Cambridge University Press
978-1-108-48053-6 — The Probability Companion for Engineering and Computer Science
Adam Prügel-Bennett
Excerpt
More Information

1

Introduction

Contents

This book is a survey of the mathematical tools and techniques in probability and is aimed primarily at scientists and engineers. The intention is to give a broad overview of the subject, starting from the very basics but covering techniques used at the forefront of research in probabilistic modelling. Before we get to some of the more advanced tools it is necessary to understand the language of probability and some of the foundational concepts.

This chapter sets up the mathematical language we need in order to discuss probabilities and their properties. Most of this consists of definitions and simple mathematics, but it is a prerequisite for talking about the more interesting tools that we meet in later chapters.

1.1 Why Probabilities?

We live in a world full of uncertainties. The toss of a coin is sufficiently uncertain that it is regularly used to decide who goes first in many sporting competitions. Yet even with coin tosses we can make strong predictions. Thus if we toss a coin 1000 times with overwhelming probability, we are likely to get between 450 and 550 heads. The mathematical language that allows us to reason under uncertainty is probability theory. This book aims to provide a broad-brush overview of the mathematical and computational tools used by engineers and scientists to make sense of uncertainties.

In some situations what we observe is the consequence of so many unobserved and uncertain events that we can make extremely precise predictions that are taken as laws of physics even though they are just statements about what is overwhelmingly probable. The field of statistical physics (aka statistical mechanics) is founded on probability. However, uncertainty is ubiquitous and often does not involve a sufficient number of events to enable precise predictions. In these situations probability theory can be important in understanding experiments and making predictions. For example, if we wish to distinguish between natural fluctuations in the weather and the effects of climate change it is vital that we can reason accurately about uncertainty. However, probability can only answer these pressing scientific questions in combination with accurate models of the world. Such models are the subject matter of the scientific disciplines. Probability theory acts as a unifying glue, allowing us to make the best possible predictions or extract the most amount of information from observations. Although probability is not a prerequisite for doing good science, in almost any discipline in science, engineering, or social science it enhances a practitioner's armoury. I hope to give a spirit of the range of applications through examples sprinkled across the text.

Becoming a researcher in any field involves developing a toolkit of techniques that can be brought out as needed to tackle new problems. To be an accomplished user, the researcher has to acquire experience through practical application of the tools. This text cannot replace that step; rather, its intention is to make new researchers aware of what probabilistic tools exist and provide enough intuition to be able to judge the usefulness of the tool. In many ways this text is my personal compilation of tricks I've learned over many years of probabilistic modelling. The subject, and consequently this book, is mathematical, and in places I go into detail, but I recommend that you skip sections when you are getting bogged down or feel you have to push on even though you don't understand everything. This is a high-level tour; when there is a technique you really want to use you can come back and spend the time necessary to master that technique.

We start slowly by carefully, defining the key concepts we use and point out possible misunderstandings. Apologies to those who find this too elementary; however, we will quickly get into more advanced material. Without any more fuss let's get started.

1.2 Events and Probabilities

It is useful to know the mathematical language and formalism of probability. There are two main reasons for this: firstly, it allows you to read and understand the literature; secondly, when you write papers or your thesis, it is necessary to be able to talk the talk. For example, if you have a quantity you are treating as a random variable, you should call it a random variable, but this also requires you to know precisely what is meant by the term.

