12.2 The Model-Comparison Approach

Zongyi Liu

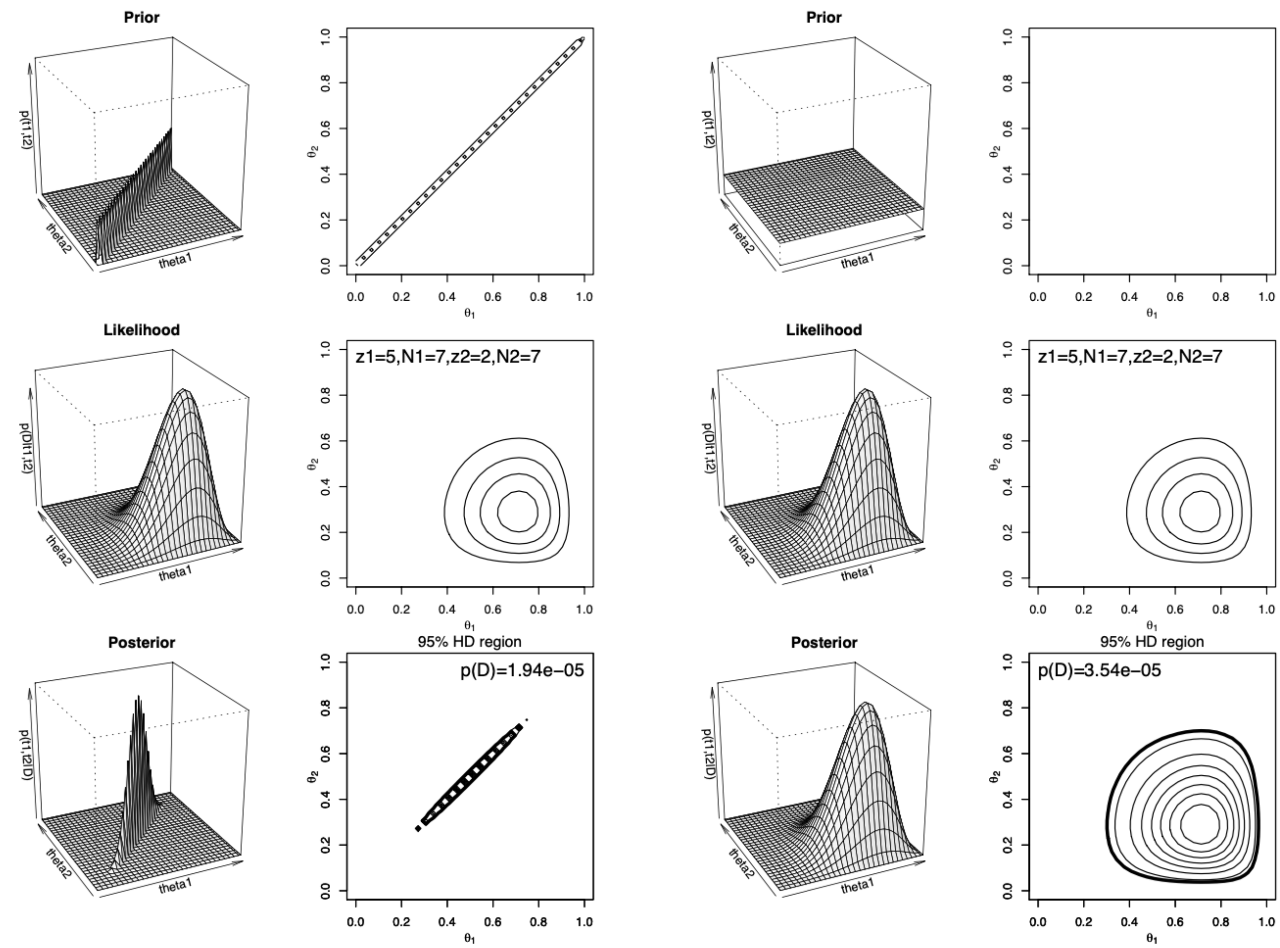
2023-06-25

12.2 The Model-Comparison Approach

The previous section posed the question, is the null value among the credible values, in terms of parameter estimation. We started with an informed prior distribution on the parameters, and then examined the posterior distribution on the parameters.

Here we need to decide which of two hypothetical priors is least unbelievable. One hypothetical prior expresses the idea that the parameter value is exactly the null value. The alternative hypothetical prior expresses the idea that the parameter could be anything.

Neither of these hypothetical priors is informed by prior knowledge. This lack of being informed is often taken as a desirable aspect of the approach, not a defect, because the method is thereby "automatic" insofar as it obviates disputes about prior knowledge.



