Solon Karapanagiotis

Which bicycle lock should I buy? A journey to decision making under uncertainty

FEBRUARY 2019

Cite this work as: Solon Karapanagiotis. Which bicycle lock should I buy? A journey to decision making under uncertainty, 2019.

Please send feedback on this book via karapanagiotissolon@gmail.com

First edition, January 2019

Contents

Course Rationale

This course is about making decisions under uncertainty. We are always uncertain about future events: will it rain tomorrow? But we still need to make a decision: should I get my umbrella when I leave home? Probability allows us to quantify our uncertainty. We all are familiar with statements like "there is 20% chance it will rain tomorrow". Still, this has not solved the original question: is it worth caring the umbrella around with me? This course will try to present a framework to give answers to such questions.

We will start with a motivating example where we will be uncertain on the best action to take and use it to calculate probabilities of events. That is, we will answer questions such as "how likely is that it will rain tomorrow?". We will go a step further and answer more complicated questions such as "how likely is that rain tomorrow given it is winter?". Finally, we will explore the framework that will allow us to make decisions: "shall I get my umbrella, or not?".

This course is aimed at students with interest in probability, decision theory and making decision in the face of uncertainty.

Motivating example

You have just bought an old bicycle for £80. You can either buy a Kryptonite lock for £20 or a cheap chain and lock for £8. You reckon it is extremely unlikely anyone would bother stealing your bike if it is locked with Kryptonite. But it is more probable if you buy a cheap chain. If the bike is stolen, you will not bother to replace it. Which lock do you buy? [1]

[1] Ian Hacking. *An Introduction to Probability and Inductive Logic Desk Examination Edition*. Cambridge University Press, 2001

This is the example we will be using throughout this handbook. Each chapter functions as a building block that will help us answer the final question.

Chapter Outline

The 1st chapter introduces tree diagrams and 2 by 2 tables. These are two ways to visualise information. They help answer questions like "How likely is that my bike will be stolen?". In other words, they help us calculate probabilities of events.

In the 2nd chapter we will differentiate between categorical and conditional probabilities. Briefly, for two events A and B, the categorical probability answers the questions: "how likely is event A occurs?". We denote this P(A). The conditional probability reads as given the event B has already happened how likely is event A will occur. We denote this P(A given B). We will see how to use the tree diagram to calculate conditional probabilities.

During the 3rd chapter we will introduce Bayes Rule. It follows simply from conditional probability and allows us to calculate the *inverse* probability, P(B given A). It is a simple formula that describes how to update the probabilities of events when given evidence - how to learn from experience.

The 4^{th} chapter will be dedicated to the concept of Value of an Act and how to choose among different acts. We will see how to quantify the gain of a consequence when an act is chosen and combine it with probabilities (Chapters 1-3) to calculate the Value of the Act. We will then be able to answer our initial question: Which lock do I buy?

The Appendix provides a more mathematically rigorous treatment of the probability concepts used in the first 3 chapters. Readers are encouraged to start reading it after going trough the first 3 chapters.

1st Chapter - The tree diagram

What is the purpose of the chapter?

- Introduce tree diagrams and 2 by 2 tables to visualise information.

Keywords: tree diagram, 2 by 2 table.

READ AGAIN THE MOTIVATING EXAMPLE. A natural question that you may have is: How do I know it is more probable for a cheap chain being snipped? This is what we will focus on in this chapter. The answer is simple, you probably have guessed. We will collect some data. I did that for you. I went to the police station and asked them exactly this question. They told me:

- there were 1000 registered bikes in the town,
- 400 bikes were stolen last year,
- out of those 400, 80 had Kryptonite locks,
- 360 out of the 600 non-stolen had Kryptonite locks.

Let's visualise this information, see Figure 1.

This is called a *tree diagram*. It has two main parts: the branches and the ends. The arrows are the branches. The purple boxes are the ends. Their job is to connect the ends. Together they form paths. Each path represents a possible outcome (e.g. stolen and Kryptonite lock). Figure 1 is read from left to right. Let's see an example. We start we 1000 bicycles and follow the Stolen branch. We see that 400 bicycles were stolen. We then follow the upper branch, which corresponds to bicycles that had Kryptonite locks. These were 80.

We can use the tree diagram to answer many questions of interest. We just need to follow the arrows.

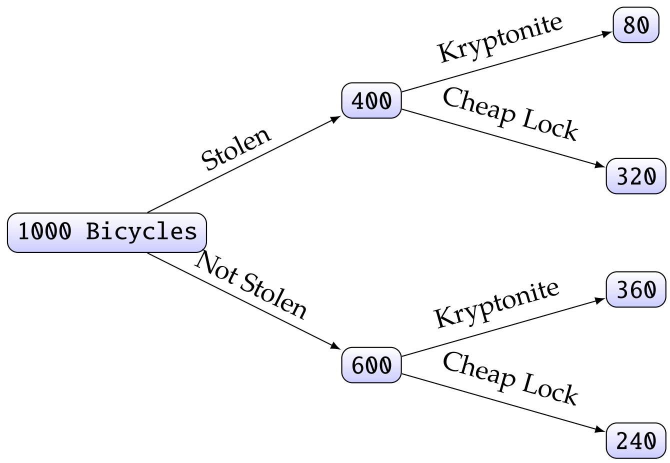


Figure 1: Tree diagram, with whole numbers.

1. How many stolen bikes had Kryptonite locks? 80.

2. How many non-stolen bikes had Kryptonite locks? 360.

3. How many bikes had Kryptonite locks? 360 + 80 = 440.

4. What proportion of the stolen bikes had Kryptonite locks? There were 400 stolen bikes. Out of which 80 had Kryptonite locks. So, the answer is $\frac{80}{400} = 0.2$.

5. What proportion of the stolen bikes had cheap locks?

What is a Tree Diagram? A tree diagram is simply a way of representing a sequence of events. Tree diagrams are particularly useful in probability since they record all possible outcomes in a clear and uncomplicated manner. Tree diagrams allow us to see all the possible scenarios of an event and calculate their probability. You just did this for question 4. Each branch in a tree diagram represents a possible scenario.

Why are they called tree diagrams? Because the look like trees (see Figure 2)! Do you see the branches?

Another way is to present the same information is to use a 2 by 2 table (see Table 1). It has two rows and two columns, hence the name. The rows describe the lock status, if the bike had Kryptonite lock or not. The columns describe the stolen status, if the bike was stolen or not. Take some time to familiarise yourself with the table and verify that it actually presents the same information as the tree-diagram (Figure 1).

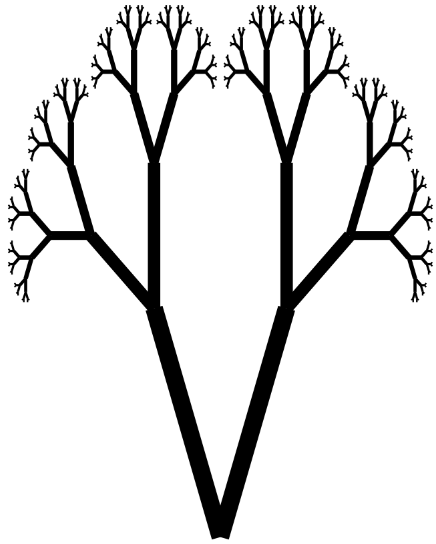


Figure 2: A tree diagram.

