BDA HM7

Piero Birello

February 2025

11.1

Consider a die that is rolled $N = 45$ times. Its six-dotted face comes out $\hat{z} = 3$ times. Under the null hypothesis $\theta = 1/6$, the probability of obtaining a number of six-dotted faces $z \leq \hat{z}$ is:

$$p(z \leq \hat{z}|N, \theta) = \sum_{z=0}^{\hat{z}} \binom{N}{z} \theta^z (1-\theta)^{N-z} = 0.04460167 \tag{1}$$

Consider now the case when we repeatedly roll the die until we obtain $z = 3$ six-dotted faces. This requires $\hat{N} = 45$ rolls. The probability of $N \geq \hat{N}$ is:

$$p(N \geq \hat{N}|z, \theta) = 1 - p(N < \hat{N}|z, \theta) = 1 - \sum_{N=z}^{\hat{N}-1} \frac{z}{N} \binom{N}{z} \theta^z (1-\theta)^{N-z} = 0.01563137 \tag{2}$$

Under both the data collection procedures, the two sided p-value is defined as twice the one sided one. In one case it is above 0.05, in the other case it is below.

11.2

The code provided can be used to estimate the 95% confidence interval for our case study. We need to identify the interval of θ values that would not be rejected by a p-value significance test. To do so, we look for two values $\theta_{low} \leq \theta_{high}$ such that their p-values are both $\simeq 0.05$. In the case of θ_{low}, extreme values of z will be those above the obtained value $\hat{z} = 3$. Hence we compute the p-value as:

$$2p(z \geq \hat{z}|N, \theta) = 2 \sum_{z=\hat{z}}^{N} \binom{N}{z} \theta^z (1-\theta)^{N-z} \tag{3}$$

For θ_{high}, extreme values are those below $\hat{z} = 3$, hence the p-value is:

$$2p(z \leq \hat{z}|N, \theta) = 2 \sum_{z=0}^{\hat{z}} \binom{N}{z} \theta^z (1-\theta)^{N-z} \tag{4}$$

Running the provided code, we get $\theta_{low} = 0.014$ and $\theta_{high} = 0.183$.
If we consider the case where the stopping criterion corresponds to reaching $z = 3$, then we can compute confidence intervals as:

```
# Confidence intervals (stopping criterion: threshold z)
highTailN = z:(N-1)
for ( theta in seq( 0.140 , 0.160 , 0.001) ) {
  show( c(
    theta ,
    2* (1 - sum( z/highTailN * choose(highTailN,z) * theta^z * (1-theta)^(highTailN-z) ))
  ))
}
lowTailN = z:N
for ( theta in seq( 0.005 , 0.020 , 0.001) ) {
  show( c(
    theta ,
    2* sum( z/lowTailN * choose(lowTailN,z) * theta^z * (1-theta)^(lowTailN-z) )
  ))
}
```

We obtain $\theta_{low} = 0.014$, $\theta_{high} = 0.154$. The confidence interval thus depends on the stopping criterion.

11.3

The given code computes the p-value for $\hat{z} = 3$, $\hat{N} = 45$ in the case in which all N's in $\mathcal{N} = \{40, ..., 50\}$ were possible and equiprobable a priori. The code computes:

$$p = \frac{1}{|\mathcal{N}|} \sum_{N \in \mathcal{N}} \sum_{z=0}^{z^{(N)}} \binom{N}{z} \theta^z (1-\theta)^{N-z} \tag{5}$$

where $z^{(N)}$ is the highest z value as extreme (as low) as $\hat{z}/\hat{N}$ given N, i.e., $z^{(N)} = max\{z | z/N \leq \hat{z}/\hat{N}\}$. We obtain $p = 0.05562808$.

12.1

Consider a coin, let $z = 7$, $N = 24$. Then if we only admit the null hypothesis $\theta_{null} = 0.5$:

$$p(D) = p(D|\theta_{null}) = \theta_{null}^z (1-\theta_{null})^{N-z} = 5.960464e-08 \tag{6}$$