12.2.1 Are the Biases of Two Coins Equal or Not?

There is a situation in which we have two coins, and we would like to infer whether their biases are equal or not. We pose the question as a model comparison, such that one model expresses the "null" hypothesis that the two biases are equal, and the other model expresses the "alternative" hypothesis that the two biases could be any combination. The two models are distinguished by their priors (because the likelihood functions are the same in both models): The null model has zero prior probability on all combinations of biases except where they are equal, while the alternative model has uniform probability on all combinations of biases. The uniform alternative is chosen because it is a convenient expression of non-informed indifference.

The graph below shows an example of this approach. The "null" model, M_{null}, is shown in the left columns. Notice that this prior is shaped as a ridge along all values for which $\theta_1 = \theta_2$. The prior gives zero probability to any point at which $\theta_1 \neq \theta_2$.

The lower row of the graph shows the posteriors resulting from two hypothetical priors.

To do model comparison, we use the implication of Bayes rule:

$$\frac{p(M_{alt}|D)}{p(M_{null}|D)} = \underbrace{\frac{p(D|M_{alt})}{p(D|M_{null})}}_{\mathbf{BF}} \frac{p(M_{alt})}{p(M_{null})}$$

The posterior plots in Figure 12.4 display the values of p(D|M) for each model. The evidence for the null model is $p(D|M_{null}) = 1.94 * 10^{-5}$, while the evidence for the alternative model is $p(D|M_{alt}) = 3.54 * 10^{-5}$. The Bayes factor therefore slightly favors the alternative prior, but not by much. Because the ratio of posterior probabilities is not very extreme, we would conclude that either model remains reasonably credible, given the data.

12.2.1.1 Formal Analytical Solution

The Bayes factor for the null and alternative models can also be computed analytically in this case. The alternative hypothesis has a uniform prior. It is uniform because the alternative hypothesis in this approach is supposed to be an "automatic" conventional prior that expresses a hypothesis complementary to the null hypothesis. A uniform prior on θ_1, θ_2 can be described as a product of beta distributions, namely, $beta(\theta 1|1,1) * beta(\theta 2|1,1)$. Therefore we can determine an exact value for $p(D|M_{alt})$ from Equation 8.5 (p. 131), which becomes

$$\begin{aligned} p(D|M_{alt}) &= \frac{B(z_1+1, N_1-z_1+1)\, B(z_2+1, N_2-z_2+1)}{B(1,1)B(1,1)} \\ &= B(z_1+1, N_1-z_1+1)\, B(z_2+1, N_2-z_2+1) \end{aligned}$$

We can also derive a formal analytical expression for the evidence for the null hypothesis, $p(D|M_{null})$. The double integral for the evidence simplifies to a single integral over $\theta = \theta 1 = \theta_2$

$$\begin{aligned} p(D|M_{null}) &= \iint d\theta_1 d\theta_2\, p(D|\theta_1, \theta_2) p(\theta_1, \theta_1) \\ &= \int d\theta\, p(D|\theta) \end{aligned}$$

We now plug in the Bernoulli likelihood function to get

$$\begin{aligned} p(D|M_{null}) &= \int d\theta\, p(D|\theta) \\ &= \int d\theta\, \theta^{(z_1)}(1-\theta)^{(N_1-z_1)}\, \theta^{(z_2)}(1-\theta)^{(N_2-z_2)} \\ &= \int d\theta\, \theta^{(z_1+z_2)}(1-\theta)^{(N_1-z_1+N_2-z_2)} \\ &= B(\,z_1+z_2+1\,,\, N_1-z_1+N_2-z_2+1\,) \end{aligned}$$

The Bayes factor for the alternative hypothesis, relative to the null hypothesis, is then the ratio of equations above

$$\frac{p(D|M_{alt})}{p(D|M_{null})} = \frac{B(z_1+1, N_1-z_1+1)\, B(z_2+1, N_2-z_2+1)}{B(\,z_1+z_2+1\,,\, N_1-z_1+N_2-z_2+1\,)}$$