Table 1: 2 by 2 table.

		Stolen		
		Yes	No	Total
Kryptonite	Yes	80	360	440
	No	320	240	560
	Total	400	600	1000

Exercise 1

1. Try answering the same questions 1-5 as in page 10 using the 2 by 2 table (Table 1). Do you get the same answers?
2. A committee of 3 members is to be formed consisting of one representative each from labour, management, and the public. If there are 3 possible representatives from labour, 2 from management, and 4 from the public, determine how many different committees can be formed.
3. Dr. No has a patient who is very sick. Without further treatment, this patient will die in about 3 months. The only treatment alternative is a risky operation. The patient is expected to live about 1 year if he survives the operation; however, the probability that the patient will not survive the operation is 0.3. Draw a

decision tree for this simple decision problem. Show all the probabilities and outcome values.

4. Time to experiment: Toss a coin 20 times. Every time it comes up heads roll a die and record the number. If it comes up tails, toss once more and record the outcome (heads or tails). Use a tree diagram to visualise the process and the results.

2nd Chapter - Conditional Probability

What is the purpose of the chapter?

- Differentiate between categorical and conditional probability.
- Use the tree diagram to calculate conditional probabilities.

Keywords: categorical probability, conditional probability.

Let's revisit the tree diagram (see Figure 3). Last time, we saw briefly how to calculate proportions. We will spend a little bit more time now.

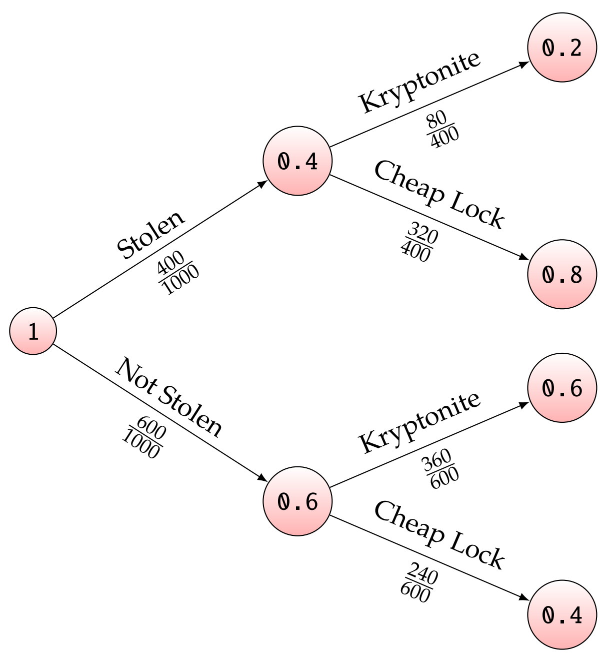


Figure 3: Tree diagram, with proportions.

Those proportions are probabilities. We express probabilities in numbers ranging between 0 and 1. You can read them as "the probability that your bike will get stolen is 0.4". I call this *categorical* probability.

Take a look at question 4 on page 10 once more. This is the probability the bike has Kryptonite locks, conditional on being stolen. We call this *conditional* probability. The most important new idea in this chapter is the probability that something happens, on condition that something else happens. Here are a few examples:

Example: Categorical vs Conditional Probability

Categorical: What is the probability that it rains tomorrow?
Conditional: What is the probability that it rains tomorrow, given that there have been rains on each of the 3 preceding days?

Categorical: What is the probability that my bike will get stolen?
Conditional: What is the probability that my bike will get stolen, given I bought a cheap lock?

You have probably heard people say things like:

- The chance of rain tomorrow is 75%.
- Based on how poorly the interview went, it is unlikely I will get the job.
- In a drawer of ten socks, 8 of them yellow, there is a twenty percent chance of choosing a sock that is not yellow.

All of these statements are about probability. We see words like "chance", "(un) likely", "probably" since we do not know for sure something will happen, but we realise there is a very good chance that it will.

Probabilities can be written as fractions, decimals or percentages on a scale from 0 to 1. We will all three ways in this course.

Notation

(It will be useful later)
Event S: bike Stolen
Event N: bike Not stolen
Even K: bike had Kryptonite lock
Even C: bike had cheap lock
Probability is represented: $P(\,)$

Examples of probability: P(bike stolen) = P(S) = 0.4 (see Figure 3)

I use the word *given* for conditional probabilities.

Example: Parking

If you park overnight near my home, and do not live on the block, you may be ticketed for not having a permit for overnight parking. The fine will be £20. But the street is only patrolled on average about once a week. What is the probability of being fined? Apparently the street is never patrolled on two consecutive nights. What is the probability of being ticketed tonight, conditional on having been ticketed on this street last night?

We will formally answer these questions in Chapters 3 and 4. For now, try to spot the categorical and conditional probabilities.

Having gained some intuition on the difference between categorical and conditional probabilities our first definition follows. It is the mathematical definition of conditional probability.

For any two events A and B, we may also define the new event (A and B), called the *intersection* of A and B, to consist of all outcomes that are both in A and in B. That is, the event (A and B) will occur only if both A and B occur. Consider the experiment of flipping two coins: if A = $\{(H,H),(H,T),(T,H)\}$ is the event that at least 1 head occurs and B = $\{(H,T),(T,H),(T,T)\}$ is the event that at least 1 tail occurs, then (A and B) = $\{(H,T),(T,H)\}$. We can then calculate the probability of this new event, that is P(A and B). Note this is the numerator of the definition of conditional probability.

Definition: Conditional Probability

When P(B) $>$ 0,

$$P(\text{A given B}) = \frac{P(\text{A and B})}{P(\text{B})}.$$

This expression reads as "the probability A occurs given B has already occurred is equal to the probability that both A and B occur divided by the probability that B occurs."* A and B can be any event you are interested in. An example will make things more clear. For instance, "what is the probability that the sidewalk is wet given that it rained earlier"? If A is "the sidewalk is wet" and B is "it rained earlier," the expression reads as "the probability the sidewalk is wet given that it rained earlier is equal to the probability that the sidewalk is wet and it rains over the probability that it rains".

*This statement implies an explicit temporal causality between A and B which may not be valid in many contexts. Try to think of some examples. In that case, an appropriate interpretation should be "$P(\text{A given B}) = P(\text{A = a given B = b})$ is the probability of A being in state a under the constraint that B is in state b".

Note the constraint $P(\text{B}) > 0$. This means P(B) must be a positive number, because we cannot divide by zero. But why is the rest of this definition sensible? Some examples will suggest why.

Example: Ice Cream

70% of your friends like Chocolate, and 35% like Chocolate AND like Strawberry. What percent of those who like Chocolate also like Strawberry? We are essentially asking for the probability that someone that likes chocolate also likes Strawberry, P(Strawberry given Chocolate).

