

Statistics

osimpson

April 24, 2021

1 Bayesian Statistics

Theorem 1. (*Bayes*)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

1.1 Parameter Estimation with Bayes Rule

Let's rewrite Bayes' rule as such:

$$P(\Theta|\text{data}) = \frac{P(\text{data}|\Theta)P(\Theta)}{P(\text{data})}, \quad (2)$$

where Θ represents parameters of interest, and data are observed samples. Our goal is to estimate Θ given the observed samples.

Let's establish some terminology that statisticians use a lot:

- $P(\Theta|\text{data})$ is the *posterior distribution*
- $P(\Theta)$ is the *prior distribution* - it represents prior intuition about the parameters without considering the data
- $P(\text{data}|\Theta)$ is the *likelihood*, the distribution of sample outcomes for all possible values of Θ

The denominator $P(\text{data})$ can be thought of as a normalizing constant to ensure that the posterior is indeed a distribution.

Now, given this information, Bayes' rule tells us that

$$P(\Theta|\text{data}) \propto P(\text{data}|\Theta)P(\Theta), \tag{3}$$

which gives us a way to estimate $P(\Theta|\text{data})$. Namely, given our assumed prior, we can update our probabilities with information from observed data. Essentially, we are updating our prior intuitions with information from our sample.

The specifics of how updates work depend on choice of prior, and the form of the posterior. We have some autonomy here as the practitioner. For example, we may choose a prior whose parametric form is retained through to the posterior. This type of prior is called a *conjugate prior*.