

# An Introduction to Probability and Simulation

Kevin Ross

2020-07-11



# Contents

<b>Preface</b>	<b>5</b>
<b>1 Introduction</b>	<b>11</b>
<b>2 What is Probability?</b>	<b>13</b>
2.1 Instances of randomness . . . . .	13
2.2 Interpretations of probability . . . . .	15
2.3 Proportional reasoning and tables of counts . . . . .	22
2.4 Working with probabilities . . . . .	23
2.5 Approximating probabilities - a brief introduction to simulation .	26
2.6 Sliding scale of probability, or Probability of what? . . . . .	28
2.7 Common misinterpretation and fallacies (e.g. outbreak of Asian disease, Utts book) . . . . .	28
2.8 Why study coins, dice, cards, and spinners? . . . . .	29
<b>3 The Language of Probability</b>	<b>31</b>
3.1 Sample space of outcomes . . . . .	31
3.2 Events . . . . .	38
3.3 Random variables . . . . .	47
3.4 Probability spaces . . . . .	57



# Preface

## Why study probability *and simulation*?

Why study probability?

- Probability is the study of uncertainty, and life is uncertain
- Probability is used in a wide variety of fields, including: statistics, physics, engineering, biology, medicine, finance, actuarial science, political science, law, sports , ...
- Many topics and problems in probability are frequently misunderstood and sometimes counter intuitive, so it's worthwhile to take a careful study
- “Probabilistic thinking” is an important component of statistical literacy (e.g. how to assess risk when making decisions)
- Probability provides the foundation for many important statistical concepts and methods such as p-values and confidence intervals

Why use **simulation** to study probability?

- Many concepts encountered in probability can seem esoteric; simulation helps make them more concrete.
- Simulation provides an effective tool for analyzing probability models and for exploring effects of changing assumptions
- Simulation can be used to check analytical solutions
- Simulation is often the best or only method for investigating many problems which are too complex to solve analytically
- Simulation-based reasoning is an important component of statistical literacy (e.g. understanding a p-value via simulation)
- Many statistical procedures employ simulation-based methods (e.g. bootstrapping)

### 0.0.1 Learning Objectives/Goals/Style??? (Better title)

- Don't skip on rigorous definitions (RV is function defined on probspace) but deemphasize mathematical computation (counting and calculus)

- Emphasize simulation
- Visualize in lots of plots
- Start multivariate relationships early
- Rely on statistical literacy
- Active learning, workbook style

## Symbulate

This book uses the Python package Symbulate (<https://github.com/dlsun/symbulate>) which provides a user friendly framework for conducting simulations involving probability models. The syntax of Symbulate reflects the “language of probability” and makes it intuitive to specify, run, analyze, and visualize the results of a simulation. In Symbulate, probability spaces, events, random variables, and random processes are symbolic objects which can be manipulated, independently of their simulated realizations. Symbulate’s consistency with the mathematics of probability reinforces understanding of probabilistic concepts. The article Ross and Sun (2019) discusses Symbulate and its features in more detail.

To install Symbulate, it is recommended that you first install the Anaconda distribution, which is a Python environment with many scientific packages installed (including all of the packages that Symbulate is built on). After installing Anaconda, the recommended way to installing Symbulate is run the command

```
!pip install symbulate
```

from inside a notebook. (The Symbulate package can also be downloaded and installed manually from the Symbulate Github repository following these instructions.)

The following command imports Symbulate during a Python session.

```
from symbulate import *
```

The Symbulate command `plot()` produces graphics. These graphics can be customized (by changing axis limits, adding titles, legends, etc) using Matplotlib, and in particular the `pyplot` method, which can be imported with

```
import matplotlib
import matplotlib.pyplot as plt
```

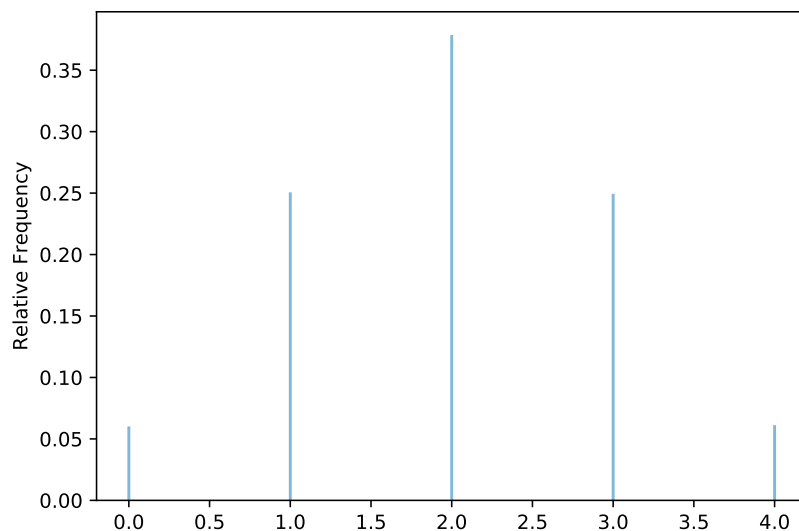
Jupyter or Google Colab notebooks provide a natural interface for Symbulate. The code in this book matches as closely as possible the commands that would be entered into cells in a notebook. However, certain commands that appear

throughout the book are needed only to properly produce the output in this book, and not if working directly in notebooks. In particular, `plt.figure()` and `plt.show()` are not needed to produce graphics in Jupyter (with the use of the `%matplotlib inline` “magic”). In addition, Jupyter automatically displays the result of the last line in a cell; `print()` is generally not needed to display output (unless you wish to format it, or it is not the last line in the cell).

For example, a code snippet that appears in this book as

```
x = RV(Binomial(4, 0.5)).sim(10000)
plt.figure()
x.plot()
plt.show()
print(x)
```

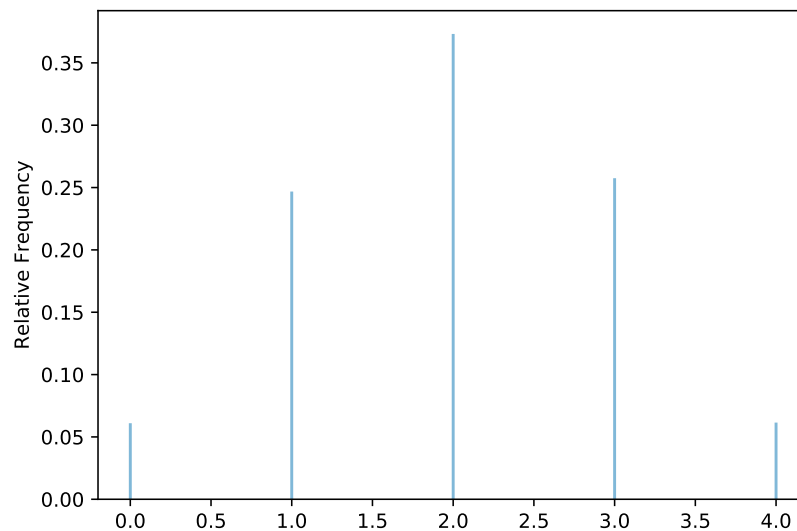
```
x = RV(Binomial(4, 0.5)).sim(10000)
# plt.figure()
x.plot()
plt.show()
```



```
print(x)
```

```
## <sympulate.results.RVResults object at 0x000000001D146CC8>
```

```
x = RV(Binomial(4, 0.5)).sim(10000)
plt.figure()
x.plot()
plt.show()
```



```
print(x)
```

```
## <symbulate.results.RVResults object at 0x000000001C307F48>
```

would be entered in a Jupyter notebook as

```
x = RV(Binomial(4, 0.5)).sim(10000)
x.plot()
x
```

## Don't do what Donny Don't does

Some of the examples and exercises in this book are labeled “Don't do what Donny Don't does”. This is a Simpson's reference. In this text, Donny represents a student who makes many of the mistakes commonly made by students studying probability. The idea of these problems is for you to learn from the common mistakes that Donny makes, by identifying why he is wrong and by helping him understand and correct his mistakes. (But be careful: sometimes Donny is right!)



## About this book

This book was written using the **bookdown** package (Xie, 2020), which was built on top of R Markdown and **knitr** (Xie, 2015). Python code was run in RStudio using the **reticulate** package.



# Chapter 1

## Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 1.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 1.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

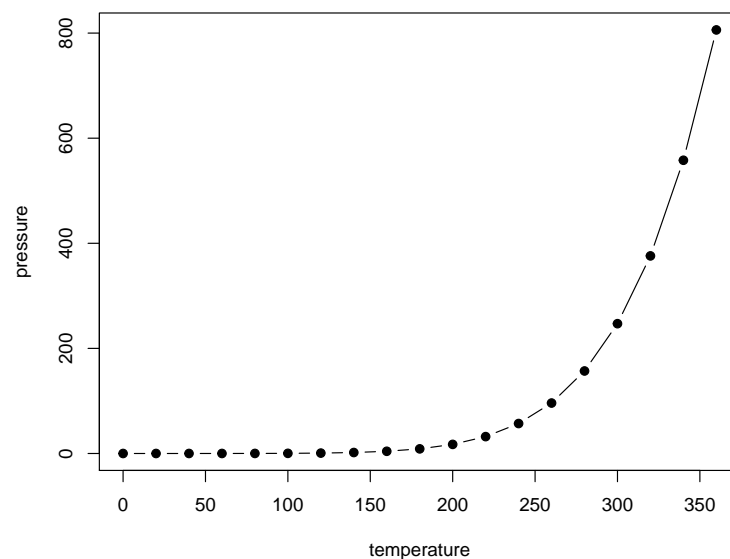


Figure 1.1: Here is a nice figure!

Table 1.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

## Chapter 2

# What is Probability?

This chapter provides a non-technical introduction to randomness and probability. Many of the topics introduced in this chapter will be covered in more detail in later chapters.

NO MATH BEYOND TWO WAY TABLES AND BASIC EXPECTED VALUES

Expand chapter 1 with more example of probability/simulation computations with as little math and terminology as possible. (e.g. Don't define RV or event or sample space.)

Add a few motivating examples with links - 538 election, hurricane, etc

### 2.1 Instances of randomness

instances of randomness - examples; include in other sections? (random sampling, random assignment, future/past, etc, physical (coin, die), quantum)

**Exercise 2.1.** Each of the following situations involves a probability. How are the various situations similar, and how are they different? What is one feature that all of the situations have in common? If you were to estimate the probability in question, how might you do it? What are some things to consider? The goal here is not to do any calculations but rather to think about, via these examples, similarities and differences of situations in which probabilities are of interest.

1. The probability that a single flip of a fair coin lands on heads.
2. The probability that in two flips of a fair coin exactly one flip lands on heads.

3. The probability that in 10000 flips of a fair coin exactly 5000 flips land on heads.
4. The probability that in 10000 flips of a fair coin “around” 5000 flips land on heads.
5. The probability you win the next Powerball lottery if you purchase a single ticket, 6-7-16-23-26, plus the Powerball number, 4. (FYI: There are roughly<sup>1</sup> 300 million possible winning number combinations.)
6. The probability you win the next Powerball lottery if you purchase a single ticket, 1-2-3-4-5, plus the Powerball number, 6.
7. The probability that someone wins the next Powerball lottery. (FYI: especially when the jackpot is large, there are hundreds of millions of tickets sold.)
8. The probability that a “randomly selected” Cal Poly student is from CA.
9. The probability that Hurricane Humberto makes landfall in the U.S.
10. The probability that the Los Angeles Chargers win the next Superbowl.
11. The probability that Donald Trump wins the 2020 U.S. Presidential Election.
12. The probability that extraterrestrial life currently exists somewhere in the universe.
13. The probability that you ate an apple on April 17, 2009.

- The subject of probability concerns *random* phenomena.
- A phenomenon is **random** if there are multiple potential outcomes, and there is **uncertainty** about which outcome will occur. *Uncertainty* is the feature that all the scenarios have in common.
- Uncertainty does not necessarily mean uncertainty about an occurrence in the future. For example, you either ate or apple or not on April 17, 2009, but you’re not certain about it. But if you tended to eat a lot of apples ten years ago, then you might give a high probability to the event that you ate one on April 17, 2009.
- Many phenomena involve physical randomness<sup>2</sup>, like flipping a coin or drawing powerballs at random from a bin.
- Statistical applications often involve the planned use of physical randomness
  - **Random selection** involves selecting a *sample* of individuals at random from a *population* (e.g. via random digit dialing).
  - **Random assignment** involves assigning individuals at random to groups (e.g. in a randomized experiment).

<sup>1</sup>The exact count is 292,201,338. We will see how to compute this number later.

<sup>2</sup>We will refer to as “random” any scenario that involves a reasonable degree of uncertainty. We’re avoiding philosophical questions about what is “true” randomness, like the following. Is a coin flip really random? If all factors that affect the trajectory of the coin were known precisely, then wouldn’t the outcome be determined? Does true randomness only exist in quantum mechanics?

- However, in many other situations randomness just vaguely reflects uncertainty.
- In any case, random does *not* mean haphazard. In a random phenomenon, while individual outcomes are uncertain, there is a *regular distribution of outcomes over a large number of (hypothetical) repetitions*.
  - In two flips of a fair coin we wouldn't necessarily see one head and one tail. But in 10000 flips of a fair coin, we would expect to see close to 5000 heads and 5000 tails.
  - We don't know who will win the next Superbowl, but we can and should certainly consider some teams as more likely to win than others. We could imagine a large number of hypothetical 2019 seasons; how often would we expect the Eagles to win? The Raiders? (Hopefully a lot for the Eagles; probably not much for the Raiders).
- Also, random does *not* necessarily mean equally likely. In a random phenomenon, certain outcomes or events might be more or less likely than others.
  - It's much more likely that a randomly selected Cal Poly student is from CA than not.
  - Not all NFL teams are equally likely to win the next Superbowl.

## 2.2 Interpretations of probability

In the previous section we encountered a variety of scenarios which involved uncertainty, a.k.a. randomness. Just as there are a few “types” of randomness, there are a few ways of interpreting probability, namely, *long run relative frequency* and *subjective probability*.

**Exercise 2.2.** Revisit the scenarios in Exercise 2.1. Now consider how “probability” is interpreted in the different scenarios. In each scenario, what does “probability” mean? How might you estimate the probability? Start to make guesses for the probabilities; are they “high” or “low”? How high or low? Again, the goal is not to do any calculations but rather to think about, via these examples, similarities and differences of situations in which probabilities are of interest.

In particular, compare

- a. Scenarios 3 and 4
- b. Scenarios 5 and 6
- c. Scenarios 6 and 7
- d. How are scenarios 1 through 8 (collectively) different from scenarios 9 through 13 (collectively)?

- The **probability** of an event is a number in the interval  $[0, 1]$  measuring the event's likelihood or degree of uncertainty.
- A probability can take any values in the continuous scale from 0% to 100%<sup>3</sup>. In particular, a probability requires much more interpretation than “is the probability greater than, less than, or equal to 50%?”
- When interpreting probabilities, be careful not to confuse “the particular” with “the general”.
  - (“The particular.”) A very specific event, surprising or not, often has low probability.
    - \* Even though in 10000 flips of a fair coin we would expect to see about 5000 heads, the probability that *exactly* 5000 out of 10000 flips are heads is fairly small (about<sup>4</sup> 0.008).
    - \* The probability that the winning powerball number is 6-7-16-23-26-(4) is exactly the same as the probability that the winning powerball number is 1-2-3-4-5-(6). Each of these sequences is just one of the roughly 300 million possible sequences, and each sequences has about a 1 in 300 million chance of being the winning number. However, many people think 6-7-16-23-26-(4) is more likely because 1-2-3-4-5-(6) “doesn’t look random”.
    - \* The probability that you get a text from your best friend at 7:43pm on Oct 12, 2019 inviting you to dinner after you’ve just ordered pizza from your favorite pizza place is probably pretty small. None of these items — getting a text, having a friend invite you to dinner, ordering pizza from your favorite pizza place — is unusual, but the chances of them all combining in this way at this particular time are fairly small.
  - (“The general.”) However, if there are many like events, their combined probability can be high.
    - \* The probability that *around* 5000 out of 10000 coin flips land on heads is fairly large. For example, if “around” is interpreted as between 4900 and 5100 (for a proportion of heads between 0.49 to 0.51) the probability is about 0.956.
    - \* The probability that the winning powerball number is an ordered sequence, like 1-2-3-4-5-(6), is extremely small<sup>5</sup>. However, the probability that the winning number is not an ordered sequence, like 6-7-16-23-26-(4), is extremely high. When interpreting probabilities, be careful not to confuse an event like “the winning number is 6-7-16-23-26-(4)” (the particular, low

<sup>3</sup>Probabilities are usually defined as decimals, but are often colloquially referred to as percentages. We’re not sticklers; we’ll refer to probabilities as decimals and as percentages.

<sup>4</sup>We will see how to compute probabilities like this one in upcoming chapters.

<sup>5</sup>If a “sequence” is defined with the powerball as the last number in the sequence, then the probability is about 21 out of 300 million. Since the powerball must be a number from 1 to 26, there are only 21 tickets out of 300 million possibilities for which the numbers are in an ordered sequence: 1-2-3-4-5-(6), 2-3-4-5-6-(7), ... 21-22-23-24-25-(26).



probability) with an event like “the winning number is not an ordered sequence” (the general, high probability).

- \* The probability that some time in the next month or so a friend invites you for dinner after you’ve already had dinner on your own is probably fairly high.
- Even if an event has extremely small probability, given enough repetitions of the random phenomenon, the probability that the event occurs on *at least one* of the repetitions is high.
  - The probability that a specific powerball ticket is the winning number is about 1 in 300 million. So if you buy a single ticket, it is extremely unlikely that *you* will win.
  - However, if hundreds of millions of powerball tickets are sold, the probability that *someone somewhere* wins is pretty high. For example, if 500 million tickets are sold then there is a roughly 80% chance that at least one ticket has the winning number (under certain assumptions).

### 2.2.1 Relative frequency

The probability that a single flip of a fair coin lands on heads is 0.5. How do we interpret this 0.5? The notation of “fairness” implies that the two outcomes, heads and tails, should be equally likely, so we have a “50/50 chance”. But how else can we interpret this 50%? One way is by considering *what would happen if we flipped the coin many times*. Now, if we would flipped the coin twice, we wouldn’t expect to necessarily see one head and one tail. And we already mentioned that if we flipped the coin 10000 times, the chances of seeing *exactly* 5000 heads is small. But in many flips, we might expect to see heads on something close to 50% of flips.

Consider Figure 2.1 below. Each dot represents a set of 10,000 fair coin flips. There are 100 dots displayed, representing 100 different sets of 10,000 coin flips each. For each set of flips, the proportion of the 10,000 flips which landed on head is recorded. For example, if in one set 4973 out of 10,000 flips landed on heads, the proportion of heads is 0.4973. The plot displays 100 such proportions. We see that only 5 of these 100 proportions are less than 0.49 or greater than 0.51. So if between 0.49 and 0.51 is considered “close to 0.5”, then yes, in 10000 coin flips we would expect the proportion of heads to be close to 0.5. (In 10000 flips, the probability of heads on between 49% and 51% of flips is 0.956, so 95 out of 100 provides a rough estimate of this probability.)

