# Bayesian Statistics, Assignment for Tuesday, Oct. 29

## Reading

Study to the end of Chapter 5 of our Bayesian statistics book.

## Review/Overview

Our authors (Donovan and Mickey) have written a friendly introductory textbook. Sometimes a friendly introduction such as theirs is too long for useful review, and also, due to the chattiness, it is easy to lose the forest for the trees. So let us recap. The new developments in Chapter 4 relative to Chapter 3 were not much. In Chapter 3, we already had Bayes' Theorem, and we had Venn diagrams as an alternative way of viewing the data in conjoint tables. In Chapter 4, the first thing we got was a rewritten version of Bayes' Theorem. The rewritten version depended on further uses of equations like

$\Pr(B) = \Pr(B \,\&\&\, A) + \Pr(B \,\&\&\, {\sim}A) = \Pr(B \mid A) * \Pr(A) + \Pr(B \mid {\sim}A) * \Pr({\sim}A)$

Review the derivation if you don't remember it.

The authors summarized the first advance in Chapter 4 at the bottom of p. 45:

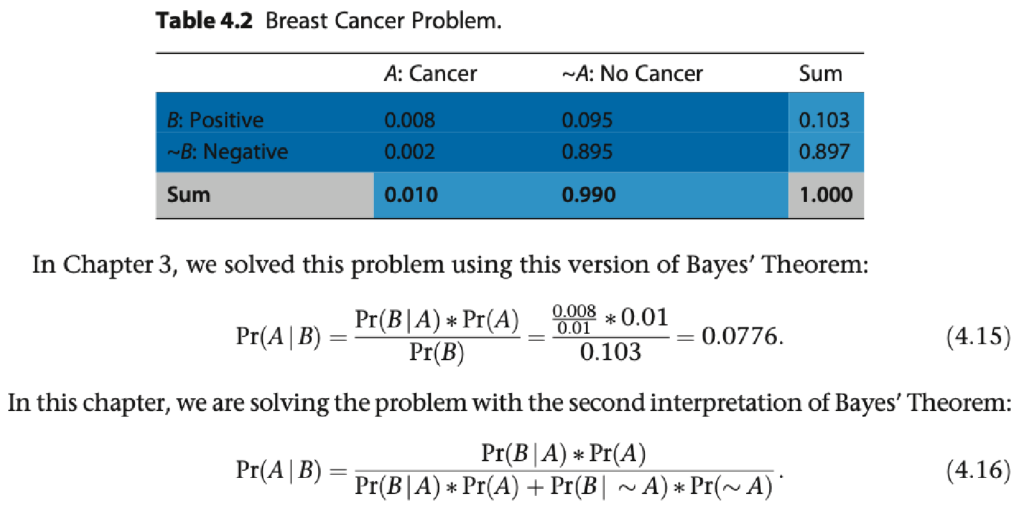


**Table 4.2** Breast Cancer Problem.

| | *A*: Cancer | ~*A*: No Cancer | Sum |
|---|---|---|---|
| *B*: Positive | 0.008 | 0.095 | 0.103 |
| ~*B*: Negative | 0.002 | 0.895 | 0.897 |
| **Sum** | **0.010** | **0.990** | **1.000** |

In Chapter 3, we solved this problem using this version of Bayes' Theorem:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) * \Pr(A)}{\Pr(B)} = \frac{\frac{0.008}{0.01} * 0.01}{0.103} = 0.0776. \tag{4.15}$$

In this chapter, we are solving the problem with the second interpretation of Bayes' Theorem:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) * \Pr(A)}{\Pr(B \mid A) * \Pr(A) + \Pr(B \mid \sim A) * \Pr(\sim A)}. \tag{4.16}$$

---

The next step in the logical development and application of Bayes' Theorem is to recognize that *A* can be anything that is true or false, and so it could be a hypothesis, *H* (maybe that is a leap of faith?). Meanwhile *B* could be some data (denoted with a script *O*, for "observations").

