12.1 Bayesian Approach to Testing a Point (Null) Hypothesis

Zongyi Liu

2023-06-25

In previous chapter we discussed about the **null hypothesis significance testing (NHST)**. This chapter is mainly deal with Bayesian approach to testing a point (null) hypothesis.

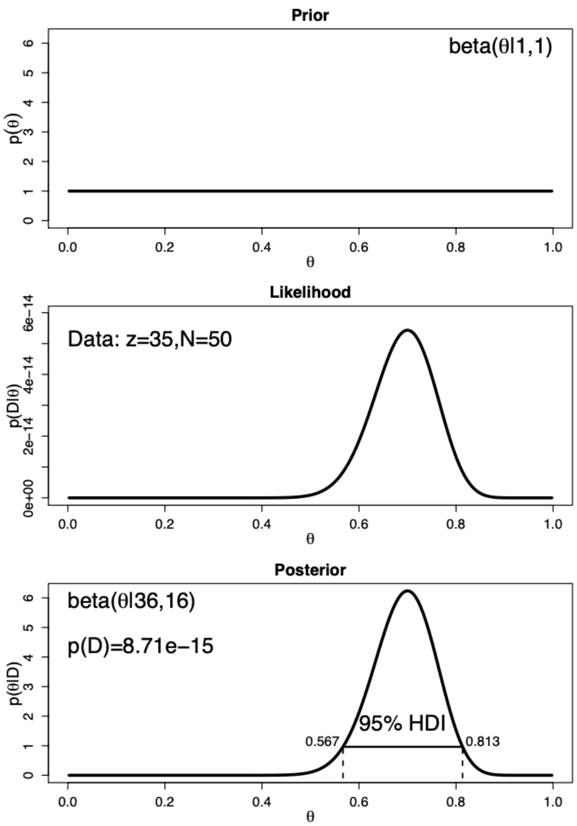
12.1 The Estimation Approach

This approach has been used in many occasions in this and other books. It goes roughly like:

> A parameter value is declared to be not credible if it lies outside the 95% HDI of the posterior distribution of that parameter. If a parameter value lies within the 95% HDI, it is said to be among the credible values.

12.1.1 Is a Null Value of a Parameter Among the Credible Values?

There are many cases in which we wanted to know whether a "coin bias" of 0.5 was credible. After learning, there were some test cases in which conflicting cue-words were displayed together. For each type of conflicting case, the researcher wanted to know if people were responding randomly, or instead had a bias away from 50/50 responding. The (fictitious but realistic) data showed that of 50 respondents, 35 gave response A whereas 15 gave response B. If the researcher begins with a uniform prior, then the posterior is a $beta(\theta|36, 16)$ distribution, which has a 95% HDI from 0.567 to 0.813, as shown in figure below.