1.2.1 Events

The standard mathematical formulation of probability considers a set of *elementary events*, Ω, consisting of all possible outcomes in the world we are considering. For example, we might want to model the situation of tossing a coin, in which case the set of outcomes are $\Omega = \{\text{heads, tails}\}$, or if we roll a dice the elementary events would be $\Omega = \{1, 2, 3, 4, 5, 6\}$. We take an *event*, A, to be a subset of the space of elementary events $A \subset \Omega$ (note that there is a distinction made between the terms 'elementary event' and 'event', although an event could be an elementary event). In rolling a dice, the event, A, might be the dice landing on a six, $A = \{6\}$, or a number greater than 3, $A = \{4, 5, 6\}$. The probability of an event is denoted by $\mathbb{P}\left(A\right)$. Probabilities take values between zero and one

$$0 \leq \mathbb{P}\left(A\right) \leq 1,$$

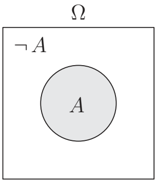
with the interpretation that $\mathbb{P}\left(A\right) = 0$ means that the event never occurs and $\mathbb{P}\left(A\right) = 1$ meaning that the event always occurs. In this *set theory* view of probabilities the probability of no event occurring is 0, i.e. $\mathbb{P}\left(\emptyset\right) = 0$ where $\emptyset = \{\}$ is the empty set. In contrast, one elementary event must happen so that $\mathbb{P}\left(\Omega\right) = 1$. For a fair coin we expect $\mathbb{P}\left(\{\text{head}\}\right) = \mathbb{P}\left(\{\text{tail}\}\right) = 1/2$.

We use the standard notation of $\cup$ and $\cap$ to denote union and intersection or sets and $\vee$ and $\wedge$ to denote 'logical or' and 'logical and'.

Talking about events gets us immediately into a linguistic dilemma. We can either consider an event to be a set of elementary events (the *set theory* point of view), or we can take it as a true–false proposition (the *propositional logic* point of view). Thus, when talking about a pair of events, A and B, we can view the event, C, of both A and B occurring as a set theoretic statement

$$C = \left\{\omega \middle| \omega \in A \text{ and } \omega \in B\right\} = A \cap B$$

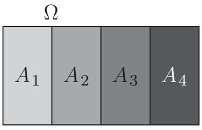
or alternatively as a logical statement $C = A \wedge B$ about predicates (both A and B are true). The event, D, that either A or B (or possibly both) are true can be viewed as a set theoretic statement, $D = A \cup B$, or as a propositional statement, $D = A \vee B$. Similarly the event, E, that A does not occur can either be written in set language as $E = A^c = \Omega - A = \{\omega | \omega \notin A\}$ or as the logical statement $E = \neg A$. Both languages have advantages. The set theoretic language makes many simple results in probability transparent that are more obscure when using the language of propositional logic. However, often when modelling a system it is much easier to think in terms of propositions. In this chapter we will tend to migrate from a language of set theory to the language of propositions.



Returning to the axioms of probability, denoting the event 'A does not occur' by $\neg A$ then

$$\mathbb{P}\left(A\right) + \mathbb{P}\left(\neg A\right) = \mathbb{P}\left(\Omega\right) = 1$$

with the intuitively clear meaning that the event will either occur or not (a coin is either a head or not a head, or a dice is either a six or not a six). If we consider a set of *exhaustive* and *mutually exclusive* events, $\{A_i | i \in I\}$, where $I \in \mathbb{N}$ is an index set (that is a set of integers that label the events) then



$$\sum_{i \in I} \mathbb{P}\left(A_i\right) = 1.$$

$\bigcup_{i \in I} A_i = A_1 \cup A_2 \cup \cdots \cup A_{|I|}$

Here *exhaustive* means that we have covered every possible outcome (i.e. $\bigcup_{i \in I} A_i = \Omega$) and *mutually exclusive* means that $A_i \cap A_j = \emptyset$ for all distinct pairs i and j. In the example of the dice, the events $\{1, 6\}$, $\{2, 5\}$, and $\{3, 4\}$ form an exhaustive and mutually exclusive set of events. When we roll a dice one of these events will occur. Note that real coins and real dice behave differently from mathematicians' coins and dice. A real coin might land on its edge, or it might roll away and get lost. Probability, like all applied mathematics, is a *model* of reality. It is the responsibility of the user of probability theory to ensure that their model captures the features that they care about.