But what if we want to be stricter about what qualifies as “close to 0.5”? You might suspect that with even more flips we would expect to observe heads on even closer to 50% of flips. Indeed, this is the case. Figure 2.2 displays the results of 100 sets of 1,000,000 fair coin flips. The pattern seems similar to Figure 2.1

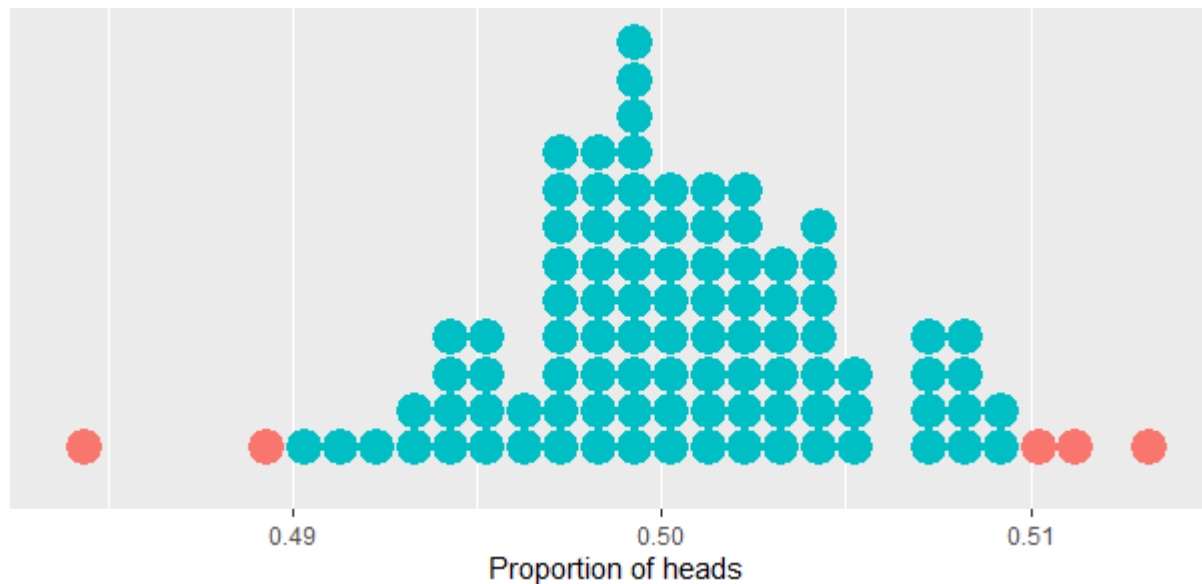


Figure 2.1: Proportion of flips which are heads in 100 sets of **10,000** fair coin flips. Each dot represents a set of **10,000** fair coin flips.

but pay close attention to the horizontal axis which covers a much shorter range of values than in the previous figure. Now 96 of the 100 proportions are between 0.499 and 0.501. So in 1,000,000 flips we would expect the proportion of heads to be between 0.499 and 0.501, pretty close to 0.5. (In 1,000,000 flips, the probability of heads on between 49.9% and 50.1% of flips is 0.955, and 96 out of 100 sets provides a rough estimate of this probability.)

In Figure 2.3 each dot represents a set of 100 millions flips. The pattern seems similar to the previous figures, but again pay close attention the horizontal axis which covers a smaller range of values. Now 96 of the 100 proportions are between 0.4999 and 0.5001. (In 100 million flips, The probability of heads on between 49.99% and 50.01% of flips is 0.977, so 96 out of 100 sets provides a rough estimate of this probability.)

The previous figures illustrate that the more flips there are, the more likely it is that we observe a proportion of flips landing on heads close to 0.5. We also see that with more flips we can refine our definition of “close to 0.5”: increasing the number of flips by a factor of 100 (10,000 to 1,000,000 to 100,000,000) seems to give us an additional decimal place of precision ( $0.5 \pm 0.01$  to  $0.5 \pm 0.001$  to  $0.5 \pm 0.0001$ .) These observations illustrate the relative frequency interpretation of probability.

- The probability of an event corresponding to the result of a random phenomenon can be interpreted as the proportion of times that the event

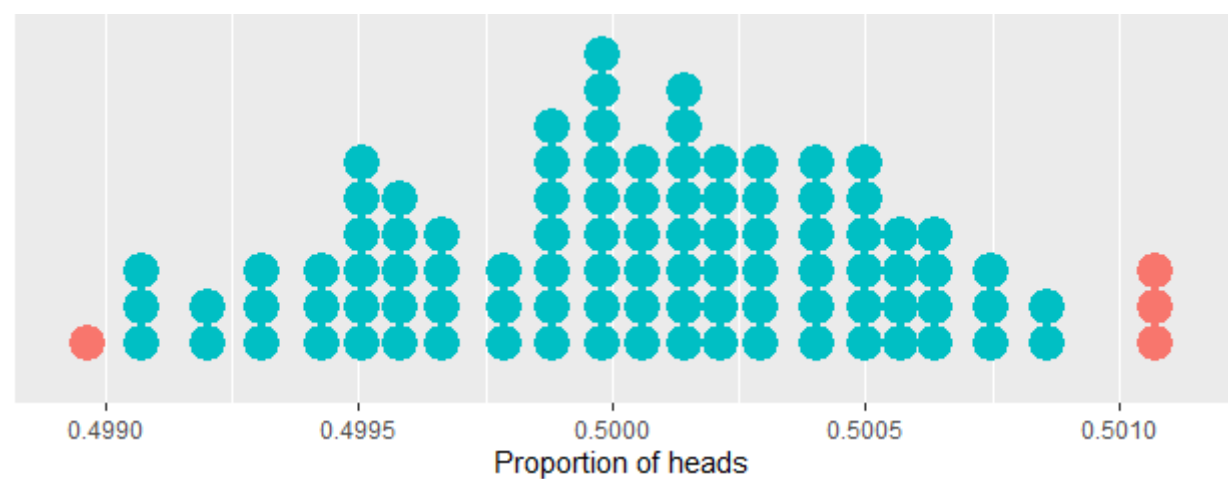


Figure 2.2: Proportion of flips which are heads in 100 sets of **1,000,000** fair coin flips. Each dot represents a set of **1,000,000** fair coin flips.

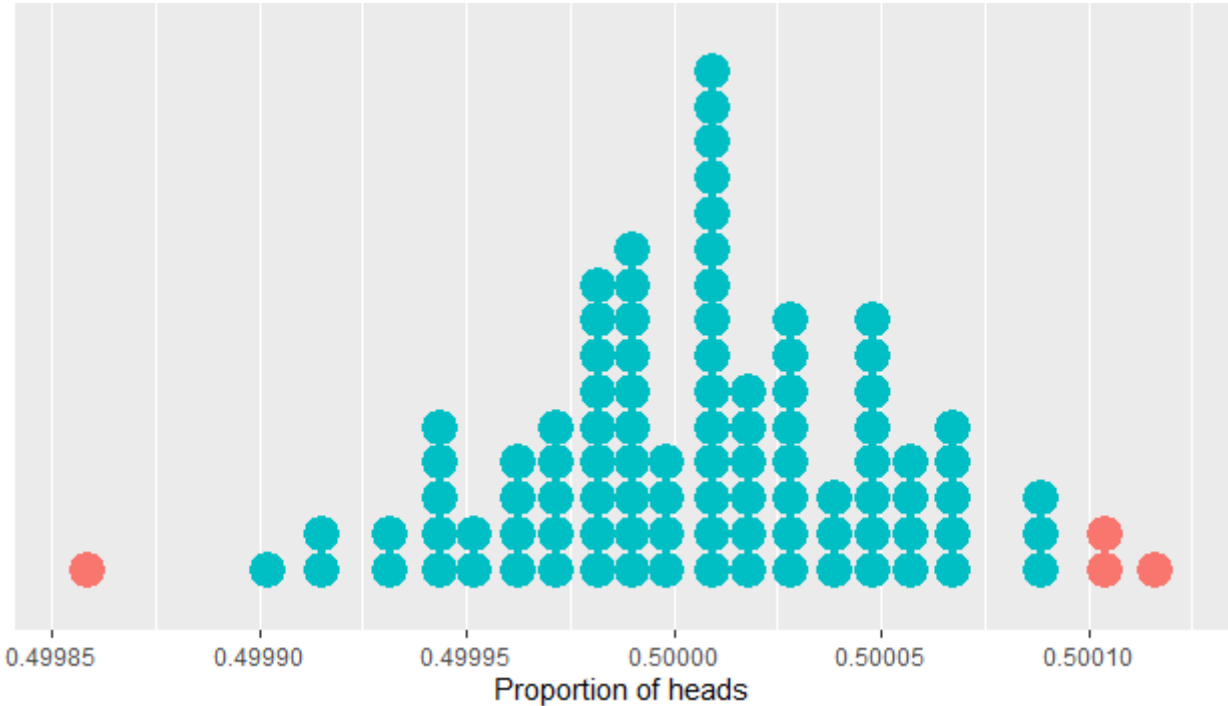


Figure 2.3: Proportion of flips which are heads in 100 sets of **100,000,000** fair coin flips. Each dot represents a set of **100,000,000** fair coin flips.

would occur in a very large number of hypothetical repetitions of the random phenomenon.

- That is, a probability can be interpreted as a **long run proportion** or **long run relative frequency**.
- This means that the probability of an event can be approximated by **simulating** the random phenomenon a large number of times and determining the proportion of simulated repetitions on which the event occurred out of the total number of repetitions of the simulation (this proportion is also called the relative frequency of the event.)
  - A **simulation** involves an artificial recreation of the random phenomenon, usually using a computer.
  - For example, if a basketball player is successful on 90% of her free throw attempts, we can simulate the player shooting a single free throw attempt by taking 10 cards and labeling 9 as “success” and 1 as “miss” then shuffling well and dealing one card.
- The **long run relative frequency** interpretation of probability can be applied when a situation can be repeated numerous times, at least conceptually, and the outcome can be observed each time.
- The relative frequency of a particular event will settle down to a single constant value after many repetitions, and that long run value is the probability of that event.
  - However, what constitutes the random phenomenon or how the simulation is conducted depends on certain assumptions. Changing those assumptions can affect probabilities of interest.
    - \* For example, if you’re interested in the probability that a die lands on 1, you need to know if it’s a four-sided die or a six-sided die, and if the die is consider “fair”.
    - \* As a more complicated example, simulating the outcome of the next Superbowl involves many assumptions

### 2.2.2 Subjective probability

The relative frequency interpretation is natural in scenarios 1 through 8 of Exercise 2.1. We can consider as repeatable situations like flipping a coin, drawing powerballs from a bin, or selecting a student at random.

On the other hand, it is difficult to conceptualize scenarios 8 through 13 of Exercise 2.1 as relative frequencies. Superbowl 2020 will only be played once, the 2020 U.S. Presidential Election will only be conducted once (we hope), and there was only one April 17, 2009 on which you either did or did not eat an apple. But while these situations are not naturally repeatable they still involve randomness (uncertainty) and it is still reasonable to assign probabilities. At this point in time, the Chargers are less likely than the Patriots (ugh) to win Superbowl 2020, Donald Trump is more likely than Dwayne Johnson to win

the U.S. 2020 Presidential Election, and if you’ve always been an “apple-a-day” person, there’s a good chance you ate one on April 17, 2009. So it still makes sense to talk about probability in uncertain, but not necessarily repeated situations.

However, the *meaning* of probability does seem different in scenarios 9 through 13 compared to 1 through 8. Consider Superbowl 2020. As of Sept 16,

- According to [fivethirtyeight.com](http://fivethirtyeight.com), the Patriots have a 19% chance of winning the Superbowl, the highest of any team, while the Chargers have a 4% chance.
- According to [footballoutsider.com](http://footballoutsider.com), the Patriots have a 26% chance of winning the Superbowl, the highest of any team, while the Chargers have a 6% chance.
- According to [playoffstatus.com](http://playoffstatus.com) the Patriots have a 7% chance of winning the Superbowl, behind the Chiefs and Packers at 9% each, while the Chargers have a 3% chance.

All three websites, as well as many others, ascribe different probabilities to the Patriots or Chargers winning. Which website, if any, is correct? In the coin flipping example, we could perform a simulation to see that the long run relative frequency is 0.5. However, simulating Superbowl 2020 involves first simulating the 2019 season to determine the playoff matchups, then simulating the playoffs to see which teams make the Superbowl, then simulating the Superbowl matchup itself. And simulating the 2019 involves simulating all the weekly matchups and potential injuries and their effects. Even just simulating a single game involves many assumptions; differences in opinions with regards to these assumptions can lead to different probabilities. For example, according to [fivethirtyeight](http://fivethirtyeight), the Chiefs have a 69% chance of beating the Ravens on Sept 22, but according to [pickingpros.com](http://pickingpros.com) it’s only 52%. Unlike in the coin flipping problem, there is no single set of rules for running the simulation, and there is no single relative frequency that determines the probability. Therefore, in a situation like forecasting the Superbowl we consider *subjective probability*.

- There are many situations where the outcome is uncertain, but it does not make sense to consider the situation as repeatable. In such situations, a **subjective (a.k.a. personal) probability** describes the degree of likelihood a given individual ascribes to a certain event.
- As the name suggests, different individuals might have different subjective (personal) probabilities for the same event.
  - In contrast to the long run relative frequency situation, in which the probability is agreed to be defined as the long run relative frequency.
- The [fivethirtyeight](http://fivethirtyeight) NFL predictions are the output of a probabilistic forecast.

- **Probabilistic forecasts** combine observed data and statistical models to make predictions.
- Rather than providing a single prediction (such as “the Patriots will win Superbowl 2020”), probabilistic forecasts provide a range of scenarios and their relative likelihoods.
- Be sure to make a distinction between *assumption* and *observation*.

## 2.3 Proportional reasoning and tables of counts

- In general, knowing probabilities of individual events alone is not enough to determine probabilities of combinations of them.
- **Two-way tables** (a.k.a. contingency tables) of counts are a useful tool for probability problems dealing with two events.
- For the purposes of constructing the table and computing related probabilities, any value can be used for the hypothetical total count (as long as you don’t round numbers in between).

Suppose we are interested in the relationship between income and college graduation rates at Cal Poly. Consider the sample space  $S = \{\text{all current Cal Poly students}\}$ . Suppose that the probability<sup>6</sup> that a (randomly selected) Cal Poly student graduates within six years is 0.75, and that the probability that a (randomly selected) Cal Poly student has a Pell grant (for low income students) is 0.2.

1. Do we have enough information to find the probability that a Cal Poly student both has a Pell Grant and graduates within six years?
2. Without knowing any other information, what is the *largest* value the probability in (a) could possibly be? Under what conditions is this value attained? Set up a two-way table corresponding to this scenario.
3. Without knowing any other information, what is the *smallest* value the probability in (a) could possibly be? Under what conditions is this value attained? Set up a two-way table corresponding to this scenario.
4. Suppose that the probability that a Cal Poly student both has a Pell Grant and graduates within six years is 0.11. Construct the corresponding two-way table.
5. Find the probability that a Cal Poly student does not have a Pell Grant and graduates within six years.
6. Donny says “75% of CP students graduate within six years while 20% of CP students have a Pell Grant. So 95% of Cal Poly students graduate within six years or have a Pell Grant.” Do you agree?
7. What would need to be true in order for the Donny’s statement in the previous part to be true?

---

<sup>6</sup>These number are estimates based on data from <https://projects.propublica.org/colleges/schools/california-polytechnic-state-university-san-luis-obispo>

## 2.4 Working with probabilities

In the previous section we saw two different interpretations of probability: relative frequency and subjective. Fortunately, the mathematics of probability work the same way regardless of the interpretation. Also, even with subjective probabilities it is helpful to consider what might happen in a simulation.

### 2.4.1 Consistency requirements

With either the relative frequency or personal probability interpretation there are some basic logical consistency requirements<sup>7</sup> which probabilities need to satisfy.

**Example 2.1.** As of Sept 18, the website [fivethirtyeight.com](http://fivethirtyeight.com) listed the following probabilities for who will win the 2019 World Series.

Houston Astros	25%
Los Angeles Dodgers	21%
New York Yankees	20%
Atlanta Braves	8%

1. Is the relative frequency or subjective probability interpretation more appropriate here?
2. According to the site, what must be the probability that the Houston Astros do *not* win?
3. According to this site, what must be the probability that one of the above four teams is the World Series champion?
4. According to this site, what must be the probability that a team other than the above four teams is the World Series champion?

*Solution.* to Example 2.1

1. It is more appropriate to think of the probabilities themselves in application as subjective. Different websites or models could reasonably assign other probabilities to the teams. However, we can still imagine the probabilities as relative frequencies, if it helps our intuition. If we think of this as a simulation, each repetition results in a World Series champion and in the long run the Astros would be the champion in 25% of repetitions.
2. Either the Astros win or they don't; if there's a 25% chance that the Astros win, there must be a 75% chance that they do not win. If we think of this as a simulation, each repetition results in either the Astros winning

<sup>7</sup>In Section 3.4, we will formalize these requirements in the axioms of probability.

or not, so if they win in 25% of repetitions, they must not win in the other 75% to account for 100% of the repetitions.

3. There is only one World Series champion, so if say the Astros win then no other team can win. Thinking again of a simulation, the repetitions in which the Astros win are distinct from those in which the Dodgers win. So if the Astros win in 25% of repetitions and the Dodgers wins in 21% repetitions, then on a total of 46% of repetitions either the Astros or Dodgers win. Adding the four probabilities, we see that the probability that one of the four teams above wins must be 74%.
4. Either one of the four teams above wins, or some other team wins. If there is a 74% chance that the winner is one of the four teams above, then there must be a 26% chance that the winner is not one of these four teams.

### 2.4.2 Odds

The words “probability”, “chance”, “likelihood”, and “odds” are colloquially treated as synonyms. However, in the mathematical language of probability, *odds* provide a different way of reporting a probability. Rather than reporting probability on a 0% to 100% scale, odds report probabilities in terms of ratios.

**Example 2.2.** In Example 2.1 the odds that the Astros win the World Series are 3 to 1 against.

1. What do you think that “3 to 1 against” means?
2. What are the odds of the Astros *not* winning?
3. What are the odds of the Yankees winning?
4. What are the odds of the Braves winning?

