winning}) = 1 - P(\text{losing}).
\]

For the first bet, the multiplication rule implies that there are a total of
\( 6 \times 6 \times 6 \times 6 = 6^4 = 1296 \) possible outcomes when we roll 4 six-sided dice. If
we are patient enough, we can list all the possibilities:

\[
egin{align*}
1, 1, 1, 1 \\
1, 1, 2, 2 \\
1, 1, 1, 3 \\
1, 1, 1, 4 \\
1, 1, 1, 5 \\
1, 1, 1, 6 \\
1, 1, 2, 1 \\
\vdots
\end{align*}
\]

On the other hand, since for each single die there are five outcomes that are
not an ace, there are \( 5^4 = 625 \) outcomes for which De Méré losses this bet. Again, we could potentially enumerate these outcomes

\[
egin{align*}
2, 2, 2, 2 \\
2, 2, 2, 3 \\
2, 2, 2, 4 \\
2, 2, 2, 5 \\
2, 2, 2, 6 \\
2, 2, 3, 2 \\
\vdots
\end{align*}
\]

The probability that De Méré wins his bet is therefore

\[
P(\text{winning first bet}) = 1 - \frac{625}{1296} = \frac{671}{1296} = 0.51775.
\]
You can corroborate this result with a simple simulation of 100,000 games:

```
> n = 100000
> die1 = sample(seq(1,6), n, replace=T)
> die2 = sample(seq(1,6), n, replace=T)
> die3 = sample(seq(1,6), n, replace=T)
> die4 = sample(seq(1,6), n, replace=T)
> sum(die1==1 | die2==1 | die3==1 | die4==1)/n

[1] 0.51961
```

For the second bet we can proceed in a similar way. As we discussed before, there are 36 equiprobable outcomes when you roll 2 six-sided dice, 35 of which are unfavorable to the bet. Therefore, there are $36^{24}$ possible outcomes when two dice are rolled together 24 times, of which $35^{24}$ are unfavorable to the player, and the probability of winning this bet is equal to

$$P(\text{winning second bet}) = 1 - \frac{35^{24}}{36^{24}} = \frac{36^{24} - 35^{24}}{36^{24}} \approx 0.49140.$$  

Again, you can verify the results of the calculation using a simulation:

```
> n = 100000
> outc = seq(1,6)
> numsneakeye = rep(0,n)
> for(i in 1:n){
+   die1 = sample(outc,24,replace=T)
+   die2 = sample(outc,24,replace=T)
+   numsneakeye[i] = sum(die1==1 & die2==1)
+ }
> sum(numsneakeye>=1)/n

[1] 0.49067
```

The fact that the probability of winning is less than 0.5 explains why De Méré was losing money! Note, however, that if he had used 25 rolls instead of 24, then the probability of winning would be $\frac{36^{25} - 35^{25}}{36^{25}} \approx 0.50553$, which would have made it a winning bet for De Méré (but not as good as the original one!).

### 1.4 Probabilities for Compound Events

A *compound event* is an event that is created by aggregating two or more simple events. For example, we might want to know what is the probability that the number selected by the roulette is black or even, or what is the probability that we draw a card from the deck that is both a spade and a number.

As the examples above suggest, we are particularly interested in two types of operations to combine events. On the one hand, the union of two events $A$ and
B (denoted by $A \cup B$) corresponds to the event that happens if either $A$ or $B$ (or both) happen. On the other hand, the intersection of two events (denoted by $A \cap B$) corresponds to the event that happens only if both $A$ and $B$ happen simultaneously. The results from these operations can be represented graphically using a Venn diagram (see Figure 1.2) where the simple events $A$ and $B$ correspond to the rectangles. In Figure 1.2(a), the combination of the areas of both rectangles corresponds to the union of the events. In Figure 1.2(b), the area with the darker highlight corresponds to the intersection of both events. The probability of the intersection of two events is sometimes called the joint probability of the two events. In the case when this joint probability is zero (i.e., both events cannot happen simultaneously), we say that the events are disjoint or mutually exclusive.

