Deriving Binomial MLE and Beta-Binomial Posterior.

Elena Moltchanova

17 Aug 2021

MLE

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$L = \log f(y|p) = \log \binom{n}{y} + y \log p + n - y \log(1-p)$$

Differentiating with respect to p:

$$\frac{dL}{dp} = \frac{y}{p} - \frac{n-y}{1-p} = 0$$

multiplying both sides by p (providing $0 < p < 1$):

$$(1-p)y - (n-y)p = y - py - np + py = y - np = 0.$$

Therefore $\hat{p} = \frac{y}{n} = \frac{\sum x_i}{n} = \bar{x}$, where x_i are the individual binary outcomes.

Differentiating for the second time:

$$\frac{d^2L}{dp^2} = -\frac{y}{p^2} - \frac{n-y}{(1-p)^2} = -\left(\frac{y}{p^2} + \frac{n-y}{(1-p)^2}\right) < 0$$

if $0 < y < n$.

** Another way is to look at the values of the likelihood at the endpoints and the candidate point:**

$$f(y|p=0) = \binom{n}{y} 0^y (1-0)^{n-y} = 0 \quad \text{for } y > 0$$

$$f(y|p=1) = \binom{n}{y} 1^y (1-1)^{n-y} = 0 \quad \text{for } y < n$$

$$f(y|p=\hat{p}) = \binom{n}{y} \hat{p}^y (1-\hat{p})^{n-y} > 0 \quad \text{for } 0 < y < n$$

Thus, the maximum is reached at $\hat{p}$.

Note, that the classical maximum likelihood estimation breaks down when $y = 0$ or $y = n$. A number of "corrections" exist in the literature. Most of them add small numbers to both, the numerator and the denominator of the y/n ratio to ensure that the estimated proportion $\hat{p}$ is such that $0 < \hat{p} < 1$.

Bayesian way

Let $y|p \sim Bin(n,p)$. Then the likelihood is

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y} \propto p^y (1-p)^{n-y}.$$

(Remember, to make use of proportionality, we will not be using multiplicative terms not containing the parameter of interest).

Let $p \sim B(a,b)$ with the p.d.f.:

$$f(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \propto p^{a-1}(1-p)^{b-1} \quad \text{for } p \in (0,1).$$

Note, that when $a = b = 1$, this becomes:

$$f(p) = \frac{1}{B(1,1)} p^{1-1}(1-p)^{1-1} = 1 \quad \text{for } p \in (0,1).$$

In other words, a uniform distribution.

We can then use Bayes' formula to derive the posterior distribution of the parameter p given data y:

$$\begin{aligned} f(p|y) &\propto f(y|p)f(p) \propto p^y(1-p)^{n-y}p^{a-1}(1-p)^{b-1} \\ &= p^{a+y-1}(1-p)^{b+n-y-1}. \end{aligned}$$

Notice, that the derived equation is a product of powers of p and $(1-p)$ respectively, which is a hallmark of the Beta density. Thus, we deduce that

$$p|y \sim B(a+y, b+n-y).$$

So, if we observe the outcomes of our binomial trials one at a time, we will just keep adding 'successes' to the first parameter, and 'failures' to the second one.