Then just changing the letters, we have another version of Equation 4.16:

$$\Pr(H \mid O) = \frac{\Pr(O \mid H) * \Pr(H)}{\Pr(O \mid H) * \Pr(H) + \Pr(O \mid \sim H) * \Pr(\sim H)}$$

The final development in Chapter 4 is to allow for many mutually-exclusive hypotheses, only one of which can be true. For example instead of just $H$ and $\sim H$, we could have three mutually-exclusive hypotheses, $H_1$, $H_2$, and $H_3$. Then we have:

$$\Pr(H_1 \mid O) = \frac{\Pr(O \mid H_1) * \Pr(H_1)}{\Pr(O \mid H_1) * \Pr(H_1) + \Pr(O \mid H_2) * \Pr(H_2) + \Pr(O \mid H_3) * \Pr(H_3)}$$

$$\Pr(H_2 \mid O) = \frac{\Pr(O \mid H_2) * \Pr(H_2)}{\Pr(O \mid H_1) * \Pr(H_1) + \Pr(O \mid H_2) * \Pr(H_2) + \Pr(O \mid H_3) * \Pr(H_3)}$$

$$\Pr(H_3 \mid O) = \frac{\Pr(O \mid H_3) * \Pr(H_3)}{\Pr(O \mid H_1) * \Pr(H_1) + \Pr(O \mid H_2) * \Pr(H_2) + \Pr(O \mid H_3) * \Pr(H_3)}$$

This is going to get clumsy if we have 15 mutually-exclusive hypotheses, so we just say there are $n$ of them and use our Σ-notation, and write:

$$\Pr(H_i \mid O) = \frac{\Pr(O \mid H_i) * \Pr(H_i)}{\sum_{j=1}^{n} \Pr(O \mid H_j) * \Pr(H_j)}$$

That is the final equation of Chapter 4 (Eq. 4.18 on p. 46). Note that $i$ and $j$ have quite different roles in this equation. The $i$ index is a “free” index, and it can be any value from 1 to $n$. The $j$ index is an “index of summation,” and it has to range over all $n$ values.

# For Problem Set 10

## The Author Problem

Multiple mutually-exclusive hypotheses is a nice touch, but for Chapter 5, Donovan and Mickey do an example with just two mutually-exclusive hypotheses: either $H$, which is going to be shorthand for “the essay was written by Hamilton,” or $\sim H$, which is going to be shorthand for “the essay was written by Madison.”

That describes the two hypotheses. Turning to the data, it is going to be the frequency (per thousand words) of “upon” in an essay.

For the Hamilton/Madison situation we just have the following two Bayes' Theorem equations:

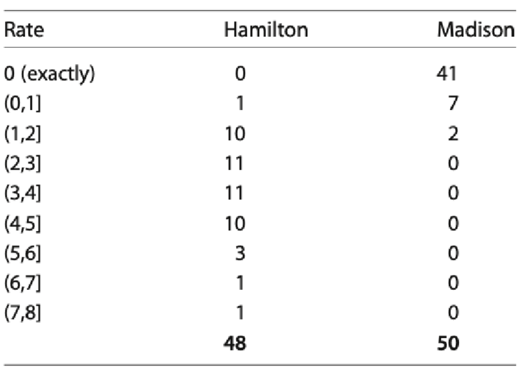
$$\Pr(H \mid O) = \frac{\Pr(O \mid H) * \Pr(H)}{\Pr(O \mid H) * \Pr(H) + \Pr(O \mid {\sim}H) * \Pr({\sim}H)}$$

$$\Pr({\sim}H \mid O) = \frac{\Pr(O \mid {\sim}H) * \Pr({\sim}H)}{\Pr(O \mid H) * \Pr(H) + \Pr(O \mid {\sim}H) * \Pr({\sim}H)}$$

The observation, *O*, is going to be one of nine possibilities: the frequency of "upon" per thousand words in an essay is 0; it is more than 0 but less than 1; it is more than 1 but less than 2; etc.; etc.; all the way up to, it is more than 7 but less than 8. There are no essays with more then 8 occurrences of "upon" per thousand words.

