2.4 THE PURPOSE OF STATISTICAL INFERENCE

How much does a particular drug affect a patient's condition? What can an average student earn after obtaining a college education? Will the Democrats win the next US presidential election? In life, we develop theories and use these to make predictions, but testing those theories is not easy. Life is complicated, and it is often impossible to exactly isolate the parts of a system which we want to examine. The outcome of history is determined by a complex nexus of interacting elements, each of which contributes to the reality that we witness. In the case of a drug trial, we may not be able to control the diets of participants and are certainly unable to control for their idiosyncratic metabolisms, both of which could impact the results we observe. There are a range of factors which affect the wage that an individual ultimately earns, of which education is only one. The outcome of the next US presidential election depends on party politics, the performance of the incumbent government and the media's portrayal of the candidates.

In life, noise obfuscates the signal. What we see often appears as an incoherent mess that lacks any appearance of logic. This is why it is difficult to make predictions and test theories about the world. It is like trying to listen to a classical orchestra which is playing on the side of a busy motorway, while we fly overhead in a plane. Statistical inference allows us to focus on the music by separating the signal from the noise. We will hear 'Nessun Dorma' played!

Statistical inference is the logical framework which we can use to trial our beliefs about the noisy world against *data*. We formalise our beliefs in models of *probability*. The models are probabilistic because we are ignorant of many of the interacting parts of a system, meaning we cannot say with certainty whether something will, or will not, occur. Suppose that we are evaluating the efficacy of a drug in a trial. Before we carry out the trial, we might believe that the drug will cure 10% of people with a particular ailment. We cannot say which 10% of people will be cured because we do not know enough about the disease or individual patient biology to say exactly whom. Statistical inference allows us to test this belief against the data we obtain in a clinical trial.

There are two predominant schools of thought for carrying out this process of inference: Frequentist and Bayesian. Although this book is devoted to the latter, we will now spend some time comparing the two approaches so that the reader is aware of the different paths taken to their shared goal.

2.5 THE WORLD ACCORDING TO FREQUENTISTS

In Frequentist (or Classical) statistics, we suppose that our sample of data is the result of one of an infinite number of exactly repeated experiments. The sample we see in this context is assumed to be the outcome of some probabilistic process. Any conclusions we draw from this approach are based on the supposition that events occur with probabilities, which represent the long-run frequencies with which those events occur in an infinite series of experimental repetitions. For example, if we flip a coin, we take the proportion of heads observed in an infinite number of throws as defining the probability of obtaining heads. Frequentists suppose that this probability actually exists, and is fixed for each set of coin throws that we carry out. The sample of coin flips we obtain for a fixed and finite number of throws is generated as if it were part of a longer (that is, infinite) series of repeated coin flips (see the left-hand panel of Figure 2.1).

In Frequentist statistics the data are assumed to be *random* and results from *sampling* from a fixed and defined *population* distribution. For a Frequentist the noise that obscures the true signal of the real population process is attributable to *sampling variation* – the fact that each sample we pick is slightly different and not exactly representative of the population.

We may flip our coin 10 times, obtaining 7 heads even if the long-run proportion of heads is $\frac{1}{2}$. To a Frequentist, this is because we have picked a slightly odd sample from the population of infinitely many repeated throws. If we flip the coin another 10 times, we will likely get a different result because we then pick a different sample.

