Introduction to Probability

We frequently make conscious or subconscious estimates of the likelihood of an event happening.
Such estimates might be based on

- instinct,
- frequency of occurrences of similar events observed in the past
- logical deduction.

In this section we use set theory to assign a measure to the likelihood of an event occurring (probability) in a way that reflects our intuition, and that adds clarity to help us with deeper calculations.

Experiments

An **experiment** is an activity or phenomenon under consideration. The experiment can produce a variety of observable results called **outcomes**. The theory of probability makes most sense in the context of activities that can be repeated or phenomena that can be observed a number of times. We call each observation or repetition of the experiment a **trial**.

Examples

▸ Rolling a six sided die and observing the number on the uppermost face is an experiment with six possible outcomes; 1, 2, 3, 4, 5 and 6.

▸ Asking a citizen of Mathland whether they intend to vote in the next election is an experiment with two possible outcomes "yes" and "no". Each citizen asked is a trial of this experiment.

Experiments

- ► Observing the amount of time it takes to find parking after you arrive at the mall is an experiment with infinitely many possible outcomes, namely any time in the interval from 0 minutes to infinity.
- ► Rolling a six sided die and observing whether the number that comes up is even or odd is an experiment with two possible outcomes; "even" and "odd".

All these experiments have some **randomness**. The outcomes cannot be determined in advance and do not follow a set pattern. Although we may not be able to predict the outcome on the next trial of an experiment, we can sometimes make a prediction about how likely each possible outcome is (based on gut, logic or past experience). This a measure of the likelihood, (**the probability**), that the outcome will occur.

Types of probability predictions

Example If I roll a six sided precision cut die, I would predict, based on a logical deduction from the symmetry of the die, that each of the 6 numbers are equally likely to be rolled. I would predict the **probability** of rolling a 6 to be $1/6$.

Example Kris Bryant's batting average in 2016 was .292. I would predict, based on observing his 2016 at-bats, that in his first at-bat in the 2017 season he will have a hit with probability around 29%.

Example I **know** that Aaron Rodgers is the best quarterback in the NFL since Joe Montana. Based on this instinct, I would predict that the Packers will win Superbowl LII with probability .9999998.

Empirical Probability

When we observe a number of trials of an experiment and record the frequency of each possible outcome, the **relative frequency** of an outcome is the proportion of times that it occurs. When we use the relative frequency as a measure of probability we are using **empirical evidence** to estimate the likelihood of an event. The **empirical probability** calculated in this way may not agree with our reasoned probability — e.g., we might reason that an younger player like Bryant is "hot" at the start of the season, so in his first at-bat in quite likely to have a hit.

Empirical Probability

Example A (Not knowing the proportions in the population) A group of 300 voters were asked who they would vote for in the next election. The following table shows the results of the poll:

Candidate	Frequency	Relative Frequency
Melinda McNulty	120	
Mark Reckless	80	
Ian Lawless	100	
Total	300	

To estimate the probability that a voter chosen at random from the voting population will vote for any given candidate, calculate the relative frequency for each candidate.

Empirical Probability

Candidate	Frequency	Relative Frequency
Melinda McNulty	120	$\frac{120}{300} = 0.4 = 40\%$
Mark Reckless	80	$\frac{80}{300} \approx 0.267 = 26.7\%$
Ian Lawless	100	$\frac{100}{300} \approx 0.333 = 33.3\%$
Total	300	100%

Empirical Probability

A random selection of a voter from the population of voters means that each voter has a chance of being selected and any two voters have equal chances of being selected (not always easy to implement when sampling from a population).

Example B (Knowing what is in the population) A city called Bigtown with 4 million residents is divided into 4 districts by a large river and a large road running perpendicular to the river. The districts are called Newtown, Oldtown, Riverton and Roadton. The number of residents in each district according to the most recent census is shown below:

Empirical Probability

Bigtown

District	Number of Residents	Relative Frequency
Newtown	500,000	
Oldtown	1,250,000	
Riverton	750,000	
Roadton	1,500,000	
Total	4,000,000	

Calculate the relative frequencies.

Bigtown

District	Number of Residents	Relative Frequency
Newtown	500,000	12.50%
Oldtown	1,250,000	31.25%
Riverton	750,000	18.75%
Roadton	1,500,000	37.50%
Total	4,000,000	100.00%

Empirical Probability

Our data suggests that if we pick a resident of Bigtown, the probability that the resident is from Newtown will be 12.50%.

Note We are using relative frequency here to estimate probability, but we are also using logic in that we are assuming that all residents are equally likely to be chosen. We will see much more about this method of reasoning in the next section.