Using the definition of conditional probability:

$$P(\text{Strawberry given Chocolate}) = \frac{P(\text{Chocolate and Strawberry})}{P(\text{Chocolate})} = \frac{0.35}{0.7} = 0.5.$$

So, 50% of your friends who like Chocolate also like Strawberry.

Example:

Let A denote the event "student is female" and let B denote the event "student is French". In a class of 100 students suppose 60 are French, and suppose that 10 of the French students are females. Find the probability that if I pick a French student, it will be a girl, that is, find P(A given B).

Since 10 out of 100 students are both French and female, then

$$P(\text{A and B}) = \frac{10}{100}.$$

Also, 60 out of the 100 students are French, so

$$P(\text{B}) = \frac{60}{100}.$$

So the required probability is

$$P(\text{A given B}) = \frac{P(\text{A and B})}{P(\text{B})} = \frac{\frac{10}{100}}{\frac{60}{100}} = \frac{1}{6} = 0.167.$$

Example: Conditional Dice

Think of a fair die. We say the outcome of a toss is even if it falls 2, 4, or 6 face up. Here is a conditional probability: P(6 given even).

In ordinary English: The probability that we roll a 6, on condition that we rolled an even number. The conditional probability of sixes, given evens.

With a fair die, we roll 2, 4, and 6 equally often. So 6 comes up a third of the time that we get an even outcome.
P(6 given even) = $\frac{1}{3}$.
This fits our definition, because,
P(6 and even) = P(6) = $\frac{1}{6}$. (why this makes sense?, i.e. why P(6 and even) = P(6) ?)
P(even) = $\frac{1}{2}$.
P(6 given even) = $\frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$.

In the previous example, it was straightforward to calculate P(6 and

A nice tutorial on how to draw probability tree diagrams can be found at http://www.statisticshowto.com/how-to-use-a-probability-tree-for-probability-qu even). Sometimes this calculation is not easy. In the next one we will make use of the tree diagram to help us.

Example: Urns

Imagine two identical urns (think of them as large vases), each containing red and green balls. Urn A has 80% red balls, 20% green, and Urn B has 60% green, 40% red. You pick an urn at random*. Is it A or B?

*This means that I cannot predict which urn you will choose. In other words, it is equally likely to pick either. More precisely, the probability that you pick urn A is equal to the probability that you pick urn B. This must then be P(A) = P(B) = 0.5.

Let's draw balls from the urn and use this information to guess which urn it is. After each draw, the ball drawn is replaced, it is put back to the urn. Hence for any draw, the probability of getting red from urn A is 0.8, and from urn B, the probability of getting red is 0.4. Using the notation we introduced before:

- P(R given A) = 0.8 (this is the conditional probability of drawing a red ball given you picked urn A),

- P(R given B) = 0.4, (this is the conditional probability of drawing a red ball given you picked urn B),

- P(A) = P(B) = 0.5, (this is the categorical probability of picking urn A or B. It is random. In other words, it is equally likely to pick either).

You draw a red ball. We want to find P(A given R), which is P(A and R)/P(R). (This is the definition of conditional probability).

You can get a red ball from either urn A or urn B. In other words, you chose urn A and draw a red ball, this is (A and R) or you chose urn B and draw a red ball, that is (B and R), the two events are independent**, so we can write

**A formal definition of independence of two events is given in the Appendix.

$$P(\text{R}) = P(\text{A and R}) + P(\text{B and R}).$$

The probability of getting urn B is 0.5; the probability of getting a red ball from it is 0.4, so that the probability of both happening is P(B and R) = P(R and B) = P(R given B) $\times$ P(B) = 0.4 $\times$ 0.5 = 0.2 (from the definition of conditional probability).

Answering the question we have used the following property: P(B and R) = P(R and B). This is part of the *commutative law* that says for two events R and B, the following is valid: (B and R) = (R and B).

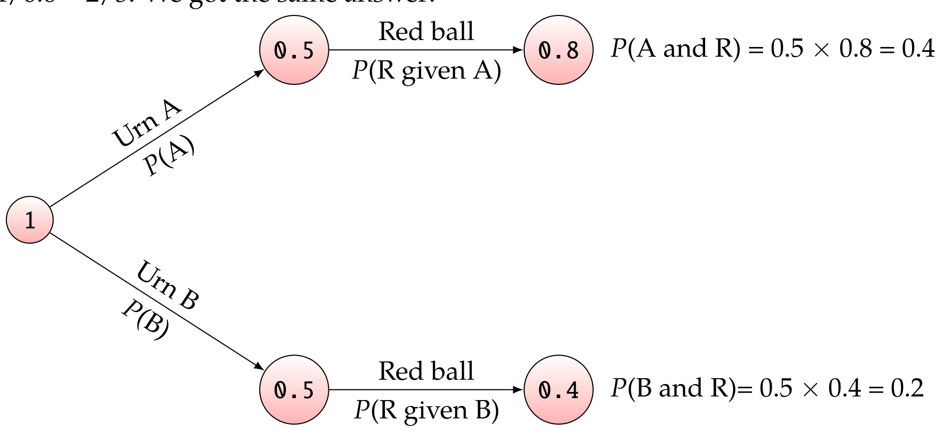
Likewise, P(A and R) = 0.8 × 0.5 = 0.4.

So, P(R) = P(A and R) + P(B and R) = 0.4 + 0.2 = 0.6.

Now we have all the ingredients we need to calculate P(A given R), which is

$$P(\text{A given R}) = \frac{P(\text{A and R})}{P(\text{R})} = \frac{0.4}{0.6} = \frac{2}{3}.$$

Let's draw the calculation. We start out with our coin and the two urns. How can we get to a red ball? There are two routes. We can toss a heads (probability 0.5), giving us urn A. Then we can draw a red ball (probability 0.8). That is the route shown here on the top branch. We can also get an R by tossing tails, going to urn B, and then drawing a red ball, as shown on the bottom branch. We get to R on one of the two branches. So the total probability of ending up with R is the sum of the probabilities at the end of each branch. Here it is 0.4 + 0.2 = 0.6. The probability of getting to an R following branch A is 0.4. Thus that part of the probability that gets you to R by A, namely P(A given R) = 0.4/0.6 = 2/3. We got the same answer!



Exercise 2

1. I have transformed our initial tree diagram (Figure 1) to a *probability tree diagram* (Figure 3), like the one above. Use it to calculate the following conditional probabilities:
   - P(Kryptonite lock given Stolen)
   - P(Cheap lock given Stolen)
   - P(Kryptonite lock given Not Stolen)

- P(Cheap lock given Not Stolen).

2. A fair die is tossed twice. Find the probability of getting a 4, 5, or 6 on the first toss and a 1, 2, 3, or 4 on the second toss.
3. Two cards are drawn from a well-shuffled ordinary deck of 52 cards. Find the probability that they are both aces if the first card is (a) replaced, (b) not replaced.
4. More practice with tree diagram and conditional probabilities: Go to `https://www.mathsisfun.com/data/probability-tree-diagrams.html`. After reading the examples, solve the questions at the bottom of the webpage.

3rd Chapter - Bayes Rule

What is the purpose of the chapter?

- Introduce Bayes Rule.

Keywords: Bayes Rule.