*Solution.* to Example 2.2

1. The probability that the Astros win is 0.25, so the probability that they do not win is 0.75. These numbers are in a 3 to 1 ratio: the probability of not winning (0.75) is 3 times greater than the probability of winning (0.25). So the odds *against* the Astros winning the World Series are 3 to 1; “against” because the Astros are less likely to win than to not win.
2. The probabilities are still in the 3 to 1 ratio, but we can say that the odds are 3 to 1 *in favor* of the Astros *not* winning.
3. The probability that the Yankees win is 0.2 and that they don’t win is 0.8, and  $0.8/0.2 = 4$ . So the odds are 4 to 1 against the Yankees winning (“against” because the Yankees are less likely to win than to not win).
4. The probability that the Braves win is 0.08 and that they don’t win is 0.92, and  $0.92/0.08 = 11.5$ . So the odds are 11.5 to 1, or 23 to 2, against the Braves winning (“against” because the Braves are less likely to win than to not win).



- The **odds** of an event is a ratio involving the probability that the event occurs and the probability that the event does not occur

$$\text{odds in favor} = \frac{\text{probability that the event occurs}}{\text{probability that the event does not occur}}$$

$$\text{odds against} = \frac{\text{probability that the event does not occur}}{\text{probability that the event occurs}}$$

- In many situations (e.g., gambling) odds are implicitly reported as odds against.
- Odds are usually expressed as whole numbers, e.g., 11 to 1, 7 to 2.

Ron and Leslie make the following bet. If Boston wins, Leslie will pay Ron \$200; if not, Ron will pay Leslie \$100. If both consider this to be a fair bet, what have they agreed that the probability that Boston wins is?

The odds of a fair bet on whether or not an event will occur imply a probability for the event

$$\text{probability that event occurs} = \frac{\text{odds in favor of the event}}{1 + \text{odds in favor of the event}}$$

### 2.4.3 Dutch book

Dutch book<sup>8</sup>

Odds give us one way to see why even subjective probabilities must follow basic logical consistency requirements.

**Example 2.3.** Donny Dont is pretty sure that the Astros are going to win the World Series. He thinks their only real competition is the Braves. The following are Donny's subjective probabilities.

Houston Astros	60%
Atlanta Braves	20%
Other	10%

---

<sup>8</sup>“Book” in the sense of a bookie taking bets

## 2.5 Approximating probabilities - a brief introduction to simulation

Here's a seemingly simple problem. Flip a fair coin four times and record the results in order. For the recorded sequence, compute *the proportion of the flips which immediately follow a H that result in H*. What value do you expect for this proportion? (If there are no flips which immediately follow a H, i.e. the outcome is either TTTT or TTTH, discard the sequence and try again with four more flips.)

For example, the sequence HHTT means the the first and second flips are heads and the third and fourth flips are tails. For this sequence there are two flips which immediately followed heads, the second and the third, of which one (the second) was heads. So the proportion in question for this sequence is  $1/2$ .

So what value do you expect for this proportion? We think it's safe to say that most people would answer  $1/2$ . Afterall, it shouldn't matter if a flip follows heads or not, right? We would expect half of the flips to land on heads regardless of whether the flip follows H, right? We'll see there are some subtleties lurking behind these questions.

To get an idea of what we would expect for this proportion, we could conduct a simulation: flip a coin 4 times and see what happens. Here are the results of a few repetitions; each repetition consists of an ordered sequence of 4 coin flips for which the proportion in question is measured. (**Flips which immediately follow H are in bold.**)

Table 2.3: Simulated outcomes for 6 sets of four flips of a fair coin.

Repetition	Outcome	Flips that follow H	H that follow H   Pro	Proportion of H following H
1	H <b>H</b> TT	2	1	0.5
2	H <b>T</b> TH	1	0	0
3	T <b>T</b> TH	0	NA	try again
4	T <b>H</b> HH	2	2	1
5	H <b>H</b> TT	2	1	0.5
6	H <b>H</b> HT	3	2	0.667

We can keep repeating the above process to investigate what happens in the long run. Rather than actually flipping coins, we use a computer to run a simulation. The table and plot below summarize the results of 1,000,000 repetitions of the simulation<sup>9</sup>. While you can't see the individual "dots" in the plot, each dot would represent a sequence of 4 coin flips (with at least one flip following a H) and the value plotted being plotted is the proportion of H following H for that sequence. The results could be summarized in a table like Table 2.3, albeit with 1,000,000 rows (after discarding rows with no flips immediately following H.)

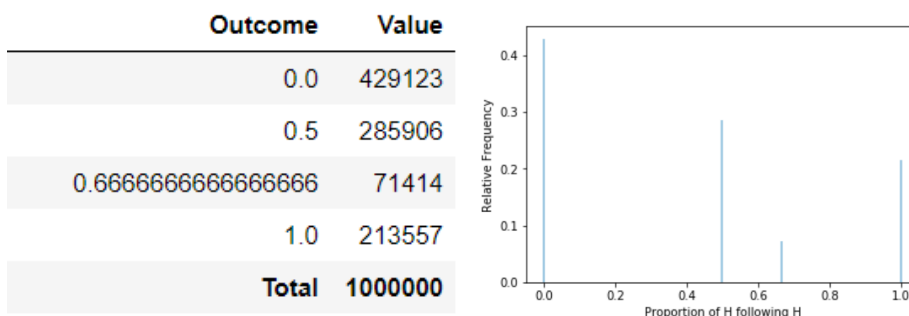


Figure 2.4: Proportion of flips immediately following Heads that result in Heads for 1,000,000 sets of 4 coin flips. (Each set has at least one flip immediately following H.) For example, the proportion in 429,123 of sets

We asked the question: *what would you expect* for the proportion of the flips which immediately follow a H that result in H? That depends on how we define what's "expected". If we are interested in the value that is most likely to occur when we flip a coin four times, then the answer is 0: we see that in the long run a little over 40% of the sets resulted in a proportion of 0, while only about 30% of sets resulted in a value of  $1/2$ . We see that the plot is not centered at  $1/2$ ; a higher percentage of repetitions resulted in a proportion below  $1/2$  than above  $1/2$ . We think that most people would find this surprising.

Another way to interpret "expected" is as "average". Just as probability can be interpreted as long run relative frequency, there is a concept called *expected value* which can be interpreted as the *long run average value*. After 1,000,000 repetitions, each involving a set of four fair coin flips, we have 1,000,000 simulated values of the proportion of H following H, as recorded in the rightmost column of Table 2.3. We could then average these values: add up the values in the column and divide by 1,000,000. It turns out that average value is 0.405, which is not  $1/2$ . Again, we think most people find this surprising.

A quick note: the term "expected value" is somewhat of a misnomer. We are *not* saying that if we flip a coin four times we would expect the proportion of H following H for that set of flips to be 0.405. In fact, the simulation shows

<sup>9</sup>Section XXX covers how to program the simulation and analyze the results.

that on any single set of four fair coin flips, the only possible values for the proportion of H following H are 0,  $1/2$ ,  $2/3$ , and 1. So in a set of four coin flips it's not possible to see a proportion of 0.405. Rather, 0.405 is the *average value of the proportion of H following H that we would expect to see in the long run over many sets of four fair coin flips*. We will return to this idea later.

We will return to this example several times throughout the book to investigate these results more closely. For now, just observe that

- The study of probability can involve some subtleties and our intuition isn't always right.
- Simulation is an effective way of investigating probability problems, and can reveal interesting and suprising patterns.

In MS coin - add statement about probability of H after H versus proportion of H after H

## 2.6 Sliding scale of probability, or Probability of what?

Here's a funny series of cartoons

What does 1/million mean?

At least one, etc.

xkcd/538 benchmarks, put in benchmarks (one in a million, a billion, etc)

probability of coincidences

## 2.7 Common misinterpretation and fallacies (e.g. outbreak of Asian disease, Utts book)

Sally Clark?

Allais paradox?

A famous study<sup>10</sup> by Kahneman and Tversky presented a group of people with the following scenario.

"Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to

---

<sup>10</sup>Source: <https://www.ncbi.nlm.nih.gov/pubmed/7455683>

combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is  $1/3$  probability that 600 people will be saved, and  $2/3$  probability that no people will be saved.

Which of the two programs would you favor?"

Which of these two options, A or B, would you choose? Of the 152 participants, a large majority (72%) favored one of the options; which one do you think it was?

A second group was presented with the same set up, but the following options instead of A/B.

If Program C is adopted 400 people will die.

If Program D is adopted there is  $1/3$  probability that nobody will die, and  $2/3$  probability that 600 people will die.

Which of these two options, C or D, would you choose? Of the 155 participants, a large majority (78%) favored one of the options; which one do you think it was?

Compute the expected number of people who die and who are saved with Program D.

Compute the expected number of people who die and who are saved with Program B.

Compare options A and C. Are these the same? How about B and D? Do the risk preferences depend on the wording of how the information is presented?

Many studies have shown that human intuition does not generally deal well with issues of uncertainty.

So it's worthwhile to take a careful study

**Example 2.4** (Monty Hall problem). Monty Hall in Chapter 1???

## 2.8 Why study coins, dice, cards, and spinners?

Many probability problems involve "toy" situations like flipping coins, rolling dice, shuffling cards, or spinning spinners. These situations might seem unexciting, or at least not very practically meaningful. However, coins and spinners and the like provide familiar, concrete situations which facilitate understanding

of probability concepts. Furthermore, simple situations often provide insight into real and complex problems. The following is just one illustration.

Many basketball players and fans alike believe in the “hot hand” phenomenon: the idea that making several shots in a row increases a player’s chances of making the next shot. However, the consensus conclusion of thirty years of studies on the hot hand, beginning with the seminal study Gilovich et al. (1985), had been that there is no statistical evidence that the hot hand in basketball is real. As a result, many statisticians regularly caution against the “hot hand fallacy”: the belief that the hot hand exists when, in reality, the degree of streaky behavior typically observed in sequential data is consistent with what would be expected simply by chance in independent trials.

The idea behind studies like Gilovich et al. (1985) is essentially the following. Consider a player who attempts 100 shots and makes 50%. If there is no hot hand, then we might expect the player to make 50% of shots both on attempts that follow hit streaks — usually considered three (or more) made attempts in a row — and on other attempts. Therefore, a success rate of 50% on both sets of attempts provides no evidence of the hot hand.

However, recent research of Miller and Sanjurjo (2018c), Miller and Sanjurjo (2018b), Miller and Sanjurjo (2018a) concludes that previous studies on the hot hand in basketball, starting with Gilovich et al. (1985), have been subject to a bias. After correcting for the bias, the authors find strong evidence in favor of the hot hand effect in basketball shooting, suggesting the hot hand fallacy is not a fallacy after all. One interesting aspect of these studies is that Miller and Sanjurjo’s methods are simulation-based.

Miller and Sanjurjo (2018c) introduced the coin flipping problem in Section 2.5 to illustrate the idea behind their research and the bias in previous studies. Consider again a player who attempts 100 shots and makes 50%. Even if there is no hot hand, Miller and Sanjurjo show that we would actually expect the player to have a shooting percentage of *strictly less than* 50% on the attempts which followed streaks, and strictly greater than 50% on the other attempts. The reason is the same as for the coin flipping problem in Section 2.5: in a fixed number of trials, the proportion of H on trials following H is expected to be less than the true probability of H, even though the trials are independent. Therefore, for the example player a success rate of 50% on both sets of attempts actually provides directional evidence in favor of the hot hand. Properly accounting for this bias leads to substantially different statistical analyses (i.e., p-values) and conclusions.

## Chapter 3

# The Language of Probability

CHANGE EVAL TO TRUE AFTER FIGURE PYTHON ERRORS OUT

This chapter introduces the fundamental terminology, objects, and mathematics of random phenomena.

### 3.1 Sample space of outcomes

A phenomenon is **random** if there are multiple potential *outcomes*, and there is *uncertainty* about which outcome will occur. Because there is wide variety in types of random phenomena, an outcome can be virtually anything:

- a sequence of coin flips
- a shuffle of a deck of cards
- the weather tomorrow in your city
- the path of a hurricane
- a sample of registered voters
- player locations over time in an NBA game (as recorded by the SportsVU system)

And on and on. In particular, an outcome does *not* have to be a number.

The first step in defining a model for a random phenomenon is to describe the possible outcomes.

**Definition 3.1.** The **sample space**, denoted  $\Omega$  (the uppercase Greek letter “omega”), is the set of all possible outcomes of a random phenomenon. An

**outcome**, denoted  $\omega$  (the lowercase Greek letter “omega”), is an element of the sample space:  $\omega \in \Omega$ .

Mathematically, the sample space  $\Omega$  is a *set* containing all possible outcomes, while an individual outcome  $\omega$  is a *point* or *element* in  $\Omega$ . The symbol  $\omega$  denotes a generic outcome, much like the symbol  $u$  in  $\sqrt{u}$  denotes a generic input to the square root function.

For example, the simplest random phenomena have just two distinct outcomes, in which case the sample space is just a set with two elements, e.g.  $\Omega = \{\text{no}, \text{yes}\}$ ,  $\Omega = \{\text{off}, \text{on}\}$ ,  $\Omega = \{0, 1\}$ ,  $\Omega = \{-1, 1\}$ . For example, the sample space for a single coin flip could be  $\Omega = \{H, T\}$ . If the coin lands on heads, we observe the outcome  $\omega = H$ ; if tails we observe  $\omega = T$ .

A random phenomenon is modeled by a *single* sample space, upon which all objects are defined. These objects, which we will encounter later, include events, random variables, stochastic processes.

Whenever possible, a sample space outcome should be defined to provide the maximum amount of information about the outcome of random phenomenon. While a random phenomenon always has a corresponding sample space, we will see that in many situations the sample space of outcomes is at best only vaguely specified and can not be feasibly enumerated.

**Example 3.1.** Roll a four-sided die<sup>1</sup> twice, and record the result of each roll in sequence as an ordered pair. For example, the outcome  $(3, 1)$  represents a 3 on the first roll and a 1 on the second; this is not the same outcome as  $(1, 3)$ .

1. Identify the sample space.
2. We might be interested in the sum of the two dice. Explain why it is still advantageous to define the sample space as in the previous part, rather than as  $\Omega = \{2, \dots, 8\}$ .

*Solution.* to Example 3.9

1. The sample space consists of 16 possible ordered pairs of rolls

$$\begin{aligned}\Omega = \{ & (1, 1), (1, 2), (1, 3), (1, 4), \\ & (2, 1), (2, 2), (2, 3), (2, 4), \\ & (3, 1), (3, 2), (3, 3), (3, 4), \\ & (4, 1), (4, 2), (4, 3), (4, 4) \}\end{aligned}$$

It is more helpful to represent the above set in a table like the following

---

<sup>1</sup>Why four-sided? Simply to make the number of possibilities a little more manageable (e.g., for in-class simulation activities). Rolling a four-sided die twice yields 16 possible pairs, while rolling a six-sided die yields 36 possible pairs.



1. Yes, we might be interested in the sum of the two dice. But we might also be interested in other things, like the larger of the two rolls, or if at least one 3 was rolled, or the first number rolled. Knowing just the sum of the dice does not provide as much information about the outcome of the random phenomenon as the sequence of individual rolls does.

**Note:** In the previous example, there was a single sample space whose outcomes represented the result of the pair of rolls. In particular, there was not a separate sample space for each of the individual rolls. Now, we could have written the sample space as the Cartesian product  $\Omega = \{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$ , where the first, respectively second, set in the product represents the result of the first, respectively second, roll. But this Cartesian product represents a single set of ordered pairs, and it is that single set which is the sample space corresponding to outcomes of the pair of rolls.

**Example 3.2.** Consider the outcome of a sequence of 4 flips of a coin.

1. Identify an appropriate sample space.
2. We might be interested in the number of heads flipped. Explain why it is still advantageous to define the sample space as in the previous part, rather than as  $\Omega = \{0, 1, 2, 3, 4\}$ .

*Solution.* to Example 3.2

1. An outcome is an ordered sequence representing the results of the four flips. For example, HTH T means heads on the first on third flips and tails on the second and fourth; this is not the same outcome as HHTT or THTH. The sample space is the following set composed of 16 distinct outcomes

$$\Omega = \{HHHH, HHHT, HHHT, HTHH, THHH, HHTT, HTHT, HTTH, \\ THHT, THTH, TTHH, HTTT, THTT, TTHT, TTTH, TTTT\}$$

2. Yes, we might be interested in the number of heads flipped, but we might also be interested in other things, such as whether the first flip was heads, the length of the longest streak of heads in a row, or the proportion of flips following H that resulted in H. Knowing just the number of heads flipped does not provide as much information about the outcome of the random phenomenon as the sequence of individual flips does.

**Note:** We reiterate what we said after Example 3.9. In the previous example, there was a single sample space whose outcomes were sequences of coin flips. In particular, there was not a separate sample space for each of the individual flips. We could have written the sample space as the Cartesian product  $\Omega = \{H, T\} \times \{H, T\} \times \{H, T\} \times \{H, T\} = \{H, T\}^4$ . But this Cartesian product represents a single set of sequences of 4 flips, and it is this single set which is the sample space corresponding to the sequence of four flips.

**Example 3.3** (Matching problem). The “matching problem” is one well known probability problem. The following version is from FiveThirtyEight.

A geology museum in California has four different rocks sitting in a row on a shelf, with labels on the shelf telling what type of rock each is. An earthquake hits and the rocks all fall off the shelf and get mixed up. A janitor comes in and, wanting to clean the floor, puts the rocks back on the shelf in random order. We might be interested in whether all the rocks are put back in the correct spot, or none are, or if the heaviest rock is put back in the correct spot.

Label the rocks 1, 2, 3, 4, and also label the spots 1, 2, 3, 4. Identify an appropriate sample space.

*Solution.* to Example 3.3

We can consider each outcome to be a particular placement of rocks in the spots. For example, the outcome 3214 (or  $(3, 2, 1, 4)$ ) represents that rock 3 is placed in spot 1, rock 2 in spot 2, rock 1 in spot 3, and rock 4 in spot 4. (We say that an outcome is a *permutation* (or reordering) of the numbers 1, 2, 3, 4.) So the sample space consists of the following 24 outcomes<sup>2</sup>.

$$\begin{aligned}\Omega = \{ & 1234, 1243, 1324, 1342, 1423, 1432 \\ & 2134, 2143, 2314, 2341, 2413, 2431 \\ & 3124, 3142, 3214, 3241, 3412, 3421 \\ & 4123, 4132, 4213, 4231, 4312, 4321 \}\end{aligned}$$

Recording outcomes in this way provides more information than if we had chosen the sample space to correspond to, for example, the number of rocks that were placed in the correct spot.

**Example 3.4** (Collector problem). So-called “collector problems” concern the following generic scenario. A set of  $n$  cards labeled  $1, 2, \dots, n$  are placed in a box (or shuffled in a deck). Cards are selected one at a time, *with replacement*, indefinitely. Identify an appropriate sample space. Selection with replacement means that a card which has been selected is returned to the box and the cards are reshuffled before the next draw is made. (“Collector problems” typically concern the number of draws needed to observe each of the  $n$  tickets. For example, roll a fair six-sided die until each of the six faces is rolled at least once, or collecting prizes until a complete set is obtained.)

