Bayesian Coin Flipping

Elements of Bayesian inference

We return to the coin flipping experiment. The ingredients of our Bayesian analysis of this situation are:

- a statistical model. We assume that our coin is modelled by a Bernoulli random variable with parameter p of returning heads. The likelihood of getting h heads in N flips is given by the binomial distribution

$$P(h|p) = \binom{N}{h} p^h (1-p)^{N-h}.$$

- a prior distribution $P(p)$. Initially, we make no assumptions about the coin, so we choose the *uniform distribution* that assigns probability density 1 to every $p \in [0, 1]$.
- some data D. We flip the coin N times and receive h heads; that's our data.

Our problem is to construct a posterior distribution $P(p|h)$ that tells us how this experiment updates our impressions about the coin.

Bayes's theorem

From our setup and Bayes's theorem:

$$P(p|h) = \frac{P(h|p)P(p)}{P(h)} = \frac{\binom{N}{h}p^h(1-p)^{N-h}}{P(h)}$$

where the denominator is

$$P(h) = \binom{N}{h}\int_{p=0}^{1} p^h(1-p)^{N-h}dp$$

The posterior

The posterior distribution, up to a constant A, is

$$P(p|h) = Ap^h(1-p)^{N-h}$$

We know from our discussion of maximum likelihood that the *most likely* value of p is h/N, the fraction of heads among all flips. This is called the *maximum a posteriori estimate* or MAP.

The posterior mean

In Bayesian inference, one often uses the *mean of the posterior distribution* as a better summary of the posterior than the point where the posterior is a maximum. To compute the mean, we need to know the constant A, which is

$$A = \frac{1}{\int_{p=0}^{1} p^h (1-p)^{N-h} dp}$$

The mean of the posterior is given by the formula

$$E[p|h] = A \int_{p=0}^{1} p^{h+1} (1-p)^{N-h} dp$$

The *Beta Integral* is the integral

$$B(a,b) = \int_{p=0}^{1} p^{a-1} (1-p)^{b-1} dp$$

and with some work one can show that

$$B(a,b) = \frac{a+b}{ab} \frac{1}{\binom{a+b}{a}}$$

Putting this all together gives the result

$$E[p|h] = \frac{h+1}{N+2}$$

Some numbers

- Given 55 heads out of 100 flips, the maximum likelihood estimate for p (and the maximum a posteriori estimate assuming a uniform prior) is $p = .55$.
- The posterior mean is $56/102 = .549$ which is a bit less.