Many mathematical texts formalise probabilities in terms of a probability space, consisting of a state space (or set of elementary events), Ω, a family of all possible events, $\mathcal{A}$, which will frequently be the set of all subsets of elementary events ($\mathcal{A} = 2^{\Omega}$, i.e. the power set of Ω) and a probability, $\mathbb{P}\left(A\right)$, associated with each event $A \in \mathcal{A}$. Thus a formal way of referring to probabilities is as a triple $(\Omega, \mathcal{A}, \mathbb{P})$. Don't be put off: this is just how mathematicians like to set things up.

What happens when the set of events are not denumerable? This would occur if the events took a continuous value, for example, *what will the temperature be tomorrow?*, or *where does a dart land?* This leads to a difficulty: the probability of any elementary event may well be zero! Worse, the family of all events, $\mathcal{A}$, can potentially become precarious, as it involves subsets of a non-denumerable sets. In over 30 years of working with probabilities I am yet to meet a case where anything precarious actually happened. To rigorously formulate probabilities in a way to avoid contradictions, even when working with the most complex of sets Andrei Kolmogorov borrowed ideas from mathematical analysis known as measure theory. If you pick up a mathematics text on probability you will get a good dose of measures (sigma), σ-fields (a generalisation of power sets), filtrations (a hierarchy of events), etc. However, don't panic. This formalism is massive overkill for nearly all situations. In most of engineering or science you will never face the pathological functions that keep mathematicians awake at night and require measure theory. I have never come across a practical application where the mechanics of measure theory was at all necessary. This is not to put down the importance of putting probability on a firm theoretical footing. However, in my experience, measure theory is neither necessary nor even helpful in applying probability to real-world problems. If you want to know more about measure theory and the type of problems which necessitate it refer to Appendix 1.A.

1.2.2 Assigning Probabilities

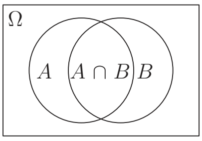
In the mathematical set-up above probabilities must have certain properties, but they are assumed to be given. One of the first tasks for engineers and scientist is to assign meaningful probabilities. How then do we do this and what do mean by probability? These questions have raised considerable debate among the philosophically minded. A seeming common-sense answer would be that it is the expected frequency of occurrence of an outcome in a large number of independent trials. Indeed, some have argued that is the only rational way of viewing probabilities. However, doubters have pointed out that there are many problems with uncertainty that are never repeated (who will win next year's Wimbledon final?). Thus, the argument goes that probabilities should be viewed as our degree of belief about something happening.

This philosophical debate, however, throws little light onto the practical question of how we should assign probabilities (however, for those interested we return to this debate in Chapter 8). Suppose we want to assign a probability to the outcome of a coin toss. If the coin has two distinct sides, then I, like most people, would happily assume that it has equal probability of being either a head or tail. Pushed on why, my first response would be that I believe that I would get heads as often as tails (a frequentist's explanation). Pushed further, I would argue that the outcome is likely to depend mainly on the speed of rotation and the time the coin has to rotate, which is beyond most people's ability to control precisely. I would find it very unlikely that the design on the faces of the coin would significantly bias the outcome. Pushed still further, I might resort to the conservation of angle momentum and small amount of air resistance or perhaps I might just shrug my shoulders and say 'that's my model of the world and I'm happy with it'. Of course, it would be possible to determine empirically the result of many coin tosses. Although I have never done this, the fact that it's not common knowledge whether heads is more likely than tails or the other way around suggests that the probability is indeed close to a half. In practice, probabilities are often assigned using a symmetry argument. That is, all outcomes at some level are considered equally likely.

Alas, one of the drawbacks of this is that the initial task of allocating probabilities often comes down to counting possible outcomes (combinatorics). Many people consequently view probability as hard and maybe even boring – personally I find combinatorics fascinating and beautiful, although I concede that it is an acquired taste. It is certainly true that combinatorics quickly becomes difficult and it is very easy to get wrong. However, it is only a very small part of probability. We are about to see that manipulating probabilities is actually relatively straightforward, and I hope this book will convince the reader that there is much, much more to probability than just counting combinations.