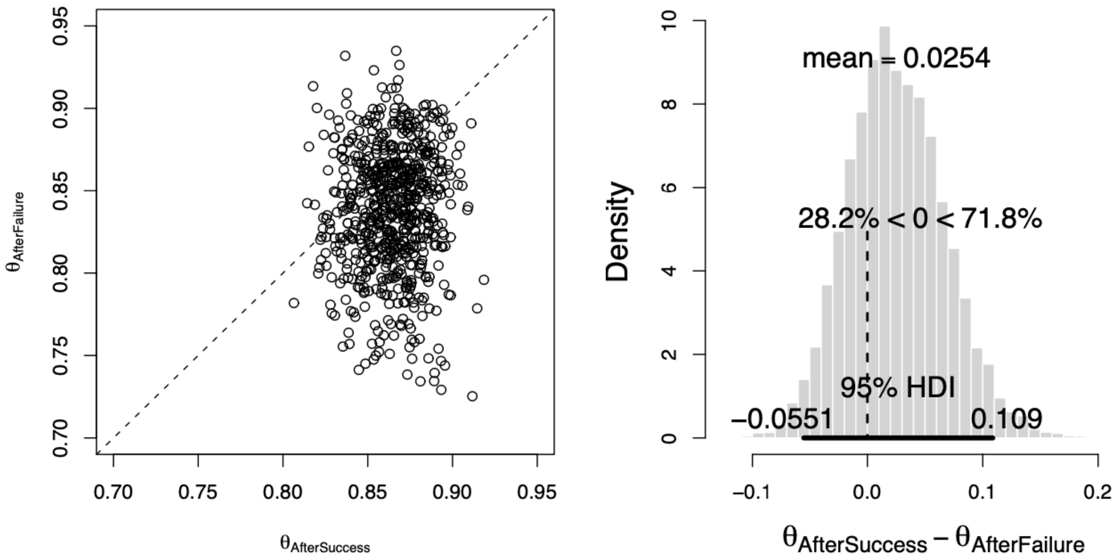


Another example, in which we considered whether a 50/50 bias was credible, was the investigation of therapeutic touch, in which researchers tested whether therapists could detect the presence of another person's hand several centimeters away. Of primary interest in the analysis was the estimate of the overall accuracy of the group of therapists.

12.1.2 Is a Null Value of a Difference Among the Credible Values?

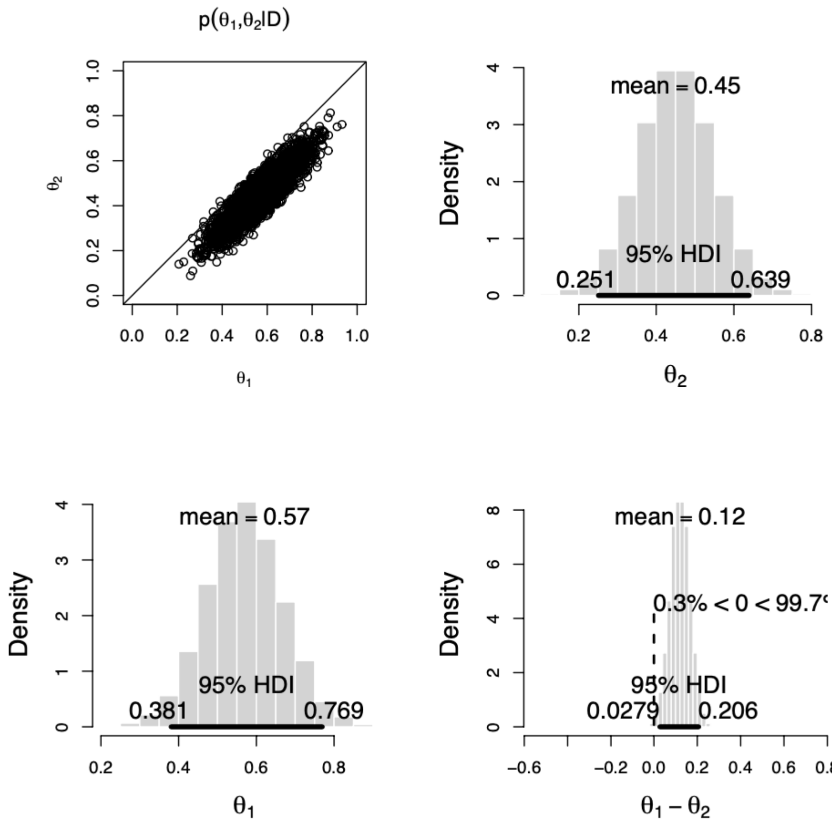
We have also already seen situations in which we wanted to know whether a non-zero difference of parameter values was credible.

The case is the Exercise 8.1, in which the posterior estimate of the difference in proportions is shown in Figure below, where it can be seen that the 95% HDI goes from -0.0551 to +0.109. The researcher would conclude that there may well be no hot hand phenomenon, but non-zero differences within the HDI are credible.