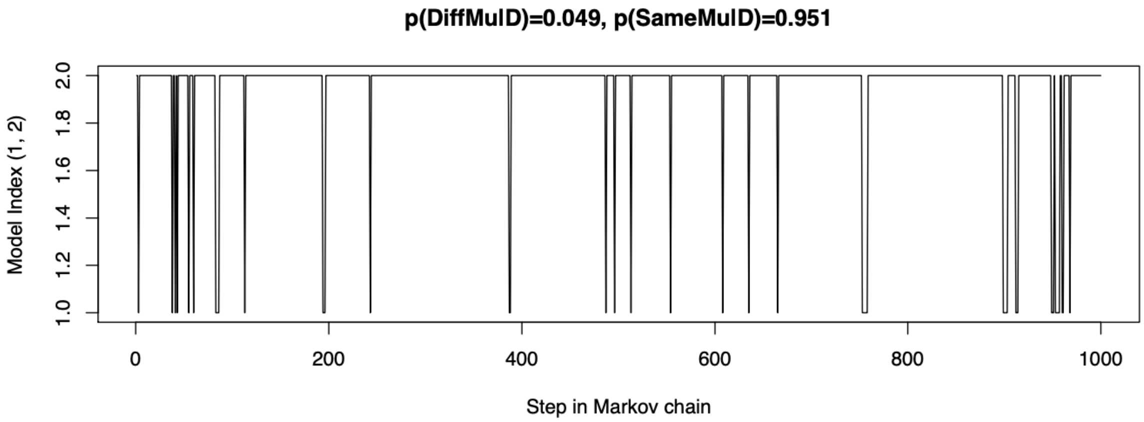
12.2.1.2 Example Application

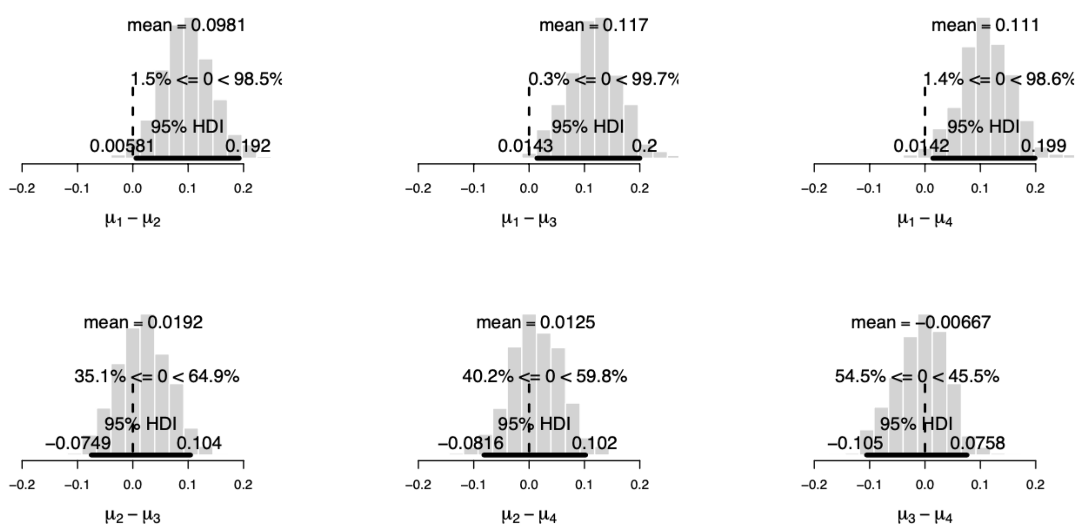
Again we reconsider the exercise 8.1, which asked whether there are more successful shots after a preceding successful shot than after a preceding failed shot.

We can use model comparison to analyze this situation. Using Equation 12.4, we obtain a Bayes factor of 0.128, which, inverted, is a Bayes factor of 7.81 in favor of the null prior. If the priors on the models are 50/50, then the posterior probability of the null model is $p(M_{null}|D) = 88.7$, and the posterior probability of the alternative model is $p(M_{alt}|D) = 11.3$. The model comparison suggests that the null model is less unbelievable than the alternative model.

However, prior belief suggests that it is extremely unlikely that there is literally zero correlation between the first and second shots of a pair of free throws in basketball. Therefore, even though the model comparison suggests that the null prior is less unbelievable than the alternative prior, what we might really want is an estimate of the difference between success after success and success after failure.

On the other hand, the direct estimation of the underlying biases did provide an estimate of the difference, as was shown in Figure 12.2, p. 242 (plot in 12.1). There we saw that a difference of zero was among the credible differences, but we also saw that the mean difference was a bit larger than zero, and we also saw explicitly that there is moderately large uncertainty in the estimate of the difference.





12.2.1.2 Are Different Groups Equal or Not?

Suppose we conduct a study that has four conditions. Suppose every participant in every condition gets the same list of twenty words to try to memorize. The ability to recall a word is modeled as a Bernoulli distribution, with probability θ_{ij} for the ith person in the jth condition. The individual recall propensity θ_{ij} depends on hyperparameters, μ_j and κ_j, that describe the overarching recall propensity in each condition, because $\theta_{ij} \sim beta(\theta_{ij}|\mu_j\kappa_j + 1, (1 - \mu_j)\kappa_j + 1)$.

We would like to know whether there is an effect of the type of music on ability to remember words. The most straight-forward way to find out is to estimate the parameters and then examine the posterior differences of the parameter estimates. The histograms in Figure 12.5 show the distributions of differences between the μ_j parameters. It can be seen that μ_1 is quite different than μ_3 and μ_4; a difference of zero falls well outside the 95% HDI intervals.