Law of large numbers

The Law of large numbers says that if an experiment is repeated many times, the relative frequency obtained for each outcome approaches the actual probability of that outcome.

This means that if we have used (sound) logic to arrive at a probability for an outcome and if we run our experiment "many" times, the relative frequency of the outcome should be "close" to our reasoned probability. This of course raises many questions about the number of trials required to get a good estimate etc... One need to explore statistics in more detail in order to answer these questions. For now we will satisfy ourselves with the knowledge that both of our methods of calculating probability agree.

Law of large numbers

Example B If we randomly selected a sample of 1,000 individuals from the population of Bigtown, how many would we expect to have from each district in our sample?

District	Relative Frequency	Sample of 1,000
Newtown	12.50%	≈ 125
Oldtown	31.25%	≈ 312
Riverton	18.75%	≈ 187
Roadton	37.50%	≈ 375
Total	100.00%	999 ≈ 1,000

Law of large numbers

By the law of large numbers, a large sample randomly chosen from the voting population should reflect the proportions of voters in the population who will vote for each candidate. However, we now face the question: "Is 300 large enough to get a good estimate of voter preferences?". Obviously the larger the sample, the better the picture of the preferences of the population, but also the more costly the polling process.

Most opinion polls tend to sample about 1000 people. We'll see in a while what's special about 1000!

Basic setup for probability — the sample space

Sample Space A sample space for an experiment is the set of all possible outcomes of the experiment. Each element of the sample space is called a **sample point** or **simple outcome** (or just an **outcome**).

Specifying the elements of your sample space is an important part of designing your experiment, since it dictates how your data will be recorded. A description of your experiment should also include a description of what should be observed. This will of course depend on what you are interested in studying.

Example If I roll a six sided die and observe the number on the uppermost face, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. On the other hand if I roll a six sided die and observe whether the number on the uppermost face is even or odd, the sample space is {even, odd}.

Basic setup for probability — the sample space

When specifying the elements of the sample space, S for an experiment, we should make sure it has the following properties:

1. Each element in the set S is a possible outcome of the experiment.
2. The list of outcomes covers all possible outcomes.
3. No two outcomes can occur on the same trial of the experiment.

Example Complete the sentences below to describe experiments leading to different sample spaces:

- Flip a pair of coins and observe ...
  $S =$
- Flip a pair of coins and observe ...
  $S =$

Basic setup for probability — the sample space

These two problems are two versions of the same thing and there are several answers.

- Flip a pair of coins and observe how many heads you get. $S = \{0, 1, 2\}$.

- Flip a pair of coins and observe how many tails you get. $S = \{0, 1, 2\}$.

- Flip a pair of coins and observe how many times you get two heads. $S = \{0, 1\}$.

- Flip a pair of coins and observe how many times you get two tails. $S = \{0, 1\}$.

- Flip a pair of coins and observe how many times you get one head and one tail. $S = \{0, 1\}$.

- And so on and so on.

Basic setup for probability — the sample space

Example Write out the sample space for the following experiments:

- ▶ Roll a pair of six sided dice and observe the minimum of the numbers on the uppermost face
  $S =$

- ▶ Roll a pair of six sided dice and observe the sum of the numbers on the uppermost face
  $S =$

- ▶ Roll a pair of six sided dice and observe the minimum of the numbers on the uppermost faces
  $S = \{1, 2, 3, 4, 5, 6\}$

- ▶ Roll a pair of six sided dice and observe the sum of the numbers on the uppermost faces
  $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Basic setup for probability — events

An event E is a subset of the sample space S.
All subsets of the sample space describe an event including
the empty set (the event that nothing happens), sets which
have a single outcome (called a **simple event**) and S itself
(the event that something happens). Sometimes we can
find a nice verbal description of the event. Given a verbal
description of an event, the corresponding subset is given
by all outcomes which fit the verbal description.

Example The experiment consists of rolling a six sided die
and observing the number on the uppermost face. The
sample space is given by

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

Find the subsets corresponding to the following verbal
description of the given events:

Basic setup for probability — events

E = the event that the number is even.

$E = \{2, 4, 6\}$

F = the event that the number is larger than two.

$F = \{3, 4, 5, 6\}$

We say that an event **occurs** if the trial yields an outcome in the event set.

For example, if we roll a 3 then F **has occurred**, but E **has not occurred**.

