From Chapter 6 'Bernoulli and Bayes' of Elementary Bayesian Statistics
Gordon Antelman, Edward Elgar Publishing Limited, Cheltenham UK, 1997.

6.1 Introduction

In the last chapter, we saw how one can find (deductive) probabilistic answers to some possible questions about future outcomes of a Bernoulli process with a known parameter p. In this chapter, we shall study Bayesian methods for making (inductive) inferences about an *uncertain* Bernoulli p on the basis of prior knowledge about p and observed data from the Bernoulli process. Classical methods of inference about p will be discussed in Chapters 11 and 12.

Let me first informally introduce and illustrate a Bayesian analysis of an uncertain Bernoulli p with a slightly modified version of a novel and useful application of work sampling discussed by Fuller (1985).

Example WORK SAMPLING—I

Suppose you, as a good up-to-date manager practicing continuous quality and productivity improvement, have some ideas on improving your own productivity. To see if these ideas have any merit, you would like to compare some "before" measure of productivity with a comparable "after" measure of productivity.

For now—we shall come back to this example several times—let us focus on just a "before" measure.

The measure to be used is the proportion of your time spent in productive work, call it p, as opposed to time spent doing something that would not have needed doing if things had been done right the first time. Examples of the latter might include searching for a misplaced document, recreating a deleted computer file, following up on a customer's complaint, or waiting past a scheduled time for a meeting to start.

Since p is not precisely known, I shall emphasize this by topping it with a tilde: $\tilde{p}$. In problems such as this, it is often said that "p is unknown". It is better to say that "p is uncertain"; from your job experience, you would really know quite a lot about $\tilde{p}$. For example, you might be almost certain that it is greater than 0.50, less than 0.90, and you might assess your odds that p is between, say, 0.60 and 0.80 to be about 9 to 1. A precise statement of these beliefs will be your prior distribution for $\tilde{p}$.

You would probably feel uncomfortable—most people do—about assessing this prior distribution, especially since there are an infinite number of states; viz., all of the values between zero and one. But, without any real loss, you can bypass the infinite-number problem by rounding the values of p to the nearest 5% or 10%, making the problem discrete. Then you have a contemplatable Bayes' box, like those discussed in Chapter 4, with the finitely many p-values as the possible "states". (When we reconsider this example later in this chapter, you will see that, with a little theory, the infinite number of p-values can almost always be handled very neatly and more easily.)

For illustration, let us suppose you choose just five possible values for p: 0.50, 0.60, 0.70, 0.80, and 0.90, and assess your prior distribution of $\tilde{p}$ to be

Table 6.1: Prior Distribution for $\tilde{p}$

p	0.50	0.60	0.70	0.80	0.90
Prob($\tilde{p}$ = p)	0.05	0.25	0.35	0.30	0.05

This prior distribution would reflect, for example, that your judgment is that there is only about one chance in 20 that p rounds to 0.50, about one chance in 20 that p rounds to 0.90, about one chance in four that it rounds to 0.60, a little more than one chance in three that it rounds to 0.70, and a little less than one chance in three that it rounds to 0.80.

Now that we have the rows of a Bayes' box (the five p-values) and a prior probability mass function (pmf) for $\tilde{p}$ (the row marginal), let us consider the columns which, in general, represent "data values". Suppose you are fitted with a beeper set to beep at random times; when the beeper beeps, you classify the task being worked on as

- W —for productive Work, or
- F —for "Fixing".

Assume that this process can be modeled by a Bernoulli process with parameter p; i.e., each beep is one Bernoulli trial and p = P(W). Recall from Chapter 5 that the Bernoulli process postulates specify that p is unchanging and that, if p were known, the trials are independent; the experiment should be designed so that these requirements are plausible. For example, the beeps should be unpredictable so you do not arrange, possibly subconsciously, to be doing productive work at the beep; they probably also should not be too close together to make the independence assumption more reasonable.

Suppose the first four trials give the data F_1, F_2, W_3, and F_4. From the assumption that p remains the same for each trial,

$$P(W_3) = p, \quad P(F_1) = P(F_2) = P(F_4) = (1 - p) = q \text{ for short}$$

and, from the independence,

$$
\begin{aligned}
P(F_1\ F_2\ W_3\ F_4) & \\
&= P(F_1) \times P(F_2) \times P(F_2) \times P(W_4) \\
&= (1 - p) \times (1 - p) \times (1 - p) \times (1 - p) \\
&= (1 - p)^3\, p \\
&= q^3\, p
\end{aligned}
$$

Partial Bayes' boxes for your assessed prior and the data $F_1F_2W_3F_4$ from four trials (a sample of size four), are given in Figure 6.1. The top box there has just the input information for Bayes' theorem: the upper-right corners are the conditional probabilities $P(FFWF \mid p) = q^3p$ of the data given the state and the row marginal is the prior pmf.