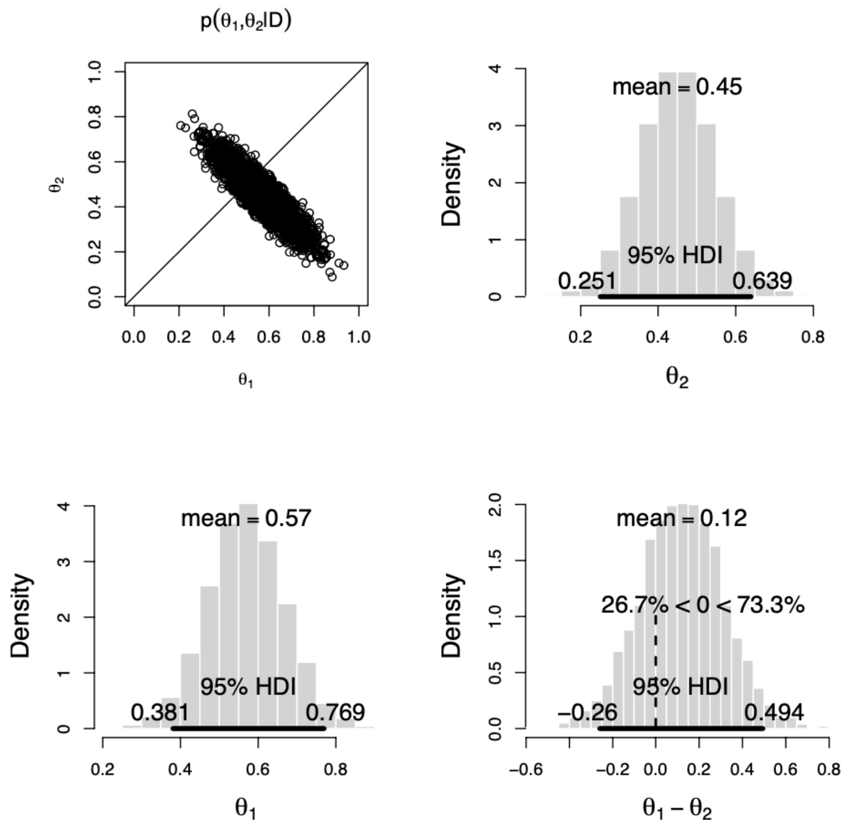


12.1.2.1 Differences of Correlated Parameters

It is important to understand that marginal distributions of single parameters do not reveal whether or not the two parameter values are different. The plot below, in its upper four panels, shows a case in which the posterior distribution for two parameter values has a strong positive correlation. Two of the panels show the marginal distributions of the single parameters. Those two marginal distributions suggest that there is a lot of overlap between the two parameters values. This is because the correlation of the two parameters are very high.



We can see a complementary case as below. Here, the marginal distributions of the single parameters are exactly the same as before: Compare the histograms of the marginal distributions, for the upper four panels and the lower four panels. Despite the fact that the marginal distributions are the same as before, the bottom right panel reveals that the difference of parameter values now straddles zero, with a difference of zero firmly in the midst of the HDI.



12.1.3 Region of Practical Equivalence (ROPE)

The estimation approach can be enhanced by including a region of practical equivalence (ROPE), which indicates a small range of values that are considered to be practically equivalent to the null value for purposes of the particular application. For example, if we wonder whether a coin is fair, for purposes of determining which team will kick off at a football game, then we want to know if the underlying bias in the coin is reasonably close to 0.50, and we don't really care if the true bias is 0.51 or 0.49, because those values are close enough for our application.

Once a ROPE is set, we make decisions according the following rule:

> A parameter value is declared to be not credible, or rejected, if its entire ROPE lies outside the 95% HDI of the posterior distribution of that parameter.

Because the ROPE and HDI can overlap in different ways, there are different decisions that can be made. In particular, we can decide to "accept" a null value:

> A parameter value is declared to be accepted for practical purposes if that value's ROPE completely contains the 95% HDI of the posterior of that parameter.

This use of a ROPE around the null value also implies that if the null value really is true, we will eventually "accept" the null value as the sample size gets large enough. This is because, as the sample size gets larger, the HDI tends to get narrower and closer to the true value. When the sample size gets very large, the HDI is almost certain to be narrow enough, and close enough to the true value, to fall entirely within the ROPE.

If we did not use a ROPE around the null value, and rejected the null value any time that it falls outside the 95% HDI, then we would incorrectly reject the null in 5% of experiments even if when null value is true.

To decide the size of ROPE, we can consider as below: if the application is a domain in which costs and benefits of decisions can be determined, then the full decision-theoretical machinery of expected utility can be brought to bear. In some domains such as medicine, expert clinicians can be interviewed, and their opinions can be translated into a reasonable consensus regarding how big of an effect is useful or important for the application. Otherwise, the ROPE might be established with somewhat arbitrary criteria, bearing in mind the key trade-off: When the ROPE is wider, there is a lower probability of falsely rejecting the null value (i.e., there is a smaller false alarm rate), but there is also a lower probability of rejecting the null value when it is false (i.e., there is a smaller hit rate).