Basic setup — Assigning Probabilities

Let $S = \{e_1, e_2, \ldots, e_n\}$ be the sample space for an experiment, where $e_1, e_2, \ldots, e_n$ are the simple outcomes. To each outcome, e_i, $1 \leq i \leq n$, in the sample space, we assign a probability which we denote by $\mathbf{P}(e_i)$. As discussed above, the probability assigned to an outcome should reflect the relative frequency with which that outcome should occur in many trials of the experiment. Note that the relative frequencies in our empirical examples added to 1 since they gave the proportion of the sample corresponding to each outcome. Likewise if we use the relative frequency model to assign probabilities, the probabilities of the outcomes in our sample space must add to 1.

Basic setup — Assigning Probabilities

Thus we respect the following basic rules when assigning probabilities to outcomes in a finite sample space $S = \{e_1, e_2, \ldots, e_n\}$:

1. $0 \leq \mathbf{P}(e_i) \leq 1, \quad 1 \leq i \leq n$
2. $\mathbf{P}(e_1) + \mathbf{P}(e_2) + \cdots + \mathbf{P}(e_n) = 1.$

Example Peter did some research on the number of pairs of shoes that the students in his dorm had (on campus). He asked each student in his dorm how many pairs of shoes he had and found that everybody had between 1 and 7 pairs of shoes. Based on his research Peter assigned probabilities to the outcomes of the following experiment:

Choose a student at random from Peter's dorm and ask how many pairs of shoes they have (on campus).

Basic setup — Assigning Probabilities

Which of the following constitutes a valid assignments of probabilities?

# Pairs	Probability
1	1/4
2	1/4
3	1/16
4	1/16
5	1/8
6	1/8
7	1/8

# Pairs	Probability
1	−1
2	1/2
3	1/2
4	1/4
5	1/4
6	1/4
7	1/4

# Pairs	Probability
1	1
2	1/4
3	1/16
4	1/16
5	1/8
6	1/8
7	1/8

The second box has a probability of -1 but all probabilities lie between 0 and 1.

The third box has all probabilities between 0 and 1 but the sum is greater than 1.

The first box has all probabilities between 0 and 1 and the sum equal 1.

Basic setup — Assigning Probabilities

Example A bag contains 5 red marbles, 10 blue marbles and 20 green marbles. You choose a marble at random from the bag and observe the color of the marble. Assign probabilities to the outcomes in the sample space for this experiment.

There are 35 marbles. Your chances of drawing a red one are $\frac{5}{35}$; your chances of drawing a blue one are $\frac{10}{35}$; and your chances of drawing a green one are $\frac{20}{35}$. Notice all probabilities are between 0 and 1 and they add up to 1.

More precisely, the sample space is $\{red, blue, green\}$ and $\mathbf{P}(red) = \frac{5}{35}$, $\mathbf{P}(blue) = \frac{10}{35}$, and $\mathbf{P}(green) = \frac{20}{35}$.

Basic setup — Assigning Probabilities

To calculate the probability of an event E, we sum the probabilities of the outcomes in the event.

Example A bag contains 5 red marbles, 10 blue marbles and 20 green marbles. You choose a marble at random from the bag and observe the color of the marble. What is the probability that the marble drawn is not red on your next trial of this experiment?

If it is not red then it is blue or green (but never both) so the probability is $\mathbf{P}(blue) + \mathbf{P}(green) = \dfrac{10}{35} + \dfrac{20}{35} = \dfrac{30}{35}$.

OR the probability is 1 minus the probability that it is red:
$1 - \mathbf{P}(red) = 1 - \dfrac{5}{35} = \dfrac{30}{35}$.

Basic setup — Assigning Probabilities

Example Peter also asked the students in his dorm how many siblings they had. The outcomes were $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. The following table shows the relative frequencies of the outcomes.

# Siblings	Rel. Freq.
0	.2
1	.2
2	.2
3	.1
4	.1
5	.1
6	.05
7	.02
8	.01
9	.01
10	.01

Basic setup — Assigning Probabilities

Using the above relative frequencies as probabilities, find
the probability that a student chosen at random from
Peter's dorm will have more than 3 siblings.

The events are 4 siblings, 5 siblings, 6 siblings, 7 siblings, 8
siblings, 9 siblings, or 10 siblings. Hence the probability of
being in this subset is
$0.1 + 0.1 + 0.05 + 0.02 + 0.01 + 0.01 + 0.01 = 0.3.$

Since every outcome is in the sample space, S, and the sum
of the probabilities of all outcomes is 1, we have $P(S) = 1$.
Since there are no outcomes in the empty set, we have
$P(\emptyset) = 0.$