9.3: Basic Probability Theory

Ideological arguments between Bayesians and frequentists notwithstanding, it turns out that people mostly agree on the rules that probabilities should obey. There are lots of different ways of arriving at these rules. The most commonly used approach is based on the work of Andrey Kolmogorov, one of the great Soviet mathematicians of the 20th century. I won't go into a lot of detail, but I'll try to give you a bit of a sense of how it works. And in order to do so, I'm going to have to talk about my pants.

9.3.1 Introducing probability distributions

One of the disturbing truths about my life is that I only own 5 pairs of pants: three pairs of jeans, the bottom half of a suit, and a pair of tracksuit pants. Even sadder, I've given them names: I call them X_1, X_2, X_3, X_4 and X_5. I really do: that's why they call me Mister Imaginative. Now, on any given day, I pick out exactly one of pair of pants to wear. Not even I'm so stupid as to try to wear two pairs of pants, and thanks to years of training I never go outside without wearing pants anymore. If I were to describe this situation using the language of probability theory, I would refer to each pair of pants (i.e., each X) as an ***elementary event***. The key characteristic of elementary events is that every time we make an observation (e.g., every time I put on a pair of pants), then the outcome will be one and only one of these events. Like I said, these days I always wear exactly one pair of pants, so my pants satisfy this constraint. Similarly, the set of all possible events is called a ***sample space***. Granted, some people would call it a "wardrobe", but that's because they're refusing to think about my pants in probabilistic terms. Sad.

Okay, now that we have a sample space (a wardrobe), which is built from lots of possible elementary events (pants), what we want to do is assign a ***probability*** of one of these elementary events. For an event X, the probability of that event P(X) is a number that lies between 0 and 1. The bigger the value of P(X), the more likely the event is to occur. So, for example, if P(X)=0, it means the event X is impossible (i.e., I never wear those pants). On the other hand, if P(X)=1 it means that event X is certain to occur (i.e., I always wear those pants). For probability values in the middle, it means that I sometimes wear those pants. For instance, if P(X)=0.5 it means that I wear those pants half of the time.

At this point, we're almost done. The last thing we need to recognise is that "something always happens". Every time I put on pants, I really do end up wearing pants (crazy, right?). What this somewhat trite statement means, in probabilistic terms, is that the probabilities of the elementary events need to add up to 1. This is known as the ***law of total probability***, not that any of us really care. More importantly, if these requirements are satisfied, then what we have is a ***probability distribution***. For example, this is an example of a probability distribution

Which.pants	Blue.jeans	Grey.jeans	Black.jeans	Black.suit	Blue.tracksuit
Label	X_1	X_2	X_3	X_4	X_5
Probability	$P(X_1)=.5$	$P(X_2)=.3$	$P(X_3)=.1$	$P(X_4)=0$	$P(X_5)=.1$

Each of the events has a probability that lies between 0 and 1, and if we add up the probability of all events, they sum to 1. Awesome. We can even draw a nice bar graph (see Section 6.7) to visualise this distribution, as shown in Figure **??**. And at this point, we've all achieved something. You've learned what a probability distribution is, and I've finally managed to find a way to create a graph that focuses entirely on my pants. Everyone wins!

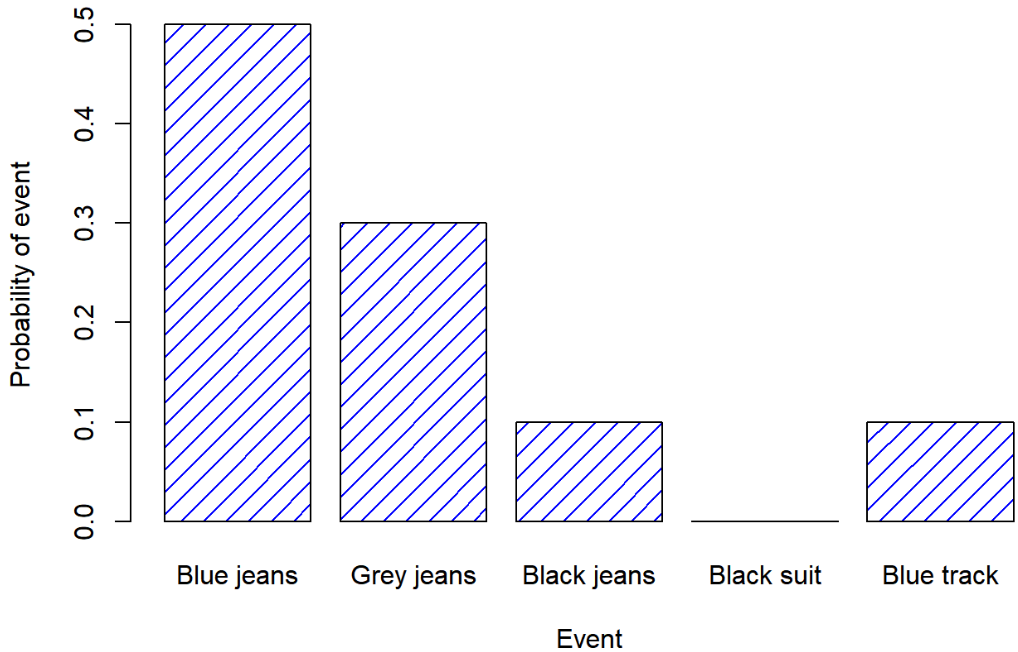

Figure 9.2: A visual depiction of the “pants” probability distribution. There are five “elementary events”, corresponding to the five pairs of pants that I own. Each event has some probability of occurring: this probability is a number between 0 to 1. The sum of these probabilities is 1.

The only other thing that I need to point out is that probability theory allows you to talk about ***non elementary events*** as well as elementary ones. The easiest way to illustrate the concept is with an example. In the pants example, it’s perfectly legitimate to refer to the probability that I wear jeans. In this scenario, the “Dan wears jeans” event said to have happened as long as the elementary event that actually did occur is one of the appropriate ones; in this case “blue jeans”, “black jeans” or “grey jeans”. In mathematical terms, we defined the “jeans” event E to correspond to the set of elementary events (X_1,X_2,X_3). If any of these elementary events occurs, then E is also said to have occurred. Having decided to write down the definition of the E this way, it’s pretty straightforward to state what the probability P(E) is: we just add everything up. In this particular case

$$P(E)=P(X_1)+P(X_2)+P(X_3)$$

and, since the probabilities of blue, grey and black jeans respectively are .5, .3 and .1, the probability that I wear jeans is equal to .9.

At this point you might be thinking that this is all terribly obvious and simple and you’d be right. All we’ve really done is wrap some basic mathematics around a few common sense intuitions. However, from these simple beginnings it’s possible to construct some extremely powerful mathematical tools. I’m definitely not going to go into the details in this book, but what I will do is list – in Table 9.1 – some of the other rules that probabilities satisfy. These rules can be derived from the simple assumptions that I’ve outlined above, but since we don’t actually use these rules for anything in this book, I won’t do so here.

Table 9.1: Some basic rules that probabilities must satisfy. You don’t really need to know these rules in order to understand the analyses that we’ll talk about later in the book, but they are important if you want to understand probability theory a bit more deeply.

English	Notation	NANA	Formula
Not A	$P(\neg A)$	=	$1-P(A)$
A or B	$P(A\cup B)$	=	$P(A)+P(B)-P(A\cap B)$
A and B	$P(A\cap B)$	=	$P(A\|B)P(B)$