Bayesian Statistics

August 8, 2022

Contents

1 Building a model

Designing a simple Bayesian model benefits from a design loop with 3 steps:

1. Data story: Motivate the model by narrating how the data might arise.

2. Update: Educate your model by feeding it the data.

3. Evaluate: All statistical models require supervision, leading to model revision.

1.1 Data story

Bayesian data analysis usually means producing a story for how the data came to be. Such a story may be descriptive, i.e. it specifies associations that can be used to predict outcomes, given observations. Or it may be causal, a theory of how some events produce other events. You can motivate your data story by trying to explain how each piece of data is born.

1.2 Update

A Bayesian model begins with one set of plausibilities assigned to each possible parameters. These are the prior plausibilities. Then, it updates them in light of the data, to produce the posterior plausibilities. This updating process is a kind of learning, called Bayesian updating.

1.3 Evaluate

The model's certainty is no guarantee that the model is a good one. Models of all sorts can be very confident about an inference, even when the model is seriously misleading.

2 Components of the model

1. The number of ways each conjecture could produce an observation
2. The accumulated number of ways each conjecture could produce the entire data
3. The initial plausibility of each conjectured cause of the data

2.1 Variables

Variables Symbols that can take on different values.

There could be variables for things we wish to infer and things we might observe. Unobserved variables are usually called parameters. We want to infer parameters from the other (observed) variables.

2.2 Definitions

We need to define the variables. In defining each, we build a model that relates the variables to one another.

2.2.1 Observed variables

Likelihood A distribution function assigned to an ovserved variable.

For the count of water W and land L, we define how plausible any combiniation of W and L would be, for a specific value of p. If we assume that each toss is independent of the other tosses and the probability of W is the same on every toss, we have

$$Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L.$$

For example, $P(W = 6, L = 4|p = 0.5) = 0.164$. This value, 0.164, is the relative number of ways to get six water, holding p at 0.5 and $N = W + L$ at nine.

2.2.2 Unobserved variables

For every parameter you intend your Bayesian machine to consider, you must provide a distribution of prior plausibility, its prior. A Bayesian machine must have an initial plausibility assignmnet for each possible value of the parameter, and these initial assignments do useful work. Finding the priors are both engineering assumptions and scientific assumption.

- engineering assumption, chosen to help the machine to learn
- scientific assumption, chosen to reflect what we know about a phenomenon

There is not law mandating we use only one prior. If you don't have a strong argument for any particular prior, then try different ones.

2.2.3 A model is born

$$W \approx \text{Binomial}(N, p)$$

where $N = W + L$.

$$p \approx \text{Uniform}(0, 1)$$

2.3 Making the model go

A Bayesian model can update all of the prior distributions to their purely logical consequences: the posterior distribution. This distribution contains the relative plausibility of different parameter values, conditional on the data and the model, denoted as $Pr(p|W, L)$.

2.3.1 Bayes' theorem

The joint probability of the data W and L and any particular value of p is

$$Pr(W, L, p) = Pr(W, L|p)Pr(p), \quad Pr(W, L, p) = Pr(p|W, L)Pr(W, L)$$

$$Pr(p|W, L) = \frac{Pr(W, L|p)Pr(p)}{Pr(W, L)}.$$

The probability of any particular value of p, considering the data, is equal to the product of the relative plausibility of the data, conditional on p and the prior plausibility of p, divided by $Pr(W, L)$, the average probability of the data (evidence/average liklihood). The job of this term (evidence) is to standardize the posterior.

$$Pr(W, L) = E_p(Pr(W, L|p)) = \int Pr(W, L|p)Pr(p)dp.$$