1.2.3 Joint and Conditional Probabilities

Probabilities of single events are somewhat boring. There is little to say about them. The interest comes when we have two or more events. To reason about this we need to set up a formalism for calculating with multiple events: a calculus of probability. It turns out that once you understand how to handle two events then the generalisation to more events is simple.



The *joint probability* of two events, A and B, both occurring is denoted $\mathbb{P}(A, B)$. If we think of events $A \subset \Omega$ and $B \subset \Omega$ being subsets of elementary events, then

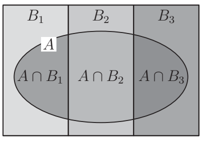
$$\mathbb{P}(A, B) = \mathbb{P}(A \cap B).$$

That is, it is the probability of the intersection of the two subsets.

Example 1.1 Rolling an Honest Dice
If we consider the probability of the event, A, of an honest dice landing on an even number ($A = \{2, 4, 6\}$) and the event, B, of the number being greater than 3 ($B = \{4, 5, 6\}$), then $\mathbb{P}(A, B) = \mathbb{P}(A \cap B) = \mathbb{P}(\{4, 6\}) = 1/3$.

From elementary set theory $A = A \cap \Omega = A \cap (B \cup \neg B) = (A \cap B) \cup (A \cap \neg B)$. However, $A \cap B$ and $A \cap \neg B$ are non-overlapping sets (that is $(A \cap B) \cap (A \cap \neg B) = \emptyset$), so that

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \neg B) = \mathbb{P}(A, B) + \mathbb{P}(A, \neg B). \tag{1.1}$$



This is sometimes known as the 'additive law of probability'. It trivially generalises to many events. If $\{B_i | i \in I\}$ forms an *exhaustive* and *mutually exclusive* set of events, so that $A = \bigcup_{i \in \mathcal{I}} A \cap B_i$, then

$$\mathbb{P}(A) = \sum_{i \in \mathcal{I}} \mathbb{P}(A \cap B_i) = \sum_{i \in \mathcal{I}} \mathbb{P}(A, B_i).$$

This is an example of the *law of total probability*. Although it is possible to formalise probability in terms of sets (we could, for example, use $\mathbb{P}(A \cap B)$ as our standard notation for the joint probabilities), when we come to modelling the real world, it is more natural to think of the events as logical (true-false) statements (or predicates). The set notation then looks rather confusing. Thus, it is more usual to think of the additive law of probability as an axiom that we can exploit when necessary.

Note, while joint probabilities are symmetric in the sense

$$\mathbb{P}(A, B) = \mathbb{P}(B, A)$$

conditional probabilities aren't

$$\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$$

unless

$$\mathbb{P}(A) = \mathbb{P}(B).$$

The second building block for reasoning about multiple events is the *conditional probability* of event A occurring *given* that event B occurs. It is denoted $\mathbb{P}(A|B)$. The conditional probability is equal to the joint probability of both events occurring divided by the probability that event B occurs

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)}. \tag{1.2}$$

It is a probability for A with all the usual properties of a probability, for example

$$\mathbb{P}(A|B) + \mathbb{P}(\neg A|B) = 1.$$

(Note that it is *not* a probability for B so that $\mathbb{P}(A|B)+\mathbb{P}(A|\neg B)$ will not generally be equal to 1.) The conditional probability is not defined if $\mathbb{P}(B) = 0$ (although this doesn't usually worry us as we tend not to care about events that will never happen). Given Equation (1.2) it might seem that conditional probabilities are secondary to joint probabilities, but when it comes to modelling real systems it is often the case that we can more easily specify the conditional probability. That is, if A depends on B in some way then $\mathbb{P}(A|B)$ is the probability of A when you know B has happened and this is often easy to model. However, it is wrong to think that conditional probabilities always express causality. If $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ are both meaningful whatever the causal relationship (e.g. $\mathbb{P}(A|B)$ is well defined even if A causes B). A consequence of Equation (1.2) is that

$$\mathbb{P}(A, B) = \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A). \tag{1.3}$$

This is sometimes known as the 'multiplicative law of probabilities'. Equations (1.1) and (1.3) provide the cornerstone to developing a calculus for reasoning about probabilities.

