0
Introduction

A life which included no improbable events would be the real statistical improbability.

Poul Anderson

It is plain that any scientist is trying to correlate the incoherent body of facts confronting him with some definite and orderly scheme of abstract relations, the kind of scheme which he can borrow only from mathematics.

G.H. Hardy

This chapter introduces the basic concepts of probability in an informal way. We discuss our everyday experience of chance, and explain why we need a theory and how we start to construct one. Mathematical probability is motivated by our intuitive ideas about likelihood as a proportion in many practical instances. We discuss some of the more common questions and problems in probability, and conclude with a brief account of the history of the subject.

0.1 Chance

My only solution for the problem of habitual accidents is to stay in bed all day. Even then, there is always the chance that you will fall out.

Robert Benchley

It is not certain that everything is uncertain.

Blaise Pascal

You can be reasonably confident that the sun will rise tomorrow, but what it will be shining on is a good deal more problematical. In fact, the one thing we can be certain of is that uncertainty and randomness are unavoidable aspects of our experience.

At a personal level, minor ailments and diseases appear unpredictably and are resolved not much more predictably. Your income and spending are subject to erratic strokes of good or bad fortune. Your genetic makeup is a random selection from those of your parents. The weather is notoriously fickle in many areas of the globe. You may decide to play cards, invest in shares, bet on horses, buy lottery tickets, or engage in one or several other forms of gambling on events that are necessarily uncertain (otherwise, gambling could not occur).

At a different level, society has to organize itself in the context of similar sorts of uncertainty. Engineers have to build structures to withstand stressful events of unknown magnitude and frequency. Computing and communication systems need to be designed to cope with uncertain and fluctuating demands and breakdowns. Any system should be designed to have a small chance of failing and a high chance of performing as it was intended. Financial markets of any kind should function so as to share out risks in an efficient and transparent way, for example, when you insure your car or house, buy an annuity, or mortgage your house.

This uncertainty is not confined to the future and events that have yet to occur; much effort is expended by scientists (and by lawyers, curiously) who seek to resolve our doubt about things that have already occurred. Of course, our ignorance of the past is perhaps not quite as pressing as our ignorance of the future because of the direction in which time's arrow seems to be running. (But the arguments about the past are, paradoxically, somewhat more bad tempered as a rule.) In addition, and maybe most annoyingly, we are not certain about events occurring right now, even among those within our direct observation. At a serious level, you can see the human genome expressing itself in everyone you know, but the mechanisms remain largely a mystery. The task of unravelling this genetic conundrum will require a great deal of probability theory and statistical analysis. At a more trivial level, illusionists (and politicians) make a handsome living from our difficulties in being certain about our own personal experience (and prospects).

It follows that everyone must have some internal concept of chance to live in the real world, although such ideas may be implicit or even unacknowledged.

These concepts of chance have long been incorporated into many cultures in mythological or religious form. The casting of lots (sortilege) to make choices at random is widespread; we are all familiar with "the short straw" and the "lucky number." The Romans, for example, had gods of chance named Fortuna and Fors, and even today we have Lady Luck. Note that if you ransack the archives of the literary response to this state of affairs, one finds it to be extremely negative:

- "Fortune, that favours fools." *Ben Jonson*
- "For herein Fortune shows herself more kind than is her custom." *William Shakespeare*, *Merchant of Venice*
- "Ill fortune seldom comes alone." *John Dryden*
- "[T]he story of my life . . . wherein I spake of most disastrous chances." *William Shakespeare*, *Othello*
- Probability is the bane of the age. *Anthony Powell, Casanova's Chinese Restaurant*

This list of reproaches to Fortune could be extended almost indefinitely; in fact, you may have expressed similar sentiments yourself (although perhaps less poetically).

Nevertheless, it is a curious feature of human nature that, despite our oft-stated deprecation of this randomness, many people seek out extra uncertainty. They enter lotteries, bet on horses, and free-climb on rockfaces of dubious stability. A huge part of the entertainment industry is geared toward supplying surprise and uncertainty. This simultaneous desire to be safe and yet at risk is an interesting trait that seems difficult to explain; fortunately, however, that is not our problem here.