THE PREVIOUS CHAPTER ended with two examples of the same form: urns and dice. The numbers were changed a bit, but the problems in each case were identical. Then we asked, what is the probability that the urn drawn was A, conditional on drawing a red ball? We asked for:
P(A given R) = ?
Similarly, in our bike theft example we asked for,
P(K given S) = ?

In Chapter 2 we solved these problems directly from the definition of conditional probability. There is an easy rule for solving problems like that. It is called *Bayes' Rule*.

Bayes' Rule is named after Thomas Bayes (1702-1761), an English minister who was interested in probability. Bayes never published what would become his most famous accomplishment; his notes were published after his death. Check https://en.wikipedia.org/wiki/Thomas_Bayes

Definition: Bayes Rule

$$P(A \text{ given } B) = \frac{P(A)P(B \text{ given } A)}{P(B)}$$

$$= \frac{P(A)P(B \text{ given } A)}{P(A)P(B \text{ given } A) + P(A^c)P(B \text{ given } A^c)}$$

For the last equality in Bayes Rule we used the following: $P(B) = P(A)P(B \text{ given } A) + P(A^c)P(B \text{ given } A^c)$. First, the notation A^c, refers to as the complement of A, to consist of all outcomes not in A. That is, A^c will occur if and only if A does not occur.

In words, the denominator of Bayes Rule states that the probability of the event B is a weighted average of the conditional probability of B given that A has occurred and the conditional probability of B given that A has not occurred. More details are given in the Appendix on page 43.

Example: Urns

Here is the urn problem from page 17. Imagine two urns, each containing red and green balls. Urn A has 80% red balls, 20% green, and Urn B has 60% green, 40% red. You pick an urn at random, and then draw balls from the urn in order to guess which urn it is. After each draw, the ball drawn is replaced back to the urn. Hence for any draw, the probability of getting red from urn A is 0.8, and from urn B it is 0.4.

- P(R given A) = 0.8
- P(R given B) = 0.4
- P(A) = P(B) = 0.5

You draw a red ball. What is P(A given R)?

Solution by Bayes' Rule:

$$P(A \text{ given } R) = \frac{P(A)P(R \text{ given } A)}{P(A)P(R \text{ given } A) + P(B)P(R \text{ given } B)} = \frac{(0.5 \times 0.8)}{(0.5 \times 0.8) + (0.5 \times 0.4)} = \frac{2}{3}. \quad (1)$$

This is the same answer as was obtained on page 17.

Take a look at the denominator of equation (1) on the left and compare it with the denominator of Bayes Rule in the previous page. Verify they are the same. That is:

$$P(A)P(R \text{ given } A) + P(B)P(R \text{ given } B)$$

$$= P(A)P(R \text{ given } A) + P(A^c)P(R \text{ given } A^c) = P(R)$$

Bayes Rule allowed us to answer the "inverse" question. We knew P(R given A) and wanted to know P(A given R). Do you see the difference? We started with the probability of getting a red ball given we chose urn A - P(R given A) - and we asked for the probability of drawing from urn A given that we drew a red ball - P(A given R).

Example: Cards

Card example from `https://brilliant.org/wiki/bayes-theorem/`

If a single card is drawn from a standard deck of playing cards, the probability that the card is a queen is 4/52, since there are 4 queens in a standard deck of 52 cards. Rewording this,

$$P(\text{Queen}) = \frac{4}{52} = \frac{1}{13}.$$

If evidence is provided (for instance, someone looks at the card) that the single card is a face card, then the probability P(Queen given Face) can be calculated using Bayes' Rule:

$$P(\text{Queen given Face}) = \frac{P(\text{Queen})P(\text{Face given Queen})}{P(\text{Face})}$$

Since every Queen is also a face card, P(Face given Queen)= 1. Since there are 3 face cards in each suit (Jack, Queen, King), the probability of a face card is P(Face) = $\frac{3}{13}$. Using Bayes Rule gives P(Queen given Face) = $\frac{1}{3}$.

Example: Spiders

A tarantula is a large, fierce-looking, and somewhat poisonous tropical spider. Once upon a time, 3% of consignments of bananas from Honduras were found to have tarantulas on them, and 6% of the consignments from Guatemala had tarantulas. 40% of the consignments came from Honduras. 60% came from Guatemala. A tarantula was found on a randomly selected lot of bananas. What is the probability that this lot came from Guatemala?

Let G = The lot came from Guatemala. P(G) = 0.6.
Let H = The lot came from Honduras. P(H) = 0.4.
Let T = The lot had a tarantula on it.

- P(T given G) = 0.06.
- P(T given H) = 0.03.

$$P(\text{G given T}) = \frac{P(\text{G})P(\text{T given G})}{P(\text{G})P(\text{T given G}) + P(\text{H})P(\text{T given H})}$$

$$P(\text{G given T}) = \frac{(0.6 \times 0.06)}{[(0.6 \times 0.06) + (0.4 \times 0.03)]} = \frac{3}{4}.$$

I will stress again the "beauty" (in my opinion) of Bayes Rule. It allowed us to transition from P(T given G) to P(G given T) in the spider example. Note these are two different things! P(T given G) answers the question: "If I knew the bananas came from Guatemala, what is the probability I find tarantulas?". If I see a tarantula on the banana this question is useless. Then, I want to know "What is the probability the bananas came from Guatemala, given I found a tarantula?" - this is P(G given T). You may not care about this, but if you owned a grocery store you would not like to order bananas from Guatemala. The probability is quite high, $\frac{3}{4}$.

One application of Bayes' Rule is in spam filtering. It refers to the automatic processing of incoming messages and categorizing them as spam (unwanted email) or not. Have you ever wonder how this works?

For more details see https://en.wikipedia.org/wiki/Naive_Bayes_spam_filtering

Example: Spam Filtering

We want to know the probability an email is spam given it contains certain words:

$$P(\text{spam given words}) = \frac{P(\text{words given spam})P(\text{spam})}{P(\text{words})}.$$

Bayes Rule allows us to predict the chance a message is really spam given the "test results" (the presence of certain words). Spam filtering based on a just the presence of certain words in the email is flawed - it's too restrictive and false positives are too great. But filtering using Bayes Rule gives us a middle ground - we use probabilities. As we analyze the words in a message, we can compute the chance it is spam (rather than making a yes/no decision). If a message has a 99% chance of being spam, it probably is.

Bayes Rule on Wikipedia `https://en.wikipedia.org/wiki/Bayes%27_theorem`

Generalisation of Bayes Rule

The same formula holds for any number of mutually exclusive and jointly exhaustive events:

$$A_1, A_2, A_3, \ldots, A_k.$$

Mutually exclusive means that only one of the events can be true. Jointly exhaustive means that at least one must be true. By extending the Bayes Rule, if P(E) > 0, and for every i, $P(A_i) > 0$, we get for any event A_k,

$$P(A_j \text{ given } E) = \frac{P(A_j)P(E \text{ given } A_j)}{P(E)}$$

$$= \frac{P(A_j)P(E \text{ given } A_j)}{\sum_{i=1}^{k}[P(A_i)P(E \text{ given } A_i)]}.$$

Here the $\sum$ (the Greek capital letter sigma) stands for the sum of the terms with subscript i. Add all the terms $[P(A_i)P(E \text{ given } A_i)]$ for $i = 1, i = 2$, up to $i = k$.