*Solution.* to Example 3.4

---

<sup>2</sup>There are 4 rocks that could potentially go in spot 1, then 3 rocks that could potentially go in spot 2, 2 to spot 3, and 1 left for spot 4. This results in  $4 \times 3 \times 2 \times 1 = 4! = 24$  possible outcomes. We will see more counting rules in XXXX.

We can let  $\Omega = \{1, \dots, n\}^\infty$ . That is, each outcome is an infinite sequence — since draws continue indefinitely — with each element in the sequence in  $\{1, \dots, n\}$ . For example,  $(3, 2, 6, 3, 4, 1, 2, 4, \dots)$  means card 3 is the first selection, card 2 the second, card 6 the third, card 3 the fourth, etc. Recording outcomes in this way, we could investigate the number of draws needed to obtain each of the  $n$  cards at least  $r$  times, for any  $r = 1, 2, \dots$  (Note that  $\Omega$  is an uncountable set.)

**Example 3.5.** Many statistical applications involve *random sampling*. For example, polling organizations often select random samples of Americans. Typically the selection involves random digit dialing: a sample of say 1000 phone numbers are randomly selected from some large bank (hundreds of millions) of phone numbers; this is the *population*. An outcome would consist of the 1000 phone numbers selected; this is the *sample*.

As an extraordinarily unrealistic, oversimplified, but concrete example, suppose the bank only contains 5 numbers, labeled  $\{1, 2, 3, 4, 5\}$ , from which 3 are selected. Describe an appropriate sample space. Note: the order in which the numbers are selected does not matter; we only care which numbers are selected.

*Solution.* to Example 3.5

An outcome consists of a *subset* of size 3 from the set  $\{1, 2, 3, 4, 5\}$ , representing the list of the 3 phone numbers selected. There are 10 distinct outcomes

$$\Omega = \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 4, 5\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 4, 5\}, \{3, 4, 5\}\}$$

In a more realistic setting, the population would consist of hundreds of millions of phone numbers, and the sample space would be composed of all possible subsets (samples) of 1000 phone numbers. Even if the order in which the numbers are selected is irrelevant, the sample space is enormous and could never be feasibly enumerated as in this oversimplified example. But the idea is the same: an outcome consists of a *subset* of numbers, and the sample space is a collection of possible subsets. (That is, the sample space is a set of sets.)

RANDOM ASSIGNMENT EXAMPLE, or save for exercise?

**Example 3.6.** Regina and Cady plan to meet for lunch between noon and 1 but they are not sure of their arrival times. We might be interested in questions involving whether they arrive within 15 minutes of one another, who arrives first, or how long the first person to arrival needs to wait for the second,

Describe an appropriate sample space. Rather than dealing with clock time, it is helpful to represent noon as time 0 and measure time as fraction of the hour after noon, so that times take values in the continuous interval  $[0, 1]$ . For example, 12:15 corresponds to a time of 0.25, 12:30 to 0.5, 12:42 to 0.7.

*Solution.* to Example 3.6

An outcome is a pair of values  $\omega = (\omega_1, \omega_2)$  corresponding to the arrival times of (Regina, Cady). For example, the outcome  $(0.5, 0.7)$  represents Regina arriving at time 0.5 (12:30) and Cady at time 0.7 (12:42), while  $(0.7, 0)$  represents Regina arriving at time 0.7 (12:42) and Cady at time 0 (noon). The sample space is  $\Omega = [0, 1] \times [0, 1] = [0, 1]^2$ , the Cartesian product  $\{(\omega_1, \omega_2) : \omega_1 \in [0, 1], \omega_2 \in [0, 1]\}$ .

In the above example, outcomes were measured on a continuous scale; any real number between 0 and 1 was a possible outcome of a single spin (of the idealized, infinitely precise spinner). Even in situations where outcomes are inherently discrete, it is often more convenient to model them as continuous. For example, if an outcome represents the annual salary in dollars of a randomly selected U.S. household, it would be more convenient to model the sample space as the continuous interval  $[0, \infty)$  rather than  $[0, m]$  where  $m$  is some large dollar amount providing an upper bound on the maximum possible salary. But we would need to be sure that  $m$  is large enough so that all possible outcomes are in the sample space  $[0, m]$ . Without knowing this bound in advance, it is convenient to just choose the unbounded interval  $[0, \infty)$  rather than  $\{0, 1, 2, \dots\}$  or  $\{0, 0.01, 0.02, \dots\}$ .

**Example 3.7.** Many games involve rolling a die or spinning a spinner like the one in Figure 3.1 which returns values between 0 and 1. A spinner can be thought of as a continuous analog of a die roll. The values in the picture are rounded to two decimal places, but consider an idealized model where the spinner is infinitely precise so that any real number between 0 and 1 is a possible outcome. The sample space corresponding to a single spin is then  $\Omega = [0, 1]$ , a continuous interval of real numbers. Now suppose the spinner is spun twice; what is an appropriate sample space?

*Solution.* to Example 3.7

An outcome is now a pair of values  $(\omega_1, \omega_2)$  corresponding to the results of the first and second spins. The sample space is  $\Omega = [0, 1] \times [0, 1]$ , where  $[0, 1] \times [0, 1]$ , also denoted  $[0, 1]^2$ , is the Cartesian product  $[0, 1] \times [0, 1] = \{(\omega_1, \omega_2) : \omega_1 \in [0, 1], \omega_2 \in [0, 1]\}$ .

In the above (idealized) example, outcomes were measured on a continuous scale; any real number between 0 and 1 was a possible outcome of a single spin (of the idealized, infinitely precise spinner). Even in situations where outcomes are inherently discrete, it is often more convenient to model them as continuous. For example, if an outcome represents the annual salary in dollars of a randomly selected U.S. household, it would be more convenient to model the sample space as the continuous interval  $[0, \infty)$  rather than  $[0, m]$  where  $m$  is some large dollar amount providing an upper bound on the maximum possible salary. But we would need to be sure that  $m$  is large enough so that all possible outcomes are in the sample space  $[0, m]$ . Without knowing this bound in advance, it is convenient to just choose the unbounded interval  $[0, \infty)$  rather than  $\{0, 1, 2, \dots\}$  or  $\{0, 0.01, 0.02, \dots\}$ .

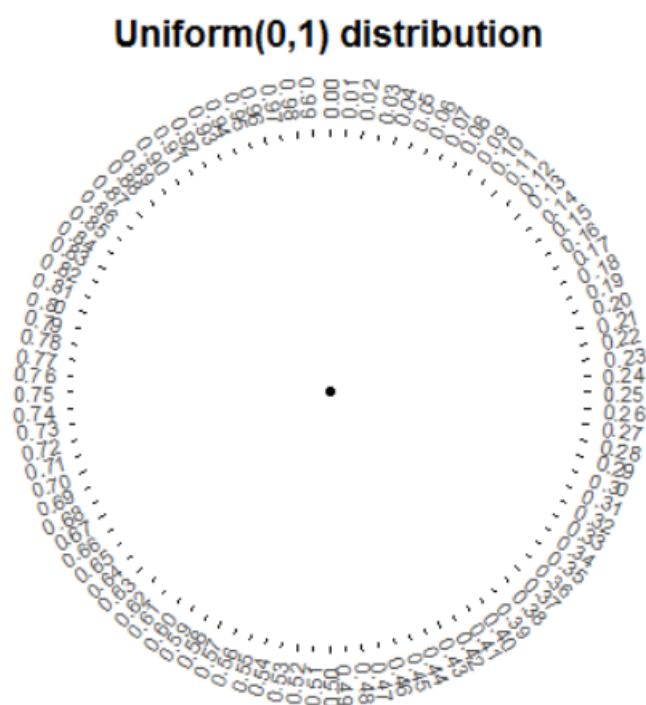


Figure 3.1: A Uniform(0, 1) spinner. The values in the picture are rounded to two decimal places, but in the idealized model the spinner is infinitely precise so that any real number between 0 and 1 is a possible outcome.

**Example 3.8.** Select a U.S. high school student and record the student’s SAT Math and Reading score. What is an appropriate sample space?

*Solution.* to Example 3.8

An outcome will be an ordered pair represent (Math, Reading) score. Technically, possible scores are 200 through 800 in increments of 10. So we could consider the sample space to be  $\{200, 210, 220, \dots, 790, 800\} \times \{200, 210, 220, \dots, 790, 800\}$ .

However, we could also model scores on a continuous scale, taking any value in the interval from 200 to 800. In this case, the sample space would be  $\Omega = [200, 800] \times [200, 800]$ . Such a specification is often more convenient mathematically.

Add somewhere something like: In practice we rarely enumerate the sample space as we did in the examples in this section. Nonetheless, there is always some underlying sample space corresponding to all possible outcomes of the random phenomenon. Even when we do not specify the sample space, it is important to consider what the sample space is. We must first determine what is possible before determining when is probable. The sample space in essence defines our denominator. Does the sample space corresponding to a single lottery ticket or some lottery ticket somewhere? (Need an example that highlights different denominators of sample space. Maybe like Oscar winner example?)

## 3.2 Events

In a random phenomenon, an “event” is something that could happen. For example, if we’re interested in tomorrow’s weather, events include

- “the high temperature is 75°F”
- “the high temperature is above 75°F”
- “it rains”
- “it does not rain”
- “it rains and the high temperature is above 75°F”
- and so on

Mathematically, an *event* is a collection of sample space outcomes that satisfy some criteria.

**Definition 3.2.** An **event**  $A$  is a *subset* of the sample space:  $A \subseteq \Omega$ . The collection of all events of interest<sup>3</sup> is denoted  $\mathcal{F}$ . If the random phenomenon

---

<sup>3</sup>For the purposes of this text,  $\mathcal{F}$  can be considered to be the set of all subsets of  $\Omega$ . Technically,  $\mathcal{F}$  is a  $\sigma$ -field of subsets of  $\Omega$ :  $\mathcal{F}$  contains  $\Omega$  and is closed under countably many elementary set operations (complements, unions, intersections). This requirement ensures that if  $A$  and  $B$  are “events of interest”, then so are  $A \cup B$ ,  $A \cap B$ , and  $A^c$ . While this level of technical detail is not needed, we prefer to introduce the idea of a “collection of events” now since a probability measure is a function whose input is an event (set) and not an outcome.

yields outcome  $\omega$ , we say “event  $A$  occurred” if  $\omega \in A$ .

- Recall that the sample space is the collection of all possible outcomes of a random phenomenon.
- An event represents those outcomes which satisfy some criteria.
- There is a single sample space, upon which all events are defined.
- An outcome in a sample space should be defined to record as much information as possible so that the occurrence or non-occurrence of all events of interest can be determined.
- Events are typically denoted with capital letters near the start of the alphabet, with or without subscripts (e.g.  $A$ ,  $B$ ,  $C$ ,  $A_1$ ,  $A_2$ )
- Events can be composed from others using basic set operations like unions ( $A \cup B$ ), intersections ( $A \cap B$ ), and complements ( $A^c$ ).
  - Read  $A^c$  as “not”  $A$ .
  - Read  $A \cap B$  as “ $A$  and  $B$ ”
  - Read  $A \cup B$  as “ $A$  or  $B$ ”. Note that unions ( $\cup$ , “or”) are always inclusive.  $A \cup B$  occurs if  $A$  occurs but  $B$  does not,  $B$  occurs but  $A$  does not, or both  $A$  and  $B$  occur.
- An event  $A$  is a set. The collection  $\mathcal{F}$  is a *collection of sets*.
- While an event is always a set, it can be a set consisting of a single outcome, or no outcomes at all (the empty set  $\emptyset$ ).
- A collection of events  $A_1, A_2, \dots$  are **disjoint** (a.k.a. mutually exclusive) if  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ . Events are disjoint if they have no outcomes in common.

As an example, consider a single roll of a four-sided die. The sample space consists of the four possible outcomes  $\Omega = \{1, 2, 3, 4\}$ . Any subset of this sample space is an event. The following table lists the collection of all events ( $\mathcal{F}$ ), and whether they occur if the single roll results in a 3 (that is, for the outcome  $\omega = 3$ ).

Event	Description	Occurs upon outcome $\omega = 3$ ?
$\emptyset$	Roll nothing (not possible)	No
$\{1\}$	Roll a 1	No
$\{2\}$	Roll a 2	No
$\{3\}$	Roll a 3	Yes
$\{4\}$	Roll a 4	No
$\{1, 2\}$	Roll a 1 or a 2	No
$\{1, 3\}$	Roll a 1 or a 3	Yes
$\{1, 4\}$	Roll a 1 or a 4	No
$\{2, 3\}$	Roll a 2 or a 3	Yes
$\{2, 4\}$	Roll a 2 or a 4	No
$\{3, 4\}$	Roll a 3 or a 4	Yes
$\{1, 2, 3\}$	Roll a 1, 2, or 3 (a.k.a. do not roll a 4)	Yes

Event	Description	Occurs upon outcome $\omega = 3$ ?
$\{1, 2, 4\}$	Roll a 1, 2, or 4 (a.k.a. do not roll a 3)	No
$\{1, 3, 4\}$	Roll a 1, 3, or 4 (a.k.a. do not roll a 2)	Yes
$\{2, 3, 4\}$	Roll a 2, 3, or 4 (a.k.a. do not roll a 1)	Yes
$\{1, 2, 3, 4\}$	Roll something	Yes

**Example 3.9.** Roll a four-sided die *twice*, and record the result of each roll in sequence. For example, the outcome  $(3, 1)$  represents a 3 on the first roll and a 1 on the second; this is not the same outcome as  $(1, 3)$ .

1. Identify the sample space.
2. Identify  $A$ , the event that the sum of the two dice is 4.
3. Identify  $B$ , the event that the sum of the two dice is at most 3.
4. The previous events consider the sum of the two dice. Explain why we don't just consider the sample space to be  $\{2, \dots, 8\}$ , so that for example  $B = \{2, 3\}$ .
5. Identify  $C$ , the event that the larger of the two rolls (or the common roll if a tie) is 3
6. Identify and interpret  $A \cap C$ .
7. Identify  $D$ , the event that the first roll is a 3.
8. Identify  $E$ , the event that the second roll is a 3.
9. Identify and interpret  $D \cap E$ .
10. Identify and interpret  $D \cup E$ .
11. If the outcome is  $(1, 3)$ , which of the events above occurred?

*Solution.* to Example 3.9

1. The sample space consists of 16 possible ordered pairs of rolls

$$\Omega = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (2, 4), \\ (3, 1), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), (4, 4)\}$$

2.  $A = \{(1, 3), (2, 2), (3, 1)\}$  is the event that the sum of the two dice is 4.
3.  $B = \{(1, 1), (1, 2), (2, 1)\}$  is the event that the sum of the two dice is at most 3.
4. We might be interested in events other than ones that involve the sum of the dice, like those in the following parts. Knowing just the sum of the dice does not provide enough information to investigate such events.
5.  $C = \{(1, 3), (2, 3), (3, 1), (3, 2), (3, 3)\}$ , the event that the larger of the two rolls (or the common roll if a tie) is 3
6.  $A \cap C = \{(1, 3), (3, 1)\}$  is the event that both the sum of the two dice is 4 and the larger of the two rolls is 3.
7.  $D = \{(3, 1), (3, 2), (3, 3), (3, 4)\}$ , the event that the first roll is a 3. Note that each outcome in the sample space consists of a pair of rolls, so we must account for both rolls in defining events, even if the event of interest



involves just the first roll. There is always a single sample space upon which all events are defined.

8.  $E = \{(1, 3), (2, 3), (3, 3), (4, 3)\}$ , the event that the second roll is a 3. Note that this is not the same event as  $D$ .
9.  $D \cap E = \{(3, 3)\}$  is the event that both rolls result in a 3. While an event is always a set, it can be a set consisting of a single outcome (or the empty set).
10.  $D \cup E = \{(3, 1), (3, 2), (3, 3), (3, 4), (1, 3), (2, 3), (4, 3)\}$  is the event that at least one of the two rolls results in a 3. Notice that the union is inclusive:  $(3, 3)$  is an element of  $D \cup E$ . But also notice that the outcome  $(3, 3)$  only appears once in the set  $D \cup E$ .
11. If the outcome is  $(1, 3)$  then events  $A, C, A \cap C, E, D \cup E$  all occur. Events  $B, D$ , and  $D \cap E$  do not occur.

**Example 3.10.** Flip a coin 4 times, and record the result of each trial in sequence. For example, HTTH means heads on the first on last trial and tails on the second and third.

1. Specify the sample space.
2. Identify  $A$ , the event that exactly 3 of the flips land on heads.
3. Identify  $B$ , the event that exactly 4 of the flips land on heads.
4. Identify  $C$ , the event that the at least 3 of the flips land on heads. How does  $C$  relate to  $A$  and  $B$ ?
5. The previous events all consider the number of heads flipped. Explain why we don't just consider the sample space to be  $\{0, 1, 2, 3, 4\}$ , so that for example  $C = \{3, 4\}$ .
6. Identify  $D$ , the event that at least 3 heads are flipped in a row.
7. Identify  $E$ , the event that the first two flips results in tails.
8. Identify  $D \cap E$ .
9. If the outcome is HHTH, which of the events above occurred?

*Solution.* to Example 3.10

1. The sample space is composed of 16 outcomes

$$\Omega = \{HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, HTTH, THHT, THTH, TTHH, HTTT, THTT, TTHT, TTTH, TTTT\}$$

2.  $A = \{HHHT, HHTH, HTHH, THHH\}$  is the event that exactly 3 of the flips land on heads
3.  $B = \{HHHH\}$ , the event that exactly 4 of the flips land on heads. While an event is always a set, it can be a set consisting of a single outcome.
4.  $C = \{HHHT, HHTH, HTHH, THHH, HHHH\}$  is the event that the at least 3 of the flips land on heads. Also  $C = A \cup B$ .
5. Yes, the previous events all consider the number of heads flipped, but we might be interested in events — like the ones in the following parts —

whose occurrence cannot be determined simply by knowing the number of heads. The sample space should always be defined in such a way to provide enough information to investigate any relevant event of interest. There is always a single sample space upon which all events are defined.

6.  $D = \{HHHT, THHH, HHHH\}$  is the event that at least 3 heads are flipped in a row. Note that  $D$  is not the same event as  $C$ .
7.  $E = \{TTHH, TTHT, TTTH, TTTT\}$  is the event that the first two flips result in tails. Note that each outcome consists of the results of 4 flips, so we must account for all 4 flips in defining events. There is always a single sample space upon which all events are defined.
8.  $D \cap E = \emptyset$  so the events  $D$  and  $E$  are disjoint; it is not possible to have at least three heads in a row when the first two flips (out of 4) are tails.
9. If the outcome is HHTH then events  $A$  and  $C$  occur. Events  $B$ ,  $D$ ,  $E$ , and  $D \cap E$  do not occur.

