5.3 Three Inferential Goals

Zongyi Liu

2023-06-11

5.3 Three Inferential Goals

5.3.1 Estimating the Binomial Proportion

The posterior distribution over θ tells us exactly how much we believe in each possible value of θ. When the posterior is a beta distribution, we can plot the detailed graph of the distribution.

This plot shows examples of posterior beta distributions.

- They use the same data, so they have the same likelihood plot
- They have different priors, thus have different posteriors
- The 95% HDI is much wider when the prior is uncertain

When $\theta = 0.5$, we can say that the coin is fair. We first establish a **region of practical equivalence (ROPE)** around the value of interest, which means a small interval such that any value within the ROPE is equivalent to the value of interest for all practical purposes.

If we want to declare a reasonably credible region, we can utilize the HDI.

There are two roles of the HDI:

- As a summary of the distribution
- For deciding whether a value of interest is credible

5.3.2 Predicting Data

The belief in the parameter values, $p($
theta), is the current posterior belief, which we can indicate as $p(\theta \| z, N)$. Because $p(y=1|\theta) = \theta$, thus:

$$
\begin{aligned}
p(y=1) &= \int d\theta \, p(y=1|\theta) \, p(\theta|z,N) \\
&= \int d\theta \, \theta \, p(\theta|z,N) \\
&= \overline{\theta}|z,N \\
&= (z+a)/(N+a+b)
\end{aligned}
$$

Thus we can say that the predicted probability of heads is just the mean of the posterior distribution over θ. The predicted probability of getting a head on the next flip is somewhere between the prior mean and the proportion of heads in the flips observed so far.

We can think about it by considering a particular prior and sequence of flips.

Let's make that concrete by considering a particular prior and sequence of flips. Sup- pose that we start with a uniform prior, i.e., $beta(\theta; 1, 1)$. We flip the coin once, and get a head. The posterior is then $beta(\theta; 2, 1)$, which has a mean of 2/3. Thus, after the first flip comes up heads, the predicted probability of heads on the next flip is 2/3. Suppose we flip the coin a second time, and again get a head. The posterior is then $beta(\theta; 3, 1)$, and the predicted probability of heads on the next flip is 3/4. Notice that even though we have flipped the coin twice and observed heads both times, we do not predict that there is 100% chance of coming up heads on the next flip, because the uncertainty of the prior beliefs is mixed with the observed data.

5.3.3 Model Comparison

We would like to use results to compare models. To do this, we need to compute the evidence, $p(D|M)$, for each model.

In the present scenario, the data D are expressed by the values z and N. When using a Bernoulli likelihood and a beta prior, then the evidence $p(D|M)$ is $p(z, N)$ and it is especially easy to compute. In Equation 5.7, the denominator showed that

$$B(a, b)p(z, N') = B(z + a, N - z + b)$$

Solving for $p(z, N)$ reveals that

$$p(z, N) = B(z + a, N - z + b)/B(a, b)$$

The lower panels of Figure 5.2 show the values of $p(z, N)$ for two different priors, for a fixed data set $z = 11$, $N = 14$. One prior is uniform, while the other prior is strongly peaked over $\theta = .50$. The data have a proportion of 1's that is not very close to .5, and therefore the prior that is peaked over 0.5 does not capture the data very well. The peaked prior has very low belief in values of θ near the data proportion, which means that $p(z, N)$ for the peaked-prior model is relatively small.