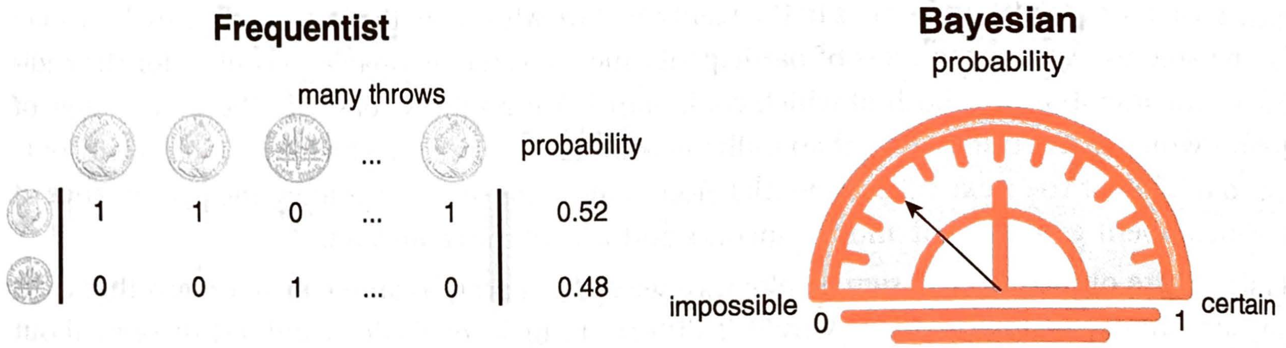

Figure 2.1 The Frequentist (left) and Bayesian (right) approaches to probability.

2.6 THE WORLD ACCORDING TO BAYESIANS

Bayesians do not imagine repetitions of an experiment in order to define and specify a probability. A probability is merely taken as a measure of certainty in a particular belief. For Bayesians the probability of throwing a 'heads' measures and quantifies our underlying belief that before we flip the coin it will land this way.

In this sense, Bayesians do not view probabilities as underlying laws of cause and effect. They are merely abstractions which we use to help express our uncertainty. In this frame of reference, it is unnecessary for events to be repeatable in order to define a probability. We are thus equally able to say, 'The probability of a heads is 0.5' or 'The probability of the Democrats winning the 2020 US presidential election is 0.75'. Probability is merely seen as a scale from 0, where we are certain an event will not happen, to 1, where we are certain it will (see the right-hand panel of Figure 2.1).

A statement such as 'The probability of the Democrats winning the 2020 US presidential election is 0.75' is hard to explain using the Frequentist definition of a probability. There is only ever one possible sample – the history that we witness – and what would we actually mean by the 'population of all possible US elections which happen in the year 2020'?

For Bayesians, probabilities are seen as an expression of subjective beliefs, meaning that they can be updated in light of new data. The formula invented by the Reverend Thomas Bayes provides the only logical manner in which to carry out this updating process. Bayes' rule is central to Bayesian inference whereby we use probabilities to express our uncertainty in parameter values after we observe data.

Bayesians assume that, since we are witness to the data, it is *fixed*, and therefore does not vary. We do not need to imagine that there are an infinite number of possible samples, or that our data are the undetermined outcome of some random process of sampling. We never perfectly know the value of an unknown parameter (for example, the probability that a coin lands heads up). This epistemic uncertainty (namely, that relating to our lack of knowledge) means that in Bayesian inference the parameter is viewed as a quantity that is probabilistic in nature. We can interpret this in one of two ways. On the one hand, we can view the unknown parameter as truly being fixed in some absolute sense, but our beliefs are uncertain, and thus we express this uncertainty using probability. In this perspective, we view the sample as a noisy representation of the signal and hence obtain different results for each set of coin throws. On the other hand, we can suppose that there is not some definitive true, immutable probability of obtaining a heads, and so for each sample we take, we unwittingly get a slightly different parameter. Here we get different results from each round of coin flipping because each time we subject our system to a slightly different probability of its landing heads up. This could be because we altered our throwing technique or started with the coin in a different position. Although these two descriptions are different philosophically, they are not different mathematically, meaning we can apply the same analysis to both.

2.7 DO PARAMETERS ACTUALLY EXIST AND HAVE A POINT VALUE?

For Bayesians, the parameters of the system are taken to vary, whereas the known part of the system – the data – is taken as given. Frequentist statisticians, on the other hand, view the unseen part of the system – the parameters of the probability model – as being fixed and the known parts of the system – the data – as varying. Which of these views you prefer comes down to how you interpret the parameters of a statistical model.

In the Bayesian approach, parameters can be viewed from two perspectives. Either we view the parameters as truly *varying*, or we view our knowledge about the parameters as imperfect. The fact that we obtain different estimates of parameters from different studies can be taken to reflect either of these two views.

In the first case, we understand the parameters of interest as varying – taking on different values in each of the samples we pick (see the top panel of Figure 2.2). For example, suppose that we conduct a blood test on an individual in two consecutive weeks, and represent the correlation between the red and white cell count as a parameter of our statistical model. Due to the many factors that affect the body's metabolism, the count of each cell type will vary somewhat randomly, and hence the parameter value may vary over time. In the second case, we view our uncertainty over a parameter's value as the reason we estimate slightly different values in different samples. This uncertainty should, however, decrease as we collect more data (see the middle panel of Figure 2.2). Bayesians are more at ease in using parameters as a means to an end – taking them not as real immutable constants, but as tools to help make inferences about a given situation.