**Example 3.11** (Matching problem). So-called “matching problems” concern the following generic scenario. A set of  $n$  cards labeled  $1, 2, \dots, n$  are placed in  $n$  boxes labeled  $1, 2, \dots, n$ , with exactly one card in each box. Typical questions of interest involve whether the number of a card matches the number of the box in which it is placed. (More colorful descriptions include returning babies at random to mothers or placing rocks at random back on a museum shelf.) Consider the case  $n = 4$ .

1. Identify a reasonable sample space.
2. Specify the event that all cards match their box.
3. Specify the event that no cards match their box.
4. Specify the event that exactly 3 cards match their box.
5. Specify the event that card 3 is placed in box 3.

*Solution.* to Example 3.11

1. We can consider each outcome to be a particular placement of cards in the boxes. For example, the outcome 3214 (or  $(3, 2, 1, 4)$ ) represents that card 3 is placed in box 1, card 2 in box 2, card 1 in box 3, and card 4 in box 4. So the sample space consists of the following 24 outcomes<sup>4</sup>.

$$\begin{aligned}\Omega = \{ & 1234, 1243, 1324, 1342, 1423, 1432 \\ & 2134, 2143, 2314, 2341, 2413, 2431 \\ & 3124, 3142, 3214, 3241, 3412, 3421 \\ & 4123, 4132, 4213, 4231, 4312, 4321 \}\end{aligned}$$

2.  $A = \{1234\}$  is the event that all cards match their box. While an event is always a set, it can be a set consisting of a single outcome (or the empty set).

---

<sup>4</sup>There are 4 cards that could potentially go in box 1, then 3 cards that could potentially go in box 2, 2 to box 3, and 1 left for box 4. This results in  $4 \times 3 \times 2 \times 1 = 4! = 24$  possible outcomes. We will see more counting rules in Chapter ??

3.  $B = \{2143, 2341, 2413, 3142, 3412, 3421, 4123, 4312, 4321\}$  is the event that no cards match their box.
4.  $\emptyset$  is the event in which exactly 3 cards match their box. There are no outcomes in which exactly 3 cards match; if three cards match, then the fourth card must necessarily match too.
5.  $C = \{1234, 1432, 2134, 2431, 4132, 4231\}$  is the event that card 3 is placed in box 3. Since our sample space consists of the placements of each of the cards, each event must be expressed in terms of these outcomes.

Remember that intervals of real numbers such as  $(a, b)$ ,  $[a, b]$ ,  $(a, b]$  are also sets, and so can also be events.

**Example 3.12.** Consider as a sample space all possible scores on the SAT Math exam (ranging from 200 to 800). Even though actual scores are multiples of 10, it is mathematically convenient to consider the sample space as  $\Omega = [200, 800]$ , the set of all real numbers between 200 and 800 (rather than  $\{200, 210, 220, \dots, 800\}$ .)

1. Identify  $A$ , the event that an SAT Math score is at most 700.
1. Identify  $B$ , the event that an SAT Math score is greater than 500.
1. Identify and interpret  $A \cap B$ .

*Solution.* to Example 3.12

1.  $A = [200, 700]$  is the event that an SAT Math score is at most 700.
2.  $B = (500, 800]$  is the event that an SAT Math score is greater than 500.
3.  $A \cap B = (500, 700]$  is the event that an SAT Math score is greater than 500 but at most 700.

**Example 3.13.** Suppose the spinner in Figure 3.1 is spun twice. Let the sample space be  $\Omega = [0, 1] \times [0, 1]$ . Represent each of the following events using set notation, but also sketch a picture. (Hint: represent the sample space as a square.)

1. Identify  $A$ , the event that the first spin is larger than the second.
2. Identify  $B$ , the event that the smaller of the two spins (or common value if a tie) is less than 0.5.
3. Identify  $C$ , the event that the sum of the two spins is less than 1.5.
4. Identify  $D$ , the event that the first spin is less than 0.4.

*Solution.* to Example 3.13

1.  $A = \{(\omega_1, \omega_2) : \omega_1 > \omega_2\}$  is the event that the first spin is larger than the second. Note: we only consider  $(\omega_1, \omega_2)$  in the sample space  $[0, 1] \times [0, 1]$ ; the conditions  $0 \leq \omega_1 \leq 1, 0 \leq \omega_2 \leq 1$  are assumed.

2.  $B = \{(\omega_1, \omega_2) : \min(\omega_1, \omega_2) < 0.5\}$  is the event that the smaller of the two spins (or common value if a tie) is less than 0.5. This is equivalent to the event that at least one of the spins is less than 0.5  $B = \{(\omega_1, \omega_2) : \omega_1 < 0.5 \text{ or } \omega_2 < 0.5\} = ([0.5, 1] \times [0.5, 1])^c$ .
3.  $C = \{(\omega_1, \omega_2) : \omega_1 + \omega_2 \geq 1.5\} = \{(\omega_1, \omega_2) : \omega_2 \geq 1.5 - \omega_1\}$  is the event that the sum of the two dice is at least 1.5.
4.  $D = \{(\omega_1, \omega_2) : \omega_1 < 0.4\}$  is the event that the first spin is less than 0.4. Remember, sample space outcomes are pairs of spins.

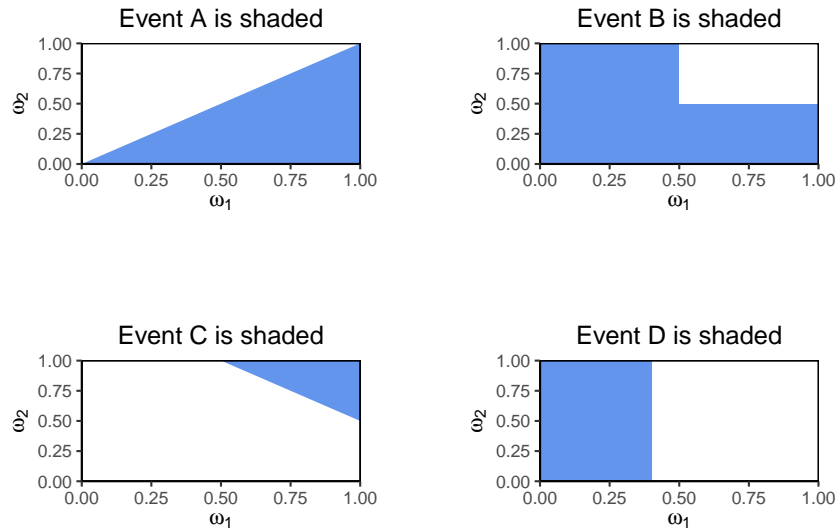


Figure 3.2: Illustration of the events in Exercise 3.13. The square represents the sample space  $\Omega = [0, 1] \times [0, 1]$ .

**Example 3.14** (Don't do what Donny Don't does.). Donny Don't is asked a series of questions involving a pair of rolls of six-sided dice, such as “what is the event that the sum of the dice is at least 10”. Donny's responses are below; explain to him what is wrong with his responses and help him see the correct answers.

1. The possible rolls are 1 through 6, so the sample space is  $\{1, 2, 3, 4, 5, 6\}$ .
2. The sum of the two dice can be 2 through 12, so the event that the sum of the two dice is at least 10 is  $\{10, 11, 12\}$ .
3. The event that the first roll is a 3 is  $\{3\}$ .
4. The event that the first roll is a 3 and the second roll is a 1 is  $\{3, 1\}$

5. Donny's sample space from the first question might correspond to what dice rolling scenario? What does  $\{3, 1\}$  represent in this scenario?

*Solution.* to Example 3.14

1. The questions involve a pair of rolls, so best to record an outcome as an ordered pair, e.g.,  $(5, 2)$  for 5 on the first roll and 2 on the second. Therefore, the sample space would be the following set of 36 possible outcomes.

$$\begin{aligned}\Omega = \{ & (1, 1), (1, 2), \dots, (1, 6), \\ & (2, 1), (2, 2), \dots, (2, 6), \\ & \vdots \quad \quad \quad \vdots \\ & (6, 1), (6, 2), \dots, (6, 6) \}\end{aligned}$$

2. Donny's answers to the first two parts are inconsistent, since there is always a single sample space. So if he says the answer to the first part is  $\{1, \dots, 6\}$ , then any event must be a subset of that sample space and his answer to the second part must be wrong. Using the sample space of 36 ordered pairs from the previous answer, the correct event that the sum of the two dice is at least 10 is

$$\{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}$$

If Donny's sample space in the first part had been  $\{2, \dots, 12\}$ , corresponding to the sum of the two dice, then his answer of  $\{10, 11, 12\}$  would have been correct. However, using such a sample space, he would not have been able to answer the remaining questions (which don't involve the sum of the rolls). There is always one sample space on which all events are defined.

3. Donny didn't take into account that an outcome is a pair of rolls. The correct event is  $\{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$ , the set of all pairs of rolls for which the first roll is 3.
4. Maybe Donny is just using bad notation here, but it sure looks like he is confusing an outcome with an event. The answer should be  $\{(3, 1)\}$ , the set containing the single outcome  $(3, 1)$ . Notice that this is not the same set as  $\{(1, 3)\}$ . (But the set  $\{3, 1\}$  is the same as the set  $\{1, 3\}$ .)
5. The sample space of  $\{1, 2, 3, 4, 5, 6\}$  could correspond to a *single* roll of a fair six-sided die. In this case, the event  $\{3, 1\}$  would be the event that the roll is either a 3 or a 1. (The set  $\{3, 1\}$  is the same as the set  $\{1, 3\}$ .)

**Example 3.15** (Don't do what Donny Don't does.). At various points in his homework, Donny Don't writes the following. Explain to Donny, both mathematically and intuitively, why each of the following symbols is nonsense. Below,  $A$  and  $B$  represent events,  $X$  and  $Y$  represent random variables.

1.  $P(A = 0.5)$

2.  $P(A + B)$
3.  $P(A) \cup P(B)$
4.  $P(X)$
5.  $P(X = A)$
6.  $P(X \cap Y)$

*Solution.* to Example 3.15

1.  $A$  is a set and 0.5 is a number; it doesn't make mathematical sense to equate them. Suppose that  $A$  is the event that it rains tomorrow. It doesn't make sense to say (as  $A = 1$  does) "it rains tomorrow equals 0.5". If we want to say "the probability that it rains tomorrow equals 0.5" we should write  $P(A) = 0.5$ .
2.  $A$  and  $B$  are sets; it doesn't make mathematical sense to add them. Suppose that  $B$  is the event that tomorrow's high temperature is above 80 degrees F. It doesn't make sense to say (as  $A + B$  does) "the sum of (it rains tomorrow) and (tomorrow's high temperature is above 80 degrees F)". If we want "(it rains tomorrow) OR (tomorrow's high temperature is above 80 degrees F)", then we need  $A \cup B$ . Union is an operation on sets; addition is an operation on numbers.
3.  $P(A)$  and  $P(B)$  are numbers; union is an operation on sets, it doesn't make mathematical sense to take a union of numbers. Since  $P(A)$  and  $P(B)$  are numbers then mathematically we can add them. But keep in mind that  $P(A) + P(B)$  is not necessarily a probability, for example if  $P(A) = 0.5$  and  $P(B) = 0.6$ . If we want "the probability that (it rains tomorrow) OR (tomorrow's high temperature is above 80 degrees F)" then the correct symbol is  $P(A \cup B)$ , which would be equal to  $P(A) + P(B)$  only if  $A$  and  $B$  were disjoint (which in the example would mean it's not possible to have a rainy day with a high temperature above 80 degrees F).
4.  $X$  is a random variable, and probabilities are assigned to events. If  $X$  is tomorrow's high temperature in degrees F then  $P(X)$  reads "the probability that tomorrow's high temperature in degrees F", which is missing any qualifying information that could define an event. We could write  $P(X > 80)$  to represent "the probability that (tomorrow's high temperature is greater than 80 degrees F)".
5.  $X$  is a random variable (a function) and  $A$  is an event (a set), and it doesn't make sense to equate these two different mathematical objects. It doesn't make sense to say (as  $X = A$  does) "tomorrow's high temperature in degrees F equals the event that it rains tomorrow".
6.  $X$  and  $Y$  are RVs (functions) and intersection is an operation on sets. If  $Y$  is the amount of rainfall in inches tomorrow then  $X \cap Y$  is attempting to say "tomorrow's high temperature in degrees F and the amount of rainfall in inches tomorrow", but this is still missing qualifying information to define a valid event for which a probability can be assigned. We could say  $P(\{X > 80\} \cap \{Y < 2\})$  to represent "the probability that (tomorrow's high temperature is greater than 80 degrees F) AND (the amount

of rainfall tomorrow is less than 2 inches)”.

### 3.3 Random variables

Statisticians use the terms *observational unit* and *variable*. Observational units are the people, places, things, etc., for which data are observed. Variables are the measurements made on the observational units. For example, the observational units in a study could be college students, while variables could be high school GPA, college GPA, SAT score, years in college, number of Statistics courses taken, etc.

In probability, an *outcome* of a random phenomenon plays a role analogous to an observational unit in statistics. The sample space of outcomes is often only vaguely defined. In many situations we are less interested in detailing the outcomes themselves and more interested in whether or not certain events occur, or with measurements that we can make for the outcomes. For example, if the random phenomenon corresponds to randomly selecting a random sample of students at a college, an outcome could be the list of students selected for the sample. But we are less interested in who the students are, and more interested in questions which involve variables, such as: what is the distribution of SAT scores? What is the relationship between high school GPA and college GPA? What is the average number of years before graduation?

Roughly, a *random variable* assigns a number to each outcome of a random phenomenon. More precisely

**Definition 3.3.** A **random variable (RV)**  $X$  is a *function* that takes an outcome in the sample space as input and returns a real number as output; that is,  $X : \Omega \mapsto \mathbb{R}$ . Random variables are typically denoted by capital letters near the end of the alphabet, with or without subscripts: e.g.  $X$ ,  $Y$ ,  $Z$ , or  $X_1$ ,  $X_2$ ,  $X_3$ , etc. The value that the random variable  $X$  assigns to the outcome  $\omega$  is denoted  $X(\omega)$ .

**Example 3.16.** Flip a coin 4 times, and record the result of each trial in sequence. For example, HTTH means heads on the first on last trial and tails on the second and third. One random variable is  $X$ , the number of heads flipped.

1. Explain why  $X$  is a random variable.
2. Evaluate each of the following:  $X(HHHH)$ ,  $X(HTHT)$ ,  $X(TTHH)$ .
3. Identify the possible values of  $X$ . Why not let the sample space just consist of this set of possible values?

*Solution.* to Example 3.16

1.  $X$  maps each outcome to a number via the function “count the number of heads”.

2.  $X(HHHH) = 4, X(HTHT) = 2, X(TTHH) = 2$ .
3. The possible values of  $X$  are 0, 1, 2, 3, 4. You might ask: if we only care about the number of heads, why bother with the coin flip sequence at all? That is, why not define the sample space as  $\{0, 1, 2, 3, 4\}$  rather than  $\{HHHH, HHHT, HHTH, \dots\}$ . The main reason why is that we often want to define many random variables on the same sample space, and study the relationships between random variables. For example, if a sample space outcome were just the number of heads, we would not be able to obtain information about the longest number of heads in a row, or the proportion of heads on trials that follow heads. Moreover we would not be able to study the relationship between random variables unless they were defined on a common sample space. As a statistics analogy, you would not be able to study the relationship between SAT scores and college GPA unless you measured both variables for the same set of observational units. (Another but less important reason is that is often convenient to work with probability spaces in which the outcomes are equally likely, as is the case for  $\{HHHH, HHHT, HHTH, \dots\}$  but not for  $\{0, 1, 2, 3, 4\}$ , for four flips of a fair coin.)

The RV itself is typically denoted with a capital letter ( $X$ ); possible values of that random variable are denoted with lower case letters ( $x$ ). Think of the capital letter  $X$  as a label standing in for a formula like “the number of heads in 4 flips of a coin” and  $x$  as a dummy variable standing in for a particular value like 3.

We are often interested in events which involve random variables. For example, the expressions  $X = x$  or  $\{X = x\}$  are shorthand for the *event*  $\{\omega \in \Omega : X(\omega) = x\}$ , the set of outcomes  $\omega$  for which  $X(\omega) = x$ . Remember that any event is a subset of the *sample space*. So objects like  $\{X = x\}$  are subsets<sup>5</sup> of  $\Omega$ .

**Example 3.17.** Flip a coin 4 times, and record the result of each trial in sequence. For example, HTTH means heads on the first on last trial and tails on the second and third. One random variable is  $X$ , the number of heads flipped.

1. Identify and interpret the event  $\{X = 3\}$ . That is, identify the outcomes  $\omega$  for which  $X(\omega) = 3$ .
2. Identify and interpret the event  $\{X = 4\}$ .
3. Identify and interpret the event  $\{X \geq 3\}$ . How is this event related to the previous two?

*Solution.* to Example 3.17

1.  $\{X = 3\} = \{HHHT, HHTH, HTHH, THHH\}$  is the event that exactly 3 of the flips land on heads. This is an event because it is a subset of the sample space.

---

<sup>5</sup>Technically, we have some collection  $\mathcal{F}$  of events of interest, and so we require sets like  $\{X \leq x\}$  to be in  $\mathcal{F}$ . This requirement is satisfied by requiring  $X$  to be an  $\mathcal{F}$ -measurable function.



2.  $\{X = 4\} = \{HHHH\}$ , the event that exactly 4 of the flips land on heads. Notice that the event is the set  $\{HHHH\}$ , which consists of the single outcome  $HHHH$ .
3.  $\{X \geq 3\} = \{HHHT, HHTH, HTHH, THHH, HHHH\}$  is the event that at least 3 of the flips land on heads. Also  $\{X \geq 3\} = \{X = 3\} \cup \{X = 4\}$ .

Recall that for a mathematical function<sup>6</sup>  $g$ , given an input  $u$ , the function returns a real number  $g(u)$ . For example, if  $g(u) = u^2$  then  $g(3) = 9$ . If the input comes from some set  $S$  (i.e.  $u \in S$ ), we often write  $g : S \mapsto \mathbb{R}$ .

Likewise, a random variable  $X$  is a function which maps each outcome  $\omega$  in the sample space  $\Omega$  to a real number  $X(\omega)$ ;  $X : \Omega \mapsto \mathbb{R}$ . For a single outcome  $\omega$ , the value  $x = X(\omega)$  is a single number; notice that  $x$  represents the *output* of the function  $X$  rather than the input. However, it is important to remember that the RV  $X$  itself is a *function*, and *not* a single number.

You are probably familiar with functions which have simple closed form formulas of their arguments:  $g(u) = 5u$ ,  $g(u) = u^2$ ,  $g(u) = e^u$ , etc. While any random variable is some function, the function is rarely specified as an explicit mathematical formula of its input  $\omega$ . Often, outcomes are not even numbers (e.g., sequences of coin flips), or only vaguely specified if at all (e.g., possible paths of a hurricane). In the coin flip example, we defined  $X$  only through the words “number of flips that land on heads”; translating even this simple situation into a formula of  $\omega$  requires some notation<sup>7</sup>.