Thus, looking at the denominator:

$$P(E) = \sum_{i=1}^{k}[P(A_i)P(E \text{ given } A_i)]$$

shows how, for given events $A_1, A_2, A_3, \ldots, A_k$, of which one and only one must occur, we can compute $P(E)$ by first conditioning on which one of the A_i occurs. That is, $P(E)$ is equal to a weighted average of $P(E \text{ given } A_i)$, each term being weighted by the probability of the event on which it is conditioned. Same story are the simpler version of Bayes Rule.

Exercise 3

1. The probability that it is Friday and that a student is absent is 0.03. Since there are 5 school days in a week, the probability that

it is Friday is 0.2. What is the probability that a student is absent given that today is Friday?

2. Calculate the P(H given T) for the spider example above. What do you conclude? If a tarantula was found on a lot, would it be more likely the bananas were coming from Honduras or Guatemala?

3. Calculate the following conditional probabilities from our motivating example:

    - P(Stolen given Kryptonite lock)
    - P(Stolen given Cheap lock)
    - P(Not Stolen given Kryptonite lock)
    - P(Not Stolen given Cheap lock).

    What do you conclude? Write a few sentences comparing the above probabilities. Focus on their interpretations.

4. A family has two children. Given that one of the children is a boy, what is the probability that both children are boys?

4th Chapter - Making decisions

What is the purpose of the chapter?

- Calculate the Expected Value of an Act.
- Introduce the Value Rule: How to choose among different acts?
- Provide guidelines how to set-up and solve decision problems.

Keywords: Expected Value of an Act, Value Rule

Uncertainty is everywhere. We are not just uncertain about what will happen, or what is true, but also when we are uncertain about what to do. Until now we used probability to quantify how uncertain we are. In this chapter we will see how we can use this to make decisions. We will see that decisions need more than probability. They are based on the value of possible outcomes.

Decisions depend on acts and consequences. For example,
Should you open a small business?
Should you take an umbrella?
Should you buy a Lotto ticket?

In each case you settle on an *act*. Doing nothing at all counts as an act.
Acts have *consequences*.
You go broke (or maybe found a great company).
You stay dry when everyone else is sopping wet (or you mislay your umbrella).
You waste a pound (or perhaps win a fortune).

Suppose you can represent the cost or benefit of a possible consequence by a number - so many pounds, perhaps. Call that number the *utility* of the consequence. We have all the ingredients to define the Expexted Value of an Act.

Definition: Expected Value of an Act

Consider an Act with just two possible consequences
Act: A
Consequences: C1, C2
Utility: U1, U2
Probabilities: P(C1 given A) and P(C2 given A).

Expected Value of A = Exp(A) = [P(C1 given A)][U1] + [P(C2 given A)][U2].

If you choose act A, the possible consequences are C1 and C2. If C1 happens you have an utility of U1. If C2 happens you have an utility of U2. Note the utility can be negative also. This means we incur a loss. Also the utility can be zero (see the example that follows). It is clear from the definition you need to multiply the utility of each consequence with its probability. Note this is a conditional probability, so you have to calculate it first. This was the purpose of the first 3 chapters.

Example: Ticket

Your aunt offers you a free lottery ticket for your birthday, but says you do not have to take it. The two possible acts are: accept, and do not accept. The value of not accepting is just 0, you don't gain or lose anything. What is the value of accepting? Suppose the lottery has 100 tickets, with a prize of £50 for the one ticket that is drawn. If you accept the ticket, there are two possible consequences:

Consequence 1: Your ticket is drawn.
Utility of Consequence 1: £50.
Probability of Consequence 1: 0.01.
Consequence 2: Your ticket is not drawn.
Utility of Consequence 2: 0.
Probability of Consequence 2: 0.99.

Exp(accepting the ticket) = (0.01)(£50) + (0.99)(0) = 50p.
Exp(not accepting) = 0.

How do you choose among possible acts? The most common decision rule is an obvious one.

Rule: Value Rule
Act so as to maximize value. Perform the action with the highest expected value.

So, accept your aunt's offers!

Example: Parking example

You are staying overnight with friends. They live on a crowded city street, where curbside parking is restricted to residents with parking decals on their cars. You have driven a car, but there is no place nearby where you can park legally without paying. There is a nearby lot that charges £3 a night. It is the middle of winter, freezing cold. It is a half-hour walk to the nearest likely free spot. Your friends say: Just park on the street. It is patrolled about only once every ten days, so the probability of getting a ticket is only 0.1. The fine is £20, and you always pay parking fines in order to renew your license. What is the expected value of parking illegally? It will be a negative value.

I: Park illegally.
L: Park in the lot.
T: You get a ticket.
P: You pay £3.
P(T given I) = 0.1.

If you valued only the cash loss (and discounted the inconvenience of getting a ticket), the benefit of parking illegally would be -£20. The value of the act, park illegally, would be:

Exp(I) = (0.1)(-£20) + (0.9)(£0) = -£2.

There are no uncertainties about parking in the lot. If you valued only the cash expense, the expected value of the act, park in the lot is -£3.

What to do?

The value rule gives clear advice. Act so as to maximise expected value. The highest value you can expect is -£2. So, park illegally. That is not the end of the matter. Some people think it is wrong to break the law-any law. Other people might think the law is not so important, but it is wrong to harm other people, for example, by taking the parking place of a local resident.

Finally, its is time to answer the question we started with: Which lock should you buy?

Example: Bicycle
Act K: Buy a Kryptonite lock.
Act L: Buy a cheap chain and lock.
Consequence S: your bike is stolen.
Consequence N: your bike is not stolen.

If you bike gets stolen you have a cost of -£80. This is the cost of the bike. You have already spent your money, so the further benefit of not losing your bike is 0.

$$\text{Exp(K)} = [P(\text{S given K})][-£80 - £20] + [P(\text{N given K})][0 - £20] = 0.18(-£100) + 0.82(-£20) = -£34.4$$

$$\text{Exp(L)} = [P(\text{S given L})][-£80 - £8] + [P(\text{N given L})][0 - £8] = 0.57(-£88) + 0.43(-£8) = -£53.6.$$

So if you want to maximise value, you should buy the Kryptonite lock.

Generalisation of Expected Value Rule

The Expected Value Rule was given for an act with only two possible consequences. This can be expended as follows. Consider an Act with n possible consequences:
Act: A
Consequences: C1, C2, . . . , Cn

Utilities: U1, U2, . . . , Un

Probabilities: P(C1 given A), P(C2 given A), . . . , P(Cn given A)

$$\text{Expected Value of A} = \text{Exp(A)} = \sum_{i=1}^{n} [\text{P}(Ci \text{ given } A)][Ui]$$

Setting up and solving a decision problem [2]

[2] Andrew Gelman and Deborah Nolan. *Teaching statistics: A bag of tricks*. Oxford University Press, 2017

The process of setting up and solving a decision problem can be broken down into 6 steps.