Extending these laws to more events is simple. The trick is to split all the events into two groups. These groups of event can be considered as single compound events. We can then apply the laws of probability given above to the compound events. Thus, for example,

$$\begin{aligned} \mathbb{P}(A|B,C) + \mathbb{P}(\neg A|B,C) &= 1 && \text{treating } B \wedge C \text{ as a single event} \\ \mathbb{P}(A,B,C) + \mathbb{P}(A,B,\neg C) &= \mathbb{P}(A,B) && \text{treating } A \wedge B \text{ as a single event.} \end{aligned}$$

We interpret $\mathbb{P}(A|B,C)$ as $\mathbb{P}(A|(B,C))$, that is the probability of A given B and C (the comma has a higher precedence – binds stronger – than the bar).

With three events there are a large number of identities between joint and conditional probabilities, e.g.

$$\begin{aligned} \mathbb{P}(A,B,C) &= \mathbb{P}(A,B|C)\,\mathbb{P}(C) = \mathbb{P}(A|B,C)\,\mathbb{P}(B|C)\,\mathbb{P}(C) \\ &= \mathbb{P}(A,C|B)\,\mathbb{P}(B) = \mathbb{P}(A|B,C)\,\mathbb{P}(C|B)\,\mathbb{P}(B) \\ &= \mathbb{P}(B,C|A)\,\mathbb{P}(A) = \mathbb{P}(C|A,B)\,\mathbb{P}(B|A)\,\mathbb{P}(A) \end{aligned}$$

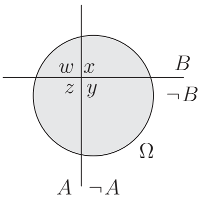
etc. This is not difficult, but some care is required to make sure that what you do is valid.

An obvious consequence of Equation (1.3) is the identity

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{1.4}$$

This formula provides a means of going from one conditional probability, $\mathbb{P}(B|A)$, to the reverse conditional probability, $\mathbb{P}(A|B)$. This seemingly innocuous equation, known as *Bayes' rule*, is the basis for one of the most powerful formalism in probabilistic inference known as the Bayesian approach. We return to this many times, particular in Chapter 8 which is devoted to Bayesian inference.

Example 1.2 Manipulating Probabilities
To understand the rules for manipulating probabilities consider two events, A and B, where the joint probabilities of these events and the negation of these events (i.e. the outcome when the event does not happen) are given by



$$\mathbb{P}(A, B) = w \qquad \mathbb{P}(A, \neg B) = z$$
$$\mathbb{P}(\neg A, B) = x \qquad \mathbb{P}(\neg A, \neg B) = y$$

where $w + x + y + z = 1$. The probabilities of events A and B are given by

$$\mathbb{P}(A) = \mathbb{P}(A, B) + \mathbb{P}(A, \neg B) = w + z$$
$$\mathbb{P}(B) = \mathbb{P}(A, B) + \mathbb{P}(\neg A, B) = w + x.$$

Then (some of) the joint probabilities are given by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(B)} = \frac{w}{w + x} \qquad \mathbb{P}(\neg A|B) = \frac{\mathbb{P}(\neg A, B)}{\mathbb{P}(B)} = \frac{x}{w + x}$$
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A, B)}{\mathbb{P}(A)} = \frac{w}{w + z} \qquad \mathbb{P}(A|\neg B) = \frac{\mathbb{P}(A, \neg B)}{\mathbb{P}(\neg B)} = \frac{z}{z + y}.$$