The Frequentist perspective is less flexible and assumes that these parameters are constant, or represent the average of a long run – typically an infinite number – of identical experiments. There are occasions when we might think that this is a reasonable assumption. For example, if our parameter represented the probability that an individual taken at random from the UK population has dyslexia, it is reasonable to assume that there is a *true*, or fixed, *population* value of the parameter in question. While the Frequentist view may be reasonable here, the Bayesian view can also handle this situation. In Bayesian statistics these parameters can be assumed fixed, but that we are uncertain of their value (here the true prevalence of dyslexia) before we measure them, and use a probability distribution to reflect this uncertainty.

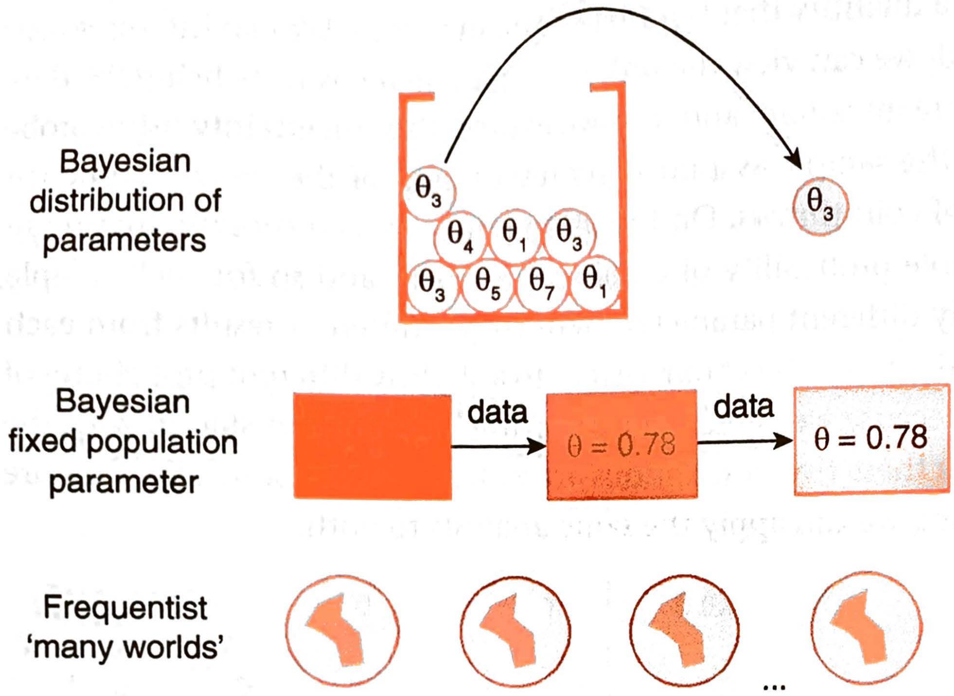

Figure 2.2 The Bayesian (top and middle) and Frequentist perspectives on parameters. In the top panel, the urn holds a large number of parameter values – a population distribution – that we sample from each time we pick a new sample. These parameters, in turn, determine the data that we obtain in our sample. The middle panel shows the Bayesian view where the uncertainty about a parameter's true value (shown in the box) decreases as we collect more data. The bottom panel represents the Frequentist view where parameters represent averages across an infinite number of exactly repeated experiments (represented by the many worlds).

But there are circumstances when the Frequentist view runs into trouble. When we are estimating parameters of a complex distribution, we typically do not view them as actually existing. Unless you view the Universe as being built from mathematical building blocks,[1] then it seems incorrect to assert that a given parameter has any deeper existence than that with which we endow it. The less restrictive Bayesian perspective here seems more reasonable.

The Frequentist view of parameters as a limiting value of an average across an infinity of identically repeated experiments (see the bottom panel of Figure 2.2) also runs into difficulty when we think about one-off events. For example, the probability that the Democrat candidate wins in the 2020 US election cannot be justified in this way, since elections are never rerun under the exact same conditions.