Instead, our task is to find a way of describing and analysing the concepts of chance and uncertainty that we intuitively see are common to the otherwise remarkably diverse examples mentioned above.

0.2 Models

> . . . and blessed are those
> whose blood and judgement are so well comingled
> that they are not a pipe for Fortune's finger
> to sound what stop she please.
>
> *W. Shakespeare*, *Hamlet*

In the preceding section, we concluded that large parts of our experience are unpredictable and uncertain. To demonstrate the effect of chance in our lives, we gave a long list of examples, and we could have made it a great deal longer were it not for lack of space and fear of boring the reader. However, to say that most things are unpredictable is to paint too negative a picture. In fact, many things are certain (death and taxes, notoriously) and even uncertain things are susceptible to judgment and insight. We learn that, in Monopoly, it is good to own the orange set of properties; we know that casinos invariably make profits; we believe that it does not really matter whether you call heads or tails when a coin is flipped energetically enough; we learn not to be on top of the mountain during a thunderstorm; and so on.

In fact, we often go further than these rough judgments and compare probabilities. Most people would agree that in roulette, black is more likely than green (the zeros); a bookmaker is more likely to show a profit than a loss on a book; the chance of a thunderstorm is greater later in the day; and so on. This is another list that could be extended indefinitely, but the point is that *because* probabilities are often comparable in this way it is natural to represent them on a numerical scale. After all, such comparisons were the principal reason for the development of numbers in the first place. It will later be shown that this numerical scale should run from 0 to 1, but we first make some general remarks.

It seems that we do share a common concept of chance because we can discuss it and make agreed statements and judgments such as those above. We therefore naturally seek to abstract these essential common features, rather than discuss an endless list of examples from first principles. This type of simple (or at least, simplified) description of a system or concept is often called a model. Agreeing that probability is a number is the first step on our path to constructing our model.

Most, perhaps all, of science conforms to this pattern; astronomy was originally developed to describe the visible movements of planets and stars; Newton's and Einstein's theories of space, time, and motion were developed to describe our perceptions of moving bodies with their mass, energy, and motion; Maxwell's equations codify the properties of electromagnetism; and so on. The first advantage of such models is their concise description of otherwise incomprehensibly complicated systems.

The second, and arguably principal, advantage and purpose of having such a model is that (if it is well chosen) it provides not only a description of the system, but also predictions about how it will behave in the future. It may also predict how it would behave in different circumstances, or shed light on its (unobserved) past behaviour.

Astronomy is one example we have mentioned; for another, consider the weather. Without a model for forecasting, your only recourse is to recall the various ways in which weather developed on the previous occasions when the situation seemed to resemble the current one. There will almost certainly be no perfect match, and identifying a "good fit" will be exceedingly time consuming or impossible.

Returning to chance and probability, we note that a primitive model for chance, used by many cultures, represents it as a supernatural entity, or god. We mentioned this in the previous section, and this procedure is, from one point of view, a perfectly reasonable and consistent model for chance. It explains the data, with no contradictions. Unfortunately, it is useless for practical purposes, such as prediction and judgment, because it is necessary that the mind of the god in question should be unpredictable and capricious, and that mind of Fortune (or whatever) is closed to us. Efforts to discover Fortune's inclination by propitiation of various kinds (sacrifice and wheedling) have met with outcomes that can at best be described as equivocal. The Romans made use of more complicated and various techniques, such as examining the behaviour of birds (augury) or casting lots (sortilege). Related modern techniques use tea leaves and astrology, but there is no evidence to suggest that any of these methods rate better than utterly useless.

Fortunately, experience over the past millennium has shown that we can do much better by using a mathematical model. This has many advantages; we mention only a few. First, a useful model must be simpler than reality; otherwise, it would be no easier to analyse than the real-life problem. Mathematical models have this stripped-down quality in abundance.

Second, mathematical models are abstract and are therefore quite unconstrained in their applications. When we define the probabilities of events in Chapter 1, and the rules that govern them, our conclusions will apply to all events of whatever kind (e.g., insurance claims, computer algorithms, crop failures, scientific experiments, games of chance; think of some more yourself).