1. Goals and decision options. The first step is to define your goals. Question to answer: "What do I want to achieve"?. In our motivating example (page 7), the goal was to decide which lock to buy.

2. Define the possible acts, the consequences of each act and their utilities. In our motivating example, the possible acts were "buy a Kryptonite lock" or "buy a cheap chain". The consequences were "the bike is stolen" or "your bike is not stolen". If the bike was stolen there is no gain but a loss of -£80 (the price of the bike) and a further loss of -£20 (if you had bought the Kryptonite lock) or -£8 (if you had bought the cheap lock). The gain of not losing the bike is 0.

3. Set up a tree diagram (like the ones in Figures 1 and 3).

4. Use the tree diagram to calculate probabilities. For this step you need to collect some data.

5. Evaluate the tree. Now you can solve the decision problem. You have all the ingredients.

6. Understanding the result. What is the recommended decision? Does it make sense?

An example using these steps follows.

Example: Gardening

You have a garden at your home. You really like gardening and decide to plant potatoes and carrots. Unfortunately, you don't have enough space for both. You have to pick one. First you need to buy the seeds. The potato seeds cost £10. The carrot seeds cost £8. You also need to water the plants: this will cost you about £20 (from the time you plant the seeds until you start harvesting). There is another problem: winter is coming and the plants may not survive. You decide to consult the internet and find a research project where the scientists compared potatoes' and carrots' resistance to cold. This is what they found: About 60% of plants die during winter. Of those that die, 30% are potatoes and 70% carrots. Of those that survive, 60% are potatoes and 40% are carrots. Now based on this information, which one one do you choose to plant?

1. The first step is to define the goal and the decision options. This is straightforward, the goal is to choose between planting potatoes or carrots.

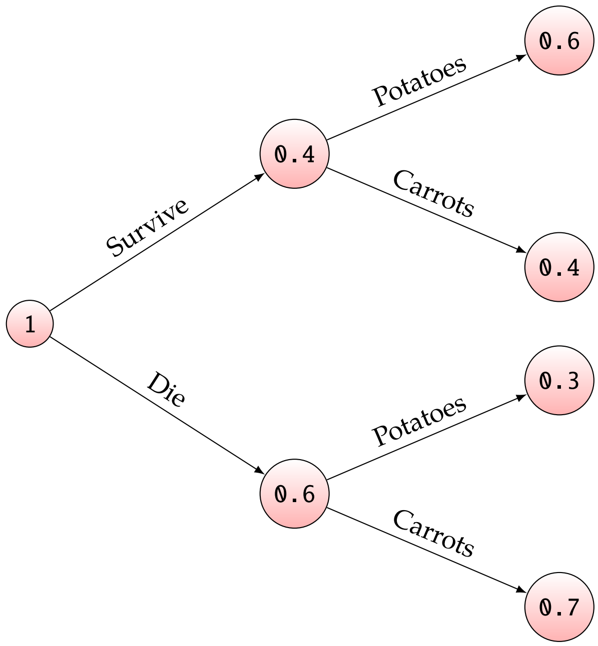
2. Step 2 requires us to define the possible acts, the consequences of each act and their gains.

   Acts:
   1. Plant potatoes.
   2. Plant carrots.

   Consequences:
   1. Plant dies during the winter.
   2. Plant survives.

   Gains:
   1. If the potatoes die you lose the £10 and the watering costs of £20.
   2. If the carrots die you lose £8 and the watering costs of £20.
   3. If either survive you have "lost" only the watering cost of £20.

3. For step 3 we need to draw a tree diagram:



4. For the 4th step we need to calculate the conditional probabilities. If we plant potatoes we need to know the probabilities that they die and survive. We need to know the *P*(Survive given Potatoes) and the *P*(Die given Potatoes). Similarly, if we plant carrots we want *P*(Survive given Carrots) and the *P*(Die given Carrots).

   Based on Bayes Rule

$$P(\text{Survive given Potatoes}) = \frac{P(\text{Survive})P(\text{Potatoes given Survive})}{P(\text{Survive})P(\text{Potatoes given Survive}) + P(\text{Die})P(\text{Potatoes given Die})}.$$

   We already have all the pieces from the tree diagram):

   - *P*(Survive) is the probability that a plant survives, this is 0.4.
   - *P*(Potatoes given Survive) is the probability that a plant that survives is potatoes: this is 0.6.
   - *P*(Dies) = 0.6.
   - *P*(Potato given Die) is the probability that a plant that dies is

potatoes: this is 0.3. So,

$$P(\text{Survive given Potatoes}) = \frac{0.4 \times 0.6}{0.4 \times 0.6 + 0.6 \times 0.3} = 0.57$$

In a similar fashion we calculate the other probabilities.

- P(Die given Potatoes) = 0.43
- P(Survive given Carrots) = 0.28
- P(Die given Carrots) = 0.72

5. Now we can calculate the Expected Value of each act:

- Exp(Plant Potatoes) = [P(Survive given Potatoes)][-£20] + [P(Die given Potatoes)][-£20-£10] = [0.57] [-£20] + [0.43] [-£20-£10] = -£24.3

- Exp(Plant Carrots) = [P(Survive given Carrots)][-£20] + [P(Die given Carrots)][-£20-£8] = [0.28][-£20] + [0.72] [-£20-£8] = -£25.8

The Value Rule says choose the act with the highest expected value. So plant potatoes!

6. Discussion: This is a somewhat expected result because the costs are quite similar. We need to spend £10 for potato seeds and £8 for carrot seeds. So we should expect our result to be driven mainly by the conditional probabilities. In fact, this was the case. The carrots are much more likely to die than the potatoes.

Exercise 4

1. Perform the calculations from the gardening example and verify you get the same results. More specifically, calculate:
    - P(Die given Potatoes).

- P(Survive given Carrots).
- P(Die given Carrots).

2. Louis plans to set up a new ice-cream business by the seaside. But where should he set up his first stall? If he sets up shop at the Beach Cafe (C), the weather won't matter. On the other hand, if he puts his stall on the beach (B), he'll do much better if it's sunny, but much worse if it rains (fewer people go to the beach when it rains). Based on the previous years' weather the probability of rain is 0.3. Both locations are equally attractive to Luis. In other words, P(C) = P(B) = 0.5. This means that Luis has no preference between B and C initially. Now decide where Louis should put his first ice-cream stall. The table below tells you how many pounds he can expect to make per day in each location:

	Rain	Sunny
Beach, B	88	268
Beach cafe, C	100	100

3. For Exercise 3 on page 11. Let $U(x)$ denote the patient's utility function, where x is the number of months to live. Assuming that $U(12) = 1.0$ and $U(0) = 0$, how low can the patient's utility for living 3 months be and still have the operation be preferred?
4. A state lottery sells tickets for a cost of £1 each. The ticket has a probability of $1/(2{,}400{,}000)$ of winning £1,000,000, and otherwise nothing.
   1. What is the expected profit of the state from each ticket sold?
   2. In the hope of increasing profits, the state considers to increase the award to £2,000,000 and to reduce the probability of winning to $1/(4{,}800{,}000)$. Do you think it's worth the trouble?

