Bayesian Modeling

Jim Albert and Jingchen Hu

2019-08-05

Contents

Chapter 1

Preface

Chapter 2

Learning About a Binomial Probability

2.1 Introduction: Thinking About a Proportion Subjectively

In previous chapters, we have seen many examples involving drawing color balls from a box. In those examples, we are given the numbers of balls of various colors in the box, and we consider questions related to calculating probabilities. For example, there are 40 white and 20 red balls in a box. If you draw two balls at random, what is the probability that both balls are white?

Here we consider a new scenario where we do not know the proportions of color balls in the box. That is, in the previous example, we only know that there are two kinds of color balls in the box, but we don't know 40 out of 60 of the balls are white (proportion of white $= 2/3$) and 20 out of the 60 of the balls are red (proportion of red $= 1/3$). How can we learn about the proportions of white and red balls? Since counting 60 balls can be tedious, how can we infer those proportions by drawing a sample of balls out of the box and observe the colors of balls in the sample? This becomes an inference question, because we are trying to infer the proportion p of the population, based on a sample from the population.

Let's continue discussing the scenario where we are told that there are 60 balls in total in a box, and the balls are either white or red. We do not know the count of balls of each of the two colors. We are given the opportunity to take a random sample of 10 balls out of these 60 balls. We are interested in the quantity p, the proportion of red balls in the 60 balls. How can we infer p, the proportion of red balls in the population (i.e. the 60 balls), based on the numbers of red and white balls we observe in the sample (i.e. the 10 balls)?

Proportions are like probabilities. Recall in Chapter 1 three views of a probability were discussed. We briefly review them here, and state the specific requirements to obtain each view.

1. The classical view: one needs to write down the sample space where each outcome is equally likely.

2. The frequency view: one needs to repeat the random experiments many times under identical conditions.

3. The subjective view: one needs to express one's opinion about the likelihood of a one-time event.

The classical view does not seem to work here, because we only know there are two kinds of color balls and the total number of balls is 60. Even if we take a sample of 10 balls, we are only going to observe the proportion of red balls in the sample. There does not seem to be a way for us to write down the sample space where each outcome is equally likely.

The frequency view would work here. One could treat the process of obtaining a sample (i.e. taking a random sample of 10 balls from the box) as an experiment, and obtain a sample proportion $\hat{p}$ from the experiment. One then could repeat the experiment many times under the same condition, get many sample proportions $\hat{p}$, and summarize all the $\hat{p}$. When one repeats the experiment enough times (a large number), one gets a good sense about the proportion p of red balls in the population of 60 balls in the box. This process is doable, but tedious, time-consuming, and prone to errors.

The subjective view perceives the unknown proportion p subjectively. It does require one to express his or her opinion about the value of p, and he or she could be skeptical and unconfident about the opinion. In Chapter 1, a calibration experiment was introduced to help one sharpen an opinion about the likelihood of an event by comparisons with opinion about the likelihood of other events. In this chapter and the chapters to follow, we introduce the key ideas and practice about thinking subjectively about unknowns and quantify one's opinions about the values of these unknowns using probability distributions.

As an example, let's think about plausible values for the proportion p of red balls. As p is a proportion, it can take any possible value between 0 and 1. In the calibration experiment introduced in Chapter 1, we focus on the scenario where only one value of p is of interest. For example, when one thinks that p is 0.5, it is saying that one's opinion about the probability of the value $p = 0.5$ is one. When we phrase it this way ("one's opinion about the probability of $p = 0.5$ is one"), it sounds like a very strong opinion, because one only allows p to take one possible value, and gives probability one of that happening. Since one typically has no thought about the exact value of the proportion p, setting one possible value for the proportion with probability one seems too strong.

Instead suppose that the proportion p can take multiple values between 0 and 1. In particular, let's consider two scenarios, in both p can take 10 different values, denoted by set A.

$$A = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\} \tag{2.1}$$

Though p can take the same 10 multiple values in both scenarios, we assign different probabilities to each possible value.

- Scenario 1:

$$f_1(A) = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1) \tag{2.2}$$

- Scenario 2:

$$f_2(A) = (0.05, 0.05, 0.05, 0.175, 0.175, 0.175, 0.175, 0.05, 0.05, 0.05) \tag{2.3}$$

To visually compare the values of two probability distributions $f_1(A)$ and $f_2(A)$, we plot $f_1(A)$ and $f_2(A)$ on the same graph as in Figure 2.1.

Figure 2.1: The same ten possible values of p, but two sets of probabilities.

Figure 2.1 labels the x-axis as the values of p (range from 0 to 1), y-axis as the probabilities (range from 0 to 1). For both panels, there are ten bars, each representing the possible values of p in the set $A = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

The probability assignment in $f_1(A)$ is called a discrete Uniform distribution, where each possible value of the proportion p is equally likely. Since there are ten possible values of p, each value gets assigned a probability of $1/10 = 0.1$. This assignment expresses the opinion that p can be any value from the set $A = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, and each value has a probability of 0.1.