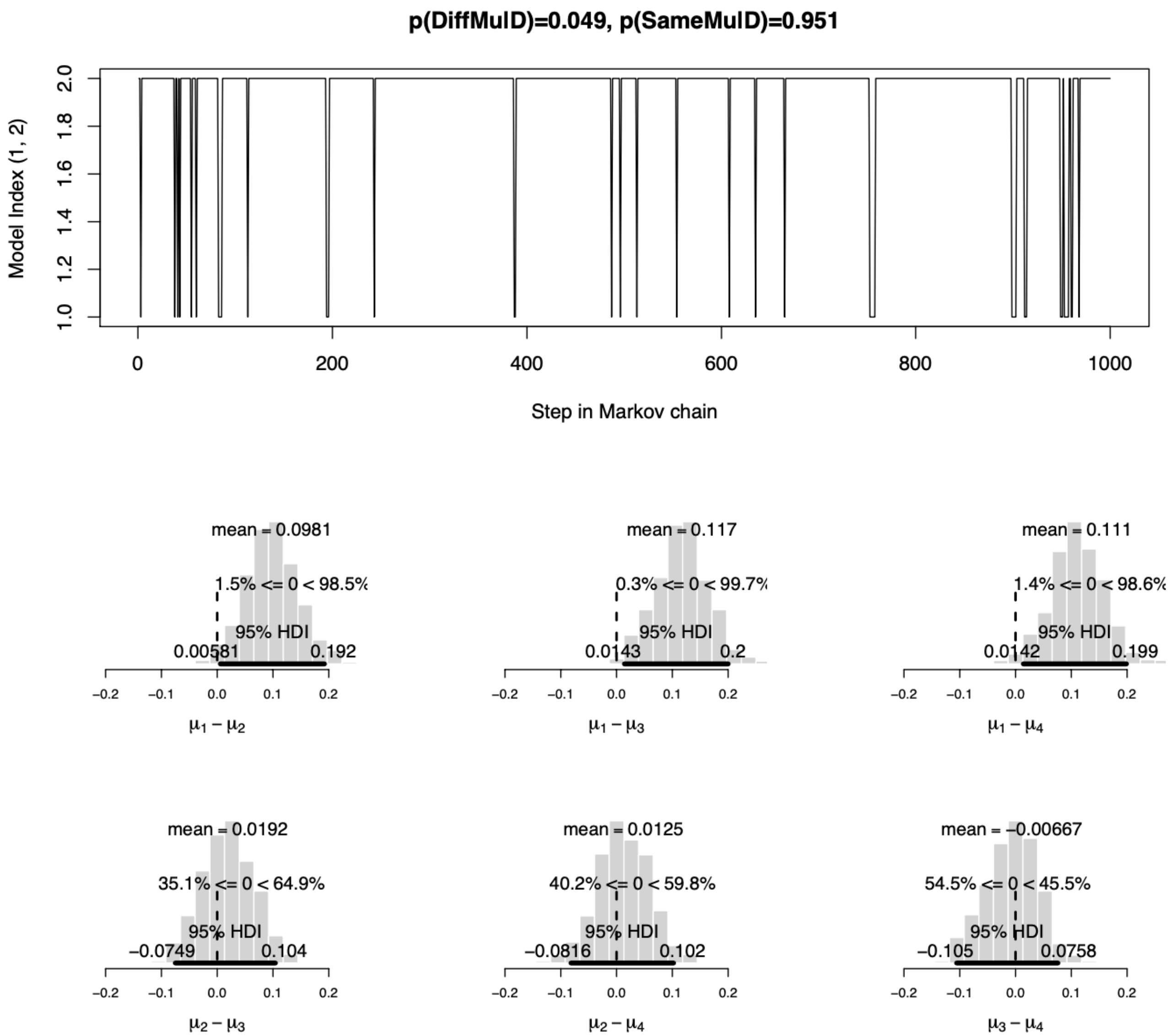


Figure 1: Figure 12.5

A model-comparison approach addresses the issue a different way. It compares the full model, which has distinct μ_j parameters for the four conditions, against a restricted model, which has a shared μ_0 parameter to describe all the conditions simultaneously. The two models have equal (50/50) prior probabilities.

In principle, we could consider all possible models formed by partitioning the four groups. For 4 groups, there are 15 distinct partitions. We could, in principle, put a prior belief on each of the 15 models, and then do a comparison of the 15 models. From the posterior probabilities of the models, we could ascertain which partition was most believable, and decide whether it is more believable that other nearly-as-believable partitions.

An explicit posterior estimate will reveal the magnitude and uncertainty of those estimates. Thus, unless we have a viable reason to believe that different group parameters may be literally identical, an estimation of distinct group parameters will tell us what we want to know, without model comparison.