Bibliography

Joseph K Blitzstein and Jessica Hwang. *Introduction to probability*. Chapman and Hall/CRC, 2014.

Andrew Gelman and Deborah Nolan. *Teaching statistics: A bag of tricks*. Oxford University Press, 2017.

Ian Hacking. *An Introduction to Probability and Inductive Logic Desk Examination Edition*. Cambridge University Press, 2001.

Sheldon Ross. *A First Course in Probability 8th Edition*. Pearson, 2009.

Murray R Spiegel, John J Schiller, and R Srinivasan. *Probability and statistics*. New York: McGraw-Hill,, 2013.

Appendix

This is a more mathematically rigorous overview of probability. We only present concepts relevant to this course. We introduce the concept of the probability of an event and then show how probabilities can be computed in certain situations. We introduce one of the most important concepts in probability theory, that of conditional probability. The importance of this concept is twofold. In the first place, we are often interested in calculating probabilities when some partial information concerning the result of an experiment is available; in such a situation, the desired probabilities are conditional. Second, even when no partial information is available, conditional probabilities can often be used to compute the desired probabilities more easily. We then present Bayes rule from a slightly non conventional point of view. We finish with the notion of independent events. As a preliminary, however, we need the concept of the sample space and the events of an experiment.

The sample space and events

Consider an experiment whose outcome is not predictable with certainty. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by S.

For example, if the outcome of an experiment consists in the determination of the sex of a newborn child, then

$$S = \{g, b\}$$

where the outcome g means that the child is a girl and b that it is a boy.

Any subset E of the sample space is known as an *event*. In other words, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred. In the preceding example, if $E = \{g\}$, then E is the event that the child is a girl. Similarly, if $F = \{b\}$, then F is the event that the child is a boy.

Finally, for any event E, we define the new event E^c, referred to as the complement of E, to consist of all outcomes in the sample space S that are not in E. That is, E^c will occur if and only if E does not occur. In our example, $E^c = \{b\}$. The only outcome that is not in $E = \{g\}$ is the event the child is a boy.

Axioms Of Probability

Consider an experiment whose sample space is S. For each event E of the sample space S, we define $P(E)$ as the probability of the event E. (P is called a probability function). We assume that $P(E)$ satisfies the following three axioms.

1. $0 \leq P(E) \leq 1$

2. $P(S) = 1$

3. For any number of mutually exclusive events $E_1, E_2, \ldots,$

   $$P(E_1 \text{ or } E_2 \text{ or } \ldots) = P(E_1) + P(E_2) + \cdots = \sum_{i=1}^{\infty} P(E_i)$$

   Mutually exclusive means that only one of the events can be true (see also side note next page). In particular, for two mutually exclusive events E_1, E_2,

   $$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$$

Axiom 1 states that the probability that the outcome of the experiment is an outcome in E is some number between 0 and 1. Axiom 2 states that, with probability 1, the outcome will be a point in the sample space S. Axiom 3 states that, for any sequence of mutually exclusive events, the probability of at least one of these events occurring is just the sum of their respective probabilities.

For instance, if our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we would have

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

On the other hand, if the coin were biased and we felt that a head were twice as likely to appear as a tail, then we would have

$$P(\{H\}) = \frac{2}{3}; P(\{T\}) = \frac{1}{3}$$

Some simple propositions

In this section, we give some simple propositions regarding probabilities.

1. $$P(E^c) = 1 - P(E)$$

   In words, the probability that an event does not occur is 1 minus the probability that it does occur. For instance, if the probability of obtaining a head on the toss of a coin is 3/8, then the probability of obtaining a tail must be 5/8.

2. $$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

   In words, this proposition gives the relationship between the probability of the *union* of two events, expressed in terms of the individual probabilities, and the probability of the *intersection* of the events.

An example: J is taking two books along on her holiday vacation. With probability 0.5, she will like the first book; with probability 0.4, she will like the second book; and with probability 0.3, she will like both books. What is the probability that she likes neither book?

Let B_1 denote the event that J likes book 1, and B_2 the event that J likes book 2. Then the probability that she likes at least one of the books is

$$P(B_1 \text{ or } B_2) = P(B_1) + P(B_2) - P(B_1 \text{ and } B_2) = 0.6$$

For any two events E and F of a sample space S, we define the new event (E or F) to consist of all outcomes that are either in E or in F or in both E and F. That is, the event (E or F) will occur if either E or F occurs. For instance, for the sex of a child example, if event $E = \{g\}$ and $F = \{b\}$, then

$$(E \text{ or } F) = \{g, b\}$$

The event (E or F) is called the *union* of the event E and the event F.

Similarly, for any two events E and F, we may also define the new event (E and F), called the *intersection* of E and F, to consist of all outcomes that are both in E and in F. That is, the event (E and F) will occur only if both E and F occur. For instance, for the sex of a child example

$$(E \text{ and } F) = \text{null}$$

This means that E and F does not contain any outcomes and hence could not occur. A child cannot be a boy and a girl at the same time (for our purposes). We call this the null event. If (E and F) = null, then E and F are said to be mutually exclusive.

Because the event that J likes neither book is the complement of the event that she likes at least one of them, we obtain the result

$$P(B_1^c \text{ and } B_2^c) = P((B_1 \text{ or } B_2)^c) = 1 - P(B_1 \text{ or } B_2) = 0.4$$

The second equality comes from DeMorgan's laws. They provide useful relationships between the three basic operations of forming unions, intersections, and complements. Imagine two event E_1 and E_2 then

$$(E_1 \text{ or } E_2)^c = E_1^c \text{ and } E_2^c$$

and

$$(E_1 \text{ and } E_2)^c = E_1^c \text{ or } E_2^c$$

These can be generalised to any number of events.

Conditional probabilities

Here we give the definition. Intuition behind it is given in chapter 2. Let E and F denote, two events. If $P(F) > 0$, then

$$P(E \text{ given } F) = \frac{P(E \text{ and } F)}{P(F)} \tag{2}$$

It is read as the (conditional) probability that E occurs given that F has occurred. Let's see an example.

A student is taking a one-hour-time-limit makeup examination. Suppose the probability that the student will finish the exam in less than x hours is $x/2$, for all $0 \leq x \leq 1$. Then, given that the student is still working after 0.75 hour, what is the conditional probability that the full hour is used?

Let L_x denote the event that the student finishes the exam in less than x hours, $0 \leq x \leq 1$, and let F be the event that the student uses the full hour. Because F is the event that the student is not finished in less than 1 hour,

$$P(F) = P(L_1^c) = 1 - P(L_1) = 0.5$$

Now, the event that the student is still working at time 0.75 is the complement of the event $L_{0.75}$, so the desired probability is obtained from

$$P(F \text{ give } L_{0.75}^c) = \frac{P(F \text{ and } L_{0.75}^c)}{P(L_{0.75}^c)} = \frac{P(F)}{1 - P(L_{0.75})} = 0.8$$

A useful expression for the probability of the intersection of an arbitrary number of events, is sometimes referred to as the *multiplication rule*.

$$P(E_1 \text{ and } E_2 \text{ and } \ldots \text{ and } E_n) = P(E_1)P(E_2 \text{ given } E_1)P(E_3 \text{ given } E_1 \text{ and } E_2)\ldots P(E_n \text{ given } E_1 \ldots E_{n-1})$$

In particular, for two events E and F,

$$P(E \text{ and } F) = P(F)P(E \text{ given } F)$$

This comes from multiplying both sides of Equation (2) by $P(F)$. In words, it states that the probability that both E and F occur is equal to the probability that F occurs multiplied by the conditional probability of E given that F occurred. It is often quite useful in computing the probability of the intersection of events.