It is more appropriate to think of a RV as a function in the sense of a scale at a grocery store which maps a fruit to its weight,  $X : \text{fruit} \mapsto \text{weight}$ . Put an apple on the scale and the scale returns a number,  $X(\text{apple})$ , the weight of the apple. Likewise,  $X(\text{orange})$ ,  $X(\text{banana})$ . The RV  $X$  is the scale itself. (This simplistic analogy assumes a sample space outcome is a single fruit. Of course, it’s even more complicated in reality since an outcome can be considered a set of fruits, so that we have for example  $X(\{2 \text{ apples}, 3 \text{ oranges}\})$ , and all fruits do not weigh the same, so that  $X(\text{this apple})$  is not the same as  $X(\text{that apple})$ .)

The RV itself is denoted with a capital letter; possible values of that random variable are denoted with lower case letters. For example  $\{X = x\}$  is shorthand for the event  $\{\omega \in \Omega : X(\omega) = x\}$ , the set of outcomes  $\omega$  for which  $X(\omega) = x$ . Think of the capital letter  $X$  as a label standing in for a formula like “the number of heads in 4 flips of a coin” and  $x$  as a dummy variable standing in for

<sup>6</sup>Throughout, we use  $g$  to denote a generic function, and reserve  $f$  to represent a probability density function. Likewise, we represent a generic function argument (or “dummy variable”) with  $u$ , since  $x$  is often used to represent possible values of a random variable  $X$ ; in the RV scenario  $x$  typically represents the *output* of the function  $X$  rather than the input (which is a sample space outcome  $\omega$ ).

<sup>7</sup>It’s easiest if we label a flip of heads as 1 and tails as 0. Represent an outcome  $\omega$  as  $(\omega_1, \omega_2, \omega_3, \omega_4)$ , where  $\omega_i \in \{0, 1\}$  is the result of the  $i$ th flip. Then  $X(\omega) = \sum_{i=1}^4 \omega_i$  represents the number of heads. For example, outcome HHTH would be represented as  $(1, 1, 0, 1)$  and  $X((1, 1, 0, 1)) = 1 + 1 + 0 + 1 = 3$ . This could be coded as `sum(omega)`.

a particular value like 3. Remember that any event is a subset of the *sample space*. So objects like  $\{X = x\}$  are subsets<sup>8</sup> of  $\Omega$ . In the scale analogy if  $X$  is weight measured in pounds, we might have  $\{X > 5\} = \{\text{watermelon, pineapple}\}$  is the set of fruits that weigh more than 5 pounds.

**Example 3.18.** Roll a four-sided die twice, and record the result of each roll in sequence. For example, the outcome  $(3, 1)$  represents a 3 on the first roll and a 1 on the second; this is not the same outcome as  $(1, 3)$ . Let  $X$  be the sum of the two dice, and let  $Y$  be the larger of the two rolls (or the common value if both rolls are the same).

1. Evaluate  $X((1, 3))$ ,  $X((4, 3))$ , and  $X((2, 2))$ .
2. Evaluate  $Y((1, 3))$ ,  $Y((4, 3))$ , and  $Y((2, 2))$ .
3. Identify and interpret  $\{X = 4\}$ .
4. Identify and interpret  $\{X = 3\}$ .
5. Identify and interpret  $\{X \leq 3\}$ .
6. Identify and interpret  $\{Y = 4\}$ .
7. Identify and interpret  $\{Y = 3\}$ .
8. Identify and interpret  $\{Y \leq 3\}$ .
9. Identify and interpret  $\{X = 4, Y = 3\}$  (that is,  $\{X = 4\} \cap \{Y = 3\}$ ).
10. Identify and interpret  $\{X = 4, Y \leq 3\}$ .
11. Identify and interpret  $\{X = 3, Y = 3\}$ .
12. Identify and interpret  $\{X \geq Y\}$ .
13. Identify the possible values of  $X$ .
14. Identify the possible values of  $Y$ .
15. Identify the possible values of the pair  $(X, Y)$ .

*Solution.* to Example 3.18

1.  $X$  is the sum of the two rolls, so  $X((1, 3)) = 4$ ,  $X((4, 3)) = 7$ , and  $X((2, 2)) = 4$ .
2.  $Y$  is the larger of the two rolls (or the common value if a tie) so  $Y((1, 3)) = 3$ ,  $Y((4, 3)) = 4$ , and  $Y((2, 2)) = 2$ .
3.  $\{X = 4\} = \{(1, 3), (2, 2), (3, 1)\}$  is the event that the sum of the two dice is 4.
4.  $\{X = 3\} = \{(1, 2), (2, 1)\}$  is the event that the sum of the two dice is 3.
5.  $\{X \leq 3\} = \{(1, 1), (1, 2), (2, 1)\}$  is the event that the sum of the two dice is at most 3.
6.  $\{Y = 4\} = \{(1, 4), (2, 4), (3, 4), (4, 4), (4, 1), (4, 2), (4, 3)\}$  is the event that the larger of the two rolls is 4.
7.  $\{Y = 3\} = \{(1, 3), (2, 3), (3, 3), (3, 1), (3, 2)\}$  is the event that the larger of the two rolls is 3.

---

<sup>8</sup>Technically, we have some collection  $\mathcal{F}$  of events of interest, and so we require sets like  $\{X \leq x\}$  to be in  $\mathcal{F}$ . This requirement is satisfied by requiring  $X$  to be an  $\mathcal{F}$ -measurable function.

8.  $\{Y \leq 3\} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$  is the event that the larger of the two rolls is at most 3. Notice that since in this example  $Y$  can only take values 1, 2, 3, 4, we have  $\{Y \leq 3\} = \{Y = 4\}^c$ .
9.  $\{X = 4, Y = 3\} \equiv \{X = 4\} \cap \{Y = 3\} = \{(1, 3), (3, 1)\}$  is the event that both the sum of the two dice is 4 and the larger of the two rolls is 3. Even though this involves two random variables, it is a single event (that is, a single subset of the sample space). There are only two outcomes for which both the sum of the two dice is 4 and the larger of the two dice is 3.
10.  $\{X = 4, Y \leq 3\} \equiv \{X = 4\} \cap \{Y \leq 3\} = \{(1, 3), (2, 2), (3, 1)\}$  is the event that both the sum of the two dice is 4 and the larger of the two rolls is at most 3. Notice that since in this example  $\{X = 4\} \subset \{Y \leq 3\}$ , we have  $\{X = 4, Y \leq 3\} = \{X = 4\}$ .
11.  $\{X = 3, Y = 3\} \equiv \{X = 3\} \cap \{Y = 3\} = \emptyset$ , since there are no outcomes for which both the sum is 3 and the larger of the two dice is 3. (If the larger of the two dice is 3, then the sum must be at least 4.)
12. The event  $\{X \geq Y\}$  represents the set of outcomes  $\{\omega : X(\omega) \geq Y(\omega)\}$ . In this example, for every possible outcome the sum of the two dice is at least as large as the larger of the two die, so  $\{X \geq Y\} = \Omega$ .
13. The possible values of  $X$  are  $\{2, 3, \dots, 8\}$
14. The possible values of  $Y$  are  $\{1, 2, 3, 4\}$
15. The possible values of the pair  $(X, Y)$  are  $\{(2, 1), (3, 2), (4, 2), (4, 3), (5, 3), (5, 4), (6, 3), (6, 4), (7, 4), (8, 4)\}$ . Notice that while, for example, 8 is a possible value of  $X$  and 1 is a possible value of  $Y$ ,  $(8, 1)$  is not a possible value of the pair  $(X, Y)$ ; it's not possible for the larger of the two dice to be 1 but their sum to be 8.

When dealing with probabilities, it is common to write  $P(X = 3)$  instead of  $P(\{X = 3\})$ , and  $P(X = 4, Y = 3)$  instead of  $P(\{X = 4\} \cap \{Y = 3\})$ ; read the comma in  $P(X = 4, Y = 3)$  as “and”. But keep in mind that an expression like “ $X = 3$ ” really represents an event  $\{X = 3\}$ , an expression which itself represents  $\{\omega \in \Omega : X(\omega) = 3\}$ , a subset of  $\Omega$ .

**Example 3.19.** Customers enter a deli and take a number to mark their place in line. When the deli opens the counter starts 0; the first customer to arrive takes number 1, the second 2, etc. We record the counter over time, continuously, as it changes as customers arrive. Time is measured in minutes after the deli opens (time 0). A sample space outcome could be represented as a path of the number of customers over time; a few such paths are illustrated in Figure 2.2. Notice the stairstep feature: a customer arrives and takes a number then the counter stays on that number for some time (the flat spots) until another customer arrives and the counter increases by one (the jumps).

There are many random variables that could be interest, including

- $N_t$ , the number of customers that have arrived by time  $t$ , where  $t \geq 0$  is minutes after time 0

- $T_j$ , the time (in minutes after time 0) at which the  $j$ th customer arrives, for  $j = 1, 2, \dots$
- $W_j$ , the “waiting” time (in minutes) between the arrival of the  $j$ th and the  $(j - 1)$ th customer.

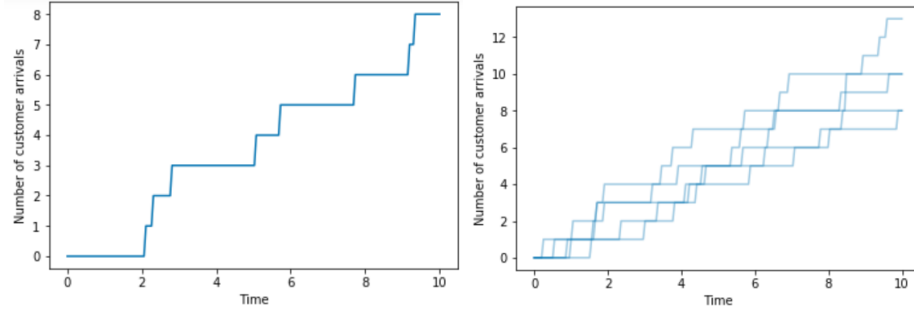


Figure 3.3: Sample space outcomes for Example 3.19. Left: a single sample path of the number of customer arrivals over time. Right: several possible paths.

For the outcome represented by the path in the plot on the left in Figure 3.3, identify (as best as you can from the plot) the value of the following random variables.

1.  $N_4$
2.  $N_{6.5}$
3.  $T_4$
4.  $T_5$
5.  $W_1$
6.  $W_5$

*Solution.* to Example 3.19

1. For this outcome  $N_4 = 3$ ; the number of customers that have arrived by time 4 is 3.
2. For this outcome  $N_{6.5} = 5$ ; the number of customers that have arrived by time 6.5 is 5. The number of customers is a whole number, but time is measured continuously (e.g., 6.5 minutes after opening).
3. For this outcome  $T_4 \approx 5.1$ . The path jumps to 4 a little after time 5, so the time at which the fourth customer arrives (when the counter jumps to 4) is about 5.1 minutes after open.
4. For this outcome  $T_5 \approx 5.9$ . The path jumps to 5 a little before time 6, so the time at which the fifth customer arrives (when the counter jumps to 5) is about 5.9 minutes after open.
5. For this outcome  $W_1 \approx 2$ .  $W_1$  is the waiting time from open until the first customer arrives, which seems to happen at about time 2 (when the counter jumps to 1).

6. For this outcome  $W_5 \approx 0.8$ .  $W_5$  is the time elapsed between the arrival of the fourth (at time 5.1) and fifth (at time 5.9) customers, which is about 0.8 minutes.

**Example 3.20** (Matching problem). So-called “matching problems” concern the following generic scenario. A set of  $n$  cards labeled  $1, 2, \dots, n$  are placed in  $n$  boxes labeled  $1, 2, \dots, n$ , with exactly one card in each box. Typical questions of interest involve whether the number of a card matches the number of the box in which it is placed. (More colorful descriptions include returning babies at random to mothers or placing rocks at random back on a museum shelf.) Consider the case  $n = 4$ . Let  $Y$  be the number of cards (out of 4) which match the number of the box in which they are placed. For  $j = 1, 2, 3, 4$ , let  $I_j = 1$  if card  $j$  is placed in box  $j$ , and let  $I_j = 0$  otherwise.

1. Evaluate  $Y(1234)$ ,  $Y(1243)$ , and  $Y(2143)$ .
2. Evaluate  $I_1(1234)$ ,  $I_1(1243)$ , and  $I_1(2143)$ .
3. Evaluate  $I_2(1234)$ ,  $I_2(1243)$ , and  $I_2(2143)$ .
4. Evaluate  $I_3(1234)$ ,  $I_3(1243)$ , and  $I_3(2143)$ .
5. Evaluate  $I_4(1234)$ ,  $I_4(1243)$ , and  $I_4(2143)$ .
6. Identify and interpret  $\{Y = 4\}$ .
7. Identify and interpret  $\{Y = 0\}$ .
8. Identify and interpret  $\{Y = 3\}$ .
9. Identify and interpret  $\{I_1 = 1\}$ .
10. Identify and interpret  $\{I_3 = 1\}$ .
11. What is the relationship between  $Y$  and the  $I_j$ 's?

*Solution.* to Example 3.20

1.  $Y$  counts the number of matches so  $Y(1234) = 4$  (all numbers match),  $Y(1243) = 2$  (1 and 2 match, 3 and 4 do not), and  $Y(2143) = 0$  (no numbers match).
2.  $I_1 = 1$  only if card 1 is in the first position, so  $I_1(1234) = 1$ ,  $I_1(1243) = 1$ , and  $I_1(2143) = 0$ .
3.  $I_2 = 1$  only if card 2 is in the second position, so  $I_2(1234) = 1$ ,  $I_2(1243) = 1$ , and  $I_2(2143) = 0$ .
4.  $I_3 = 1$  only if card 3 is in the third position, so  $I_3(1234) = 1$ ,  $I_3(1243) = 0$ , and  $I_3(2143) = 0$ .
5.  $I_4 = 1$  only if card 4 is in the fourth position, so  $I_4(1234) = 1$ ,  $I_4(1243) = 0$ , and  $I_4(2143) = 0$ .
6.  $\{Y = 4\} = \{1234\}$  is the event that all 4 cards match. Notice that the event is the set  $\{1234\}$ , which consists of the single outcome 1234.
7.  $\{Y = 0\} = \{2143, 2341, 2413, 3142, 3412, 3421, 4123, 4312, 4321\}$  is the event that none of the cards match
8.  $\{Y = 3\} = \emptyset$  is the event in which exactly 3 cards match their box. There are no outcomes in which exactly 3 cards match; if three cards match, then the fourth card must necessarily match too.

9.  $\{I_1 = 1\} = \{1234, 1243, 1324, 1342, 1423, 1432\}$  is the event that card 1 is placed in box 1. Since our sample space consists of the placements of each of the cards, each event must be expressed in terms of these outcomes.
10.  $\{I_3 = 1\} = \{1234, 1432, 2134, 2431, 4132, 4231\}$  is the event that card 3 is placed in box 3. Since our sample space consists of the placements of each of the cards, each event must be expressed in terms of these outcomes.
11.  $Y = I_1 + I_2 + I_3 + I_4$ .  $Y$  represents the total count of matches. The sum  $I_1 + I_2 + I_3 + I_4$  starts with the first box and adds 1 to the count if there is a match (and 0 otherwise), and continues for each of the boxes, resulting in the total count of matches. For example,  $Y(1243) = 2 = 1 + 1 + 0 + 0 = I_1(1243) + I_2(1243) + I_3(1243) + I_4(1243)$ . Notice that all the random variables are defined on the same probability space; that is, they have the same inputs. We expand on this idea further in the next subsection.

Random variables that only take two possible values, 0 and 1, (like  $I_1$  in the previous example) have a special name.

**Definition 3.4.** An **indicator (a.k.a. Bernoulli)** RV can take only the values 0 or 1. If  $A$  is an event then the indicator RV  $I_A$  is defined as

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A \end{cases}$$

That is,  $I_A$  equals 1 if event  $A$  occurs, and  $I_A$  equals 0 if event  $A$  does not occur. Indicators provide the bridge between events and random variables. While simple, they can be very useful. For example, representing a count as a sum of indicator RVs as in Exercise 3.20 is a very common and useful strategy.

### 3.3.1 Transformations of random variables

**A function of a random variable is also a random variable.** That is, if  $X$  is a random variable and  $g : \mathbb{R} \mapsto \mathbb{R}$  is a function, then  $Y = g(X)$  is a random variable<sup>9</sup>.

For example, suppose a sample space consists of a set of circles of various sizes. If  $X$  is a random variable representing the radius of a circle selected from this sample space, then  $Y = \pi X^2$  is a random variable representing the area of a circle selected from this sample space. Here we can write  $Y = g(X)$  with  $g(u) = \pi u^2$ .

**Sums and products, etc., of random variables defined on the same probability space are random variables.** That is, if random variables  $X$  and  $Y$  are defined on the same probability space then  $X + Y$ ,  $X - Y$ ,  $XY$ , and

---

<sup>9</sup> $Y(\omega) = g(X(\omega))$  so  $Y$  maps  $\Omega$  to  $\mathbb{R}$  via the composition of the functions  $g$  and  $X$ ; that is,  $Y = g \circ X$

$X/Y$  are also random variables. Similarly, it is possible to make comparisons such as  $X \geq Y$  for random variables defined on the same probability space. The following example emphasizes what we mean by “defined on the same probability space”.

**Example 3.21.** Roll a four-sided die twice, and record the result of each roll in sequence. For example, the outcome  $(3, 1)$  represents a 3 on the first roll and a 1 on the second; this is not the same outcome as  $(1, 3)$ . Let  $X_1$  be the result of the first roll, and  $X_2$  the result of the second.

1. Evaluate  $X_1((3, 1))$  and  $X_2((3, 1))$ .
2. If an outcome is represented by  $\omega = (\omega_1, \omega_2)$  (e.g.  $(3, 1)$ ), specify the functions that define the random variables  $X_1$  and  $X_2$ .
3. Is  $X = X_1 + X_2$  a valid random variable? If so, identify the function that defines it.
4. Evaluate  $X((3, 1))$ .

*Solution.* to Example 3.21

1.  $X_1((3, 1)) = 3$  and  $X_2((3, 1)) = 1$ .
2. Recall that there is a single sample space corresponding to the pairs of rolls, rather than a separate sample space for each of the individual rolls. Therefore, random variables need to be defined on the sample space corresponding to pairs of rolls.  $X_1((\omega_1, \omega_2)) = \omega_1$  and  $X_2((\omega_1, \omega_2)) = \omega_2$ . That is,  $X_1$  maps an ordered pair to its first coordinate, and  $X_2$  to its second.
3. Yes,  $X = X_1 + X_2$  is a valid random variable. For each outcome  $(\omega_1, \omega_2)$ ,  $X$  returns a number:  $X((\omega_1, \omega_2)) = X_1((\omega_1, \omega_2)) + X_2((\omega_1, \omega_2)) = \omega_1 + \omega_2$
4.  $X((3, 1)) = X_1((3, 1)) + X_2((3, 1)) = 3 + 1 = 4$ .