An example of the 'addition law of probability' is

$$\mathbb{P}(A|B) + \mathbb{P}(\neg A|B) = \frac{w}{w + x} + \frac{x}{w + x} = 1$$

and the 'multiplicative law of probability' is

$$\mathbb{P}(A, B) = \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$$
$$w = \frac{w}{w + x}(w + x) = \frac{w}{w + z}(w + z).$$

Note, however, that

$$\mathbb{P}(A|B) + \mathbb{P}(A|\neg B) = \frac{w}{w + x} + \frac{z}{z + y} \neq 1 \quad \text{(in general)}.$$

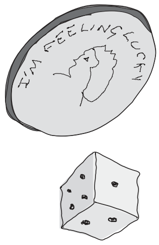
The laws of probability are very simple, but it is very easy to get confused about exactly what terms are what. Thus care is necessary.

1.2.4 Independence

Two events, A and B, are said to be *independent* if

$$\mathbb{P}(A, B) = \mathbb{P}(A)\,\mathbb{P}(B). \tag{1.5}$$

Note that independence does *not* imply $A \cap B = \emptyset$, (i.e. $\mathbb{P}(A, B) = 0$) – which says rather that event A and B cannot both happen, i.e. they are mutually exclusive. Independence is a rather more subtle, but nevertheless a strong statement about two events. Its utility is that it means that you can treat the events in isolation. From Equations (1.3) and (1.5) it follows that for independent events $\mathbb{P}\left(A|B\right) = \mathbb{P}\left(A\right)$ – i.e. the probability of event A happening is blind to whether event B happens. Equation (1.5) shows that independence is a symmetric relation and is well defined even when $\mathbb{P}\left(A\right) = 0$ or $\mathbb{P}\left(B\right) = 0$ (where the conditional probability is not defined). Using the formula $\mathbb{P}\left(A|B\right) = \mathbb{P}\left(A\right)$ as a definition of independence doesn't explicitly show the symmetry and might not be applicable. However, it is very often how we would use independence.

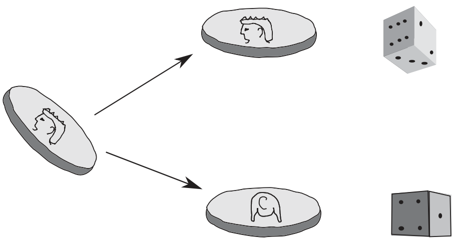


If two events are causally independent (e.g. the event of tossing heads and rolling a 6), then they will be probabilistically independent

$$\mathbb{P}\left(\text{Coin} = H, \text{Dice} = 6\right) = \mathbb{P}\left(\text{Coin} = H\right) \times \mathbb{P}\left(\text{Dice} = 6\right).$$

However, probabilistic independence is a mathematical relationship $\mathbb{P}\left(A, B\right) = \mathbb{P}\left(A\right)\mathbb{P}\left(B\right)$, which doesn't require A and B to be causally independent.

Example 1.3 Probabilistic Independence
Consider the (clearly manufactured) situation where we toss a coin with a probability p of getting heads. If we get heads, then we choose an honest dice which we throw. Otherwise we choose a biased dice $\mathbb{P}\left(D = 1\right) = \mathbb{P}\left(D = 2\right) = \mathbb{P}\left(D = 3\right) = 1/12$, $\mathbb{P}\left(D = 4\right) = \mathbb{P}\left(D = 5\right) = \mathbb{P}\left(D = 6\right) = 1/4$, where D denotes the number rolled.