If we consider a narrow prior distribution centered in θ_{null}, such as $p(\theta) = Beta(2000, 2000)$:

$$p(D) = \int d\theta p(D|\theta) p(\theta) = 6.02e-08 \tag{7}$$

see Figure 1. This is slightly larger than the previous result, as we would expect, since an interval of θ values is allowed instead of one only, even though it's a narrow one. If we consider a nearly Haldane prior, $Beta(0.01, 0.01)$, we get $p(D) = 2.87e-09$, see Figure 2. This model is less likely than the previous ones. In particular, the Bayes Factor is $BF = p(D|Haldane)/p(D|\theta_{null}) = 0.048 < 1/3$. Alternatively, the approximate Bayes Factor for the null hypothesis can be computed following the Savage-Dickey method, as the ratio between the prior probability within the ROPE and the posterior probability within the rope $0.08/1.67 = 0.048$. This matches the previous result.
If one uses a mildly informed prior like $Beta(2, 4)$, the result is $p(D|informed) = 2.22e-07$, $BF = 3.72$. Or, according to the Savage-Dickey method, it is $BF \simeq 5/1.38 = 3.62$. This is considerably larger than under the Haldane prior, still HDI's are similar in the two cases, ranging from 0.122 to 0.471 for the Haldane, and from 0.145 to 0.462 for the Beta. More generally, posterior distributions are similar. We may say that posterior

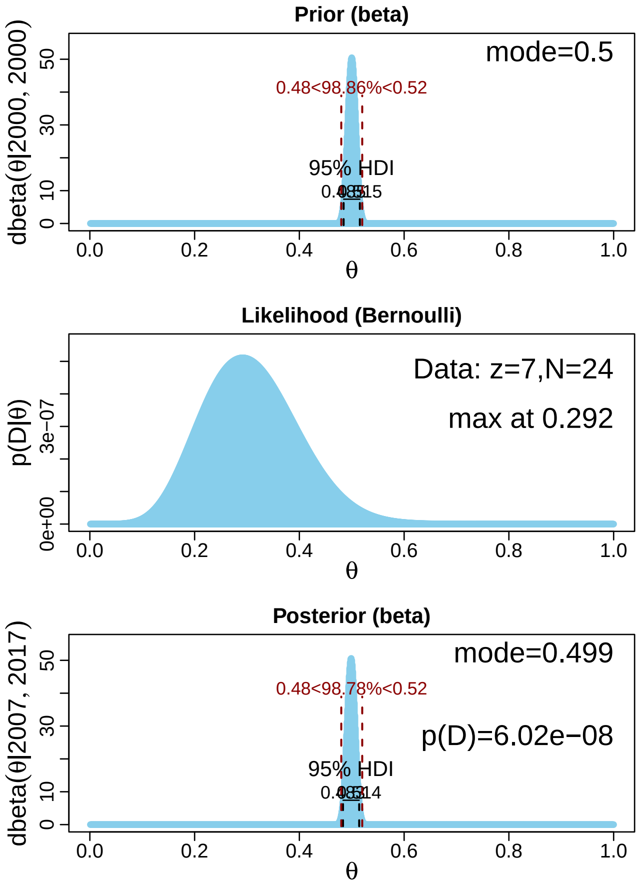


Figure 1: Exercise 12.1 B.

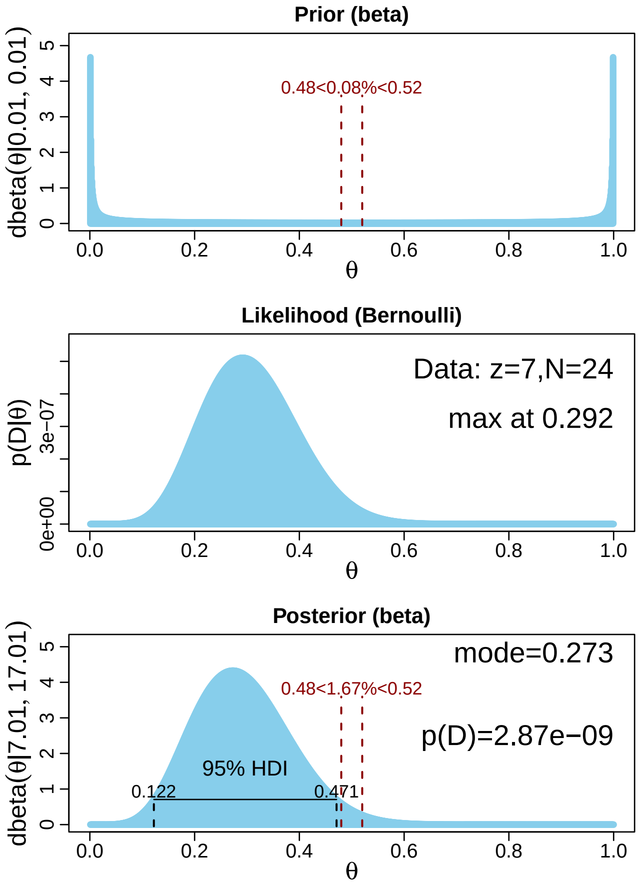


Figure 2: Exercise 12.1 C.