The above example probably seems like notational overkill. And of course we could have defined  $X$  directly as  $X((\omega_1, \omega_2)) = \omega_1 + \omega_2$  without the need for  $X_1, X_2$ . But we introduced the example to emphasize that it only makes sense to add random variables if they are defined on the same sample space. Remember that adding two random variables involves adding two *functions*, in the way that the function  $g = g_1 + g_2$  is defined by  $g(u) = g_1(u) + g_2(u)$ . It only makes sense to add two functions together if they have the same inputs.

For example, consider the random variable  $X$  from Example 3.16 and the random variable  $Y$  from Example 3.20. It wouldn't make much practical sense to consider  $X + Y$ , but it would make no mathematical sense. How would  $X + Y$  even be defined?  $X$  is defined for sequences of coin flips like HHTH, while  $Y$  is defined for the shuffles in the matching problem like 2143. If you attempted to add  $X$  and  $Y$ , which outcomes would go together? Would you add  $X(HHTH)$  to  $Y(2143)$ ? Why not add  $X(HHTH)$  to  $Y(1234)$ ? Again, adding two random variables involves adding two functions, and it doesn't make sense to add those functions if they have different inputs.

As a more practical example, if  $X$  represents SAT math score and  $Y$  represents SAT verbal score then we might be interested in the total score  $X+Y$ . Requiring  $X$  and  $Y$  to be defined on the same probability space is like requiring the scores to be measured for the same students. For example, Antwan has both a Math score  $X(\text{Antwan})$  and a Verbal score  $Y(\text{Antwan})$ , so we can consider the total score  $(X+Y)(\text{Antwan}) = X(\text{Antwan}) + Y(\text{Antwan})$ . (In statistical terms, the variables are measured for the same observational units.) It wouldn't make sense to add SAT Math scores from one set of students to SAT verbal scores for a different set of students; for example  $X(\text{Antwan}) + Y(\text{Maria})$  makes no sense.

**Example 3.22.** Flip a fair coin four times and record the results in order, e.g. HHTT means two heads followed by two tails. Recall that in Section 2.5 we considered the *proportion of the flips which immediately follow a H that result in H*. Remember that we do not consider this proportion if no flips follow a H, i.e. the outcome is either TTTT or TTTH.

Let:

- $Z$  be the number of flips immediately following H.
  - $Y$  be the number of flips immediately following H that result in H.
  - $X$  be the proportion of flips immediately following H that result in H.
1. Is  $X$  a random variable? How does it relate to  $Y$  and  $Z$ ?
  2. For each of the possible outcomes in the sample space, find the value of  $(Z, Y, X)$ .

*Solution.* to Example 3.22

1. Yes,  $X$  is a random variable because it maps each coin flip sequence to the value of proportion of the flips which immediately follow a H that result in H for that sequence; see the table below. Also,  $X = Y/Z$ . Technically  $Y$  and  $Z$  are not defined for outcomes TTTT and TTTH, but we're ignoring those sequences for the purposes of investigating the proportion of the flips which immediately follow a H that result in H.
2. In the outcomes below, the **flips which follow head are in bold**.

Table 3.2: Possible values of (1)  $Z$ , the number of flips immediately following H, (2)  $Y$ , the number of flips immediately following H that result in H, and (3)  $X$ , the proportion of flips immediately following H that result in H, for four flips of a fair coin.

Outcome ( $\omega$ )	$Z$	$Y$	$X = Y/Z$
<b>HHHH</b>	3	3	1
<b>HHHT</b>	3	2	2/3
<b>HHTH</b>	2	1	1/2



Outcome ( $\omega$ )	$Z$	$Y$	$X = Y/Z$
H $\overline{T}$ H $\overline{H}$	2	1	1/2
T $\overline{H}$ H $\overline{H}$	2	2	1
H $\overline{H}$ T $\overline{T}$	2	1	1/2
H $\overline{T}$ H $\overline{T}$	2	0	0
H $\overline{T}$ T $\overline{H}$	1	0	0
T $\overline{H}$ H $\overline{T}$	2	1	1/2
T $\overline{H}$ T $\overline{H}$	1	0	0
T $\overline{T}$ H $\overline{H}$	1	1	1
H $\overline{T}$ T $\overline{T}$	1	0	0
T $\overline{H}$ T $\overline{T}$	1	0	0
T $\overline{T}$ H $\overline{T}$	1	0	0
T $\overline{T}$ T $\overline{H}$	0	not defined	not defined
T $\overline{T}$ T $\overline{T}$	0	not defined	not defined

### 3.4 Probability spaces

In the previous sections we defined outcomes, events, and random variables, the main mathematical objects associated with a random phenomenon. But we haven't actually computed any probabilities yet. As we saw in Section 2.4, there are some basic logical consistency requirements that probabilities must satisfy. These requirements are formalized in the following.

**Definition 3.5.** A **probability space** is a triple  $(\Omega, \mathcal{F}, P)$  where

- $\Omega$  is a sample space of outcomes
- $\mathcal{F}$  is a collection of events of interest<sup>10</sup>  $A \subseteq \Omega$
- $P$  is a **probability measure** which assigns a probability  $P(A)$  to events  $A \in \mathcal{F}$ . A probability measure satisfies the following three axioms
  - $P(\Omega) = 1$
  - For all events  $A \in \mathcal{F}$ ,  $0 \leq P(A) \leq 1$
  - (*Countable additivity.*) If events  $A_1, A_2, \dots \in \mathcal{F}$  are *disjoint* (a.k.a. *mutually exclusive*) — that is  $A_i \cap A_j = \emptyset$  for all  $i \neq j$  — then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

The requirement  $0 \leq P(A) \leq 1$  makes sense in light of the relative frequency interpretation: an event  $A$  can not occur on more than 100% of repetitions or less than 0% of repetitions of the random phenomenon.

<sup>10</sup>Technically,  $\mathcal{F}$  is a  $\sigma$ -field of subsets of  $\Omega$ :  $\mathcal{F}$  contains  $\Omega$  and is closed under countably many elementary set operations (complements, unions, intersections). While this level of technical detail is not needed, we prefer to refer to a probability space as a triple to emphasize that probabilities are assigned directly to *events* rather than just outcomes.

The requirement that  $P(\Omega) = 1$  just ensures that the sample space accounts for all of the possible outcomes. If outcome  $\omega$  is observed, then event  $A$  occurs if  $\omega \in A$ . If  $P(\Omega) < 1$  then it would be possible to observe outcomes  $\omega \notin \Omega$ ; but this violates the requirement that  $\Omega$  is the set of all possible outcomes. Basically,  $P(\Omega) = 1$  says that on any repetition of the random phenomenon, “something has to happen”. If  $\Omega$  is a countable set, countable additivity and  $P(\Omega) = 1$  imply that probability of all the outcomes must add up to 1. For example, in Example 2.1  $P(\Omega) = 1$ , together with countable additivity, is what requires that the probability that a team other than those four teams win to be 26%.

Countable additivity is best understood through a diagram with areas representing probabilities, as in the figure below which represents two events (yellow / and blue \). On the left, there is no “overlap” between areas so the total area is the sum of the two pieces; this depicts countable additivity for two disjoint events. On the right, there is overlap between the two areas, so simply adding the two areas “double counts” the intersection (green  $\times$ ) and does not result in the correct total area. Countable additivity applies to any number of events, as long as there is no “overlap”.

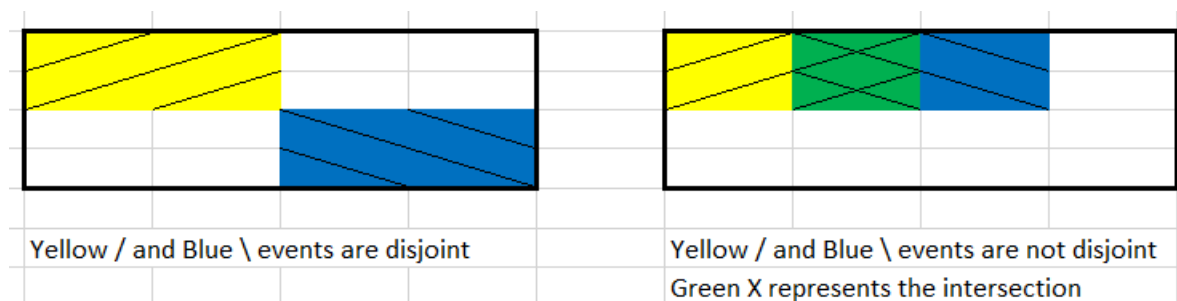


Figure 3.4: Illustration of countable additivity for two events. The events in the picture on the left are disjoint, but not on the right.

In Example 2.1, the events  $A$  = “the Astros win the 2019 World Series” and  $D$  = “the Dodgers win the 2019 World Series” are disjoint  $A \cap D = \emptyset$ ; in a single World Series, both teams cannot win. Therefore, the probability of  $A \cup D$ , the event that either the Astros or the Dodgers win, must be 46%.

The three axioms of a probability measure are minimal logical consistency requirements that must be satisfied by any probability model. There are also many physical aspects of the random phenomenon or assumptions (e.g. “fairness”, independence, conditional relationships) that must be considered when determining a reasonable probability measure for a particular situation. Often,  $P$  is defined implicitly through modeling assumptions, and probabilities of events follow from the axioms and related properties.

Many other properties follow from the axioms. The main “meat” of the axioms

is countable additivity. Thus, the key to many proofs of probability properties is to write relevant events in terms of disjoint events.

**Theorem 3.1** (Properties of a probability measure.). *Complement rule*<sup>11</sup>. For any event  $A$ ,  $P(A^c) = 1 - P(A)$ . In particular, since  $\Omega^c = \emptyset$ ,  $P(\emptyset) = 0$ .

*Subset rule*<sup>12</sup>. If  $A \subseteq B$  then  $P(A) \leq P(B)$ .

*General addition rule for two events*<sup>13</sup>. If  $A$  and  $B$  are any two events

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

*Law of total probability*. If  $B_1, \dots, B_k$  are disjoint with  $B_1 \cup \dots \cup B_k = \Omega$ , then

$$P(A) = \sum_{i=1}^k P(A \cap B_i)$$

The key to the proofs is to represent relevant events in terms of disjoint events and use countable additivity (and the other axioms).

Note:  $A \cup B$  is inclusive so we *do* want to count the possibility of both,  $A \cap B$ . The problem with simply adding  $P(A)$  and  $P(B)$  is that their sum *double counts*  $A \cap B$ . We do want to count the outcomes that satisfy both  $A$  and  $B$ , but we only want to count them once. Subtracting  $P(A \cap B)$  in the general addition rule for two events corrects for the double counting.

For example, consider the picture on the right in Figure 3.4. Suppose each rectangular cell represents a distinct outcome; there are 16 outcomes in total. Assume the outcomes are equally likely, each with probability  $1/16$ . Let  $A$  represent the yellow / event which has probability  $4/16$  and let  $B$  represent the blue \ event which has probability  $4/16$ . Then  $P(A \cup B) = 6/16$ , since there are 6 outcomes which satisfy either event  $A$  or  $B$  (or both). However, simply adding  $P(A) + P(B)$  yields  $8/16$  because the two outcomes that satisfy the green event  $A \cap B$  are counted both in  $P(A)$  and  $P(B)$ . So to correct for this double counting, we subtract out  $P(A \cap B)$ :

$$P(A) + P(B) - P(A \cap B) = 4/16 + 4/16 - 2/16 = 6/16 = P(A \cup B)$$

Warning: The general addition rule for more than two events is more complicated<sup>14</sup>; see the inclusion-exclusion principle.

<sup>11</sup>Proof: Since  $\Omega = A \cup A^c$  and  $A$  and  $A^c$  are disjoint the axioms imply that  $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ .

<sup>12</sup>Proof. If  $A \subseteq B$  then  $B = A \cup (B \cap A^c)$ . Since  $A$  and  $(B \cap A^c)$  are disjoint,  $P(B) = P(A) + P(B \cap A^c) \geq P(A)$ .

<sup>13</sup>The proof is easiest to see by considering a picture like the one Figure 3.4 .

<sup>14</sup>For three events,  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$ .

In the law of total probability<sup>15</sup> the events  $B_1, \dots, B_k$ , which represent “cases”, form a *partition* of the sample space; each outcome  $\omega \in \Omega$  lies in exactly one of the  $B_i$ . The law of total probability says that we can interpret the “overall” probability  $P(A)$  by summing the probability of  $A$  in each “case”  $P(A \cap B_i)$ .

**Exercise 3.1.** Consider a Cal Poly student who frequently has blurry, bloodshot eyes, generally exhibits slow reaction time, always seems to have the munchies, and disappears at 4:20 each day. Which of the following events,  $A$  or  $B$ , has a higher probability? (Assume the two probabilities are not equal.)

- $A$ : The student has a GPA above 3.0.
- $B$ : The student has a GPA above 3.0 and smokes marijuana regularly.

**Warning!** Your psychological judgment of probabilities is often inconsistent with the mathematical logic of probabilities.

**Example 3.23** (Don’t do what Donny Don’t does.). At various points in his homework, Donny Don’t writes the following. Explain to Donny, both mathematically and intuitively, why each of the following symbols is nonsense. Below,  $A$  and  $B$  represent events,  $X$  and  $Y$  represent random variables.

1.  $P(A = 0.5)$
2.  $P(A + B)$
3.  $P(A) \cup P(B)$
4.  $P(X)$
5.  $P(X = A)$
6.  $P(X \cap Y)$

*Solution.* to Example 3.23

1.  $A$  is a set and 0.5 is a number; it doesn’t make mathematical sense to equate them. Suppose that  $A$  is the event that it rains tomorrow. It doesn’t make sense to say (as  $A = 1$  does) “it rains tomorrow equals 0.5”. If we want to say “the probability that it rains tomorrow equals 0.5” we should write  $P(A) = 0.5$ .
2.  $A$  and  $B$  are sets; it doesn’t make mathematical sense to add them. Suppose that  $B$  is the event that tomorrow’s high temperature is above 80 degrees F. It doesn’t make sense to say (as  $A + B$  does) “the sum of (it rains tomorrow) and (tomorrow’s high temperature is above 80 degrees F)”. If we want “(it rains tomorrow) OR (tomorrow’s high temperature is above 80 degrees F)”, then we need  $A \cup B$ . Union is an operation on sets; addition is an operation on numbers.

---

<sup>15</sup>We will see a different expression of the law of total probability, involving conditional probabilities, in Section ??.

3.  $P(A)$  and  $P(B)$  are numbers; union is an operation on sets, it doesn't make mathematical sense to take a union of numbers. Since  $P(A)$  and  $P(B)$  are numbers then mathematically we can add them. But keep in mind that  $P(A) + P(B)$  is not necessarily a probability, for example if  $P(A) = 0.5$  and  $P(B) = 0.6$ . If we want "the probability that (it rains tomorrow) OR (tomorrow's high temperature is above 80 degrees F)" then the correct symbol is  $P(A \cup B)$ , which would be equal to  $P(A) + P(B)$  only if  $A$  and  $B$  were disjoint (which in the example would mean it's not possible to have a rainy day with a high temperature above 80 degrees F).
4.  $X$  is a random variable, and probabilities are assigned to events. If  $X$  is tomorrow's high temperature in degrees F then  $P(X)$  reads "the probability that tomorrow's high temperature in degrees F", which is missing any qualifying information that could define an event. We could write  $P(X > 80)$  to represent "the probability that (tomorrow's high temperature is greater than 80 degrees F)".
5.  $X$  is a random variable (a function) and  $A$  is an event (a set), and it doesn't make sense to equate these two different mathematical objects. It doesn't make sense to say (as  $X = A$  does) "tomorrow's high temperature in degrees F equals the event that it rains tomorrow".
6.  $X$  and  $Y$  are RVs (functions) and intersection is an operation on sets. If  $Y$  is the amount of rainfall in inches tomorrow then  $X \cap Y$  is attempting to say "tomorrow's high temperature in degrees F and the amount of rainfall in inches tomorrow", but this is still missing qualifying information to define a valid event for which a probability can be assigned. We could say  $P(\{X > 80\} \cap \{Y < 2\})$  to represent "the probability that (tomorrow's high temperature is greater than 80 degrees F) AND (the amount of rainfall tomorrow is less than 2 inches)".

### 3.4.1 Probability measures in a dice rolling example

A probability measure is a *set function*:  $P : \mathcal{F} \mapsto [0, 1]$  takes as an input an event (set)  $A \in \mathcal{F}$  and returns as an output a number  $P(A) \in [0, 1]$ . Sometimes  $P(A)$  is defined explicitly for an event  $A$  via a formula.

As an example, consider a single roll of a four-sided die. The sample space consists of the four possible outcomes  $\Omega = \{1, 2, 3, 4\}$ . Let's first assume that the die is fair, so all four outcomes are equally likely, each with probability<sup>16</sup>  $1/4$ . Given that the probability of each outcome<sup>17</sup> is  $1/4$ , countable additivity implies

<sup>16</sup>That the probability of each outcome must be  $1/4$  when there are four *equally likely* outcomes follows from the axioms, by writing  $\{1, 2, 3, 4\} = \{1\} \cup \{2\} \cup \{3\} \cup \{4\}$ , a union of disjoint sets, and applying countable additivity and  $P(\Omega) = 1$ .

<sup>17</sup>A probability measure is always defined on sets. When we say loosely "the probability of an outcome  $\omega$ " we really mean the probability of the event consisting of the single outcome  $\{\omega\}$ . In this example  $P(\{\omega\}) = 1/4$  for  $\omega \in \{1, 2, 3, 4\}$ .

$$P(A) = \frac{\text{number of elements in } A}{4}, \quad P \text{ assumes a fair four-sided die}$$

The following table lists all the possible events, and their probability under the assumption of equally likely outcomes.

Table 3.3: All possible events associated with a single roll of a four-sided die, and their probabilities assuming the die is fair.

Event	Description	Probability of event assuming equally likely outcomes
$\emptyset$	Roll nothing (not possible)	0
$\{1\}$	Roll a 1	1/4
$\{2\}$	Roll a 2	1/4
$\{3\}$	Roll a 3	1/4
$\{4\}$	Roll a 4	1/4
$\{1, 2\}$	Roll a 1 or a 2	2/4
$\{1, 3\}$	Roll a 1 or a 3	2/4
$\{1, 4\}$	Roll a 1 or a 4	2/4
$\{2, 3\}$	Roll a 2 or a 3	2/4
$\{2, 4\}$	Roll a 2 or a 4	2/4
$\{3, 4\}$	Roll a 3 or a 4	2/4
$\{1, 2, 3\}$	Roll a 1, 2, or 3 (a.k.a. do not roll a 4)	3/4
$\{1, 2, 4\}$	Roll a 1, 2, or 4 (a.k.a. do not roll a 3)	3/4
$\{1, 3, 4\}$	Roll a 1, 3, or 4 (a.k.a. do not roll a 2)	3/4
$\{2, 3, 4\}$	Roll a 2, 3, or 4 (a.k.a. do not roll a 1)	3/4
$\{1, 2, 3, 4\}$	Roll something	1

The above assignment satisfies all the axioms and so it represents a valid probability measure. But assuming that the outcomes are equally likely requires much more than the basic logical consistency requirements of the axioms. There are many other possible probability measures, like in the following.

