some loose notes for basic introduction to Bayesian statistics

Aleksandra Kaszowska

What are we doing here?

Bayesian reasoning, and Bayesian statistics, are a different way of looking at theories and data. That difference starts with the philosophical disagreement of how do we define probability, and ends with wildly different mathematical approaches to analysis.

They are not new – Bayes theorem was first published in 1761, and for comparison Fisher published his first work on statistical methods in 1925. But they are new to us in this class today.

Bayesian reasoning is coming into psychology now for two reasons:

1) The general dissatisfaction with frequentist approaches, driven by the replication crisis

2) Software packages and tools for Bayesian analysis are becoming more widely available

This one lecture on Bayesian inference is supposed to provide you with a very basic introduction, so that when you see Bayesian inference in action, you have a basic idea of what is happening there. But nobody is expecting you to actually do Bayesian analysis after today.

Why are Bayesians cool?

Prior means "before more evidence". Posterior means "after more evidence".

For example:

- Prior: I think there is no such thing as a "bad" pizza topping
- More evidence: trying out everything in my pantry as a pizza topping
- Posterior: I tasted oranges on a pizza and I no longer thing there is no such thing as a "bad" pizza topping

What is Bayesian inference?

In this example, A is: I will teach statistics today, and B is: I will drink beer tonight.

If I assume teaching statistics and drinking beer are two independent events, then probability of me teaching statistics has no bearing on the probability of me drinking beer.

Product rule: I will drink beer tonight regardless of whether I taught statistics or not. If the probability of me teaching statistics and drinking beer on the same day is the product (an outcome of multiplication) of the probability of me teaching statistics, and probability of me drinking beer tonight.

Sum rule: If I have two options – I will either teach statistics today or I will not; and two beer options – I will either drink beer tonight or not; then my total probability of drinking beer is a sum of probability of me drinking beer after teaching statistics and me drinking beer without teaching statistics.

Bayesian probability of a hypothesis being true given the observed data

This equation is saying that the probability of hypothesis being true in light of observed data is a ratio of two things:

1) Product of multiplication of two probabilities: Probability that my hypothesis is true, and probability of observing data X given the hypothesis is true

2) And sum of all probabilities of observing data X: so observing X when my hypothesis is true, and observing X when my hypothesis is not true

And if I have only two hypotheses...

Hypothesis M: people drink beer after they teach statistics

Data X: Aleks drank beer tonight

So then we are left with Bayesian probabilities of two hypotheses:

1) People drink beer after teaching statistics given that/because Aleks drank beer tonight

2) People do not drink beer after teaching statistics given that/although Aleks drank beer tonight

The ratio of those two hypotheses is a quantification of how much more we believe in one over the other. So now that we know that Aleks taught statistics today and also drunk beer tonight, here we have three to one odds that people drink because they teach statistics.

Bayes factor is the ratio of probability that observed data is true given the hypothesis, compared to probability of observed data being true given not-hypothesis. So probability of Aleks drinking beer because people drink beer after teaching statistics, and probability of Aleks drinking beer not because people drink beer after teaching statistics.

Interpreting the Bayes factor

Notation: BF10 is a ratio of evidence in favor of hypothesis being true divided by evidence in favor of hypothesis not being true (event happening or not happening).

If we have more belief in H1_1 than in H1_0, then we get numbers above 1; so if we have more belief in the hypothesis that teaching statistics makes people drink beer than in the hypothesis that teaching statistics does not make people drink beer, then we get numbers above 1.

If we have more belief in H1_0 than in H1_1, then we get numbers below 1; so if we have more belief in event not happening than event happening, then we get numbers below 1. If we have more belief that teaching statistics does not make people drink beer compared to teaching statistics making people drink beer.

You can flip it too: that's why the little 10 or 01 matter a lot.

Bayes factor vs posterior probabilities

Bayes factor: how much did Aleks drinking beer tonight prove the point that people drink beer after teaching statistics?

Posterior probabilities: how much do we now think that people drink beer because they teach statistics after we saw Aleks in a bar, compared to our belief that people do not drink beer after teaching statistics given that we saw Aleks in a bar?

It is possible that in our experiment, Bayes factor strongly implied that Aleks drinks beer because she teaches statistics. But it is still possible that the posterior probability is higher: the total evidence we have that teaching statistics does not drive people to drink. For example, because we have never before seen other instructors drinking beer after they taught statistics.