Bayes Formula

Let E and F be events. We may express E as

$$E = [(E \text{ and} F) \text{ or } (E \text{ and } F^c)]$$

for, in order for an outcome to be in E, it must either be in both E and F or be in E but not in F. As $(E \text{ and } F)$ and $(E \text{ and } F^c)$ are clearly mutually exclusive, we have, by Axiom 3,

$$P(E) = P(E \text{ and } F) + P(E \text{ and } F^c)$$

$$= P(E \text{ given } F)P(F) + P(E \text{ given } F^c)P(F^c)$$

$$= P(E \text{ given } F)P(F) + P(E \text{ given } F^c)[1 - P(F)]$$

In words, it states that the probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred—each conditional probability being given as much weight as the event on which it is conditioned has of occurring. This is an extremely useful formula, because its use often enables us to determine the probability of an event by first "conditioning" upon whether or not some second event has occurred. That is, there are many instances in which it is difficult to compute the probability of an event directly, but it is straightforward to compute it once we know whether or not some second event has occurred. We illustrate this idea with two examples.

1^{st} example: An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas this probability decreases to .02 for a person who is not accident prone. If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

We shall obtain the desired probability by first conditioning upon whether or not the policyholder is accident prone. Let A_1 denote the event that the policyholder will have an accident within a year of purchasing the policy, and let A denote the event that the policyholder is accident prone. Hence, the desired probability is given by

$$P(A_1) = P(A_1 \text{ given } A)P(A) + P(A_1 \text{ given } A^c)P(A^c) = (0.4)(0.3) + (0.2)(0.7) = 0.26$$

2^{nd} example: At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose, however, that a new piece of evidence which shows that the criminal has a certain characteristic (such as left-handedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?

Letting G denote the event that the suspect is guilty and C the event that he possesses the characteristic of the criminal, we have

$$P(G \text{ given } C) = \frac{P(G \text{ and } C)}{P(C)}$$

$$= \frac{P(C \text{ given } G)P(G)}{P(C \text{ given } G)P(G) + P(C \text{ given } G^c)P(G^c)}$$

$$= 1(0.6)1(0.6) + (0.2)(0.4) = 0.882$$

Essentially the second equality is Bayes Rule as given in chapter 3. It should be obvious by the construction that Bayes Rule is nothing more than a reformulation of the definition of conditional probability.

where we have supposed that the probability of the suspect having the characteristic if he is, in fact, innocent is equal to 0.2, the proportion of the population possessing the characteristic.

Independent events

The previous examples of this chapter show that $P(E \text{ given } F)$, the conditional probability of E given F, is not generally equal to $P(E)$, the unconditional probability of E. In other words, knowing that F has occurred generally changes the chances of E's occurrence. In the special cases where $P(E \text{ given } F)$ does in fact equal $P(E)$, we say that E is independent of F. That is, E is independent of F if knowledge that F has occurred does not change the probability that E occurs. Since

$$P(E \text{ given } F) = \frac{P(E \text{ and } F)}{P(F)},$$

it follows that E is independent of F if

$$P(E \text{ and } F) = P(E)P(F) \tag{3}$$

The fact that this equation (3) is symmetric in E and F shows that whenever E is independent of F, F is also independent of E. We thus have the following definition: Two events E and F are said to be independent if equation (3) holds. Two events E and F that are not independent are said to be dependent. This can be generalised to more than two events.

Example: A card is selected at random from an ordinary deck of 52 playing cards. If E is the event that the selected card is an ace and F is the event that it is a spade, then E and F are independent. This follows because $P(E \text{ and } F) = \frac{1}{52}$, whereas $P(E) = \frac{4}{52}$ and $P(F) = \frac{13}{52}$.

Conditional probabilities satisfy all of the properties of ordinary probabilities. So $P(E \text{ given } F)$ satisfies the three axioms of a probability.

1. $0 \leq P(E \text{ given } F) \leq 1$
2. $P(S \text{ given } F) = 1$
3. For any numner of mutually exclusice events $E_1, E_2, \ldots$

$$P((E_1 \text{ given } F) \text{ or } (E_2 \text{ give } F), \text{ or } \ldots) = \sum_{i=1}^{\infty} P(E_i \text{ given } F)$$

Further reading

For a more detailed treatment of probability we refer to chapters 1-3 from [3]. Another excellent textbook is by [4] (chapters 1 and 2). The book of [5] is less thorough but useful because it contains many solved exercises.

[3] Sheldon Ross. *A First Course in Probability 8th Edition*. Pearson, 2009
[4] Joseph K Blitzstein and Jessica Hwang. *Introduction to probability*. Chapman and Hall/CRC, 2014
[5] Murray R Spiegel, John J Schiller, and R Srinivasan. *Probability and statistics*. New York: McGraw-Hill,, 2013

Exercise 5

1. A box contains 3 marbles: 1 red, 1 green, and 1 blue. Consider an experiment that consists of taking 1 marble from the box and then replacing it in the box and drawing a second marble from the box. Describe the sample space. Repeat when the second marble is drawn without replacing the first marble.
2. A total of 28 percent of American males smoke cigarettes, 7 percent smoke cigars, and 5 percent smoke both cigars and cigarettes.
    1. What percentage of males smokes neither cigars nor cigarettes?
    2. What percentage smokes cigars but not cigarettes?
3. Suppose that you are playing blackjack against a dealer. In a freshly shuffled deck, what is the probability that neither you nor the dealer is dealt a blackjack?
4. Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?
5. Two cards are randomly chosen without replacement from an ordinary deck of 52 cards. Let B be the event that both cards are aces, let As be the event that the ace of spades is chosen, and let A be the event that at least one ace is chosen. Find
    1. $P(B \text{ given } As)$
    2. $P(B \text{ given } A)$

6. An ectopic pregnancy is twice as likely to develop when the pregnant woman is a smoker as it is when she is a nonsmoker. If 32 percent of women of childbearing age are smokers, what percentage of women having ectopic pregnancies are smokers?
7. Ninety-eight percent of all babies survive delivery. However, 15 percent of all births involve Cesarean (C) sections, and when a C section is performed, the baby survives 96 percent of the time. If a randomly chosen pregnant woman does not have a C section, what is the probability that her baby survives?
8. Suppose that each child born to a couple is equally likely to be a boy or a girl, independently of the sex distribution of the other children in the family. For a couple having 5 children, compute the probabilities of the following events:
    1. All children are of the same sex.
    2. The 3 eldest are boys and the others girls.
    3. Exactly 3 are boys.
    4. The 2 oldest are girls.
    5. There is at least 1 girl.