**Example 3.24.** Now consider a single roll of a four-sided die, but suppose the die is weighted so that the outcomes are no longer equally likely (but each outcome is still possible). Suppose that the probability of event  $\{2, 3\}$  is 0.5, of event  $\{3, 4\}$  is 0.7, and of event  $\{1, 2, 3\}$  is 0.6. Complete a table, like the one for the equally likely outcome scenario above, listing the probability of each event for this particular weighted die. In what particular way is the die weighted? That is, what is the probability of each the four possible outcomes?

*Solution.* to Example 3.24

Since the probability of not rolling a 4 is 0.6, the probability of rolling a 4 must be 0.4. Since  $\{3, 4\} = \{3\} \cup \{4\}$ , a union of disjoint sets, the probability of rolling a 3 must be 0.3. Similarly, the probability of rolling a 2 must be 0.2, and

the probability of rolling a 1 must be 0.1. The table below list the probabilities of the possible events for this particular weighted die.

Table 3.4: All possible events associated with a single roll of a fair-sided die, and their probabilities assuming the die is weighted: roll a 1 with probability 1/10, 2 with probability 2/10, 3 with probability 3/10, 4 with probability 4/10.

Event	Description	Probability of event assuming a particular weighted die
$\emptyset$	Roll nothing (not possible)	0
$\{1\}$	Roll a 1	0.1
$\{2\}$	Roll a 2	0.2
$\{3\}$	Roll a 3	0.3
$\{4\}$	Roll a 4	0.4
$\{1, 2\}$	Roll a 1 or a 2	0.3
$\{1, 3\}$	Roll a 1 or a 3	0.4
$\{1, 4\}$	Roll a 1 or a 4	0.5
$\{2, 3\}$	Roll a 2 or a 3	0.5
$\{2, 4\}$	Roll a 2 or a 4	0.6
$\{3, 4\}$	Roll a 3 or a 4	0.7
$\{1, 2, 3\}$	Roll a 1, 2, or 3 (a.k.a. do not roll a 4)	0.6
$\{1, 2, 4\}$	Roll a 1, 2, or 4 (a.k.a. do not roll a 3)	0.7
$\{1, 3, 4\}$	Roll a 1, 3, or 4 (a.k.a. do not roll a 2)	0.8
$\{2, 3, 4\}$	Roll a 2, 3, or 4 (a.k.a. do not roll a 1)	0.9
$\{1, 2, 3, 4\}$	Roll something	1

The symbol  $P$  is more than just shorthand for the word “probability”.  $P$  denotes the underlying probability measure, which represents all the assumptions about the probability model. Changing assumptions results in a change of the probability measure. We often consider several probability measures for the same sample space and collection of events; these several measures represent different sets of assumptions and different probability models.

In the dice example above, suppose  $P$  represents the probability measure corresponding to the assumption of a fair die (equally likely outcomes). With this measure  $P(A) = 2/4$  for  $A = \{1, 2\}$ . Now let  $Q$  represent the probability measure corresponding to the assumption of the weighted die; then  $Q(A) = 0.3$ . The outcomes and events are the same in both scenarios, because both scenarios involve a four sided-die. What is different is the probability measure that assigns probabilities to the events. One scenario assumes the die is fair while the other assumes the die has a particular weighting, resulting in two different probability measures.

Both probability measures in the dice example could be written as explicit set functions: for an event  $A$

$$P(A) = \frac{\text{number of elements in } A}{4}, \quad P \text{ assumes a fair four-sided die}$$

$$Q(A) = \frac{\text{sum of elements in } A}{10}, \quad Q \text{ assumes a particular weighted four-sided die}$$

We provide the above descriptions to illustrate that a probability measure operates on sets. However, in many situations there does not exist a simple closed form expression for the set function defining the probability measure which maps events to probabilities.

Perhaps the concept of multiple potential probability measures is easier to understand in a subjective probability situation. For example, each model that is used to forecast the 2019 NFL season corresponds to a probability measure which assigns probabilities to events like “the Eagles win the 2019 Superbowl”. Different sets of assumptions and models can assign different probabilities for the same events.

- A single probability measure corresponds to a particular set of assumptions about the random phenomenon.
- There can be many probability measures defined on a single sample space, each one corresponding to a different probabilistic model.
- Probabilities of events can change if the probability measure changes.

### 3.4.2 Uniform probability measures $\#\{\text{sec-uniform-prob}\}$

Sometimes  $P(A)$  is defined explicitly for an event  $A$  via a formula. For example, in the case of a finite sample space with **equally likely outcomes**,

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega} \quad \text{when outcomes are equally likely}$$

**Example 3.25.** Flip a coin 4 times, and record the result of each trial in sequence. For example, HTTH means heads on the first on last trial and tails on the second and third. The sample space consists of 16 possible outcomes. One choice of probability measure  $P$  corresponds to assuming the 16 possible outcomes are equally likely. (This assumes that the coin is fair and the flips are independent.)

1. Specify the probability of each individual outcome, e.g.  $\{HHTH\}$ .
2. Find  $P(A)$ , where  $A$  is the event that exactly 3 of the flips land on heads.
3. Find  $P(B)$ , where  $B$  is the event that exactly 4 of the flips land on heads.



4. Find and interpret  $P(A \cup B)$ .
5. Find  $P(E_1)$ , the event that the first flip results in heads.
6. Find  $P(E_2)$ , the event that the second flip results in heads.
7. Find and interpret  $P(E_1 \cup E_2)$ .
8. We assumed the 16 outcomes are equally likely. Do the axioms require this assumption?

*Solution.* to Example 3.25

1. The sample space is composed of 16 outcomes which are assumed to be equally likely, so the probability of each outcome is  $1/16$ .
2.  $A = \{HHHT, HHHT, HTHH, THHH\}$  is the event that exactly 3 of the flips land on heads. Since  $A$  consists of 4 distinct equally likely outcomes,  $P(A) = 4/16$ .
3.  $B = \{HHHH\}$ , an event consisting of a single outcome, so  $P(B) = 1/16$ .
4. Directly,  $A \cup B = \{HHHT, HHHT, HTHH, THHH, HHHH\}$ , so  $P(A \cup B) = 5/16$ . Also  $A$  and  $B$  are disjoint, so  $P(A \cup B) = P(A) + P(B) = 4/16 + 1/16 = 5/16$ .
5. Intuitively this is  $1/2$ , but sample space outcomes consist of a sequence of four coin flips, so we should define the proper event.

$$E_1 = \{HHHH, HHHT, HHHT, HTHH, HHTT, HTHT, HTTH, HTTT\}$$

So  $P(E_1) = 8/16 = 1/2$ .

6. Similar to the previous part.

$$E_2 = \{HHHH, HHHT, HHHT, THHH, HHTT, THHT, THTH, THTT\}$$

So  $P(E_2) = 8/16 = 1/2$ .

7.  $E_1 \cup E_2$  is the event that at least one of the first two flips is heads:

$$E_1 \cup E_2 = \{HHHH, HHHT, HHHT, HTHH, THHH, HHTT, HTHT, HTTH, THHT, THTH, HTTT, THTT\}$$

So  $P(E_1 \cup E_2) = 12/16$ . Note that  $E_1$  and  $E_2$  are not disjoint, so we cannot just add their probabilities. But we can use the general addition rule for two events.

$$E_1 \cap E_2 = \{HHHH, HHHT, HHHT, HHTT\}$$

So  $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = 8/16 + 8/16 - 4/16 = 12/16$ .

8. No, the axioms do not require equally likely outcomes. If, for example, the coin were biased in favor of landing on Heads, we would want a different probability measure.

Probabilities are always defined for events (sets) but to shorten notation, it is common to write  $P(X = 3)$  instead of  $P(\{X = 3\})$ , and  $P(X = 4, Y = 3)$  instead

of  $P(\{X = 4\} \cap \{Y = 3\})$ . But keep in mind that an expression like “ $X = 3$ ” really represents an event  $\{X = 3\}$ , an expression which itself represents  $\{\omega \in \Omega : X(\omega) = 3\}$ , a subset of  $\Omega$ .

Probabilities involving multiple events, such as  $P(A \cap B)$  or  $P(X = 4, Y = 3)$ , are often called **joint probabilities**.

**Example 3.26.** Roll a four-sided die twice, and record the result of each roll in sequence. For example, the outcome  $(3, 1)$  represents a 3 on the first roll and a 1 on the second; this is not the same outcome as  $(1, 3)$ . One choice of probability measure corresponds to assuming that the die is fair and that the 16 possible outcomes are equally likely. Let  $X$  be the sum of the two dice, and let  $Y$  be the larger of the two rolls (or the common value if both rolls are the same).

1. Find the probability of each individual outcome, e.g.,  $\{(3, 1)\}$ .
2. Find  $P(A)$  where  $A$  is the event the the sum of the dice is 4.
3. Find  $P(B)$  where  $B$  is the event the the sum of the dice is at most 3.
4. Find  $P(\{Y = 3\})$  a.k.a.  $P(Y = 3)$ .
5. Find  $P(\{Y \leq 3\})$  a.k.a.  $P(Y \leq 3)$ .
6. Find  $P(\{X = 4\} \cap \{Y = 3\})$  a.k.a.  $P(X = 4, Y = 3)$  a.k.a.  $P(\{(X, Y) = (4, 3)\})$ .
7. Are the values of  $X$  equally likely? The values of  $Y$ ? The values of  $(X, Y)$ ?

*Solution.* to Example 3.26

1. Since 16 distinct, equally likely outcomes comprise the total probability of 1, the probability of each individual outcome must be  $1/16$ .
2. The event that the sum of the dice is 4 is  $A = \{(1, 3), (2, 2), (3, 1)\}$  which can be written as a union of three disjoint sets  $A = \{(1, 3)\} \cup \{(2, 2)\} \cup \{(3, 1)\}$ . Therefore,  $P(A) = P(\{(1, 3)\}) + P(\{(2, 2)\}) + P(\{(3, 1)\}) = 1/16 + 1/16 + 1/16 = 3/16 = 0.1875$ .
3. The event the the sum of the dice is at most 3 is  $B = \{(1, 1), (1, 2), (2, 1)\}$ , and so similarly to the previous part,  $P(B) = 3/16 = 0.1875$ .
4. Remember that  $\{Y = 3\}$  represents the event that that larger of the two rolls is 3,  $\{Y = 3\} = \{(1, 3), (2, 3), (3, 3), (3, 1), (3, 2)\}$ , an event which can be written as the union of 5 disjoint events each having probability  $1/16$ . Then  $P(\{Y = 3\}) = 5/16 = 0.3125$ .
5.  $\{Y \leq 3\} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$  so similar to the previous parts  $P(Y \leq 3) = 9/16 = 0.5625$ .
6.  $\{X = 4\} \cap \{Y = 3\} = \{(3, 1), (1, 3)\}$  so  $P(\{X = 4\} \cap \{Y = 3\}) = 2/16 = 0.125$
7. The values of  $X$  are not equally likely; for example,  $P(\{X = 4\}) = 3/16$  but  $P(\{X = 2\}) = P(\{(1, 1)\}) = 1/16$ . The values of  $Y$  are also not equally likely; for example,  $P(\{Y = 3\}) = 5/16$  but  $P(\{Y = 1\}) = P(\{(1, 1)\}) = 1/16$ . And the values of  $(X, Y)$  pairs are also not equally likely; for example,  $P(\{(X, Y) = (4, 3)\}) = 2/16$  but  $P(\{(X, Y) = (1, 1)\}) = 1/16$ . So

just because the underlying outcomes of the sample space are equally likely does not necessarily imply that everything of interest is equally likely as well.

**Example 3.27** (Matching problem). Recall the “matching problem”. A set of  $n$  cards labeled  $1, 2, \dots, n$  are placed in  $n$  boxes labeled  $1, 2, \dots, n$ , with exactly one card in each box. Consider the case  $n = 4$ . Let  $Y$  be the number of cards (out of 4) which match the number of the box in which they are placed.

We can consider as the sample space the possible ways in which the cards can be distributed to the four boxes. For example, 3214 represents that card 3 is returned (wrongly) to the first box, card 2 is returned (correctly) to the second box, etc. So the sample space consists of the following 24 outcomes<sup>18</sup>, which we will assume are equally likely.

---

1234	1243	1324	1342	1423	1432	2134	2143
2314	2341	2413	2431	3124	3142	3214	3241
3412	3421	4123	4132	4213	4231	4312	4321

---

1. What are the possible values of  $Y$ ? Find  $P(Y = y)$  for each possible value of  $y$ .
2. Let  $B$  be the event that at least one card is placed in the correct box. Find  $P(B)$ .
3. Let  $B_1$  be the event that card 1 is placed in box 1, and define  $B_2, B_3, B_4$  similarly. Represent the event  $B$  in terms of  $B_1, B_2, B_3, B_4$ .
4. Find  $P(B_1)$ . Also find  $P(B_2), P(B_3), P(B_4)$ .
5. Find  $P(B_1 \cap B_2 \cap B_3 \cap B_4)$ .
6. Is it possible to compute  $P(B_1 \cup B_2 \cup B_3 \cup B_4)$  based on just the five probabilities from the previous two parts?

*Solution.* to Example 3.27

The following table evaluates  $Y$  for each of the possible outcomes.

---

<sup>18</sup>There are 4 cards that could potentially go in box 1, then 3 cards that could potentially go in box 2, 2 in box 3, and 1 left for box 4. This results in  $4 \times 3 \times 2 \times 1 = 4! = 24$  possible outcomes.

Y(1234)=4	Y(1243)=2	Y(1324)=2	Y(1342)=1	Y(1423)=1	Y(1432)=2	Y(2134)=2	Y(2143)=1
Y(2314)=1	Y(2341)=0	Y(2413)=0	Y(2431)=1	Y(3124)=1	Y(3142)=0	Y(3214)=2	Y(3241)=1
Y(3412)=0	Y(3421)=0	Y(4123)=0	Y(4132)=1	Y(4213)=1	Y(4231)=2	Y(4312)=0	Y(4321)=1

- 
1. The possible values of  $Y$  are 0, 1, 2, 4.  $Y$  cannot be 3, since if 3 cards match, then the fourth card must necessarily match too.  $P(Y = 0) = 9/24$ ,  $P(Y = 1) = 8/24$ ,  $P(Y = 2) = 6/24$ ,  $P(Y = 4) = 1/24$ .
  2.  $P(B) = P(Y \geq 1) = 1 - P(Y = 0) = 15/24 = 0.625$ .
  3.  $B = B_1 \cup B_2 \cup B_3 \cup B_4$ .
  4. Intuitively,  $P(B_1) = 1/4$  since card 1 is equally likely to be placed in any of the 4 boxes. In terms of the sample space outcomes,  $B_1 = \{1234, 1243, 1243, 1324, 1342, 1423, 1432\}$ , so  $P(B_1) = 6/24 = 1/4$ . Also  $P(B_2) = P(B_3) = P(B_4) = 6/24$ .
  5.  $P(B_1 \cap B_2 \cap B_3 \cap B_4) = P(1234) = 1/24$ .
  6. Nope. The events are not disjoint, so you can't just add their probabilities. Also note that  $P(B_1 \cup B_2 \cup B_3 \cup B_4) \neq P(B_1) + P(B_2) + P(B_3) + P(B_4) - P(B_1 \cap B_2 \cap B_3 \cap B_4)$ . As we mentioned previously, the general addition rule is complicated for more than two events.

For finite sample spaces with equally likely outcomes, computing the probability of an event reduces to counting the number of outcomes that satisfy the event. The continuous analog of equally likely outcomes is a **uniform probability measure**. When the sample space is uncountable, size is measured continuously (length, area, volume) rather than discretely (counting).

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{size of } A}{\text{size of } \Omega} \quad \text{for a uniform probability measure } P$$

**Example 3.28.** Suppose the spinner in Figure 3.1 is spun twice. Let the sample space be  $\Omega = [0, 1] \times [0, 1]$  and let  $P$  be the uniform probability measure. Find the probability of each of the following events. (Hint: recall that these events are depicted in Figure 3.2)

1.  $A$ , the event that the first spin is larger than the second.
2.  $B$ , the event that the smaller of the two spins (or common value if a tie) is less than 0.5.
3.  $C$ , the event that the sum of the two spins is less than 1.5.
4.  $D$ , the event that the first spin is less than 0.4.

*Solution.* to Example 3.28

1.  $P(A) = 0.5$ . The shaded triangle makes up half the area of the square. This should make sense because the first spin should be equally likely to be the larger of the two spins.
2.  $P(B) = 0.75$ .
3.  $P(C) = 0.875$ . The unshaded triangle, representing  $C^c$ , has area  $(1/2)(0.5)(0.5) = 0.125$ .
4.  $P(D) = 0.4$ .

ADD plots of shaded rectangles

### 3.4.3 Non-uniform probability measures

For countable sample spaces, the probability measure is often defined by specifying the probability of each individual outcome. Then the probability of an event is obtained by summing the probabilities of the outcomes which comprise the event. Such was the case in Example 3.24; we could have specified the probability measuring by providing the probability of each face (1 with probability 0.1, ..., 4 with probability 0.4) and then obtained probabilities of all the events in Table 3.4 by adding the appropriate outcome probabilities.

For uncountable sample spaces, integration typically plays the analogous role that summation plays for countable sample spaces.

**Example 3.29.** Consider the sample space  $\Omega = [0, \infty)$  with a probability measure<sup>19</sup> defined by

$$P(A) = \int_A e^{-u} du, \quad A \subseteq [0, \infty).$$

1. Verify that  $P(\Omega) = 1$ .
2. Compute  $P(A)$  for  $A = [0, 1]$ .
3. Without integrating again, compute  $P(B)$  for  $B = (1, \infty)$ .
4. Compute  $P(C)$  for  $C = [0, 1] \cup (2, 4)$ .

*Solution.* to Example 3.29

---

<sup>19</sup>This defines the Exponential(1) distribution; see Section ??.



# Bibliography

- Gilovich, T. D., Vallone, R. P., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3):295–314.
- Miller, J. B. and Sanjurjo, A. (2018a). A cold shower for the hot hand fallacy: Robust evidence that belief in the hot hand is justified.
- Miller, J. B. and Sanjurjo, A. (2018b). Is it a fallacy to believe in the hot hand in the nba three-point contest?
- Miller, J. B. and Sanjurjo, A. (2018c). Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica*, 86(6):2019–2047.
- Ross, K. and Sun, D. L. (2019). Symbulkate: Simulation in the language of probability. *Journal of Statistics Education*, 27(1):12–28.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.
- Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.20.1.