Let A be the event of getting tails, and B be the event of getting either a 1 or 6. The probability of getting tails is $\mathbb{P}\left(A\right) = 1 - p$ (depending on the bias of the coin). A simple calculation shows

$$\mathbb{P}\left(A, B\right) = (1-p) \times \left(\frac{1}{12} + \frac{1}{4}\right) = \frac{1}{3}(1-p)$$
$$\mathbb{P}\left(\neg A, B\right) = p \times \left(\frac{1}{6} + \frac{1}{6}\right) = \frac{1}{3}p$$

so that $\mathbb{P}\left(B\right) = \mathbb{P}\left(A, B\right) + \mathbb{P}\left(\neg A, B\right) = 1/3$. Thus, $\mathbb{P}\left(A, B\right) = (1-p)/3 = \mathbb{P}\left(A\right)\mathbb{P}\left(B\right)$; so the events are independent, though they are clearly not causally independent (insofar as which dice we roll depends on the outcome of event A).

We can generalise the idea of independence to a family of events $\{A_i | i \in I\}$ (for some index set $I \subset \mathbb{N}$). The family of events are said to be independent if *for all* $I' \subseteq I$

$$\mathbb{P}\left(\bigwedge_{i \in I'} A_i\right) = \prod_{i \in I'} \mathbb{P}\left(A_i\right)$$

where the left-hand side denotes the joint probability of all events A_i for which $i \in I'$. This is a much stronger statement than *pairwise independent* (i.e. each pair of events are independent). It is possible for a family of events to be pairwise independent without itself being independent.

Example 1.4 Eight-Sided Dice
Consider an eight-sided honest dice so that $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and consider the family of events $A = \{1, 2, 3, 4\}$, $B = \{2, 4, 6, 8\}$, and $C = \{2, 4, 5, 7\}$. Since the dice is honest $\mathbb{P}\left(A\right) = \mathbb{P}\left(B\right) = \mathbb{P}\left(C\right) = 1/2$. Similarly,

$$\begin{aligned}
\mathbb{P}\left(A, B\right) &= \mathbb{P}\left(\{2, 4\}\right) = 1/4 = \mathbb{P}\left(A\right) \mathbb{P}\left(B\right) \\
\mathbb{P}\left(A, C\right) &= \mathbb{P}\left(\{2, 4\}\right) = 1/4 = \mathbb{P}\left(A\right) \mathbb{P}\left(C\right) \\
\mathbb{P}\left(B, C\right) &= \mathbb{P}\left(\{2, 4\}\right) = 1/4 = \mathbb{P}\left(B\right) \mathbb{P}\left(C\right)
\end{aligned}$$

so that the events are pairwise independent, but

$$\mathbb{P}\left(A, B, C\right) = \mathbb{P}\left(\{2, 4\}\right) = \tfrac{1}{4} \qquad \mathbb{P}\left(A\right) \mathbb{P}\left(B\right) \mathbb{P}\left(C\right) = \tfrac{1}{8}$$

Thus the family of events are not independent.

Often we meet events, A and B, which depend on each other through some intermediate event, C. We define events as being *conditionally independent* if

$$\mathbb{P}\left(A, B | C\right) = \mathbb{P}\left(A | C\right) \mathbb{P}\left(B | C\right).$$

Example 1.5 Plumbers and Cold Weather
Consider the case where, if it is very cold, a pipe might freeze and burst, and I am highly likely to call a plumber. Thus, the event of it being cold and calling a plumber are dependent on each other so that

$$\mathbb{P}\left(\text{cold, call plumber}\right) \neq \mathbb{P}\left(\text{cold}\right) \mathbb{P}\left(\text{call plumber}\right).$$

However, they are linked (at least in my simplified version of the world) through the burst pipe. If I know the pipe is burst, then it is irrelevant whether it is cold or not. In this example, the events of calling a plumber and it being cold are conditionally independent given the event that we have a burst pipe

$$\begin{aligned}
\mathbb{P}\left(\text{cold, call plumber} | \text{burst pipe}\right) = \mathbb{P}\left(\text{cold} | \text{burst pipe}\right) \\
\mathbb{P}\left(\text{call plumber} | \text{burst pipe}\right).
\end{aligned}$$