Third, the great majority of practical problems about chance deal with questions that either are intrinsically numerical or can readily be rephrased in numerical terms. The use of a mathematical model becomes almost inescapable.

Fourth, if you succeed in constructing a model in mathematical form, then all the power of mathematics developed over several thousand years is instantly available to help you use it. Newton, Gauss, and Laplace become your (unpaid) assistants, and aides like these are not to be lightly discarded.

In the next section, therefore, we begin our construction of a mathematical model for chance. It turns out that we can make great progress by using the simple fact that our ideas about probability are closely linked to the familiar mathematical ideas of proportion and ratio.

Finally, we make the trivial point that, although the words chance, likelihood, probability, and so on mean much the same in everyday speech, we will only use one of these. What follows is thus a theory of probability.

0.3 Symmetry

Blind Fortune still bestows her gifts on such as cannot use them.
Ben Jonson

We begin with some basic ideas and notation. Many occurrences of probability appear in everyday statements such as:

The probability of red in (American) roulette is $\frac{18}{38}$.
The probability of a head when you flip a coin is 50%.
The probability of a spade on cutting a pack of cards is 25%.

Many other superficially different statements about probability can be reformulated to appear in the above format. This type of statement is in fact so frequent and fundamental that we use a standard abbreviation and notation for it. Anything of the form

the probability of A is p

will be written as:

$$\mathbf{P}(A) = p.$$

In many cases, p may represent an adjective denoting quantity, such as "low" or "high." In the examples above, A and p were, respectively,

$$\begin{aligned} A &\equiv \text{red}, & p &= 18/38 \\ A &\equiv \text{heads}, & p &= 50\% \\ A &\equiv \text{spade}, & p &= 25\%. \end{aligned}$$

You can easily think of many similar statements. Our first urgent question is, where did those values for the probability p come from? To answer this, let us consider what happens when we pick a card at random from a conventional pack. There are 52 cards, of which 13 are spades. The implication of the words "at random" is that any card is equally likely to be selected, and the proportion of the pack comprising the spade suit is $13/52 = \frac{1}{4}$. Our intuitive feelings about symmetry suggest that the probability of picking a spade is directly proportional to this fraction, and by convention we choose the constant of proportionality to be unity. Hence,

$$\mathbf{P}(\text{spade}) = \frac{1}{4} = 25\%.$$

Exactly the same intuitive interpretation comes into play for any random procedure having this kind of symmetry.

Example: American Roulette These wheels have 38 compartments, of which 18 are red, 18 are black, and two are green (the zeros). If the wheel has been made with equal-size compartments (and no hidden magnets, or subtle asymmetries), then the ball has 18 chances to land in red out of the 38 available. This suggests

$$\mathbf{P}(\text{red}) = 18/38. \qquad \bullet$$

In the case of a fair coin, of course, there are only two equally likely chances to $\mathbf{P}(\text{Head}) = 50\%$ and $\mathbf{P}(\text{Tail}) = 50\%$. This particular case of equal probabilities has passed into the language in the expression a "fifty-fifty" chance (first used in print by P.G. Wodehouse in his novel *The Little Nugget*).

In general, this argument (or expression of our intuition) leads to the following definition of probability. Suppose that some procedure with a random outcome has n distinct possible outcomes, and suppose further that by symmetry (or by construction or supposition) these outcomes are equally likely. Then if A is any collection of r of these outcomes, we define

(1) $$\mathbf{P}(A) = \frac{r}{n} = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}.$$

Note that in this case we must have

(2) $$0 \leq \mathbf{P}(A) \leq 1,$$

because $0 \leq r \leq n$. Furthermore, if A includes all n possible outcomes, then $\mathbf{P}(A) = 1$. At the other extreme, $\mathbf{P}(A) = 0$ if A contains none of the possible outcomes.

Here is another simple example.