In many cases, the probabilities of compound events can be computed directly from the sample space by carefully counting favorable cases. However, in other cases, it is easier to compute them from simpler events. Just as there is

![Figure 1.2 Venn diagram for the (a) union and (b) intersection of two events.](image-url)
a rule for probability of two events happening together, there is a second rule for the probability of two alternative events (e.g., the probability of obtaining an even number or a 2 when rolling a die), which is sometimes called the Addition Rule of probability:

For any two events, \[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

Figure 1.3 presents a graphical representation of two events using Venn diagrams; it provides some hints at why the formula takes this form. If we simply add \( P(A) \) and \( P(B) \), the darker region (which corresponds to \( P(A \cap B) \)) is counted twice. Hence, we need to subtract it once in order to get the right result. If two events are mutually exclusive (i.e., they cannot occur at the same time, which means that \( P(A \cap B) = 0 \)), this formula reduces to \( P(A \cup B) = P(A) + P(B) \).

Similar rules can be constructed to compute the joint probability of two events, \( P(A \cap B) \). For the time being, we will only present the simplified Multiplication Rule for the probability of independent events. Roughly speaking, this rule is appropriate for when knowing that one of the events occur does not affect the probability that the other will occur.

For any two independent events, \[ P(A \cap B) = P(A)P(B). \]

In Chapter 5, we cover the concept of independent events in more detail and present more general rules to compute the joint probabilities.

Figure 1.3 Venn diagram for the addition rule.
1.5 Exercises

1. A man has 20 shirts and 10 ties. How many different shirt-and-tie combinations can he make?

2. If you have 5 different pants, 12 different shirts, and 3 pairs of shoes, how many days can you go without repeating the same outfit?

3. A fair six-sided die is rolled 1,000,000 times and the number of rolls that turn out to be either a 1 or a 5, \( x_{1,5} \), are recorded. From the law of large numbers, what is the approximate value for \( x_{1,5} \) that you expect to see?

4. A website asks users to choose eight-letter usernames (only alphabetic characters are allowed, and no distinction is made between lower- and upper-case letters). How many distinct usernames are possible for the website?

5. Provide two examples of experiments for which the probability of the outcomes can only be interpreted from a subjective perspective. For each one of them, justify your choice and provide a value for such probability.

6. In how many ways can 13 students be lined up?

7. Re-write the following probability using the addition rule of probability: \( P(\text{obtaining a 5 or a 6 when rolling a six-sided die}) \).

8. Re-write the following probability using the addition rule of probability: \( P(\text{obtaining a total sum of 5 or an even sum when rolling a 2 six-sided dice}) \).

9. Re-write the following probability using the rule of probability for complementary events: \( P(\text{obtaining at least a 2 when rolling a six-sided die}) \).

10. Re-write the following probability using the rule of probability for complementary events: \( P(\text{obtaining at most a 5 when rolling a six-sided die}) \).

11. Consider rolling a six-sided die. Which probability rule can be applied to the following probability

\[ P(\text{obtaining a number higher than 2 or a number smaller than 5}) \]
12. What is the probability of obtaining at least two heads when flipping a coin three times? Which probability rule was used in your reasoning?

13. Explain what is wrong with each of the following arguments.
   (a) First argument:
       • In 1 roll of a six-sided die, I have $\frac{1}{6}$ of a chance to get an ace.
       • So in 4 rolls, I have $4 \times \frac{1}{6} = \frac{2}{3}$ of a chance to get at least one ace.
   (b) Second argument:
       • In 1 roll of a pair of six-sided dice, I have $\frac{1}{36}$ of a chance to get a double ace.
       • So in 24 rolls, I have $24 \times \frac{1}{36} = \frac{2}{3}$ of a chance to get at least one double ace.

14. What is the probability that, in a group of 30 people, at least two of them have the same birthday. Hint: Start by computing the probability that no two people have the same birthday.

15. [R] Write a simulation that allows you to estimate the probability in the previous problem.

16. [R] Modify the code for the second De Méré bet to verify that if 25 rolls are involved instead of 24 then you have a winning bet.