Bayes's theorem and logistic regression

Allen B. Downey

July 14, 2025

Abstract

My two favorite topics in probability and statistics are Bayes's theorem and logistic regression. Because there are similarities between them, I have always assumed that there is a connection. In this note, I demonstrate the connection mathematically, and (I hope) shed light on the motivation for logistic regression and the interpretation of the results.

1 Bayes's theorem

I'll start by reviewing Bayes's theorem, using an example that came up when I was in grad school. I signed up for a class on Theory of Computation. On the first day of class, I was the first to arrive. A few minutes later, another student arrived. Because I was expecting most students in an advanced computer science class to be male, I was mildly surprised that the other student was female. Another female student arrived a few minutes later, which was sufficiently surprising that I started to think I was in the wrong room. When another female student arrived, I was confident I was in the wrong place (and it turned out I was).

As each student arrived, I used the observed data to update my belief that I was in the right place. We can use Bayes's theorem to quantify the calculation I was doing intuitively.

I'll us H to represent the hypothesis that I was in the right room, and F to represent the observation that the first other student was female. Bayes's theorem provides an algorithm for updating the probability of H:

$$\mathrm{P}(H|F) = \mathrm{P}(H)\ \frac{\mathrm{P}(F|H)}{P(F)}$$

Where

- $\mathrm{P}(H)$ is the prior probability of H before the other student arrived.
- $\mathrm{P}(H|F)$ is the posterior probability of H, updated based on the observation F.
- $\mathrm{P}(F|H)$ is the likelihood of the data, F, assuming that the hypothesis is true.
- $P(F)$ is the likelihood of the data, independent of H.

Before I saw the other students, I was confident I was in the right room, so I might assign $\mathrm{P}(H)$ something like 90%.

When I was in grad school most advanced computer science classes were 90% male, so if I was in the right room, the likelihood of the first female student was only 10%. And the likelihood of three female students was only 0.1%.

If we don't assume I was in the right room, then the likelihood of the first female student was more like 50%, so the likelihood of all three was 12.5%.

Plugging those numbers into Bayes's theorem yields $\mathrm{P}(H|F) = 0.64$ after one female student, $\mathrm{P}(H|FF) = 0.26$ after the second, and $\mathrm{P}(H|FFF) = 0.07$ after the third.

2 Logistic regression

Logistic regression is based on the following functional form:

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$$

where the dependent variable, p, is a probability, the xs are explanatory variables, and the βs are coefficients we want to estimate. The logit function is the log-odds, or

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

When you present logistic regression like this, it raises three questions:

- Why is $\text{logit}(p)$ the right choice for the dependent variable?
- Why should we expect the relationship between $\text{logit}(p)$ and the explanatory variables to be linear?
- How should we interpret the estimated parameters?

The answer to all of these questions turns out to be Bayes's theorem. To demonstrate that, I'll use a simple example where there is only one explanatory variable. But the derivation generalizes to multiple regression.

On notation: I'll use $\mathrm{P}(H)$ for the probability that some hypothesis, H, is true. $\mathrm{O}(H)$ is the odds of the same hypothesis, defined as

$$\mathrm{O}(H) = \frac{\mathrm{P}(H)}{1 - \mathrm{P}(H)}$$

I'll use $\mathrm{LO}(H)$ to represent the log-odds of H:

$$\mathrm{LO}(H) = \ln \mathrm{O}(H)$$

I'll also use LR for a likelihood ratio, and OR for an odds ratio. Finally, I'll use LLR for a log-likelihood ratio, and LOR for a log-odds ratio.

3 Making the connection

To demonstrate the connection between Bayes's theorem and logistic regression, I'll start with the odds form of Bayes's theorem. Continuing the previous example, I could write

$$\mathrm{O}(H|F) = \mathrm{O}(H)\ \mathrm{LR}(F|H) \tag{1}$$

where

- $\mathrm{O}(H)$ is the prior odds that I was in the right room,
- $\mathrm{O}(H|F)$ is the posterior odds after seeing one female student,
- $\mathrm{LR}(F|H)$ is the likelihood ratio of the data, given the hypothesis.