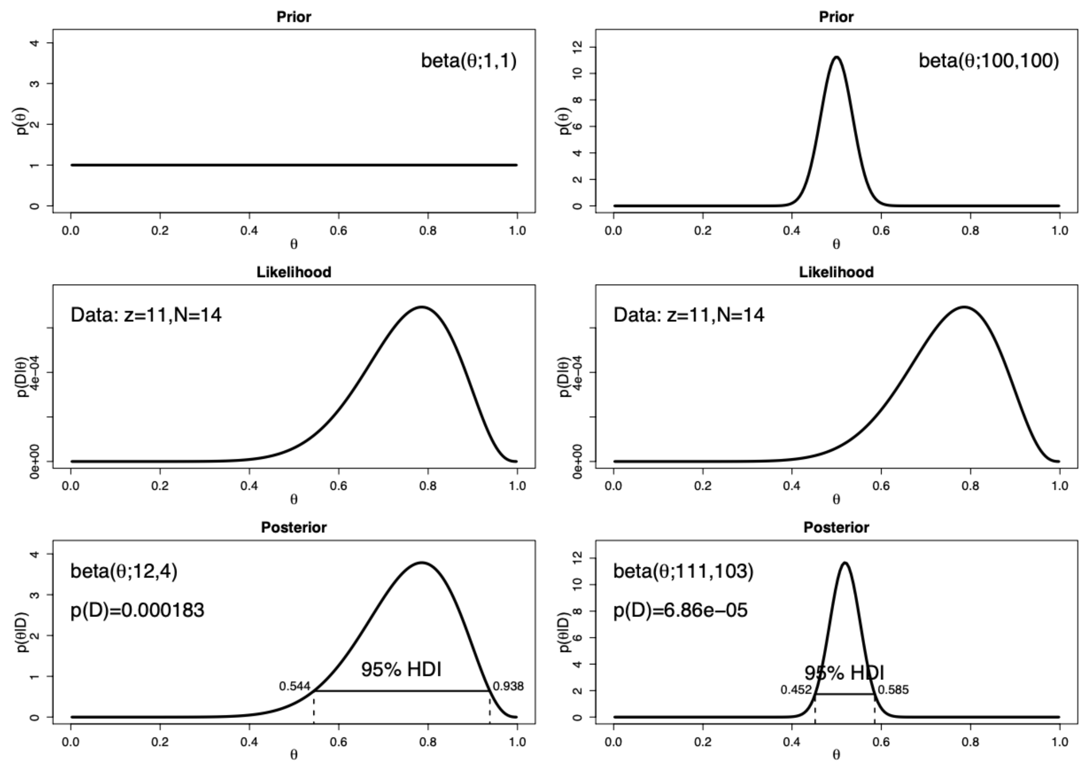

Figure 5.2: Two different beta priors, updated using the same data. (The R code that generated these graphs is in Section 5.5.1)

For another case in which half the flips were heads ($z = 7$ and $N = 14$), we can see the plot below:

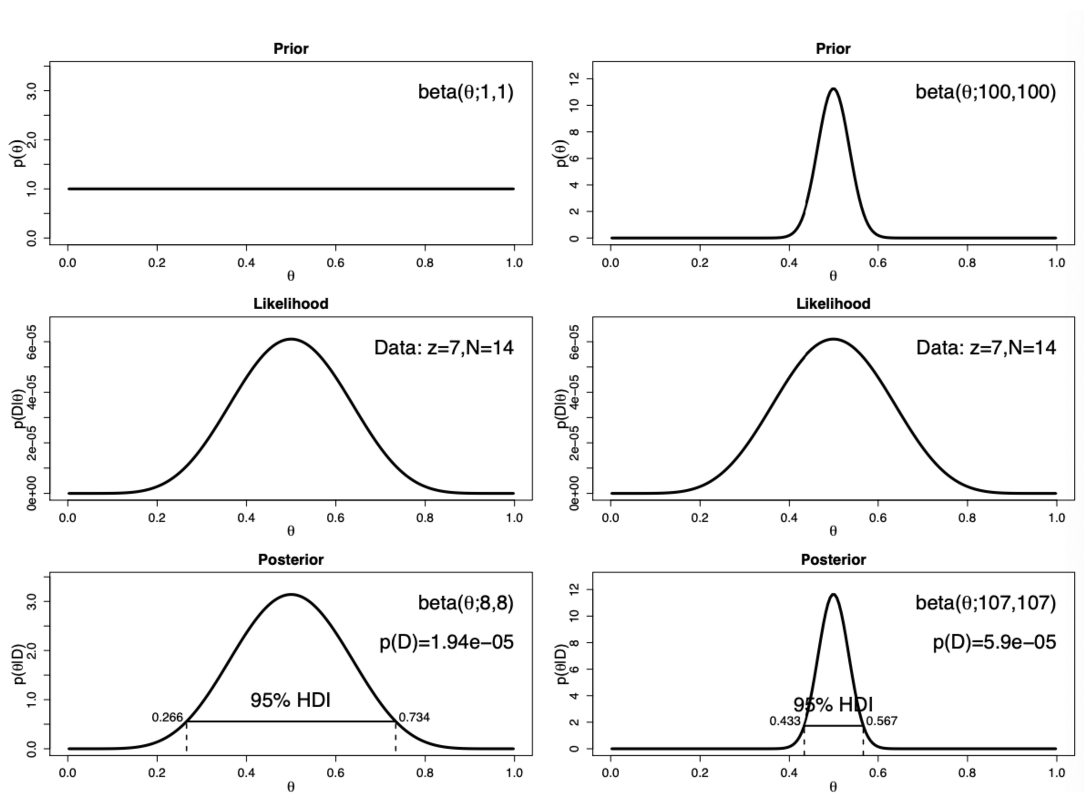

Figure 5.3: Two different beta priors, updated using the same data ($z = 7, N = 14$). The evidences are denoted by $p(D)$ (with M suppressed) in the lower panels of each column, and they favor the prior that is peaked over $\theta = 0.5$. Contrast with Figure 5.2, for which the evidences favored the uniform prior.

We prefer the model with the higher value of $p(D|M)$, but the preference is not absolute. A tiny advantage in $p(D|M)$ should not be translated into a strong preference for one model over another. After all, the data themselves are just a random sample from the world, and they could have been somewhat different. It is only when the relative magnitudes of $p(D|M)$ are very different that we can feel confident in a preference of one model over another.

5.3.3.1 Is the Best Model a Good Model?

If we have two models and some data, and find that the evidence for one model is much better than the other one. If the prior beliefs were equal, then the posterior beliefs.

However, the model comparison just told us models' **relative believabilities**, not **absolute believabilities**.