(3) Example: Die With the probability of any event now defined by (1), it is elementary to find the probability of any of the events that may occur when we roll a die. The number shown may be (for example) even, odd, prime or perfect, and we denote these events by A, B, C and D respectively. Here $n = 6$, and for $A = \{2 \text{ or } 4 \text{ or } 6\}$ we have $r = 3$. The probability that it shows an even number is

$$\mathbf{P}(A) = \mathbf{P}(\{2 \text{ or } 4 \text{ or } 6\}) = \frac{3}{6} = \frac{1}{2}.$$

Likewise, and equally trivially, we find that

$$\mathbf{P}(\text{odd}) = \mathbf{P}(B) = \mathbf{P}(\{1 \text{ or } 3 \text{ or } 5\}) = \frac{1}{2}$$

$$\mathbf{P}(\text{prime}) = \mathbf{P}(C) = \mathbf{P}(\{2 \text{ or } 3 \text{ or } 5\}) = \frac{1}{2}$$

$$\mathbf{P}(\text{perfect}) = \mathbf{P}(D) = \mathbf{P}(\{6\}) = \frac{1}{6}.$$

These values of the probabilities are not inconsistent with our ideas about how the symmetries of this die should express themselves when it is rolled. ●

This idea or interpretation of probability is very appealing to our common intuition. It is first found nascent in a poem entitled "De Vetula," which was widely distributed in manuscript form from around 1250 onward. It is, of course, extremely likely that this idea of probability had been widespread for many years before then. In succeeding years, most probability calculations during the Renaissance and the ensuing scientific revolution take this framework for granted.

However, there are several unsatisfactory features of this definition: first, there are plenty of random procedures with no discernible symmetry in the outcomes; and second, it is worrying that we do not need actually to roll a die to say that chance of a six is $\frac{1}{6}$. Surely, actual experiments should play some part in shaping and verifying our theories about the physical world? We address this difficulty in the next section.

0.4 The Long Run

Nothing is more certain than incertaincies
Fortune is full of fresh variety
Constant in nothing but inconstancy
Richard Barnfield

Suppose that some random procedure has several possible outcomes that are not necessarily equally likely. How can we define the probability $\mathbf{P}(A)$ of any eventuality A of interest? For example, suppose the procedure is the rolling of a die that is suspected to be weighted, or even clearly asymmetrical, in not being a perfect cube. What now is the probability of a six?

There is no symmetry in the die to help us, but we can introduce symmetry another way. Suppose you roll the die a large number n of times, and let $r(n)$ be the number of sixes shown. Then (provided the rolls were made under similar conditions) the symmetry between the rolls suggests that (at least approximately)

$$\mathbf{P}(\text{six}) = \frac{r(n)}{n} = \frac{\text{number of sixes}}{\text{number of rolls}}.$$

Furthermore, if you actually obtain an imperfect or weighted die and roll it many times, you will find that as n increases the ratio $\frac{r(n)}{n}$ always appears to be settling down around some asymptotic value. This provides further support for our taking $\frac{r(n)}{n}$ as an approximation to $\mathbf{P}(\text{six})$.

Of course, this procedure can only ever supply an approximation to the probability in question, as the ratio $r(n)/n$ changes with n. This is the sort of price that we usually pay when substituting empiricism for abstraction. There are other possible eventualities that may also confuse the issue; for example, if told that a coin, in 1 million flips, showed 500,505 heads and 499,495 tails, you would probably accept it as fair, and you would set $\mathbf{P}(\text{head}) = \frac{1}{2}$. But suppose you were further informed that all the heads formed a run preceding all the tails; would you now be quite so confident? Such a sequence might occur, but our intuition tells us that it is so unlikely as to be irrelevant to this discussion. In fact, routine gaming and other experience bears out our intuition in the long run. That is why it is our intuition; it relies not only on our own experience, but also on our gambling predecessors. You can believe in the symmetry and long-run interpretations for chances in roulette without ever having spun the wheel or wagered on it (the author has done neither).

