Rethinking statistics

2022-08-30

Contents

Reading:

Required Reading (everyone):

- Statistical Rethinking, Ch. 1–2 ("The Golem of Prague" and "Small Worlds and Large Worlds").

Reading Notes:

This book has two kinds of optional sections that you don't have to read, but which can enrich your experience of it (see p. xiv in the prologue):

1. **Rethinking sections** are highlighted in blue and begin with "**Rethinking:**. These sections provide broader context for the material in the chapter, connecting it to other approaches to statistical analysis, provide historical background for where the concepts you're studying came from, or warn you about ways that many people misunderstand the concepts in the chapter. These are good to read, but you can read them quickly, since they're supplementary to the main material of the chapter. They're particularly useful if you want a deeper understanding of the material.

2. **Overthinking sections** are printed in a smaller typeface, set off from the text with a heavy horizontal black line and begin with "**Overthinking:**". You can skip these sections, but they're useful if you want to get a deeper understanding of the detailed mathematics and computational methods we use. McElreath intends for you to skip these sections the first time you read the book, but they can be useful on second or later readings, when you have a clearer sense of how they fit into the broader context of the book.

Chapter 1, "The Golem of Prague" is mostly a philosophical overview of why we need to rethink our approach to statistics, and also a roadmap of the book. It will give you a sense of the approach the book takes and how the different sections you will be reading will fit together.

Chapter 2, "Small Worlds and Large Worlds" introduces the fundamental concepts of Bayesian statistics and statistical modeling. One reason I find Bayesian methods so useful is that they naturally encourage you to think about staitstical problems as mathematical models of the actual processes that generated your data. The chapter begins with a very nice example of how you can use a globe to estimate the fraction of the earth's surface covered by water, and walks you through constructing a Bayesian statistical model. Section 2.3, "Components of the model" is very important in laying out the structure that a Bayesian model must have and section 2.4 ("Making the model go") connects that structure to the formal statement of Bayes's theorem (p. 37). Bayes's theorem is very straightforward and simple, but applying it in the real world requires us to solve integrals that can be very difficult. The chapter then describes ways that we can use computational methods, such as grid sampling and Monte Carlo sampling to estimate the value of integrals that we can't solve by mathematical analysis.

As you read Chapter 2, think about the disctinction between "Small worlds" that describe the self-contained world of a statistical model and "Large worlds" in which we connect the small world of the model to the larger world in which we use the model.

Chapter 2 starts with a practical example of using statistical sampling to estimate a probability distribution. First, we look at how we can draw inferences about marbles in a bag, based on a sample of marbles drawn from the bag. Every time we draw another marble from the bag, we update our estimate of the probabilities for different numbers of marbles being white or blue.

Next, we formalize this process of updating probabilities using Bayes's theorem. This theorem is at the heart of the first third of the semester, and you should read it carefully enough to understand the different parts of the equation:

$$P(x|D) = \frac{P(D|x) \times P(x)}{P(D)},$$

where

- The general notation $P(a|b)$ is a *conditional probability*: The probability of a, given b (e.g., you observe b, and this evidence changes what you know about the probability distribution for a)..
- $P(a|D)$ is the *posterior* probability distribution for a: The updated probability distribution for the parameter or variable a, incorporating the new information you get from observing data D.
- $P(D|a)$ is the *likelihood*: The probability distribution for that you would observe data D depending on what a is.
- $P(a)$ is the *prior* probability distribution for a: The probability distribution for a before observing the new data D
- $P(D)$ is the *evidence* (the book doesn't use this name, but it's pretty standard in Bayesian statistics): It's the probability that you'd observe data D regardles what the value of a is.

For instance, if D is the number of lightning strikes in a year, the likelihood might be a Poisson distribution with rate a: $P(D|a) = \text{Poisson}(D, a)$

The book then gives a practical example of Bayesian updating, in whcih you estimate the proportion of the earth that's covered by water by repeatedly tossing a globe in the air and catching it, and updating your estimate of the proportion based on whether your index finger rests on water or land when you catch it.

It's easy to do Bayesian updating in this simple case, but for more complicated statistics that we'll encounter in our research, the math quickly becomes too difficult for us to solve the equations analytically, so we have to rely on computational approximations. This book will explore three kinds of approximations:

1. Grid sampling
2. Quadratic approximation
3. Monte Carlo sampling

We won't get to Monte Carlo sampling until Chapter 8, so for now we'll focus on grid sampling and the quadratic approximation.

One important distinction between these methods is that the quadratic approximation is a *parametric* method, in which the prior, likelihood, and posterior are represented by mathematical functions that are characterized by parameters (e.g., a Poisson distribution is characterized by a *rate* parameter and a Normal distribution is characterized by two parameters: *mean* and *standard deviation*)., whereas sampling methods (grid and Monte Carlo) represent the posterior non-parametrically, as a bunch of samples. The difference is like the difference between plotting a function and plotting a histogram. *Parametric* methods are like plotting the values of a function and *nonparametric* methods are like plotting a histogram of the relative frequency of different observed values. When the actual posterior has a clean functional form, *parametric* methods can be very useful, but often the posterior is more complicated and doesn't have a nice fuinctional form, so sampling methods are very widely used in Bayesian statistics.

In calss, I will discuss Bayes's theorem and the diffferent ways we can apply it, and the differences between sampling and quadratic approximation methods.