The lower box of Figure 6.1 completes Bayes' theorem for these data, resulting in the posterior probabilities in the lower-left corners. The column sum 0.021305 is the marginal probability of observing the data FFWF—it is a weighted average of the upper-right-corner numbers with weights given by the prior probabilities; this marginal probability is the predictive probability of FFWF.

Because each trial has two possible outcomes, two trials would have 2 × 2 = 4 outcomes, three trials would have 2 × 2 × 2 = 8 outcomes, and four trials have 24 = 16 outcomes. Thus, a full Bayes' box for four trials would have 16 columns; but if you are only interested in the posterior for the data observed, you need only fill out the one column corresponding to that data.

Figure 6.2 is a picture showing the effects of the data FFWF on the prior distribution. The open squares give the prior probabilities and the solid circles the posterior probabilities. Three F's in four trials increase your probabilities for the two smaller possible p-values and decrease your probabilities for the three larger p-values.

The sample *alone* most strongly supports a value for p of 0.25 (one W in four trials); had the prior included a value of p of 0.25, the (relative) increase in going from prior to posterior would have been greatest for that value.

For the assumed prior, in which only p's of 0.50, 0.60, 0.70, 0.80, or 0.90 are considered, the sample evidence FFWF in favor of a p near 0.25 can only push up the posterior probabilities for the nearest possible p-values—0.50 and 0.60. This can be seen from either Figure 6.1 or Figure 6.2. (The seemingly harder consideration of all possible p's between zero and one will handle this kind of situation more logically.)

From Chapter 6 'Bernoulli and Bayes' of Elementary Bayesian Statistics

Gordon Antelman, Edward Elgar Publishing Limited, Cheltenham UK, 1997.

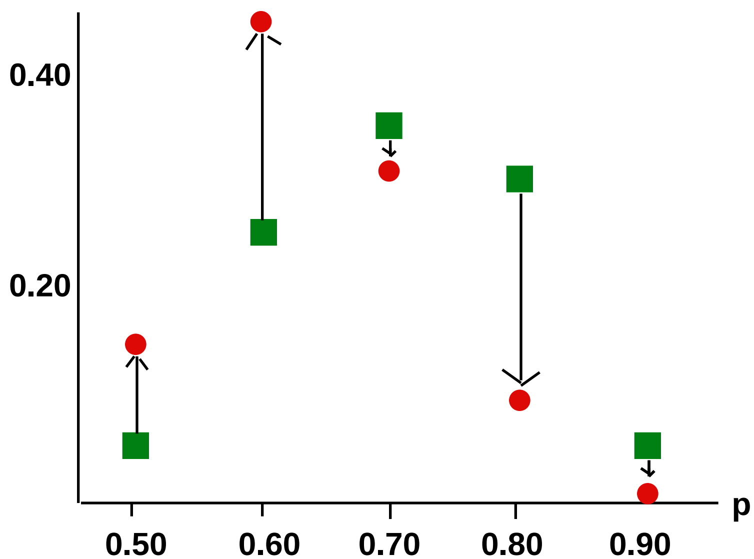
Right: Figure 6.1: Partial Bayes' Box for Three F's and One W

column	meaning / derived from...
p:	**p: proportion of time being productive**
prior prob:	**Prior Probability for each value of p**
Likelihood :	**Prob(data | proportion p)**
product:	**Prior Prob** × **Likelihood**
Post. prob	**Posterior probability** (product divided by sum of products)

Data Value FFWF ... ...

p		Likelihood	Prior Prob
0.50		$(0.5)^3(0.5)$	0.05
0.60		$(0.4)^3(0.6)$	0.25
0.70		$(0.3)^3(0.7)$	0.35
0.80		$(0.2)^3(0.8)$	0.30
0.90		$(0.1)^3(0.9)$	0.05
			1

Below: Figure 6.2 (Prior (Green Squares) and Posterior (Red Circles) probability mass functions for Three F's and One W



Data Value FFWF ... ...

p	Post. Prob	product	Likelihood	Prior Prob
0.50	0.147	0.003125	$(0.5)^3(0.5)$	0.05
0.60	0.451	0.009600	$(0.4)^3(0.6)$	0.25
0.70	0.310	0.006615	$(0.3)^3(0.7)$	0.35
0.80	0.090	0.001920	$(0.2)^3(0.8)$	0.30
0.90	0.002	0.000045	$(0.1)^3(0.9)$	0.05
	1.000	0.021305		1