The likelihood ratio of the data is:

$$\mathrm{LR}(F|H) = \frac{\mathrm{P}(F|H)}{\mathrm{P}(F|\neg H)}$$

where $\neg H$ means H is false.

Noticing that logistic regression is expressed in terms of log-odds, my next move is to write the log-odds form of Bayes's theorem by taking the log of Eqn 1:

$$\mathrm{LO}(H|F) = \mathrm{LO}(H) + \mathrm{LLR}(F|H) \tag{2}$$

If the first student to arrive had been male, we would write

$$\mathrm{LO}(H|M) = \mathrm{LO}(H) + \mathrm{LLR}(M|H)$$

Or more generally if we use X as a variable to represent the sex of the observed student, we would write

$$\mathrm{LO}(H|X) = \mathrm{LO}(H) + \mathrm{LLR}(X|H) \tag{3}$$

I'll assign $X = 0$ if the observed student is female and $X = 1$ if male. Then I can write:

$$\mathrm{LLR}(X|H) = \begin{cases} \mathrm{LLR}(F|H) & \text{if } X = 0 \\ \mathrm{LLR}(M|H) & \text{if } X = 1 \end{cases}$$

Or we can collapse these two expressions into one by using X as a multiplier:

$$\mathrm{LLR}(X|H) = \mathrm{LLR}(F|H) + X[\mathrm{LLR}(M|H) - \mathrm{LLR}(F|H)] \tag{4}$$

4 Odds ratios

The next move is to recognize that the part of Eqn 4 in brackets is the log-odds ratio of H. To see that, we need to look more closely at odds ratios.

Odds ratios are often used in medicine to describe the association between a disease and a risk factor. In the example scenario, we can use an odds ratio to express the odds of the hypothesis H if we observe a male student, relative to the odds if we observe a female student:

$$\mathrm{OR}_X(H) = \frac{\mathrm{O}(H|M)}{\mathrm{O}(H|F)}$$

I'm using the notation OR_X to represent the odds ratio associated with the variable X.

Applying Bayes's theorem to the top and bottom of the previous expression yields

$$\mathrm{OR}_X(H) = \frac{\mathrm{O}(H)\ \mathrm{LR}(M|H)}{\mathrm{O}(H)\ \mathrm{LR}(F|H)} = \frac{\mathrm{LR}(M|H)}{\mathrm{LR}(F|H)}$$

Taking the log of both sides yields

$$\mathrm{LOR}_X(H) = \mathrm{LLR}(M|H) - \mathrm{LLR}(F|H) \tag{5}$$

This result should look familiar, since it appears in Eqn 4.

5 Conclusion

Now we have all the pieces we need; we just have to assemble them. Combining Eqns 4 and 5 yields

$$\mathrm{LLR}(H|X) = \mathrm{LLR}(F) + X\ \mathrm{LOR}(X|H) \tag{6}$$

Combining Eqns 3 and 6 yields

$$\mathrm{LO}(H|X) = \mathrm{LO}(H) + \mathrm{LLR}(F|H) + X\ \mathrm{LOR}(X|H) \tag{7}$$

Finally, combining Eqns 2 and 7 yields

$$\mathrm{LO}(H|X) = \mathrm{LO}(H|F) + X\ \mathrm{LOR}(X|H)$$

We can think of this equation as the log-odds form of Bayes's theorem, with the update term expressed as a log-odds ratio. Let's compare that to the functional form of logistic regression:

$$\mathrm{logit}(p) = \beta_0 + X\beta_1$$

The correspondence between these equations suggests the following interpretation:

- The predicted value, $\mathrm{logit}(p)$, is the posterior log odds of the hypothesis, given the observed data.
- The intercept, β_0, is the log-odds of the hypothesis if $X = 0$.

- The coefficient of X, β_1, is a log-odds ratio that represents odds of H when $X = 1$, relative to when $X = 0$.

This relationship between logistic regression and Bayes's theorem tells us how to interpret the estimated coefficients. It also answers the question I posed at the beginning of this note: the functional form of logistic regression makes sense because it corresponds to the way Bayes's theorem uses data to update probabilities.