We first look at the 98 essays with known authors. Of those, 48 are known to be authored by Hamilton, and 50 are known to be authored by Madison. The observations for each of the 98 essays are summarized in a table:

**Table 5.2**



| Rate | Hamilton | Madison |
|---|---|---|
| 0 (exactly) | 0 | 41 |
| (0,1] | 1 | 7 |
| (1,2] | 10 | 2 |
| (2,3] | 11 | 0 |
| (3,4] | 11 | 0 |
| (4,5] | 10 | 0 |
| (5,6] | 3 | 0 |
| (6,7] | 1 | 0 |
| (7,8] | 1 | 0 |
| | **48** | **50** |

and in a figure:

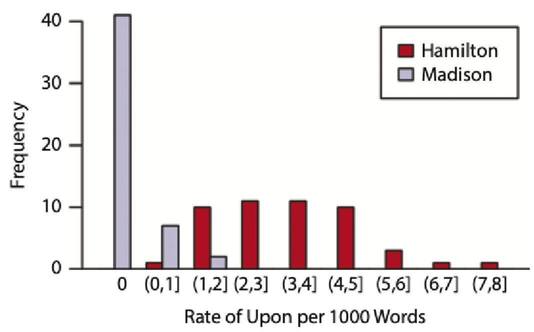


**Figure 5.3** Hamilton's and Madison's rates of "upon".

Donovan and Mickey use the sheer prolificness (48/98 of the total essays vs. 50/98 of the total essays, which is actually quite similar) to establish the “prior.” The prior is what we believe is probable if we have no additional data.

Now we add the data. In Paper 54 of the Federalist papers, whose authorship is or was unknown, there are 2 occurrences of “upon.” It is 2008 words long. That means that the frequency per thousand words is:

$$\frac{2}{2008} * 1000 = 0.996$$

which puts it in the [0, 1) observation bin.

What I want you to do for Problem Set 9 is to repeat the calculation on pp. 56-57. In other words, derive a new version of Eq. 5.20, and Eq. 5.21, and make a new version of Figure 5.5.

What is going to be different?!? I want you to assume that the essay had merely 16 fewer words (but still had 2 occurrences of “upon”). That means that the frequency per thousand words is:

$$\frac{2}{1992} * 1000 = 1.004$$

This puts it in a different observation bin!! It is now in the (1, 2] bin. You will see that you get a very different result when you rework the calculation with Paper 54 in the (1, 2] bin.

Allow me to summarize. I want you to start with:

$$\Pr(H \mid \text{observed frequency is in bin } (1, 2]) = \frac{\Pr(\text{observed frequency is in bin } (1,2] \mid H) * \Pr(H)}{\Pr(\text{observed frequency is in bin } (1,2] \mid H) * \Pr(H) + \Pr(\text{observed frequency is in bin } (1,2] \mid {\sim}H) * \Pr({\sim}H)}$$

$$\Pr({\sim}H \mid \text{observed frequency is in bin } (1, 2]) = \frac{\Pr(\text{observed frequency is in bin } (1,2] \mid {\sim}H) * \Pr({\sim}H)}{\Pr(\text{observed frequency is in bin } (1,2] \mid H) * \Pr(H) + \Pr(\text{observed frequency is in bin } (1,2] \mid {\sim}H) * \Pr({\sim}H)}$$

I want you to use 48/98 as your prior Pr($H$) and 50/98 as your prior Pr($\sim H$). I want you to get the other things you need, such as Pr(observed frequency is in bin (1 × 2] | $H$) by examining Table 5.2, and just to get you started, I will tell you that I get 10/48 for that particular probability. Get final results good to two decimal places and then make a Figure like Figure 5.5.

If having just 16 fewer words changes the result so much, then this whole calculation is fishy. I can explain why in person (this problem set writeup is already long enough). I hope you have, are having, or have had an awesome break. See you on Oct. 29.