This idea of the long run can clearly be extended to any random procedure that it is possible to repeat an arbitrary number n of times under essentially identical conditions. If A is some possible result and A occurs on $r(n)$ occasions in n such repetitions, then we say that

(1) $$\mathbf{P}(A) \approx r(n)/n.$$

Thus, evaluating the ratio $r(n)/n$ offers a way of measuring or estimating the probability $\mathbf{P}(A)$ of the event A. This fact was familiar to gamblers in the Renaissance, and presumably well before then. Cardano in his *Book on Games of Chance* (written around 1520) observes that "every die, even if it is acceptable, has its favoured side." It may be assumed that gamblers noticed that in the long run small biases in even a well-made die will be revealed in the empirical proportions of successes for the six faces.

A similar empirical observation was recorded by John Graunt in his book *Natural and Political Observations Made Upon the Bills of Mortality* (1662). He found that in a large number of births, the proportion of boys born was approximately $\frac{14}{27}$. This came as something of a surprise at the time, leading to an extensive debate. We simply interpret the observation in this statement: the probability of an as yet unborn child being male is approximately $\frac{14}{27}$. (Note that this empirical ratio varies slightly from place to place and time to time, but it always exceeds $\frac{1}{2}$ in the long run.)

Once again, we stress the most important point that $0 \leq r(n)/n \leq 1$, because $0 \leq r(n) \leq n$. It follows that the expression (1) always supplies a probability in [0, 1]. Furthermore, if A is impossible (and hence never occurs), $r(n)/n = 0$. Conversely, if A is certain (and hence always occurs), $r(n)/n = 1$.

0.5 Pay-Offs

Probability is expectation, founded upon partial knowledge.
George Boole

In reading the previous two sections, the alert reader will have already made the mental reservation that many random procedures are neither symmetrical nor repeatable. Classic examples include horse races, football matches, and elections. Nevertheless, bookmakers and gamblers seem to have no qualms about quoting betting odds, which are essentially linked to probabilities. (We explore this connection in more depth in Sections 0.7 and 1.6.) How is it possible to define probabilities in these contexts?

One possible approach is based on our idea of a "fair value" of a bet, which in turn is linked to the concept of mathematical averages in many cases. An illustrative example of great antiquity, and very familiar to many probabilists, is provided by the following problem.

A parsimonious innkeeper empties the last three glasses of beer (worth 13 each) from one barrel and the last two glasses (worth 8 each) from another, and mixes them in a jug holding five glasses. What is the "fair" price for a glass from the jug? The innkeeper calculates the value of the beer in the jug as $3 \times 13 + 2 \times 8 = 55$, and divides by 5 to obtain his "fair" price of 11 each, although discriminating customers may well have another view about that.

A similar idea extends to random situations. For example, suppose the benevolent but eccentric uncle of Jack and Tom flips a fair coin; if it shows heads then Jack gets \$1, if it shows tails then Tom gets \$1. If this is worth \$$p$ to Jack, it is worth the same to Tom. Because the uncle certainly parts with \$1, we have $2p = 1$, which is to say that a "fair" price for Jack to sell his share of the procedure *before* the uncle flips the coin is \$$\mu$, where

$$\mu = \frac{1}{2}.$$

More generally, if the prize at stake is $\$d$, then a fair price for either of their expected gifts is $\$d/2$. More generally still, by a similar argument, if you are to receive $\$d$ with probability p, then a fair price for you to sell this uncertain reward *in advance* of the experiment is $\$dp$.

But this argument has a converse. Suppose you are to receive \$1 if some event A occurs and nothing otherwise. Further suppose that you and Tom agree that he will give you $\$p$ before the experiment occurs, and that he will get your reward whatever it may be. Then, in effect, you and Tom have estimated and agreed that (at least approximately)

$$p = \mathbf{P}(A).$$

To see this, note that if you believed $\mathbf{P}(A)$ were larger than p you would hold out for a higher price, and Tom would not pay as much as p if he believed $\mathbf{P}(A)$ were less than p.