The probability assignment in $f_2(A)$ is also discrete, however, we do not see a Uniform distribution pattern of the probabilities across the board. What we see is that the probabilities of the first three values (0.1, 0.2, and 0.3) and last three (0.8, 0.9, and 1.0) values of p are each 1/3.5 of that of the middle four (0.4, 0.5, 0.6, and 0.7) values. The shape of the bins reflects the opinion that the middle values of p are 3.5 times as likely as the extreme values of p.

Both sets of probabilities follow the three probability axioms in Chapter 1. One sees that within each set,

1. Each probability is nonnegative;
2. The sum of the probabilities is 1;
3. The probability of mutually exclusive values is the sum of probability of each value, e.g. probability of $p = 0.1$ or $p = 0.2$ is $0.1 + 0.1$ in $f_1(A)$, and $0.05 + 0.05$ in $f_2(A)$.

In this introduction, we have presented a way to think about proportions subjectively. We have introduced a way to allow multiple values of p, and perform probability assignments that follow the three probability axioms. One probability distribution expresses a unique opinion about the proportion p.

To answer our inference question "what is the proportion of red balls in the box", we will take a random sample of 10 balls, and use the observed proportion of red balls in that sample to sharpen and update our belief about p. Bayesian inference is a formal method for implementing this way of thinking and problem solving, including three general steps.

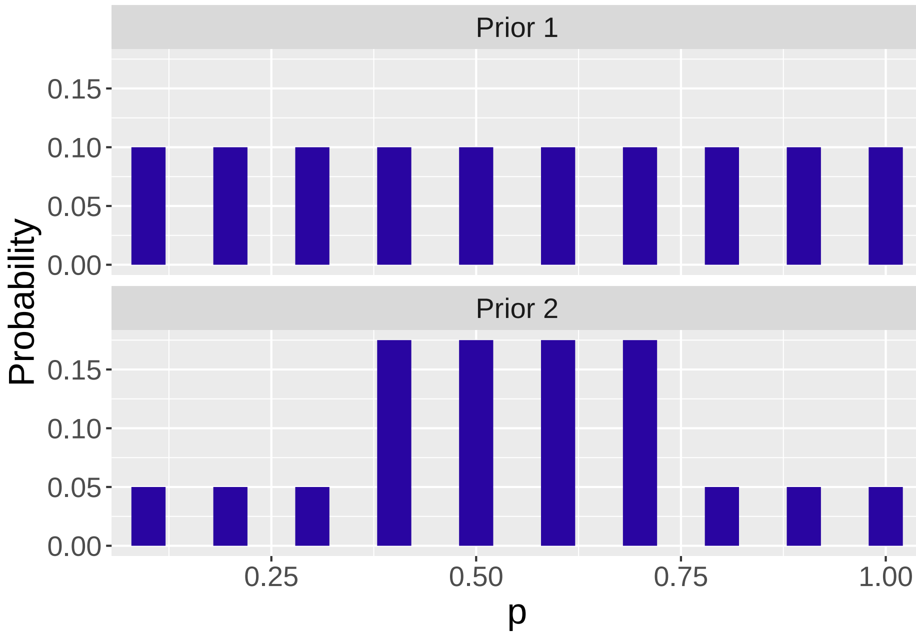
- Step 1: {(**Prior)}: express an opinion about the location of the proportion p before sampling.**
- **Step 2: {(Data/Likelihood)}: take the sample and record the observed proportion of red balls.**
- **Step 3: {(Posterior)}:use Bayes' rule to sharpen and update the previous opinion about p given the information from the sample.**

As indicated in the parentheses, the first step "Prior" constructs {*prior} opinion about the quantity of interest, and a probability distribution is used (like $f_1(A)$ and $f_2(A)$ earlier) to quantify the prior opinion. The name "prior" indicates that the opinion should be formed before collecting any data.*

The second step "Data" is the process of data collection, where the quantity of interest is observed in the collected data. For example, if our 10-ball sample contains 4 red balls and 6 white balls, the observed proportion of red balls is $4/10 = 0.4$. *Informally, how does this information help us sharpen one's opinion about p? Intuitively one would give more probability to* $p = 0.4$, *but it is unclear how the probabilities would be redistributed among the 10 values in A. Since the sum of all probabilities is 1, is it possible that some of the larger proportion values, such as* $p = 0.9$ *and* $p = 1.0$, *will receive probabilities of zero? To address these questions, the third step is needed.*

The third step "Posterior" combines one's prior opinion and the collected data to update one's opinion about the quantity of interest. Just like the example of observing 4 red balls in the 10-ball sample, one needs a structured way of updating the opinion from prior to posterior.

Throughout this chapter, the entire inference process will be described for learning about a proportion p. This chapter will discuss how to express prior opinion that matches with one's belief, how to extract information from the data/likelihood, *and how to update our opinion to its posterior, combining our prior and information from the data/likelihood in a principled way.*



Chapter 3

Literature

Here is a review of existing methods.

Chapter 4

Methods

We describe our methods in this chapter.

Chapter 5

Applications

Some significant *applications are demonstrated in this chapter.*

5.1 Example one

5.2 Example two

Chapter 6

Final Words

We have finished a nice book.