Thus, probabilities can be defined, at least implicitly, whenever people can agree on a fair price. Note that $0 \le p \le 1$ in every case. Finally, we observe that the idea of fair price can be extended and turns out to be of great importance in later work. For example, suppose you roll a fair five-sided die; if faces 1, 2, or 3 turn up you win \$13, if faces 4 or 5 turn up you win \$8. What is a fair price for one roll? Essentially the same arithmetic and ideas as we gave above for the parsimonious innkeeper reveals the fair price (or value) of one roll to be \$11 because

$$13 \times \frac{3}{5} + 8 \times \frac{2}{5} = 11.$$

We will meet this concept again under the name "expectation."

0.6 Introspection

Probability is a feeling of the mind.
Augustus de Morgan

In the previous three sections, we defined probabilities, at least approximately, in situations with symmetry, or when we could repeat a random procedure under essentially the same conditions, or when there was a plausible agreed "fair price" for a resulting prize. However, there are many asymmetric and non-repeatable random events, and for many of these we would feel distinctly unsatisfied, or even unhappy, with the idea that their probabilities should be determined by gamblers opening a book on the question. For example, what is the probability that some accused person is guilty of the charge? Or the probability of life on another planet? Or the probability of you catching a cold this week? Or the probability that Shakespeare wrote some given sonnet of doubtful provenance? Or the probability that a picture called "Sunflowers" is by van Gogh? Or the probability that Riemann's hypothesis is true? (It asserts that all the nontrivial zeros of a certain function have real part $+\frac{1}{2}$.) Or the probability that π^e is irrational? In this last question, would the knowledge that e^π is irrational affect the probability in your judgment?

In these questions, our earlier methods seem more or less unsatisfactory, and, indeed, in a court of law you are forbidden to use any such ideas in deciding the probability of guilt of an accused. One is led to the concept of probability as a "degree of belief."

If we assign a probability to any of these eventualities, then the result must of necessity be a personal or subjective assessment. Your figure need not be the same as mine or anyone else's, and any probability so obtained is called subjective or personal.

In fact, there is a strong (although perhaps minority) body of opinion that maintains that all probabilities are subjective (cf the remark of A. de Morgan at the head of this section). They argue that appeals to symmetry, the long run, or fair value, merely add a spurious objectivity to what is essentially intuition, based on personal experience, logical argument, and experiments (where these are relevant). The examples above are then simply rather trivial special cases of this general definition.

According to this approach, a probability is a measure of your "degree of belief" in the guilt of the accused, the truth of some assertion, or that a die will show a six. In the classical case based on dice, etc., this belief rests on symmetry, as it does in the "long-run" relative-frequency interpretation. The "fair price" approach uses the beliefs of all those concerned in fixing such a price. Furthermore, the approach via degrees of belief allows (at least in principle) the possibility that such probabilities could be determined by strictly logical statements relating what we know for sure to the uncertain eventuality in question. This would be a kind of inductive probability logic, an idea that was first suggested by Leibniz and later taken up by Boole. During the past century, there have been numerous clever and intriguing books about various approaches to establishing such an axiomatic framework.

This argument is clearly seductive, and for all the above reasons it is tempting to regard all types of probability as a "feeling of the mind," or as a "degree of belief." However, there are several drawbacks. First, in practice, it does not offer a wholly convincing and universally accepted way of defining or measuring probability, except in the cases discussed above. Thus, the alleged generality is a little artificial because different minds feel differently.

Second, setting this more general idea in a formal framework requires a great deal more effort and notation, which is undesirable for a first approach. Finally, it is in any case necessary for practical purposes to arrange things so that the rules are the same as those obtained from the simpler arguments that we have already outlined.

For these reasons, we do not further pursue the dream of a universal interpretation of probability; instead, we simply note this remark of William Feller:

All definitions of probability fall short of the actual practice.

0.7 FAQs

> Neither physicists nor philosophers can give any convincing account of what "physical reality" is.
>
> *G.H. Hardy*

> "What is the meaning of it, Watson?" said Holmes, solemnly, as he laid down the paper. . . . "It must tend to some end, or else our universe is ruled by chance, which is unthinkable."
>
> *A. Conan Doyle, The Adventure of the Cardboard Box*

In the preceding sections, we agree to develop a mathematical theory of probability and discuss some interpretations of probability. These supply definitions of probability in