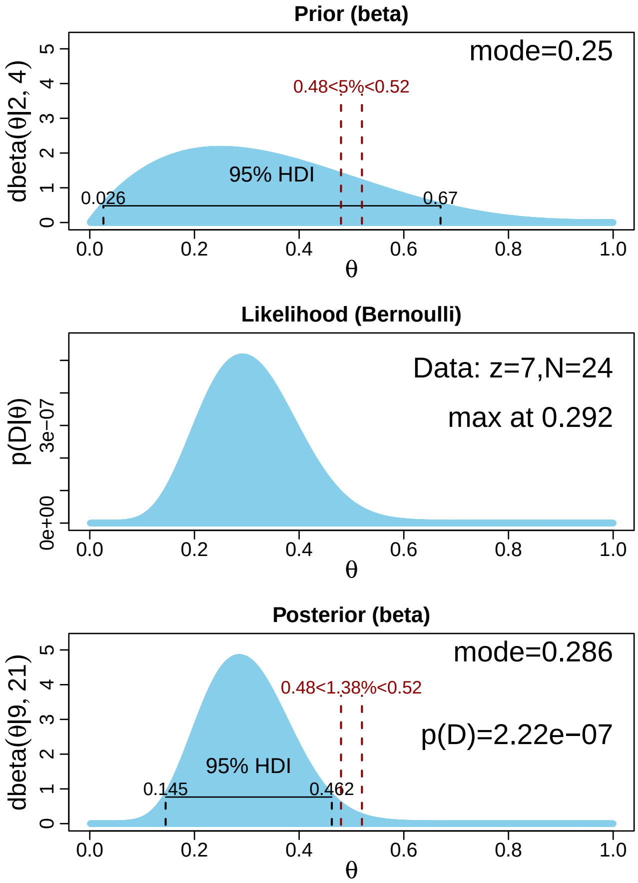


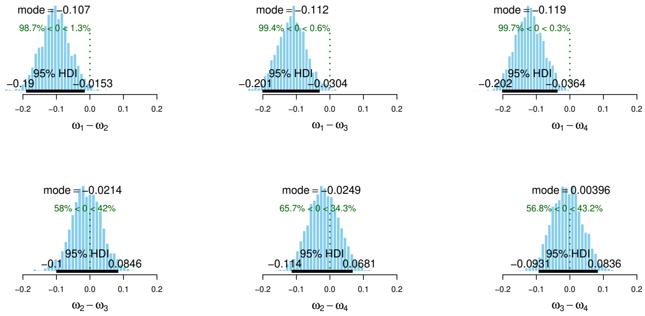
Figure 3: Exercise 12.1 E

parameter estimation is a more informative and safe method to test the null hypothesis, with respect to model comparison. If we want to adopt model comparison, a mildly informed prior may be a better choice than an Haldane prior, unless we actually believe that the coin is likely to always land on the same one among the two faces.

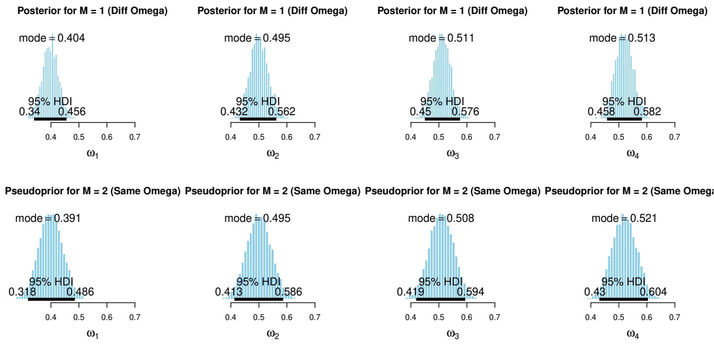
12.2

Figures in Figure 4 show us posterior estimates and a trace plot of two different models used to describe the recall abilities of a set of subjects under four different external conditions. The first model allows for distinct ω_j, one per each experimental condition, while the second model sets $\omega_j = \omega_0 \, \forall j$. We see that the second model would be preferred under a model selection approach. However, the fact that a parsimonious model not accounting for differences in ω_j is preferred does not imply that differences between all parameter pairs are negligible. From Figure 4a we actually see that posterior estimates of the differences $(\omega_1 - \omega_2)$, $(\omega_1 - \omega_3)$, $(\omega_1 - \omega_4)$ almost entirely lie on the semi-axis of negative values. This signals that ω_1 is significantly smaller than the remaining ω's, i.e., participants' recalling ability is worse under condition 1.

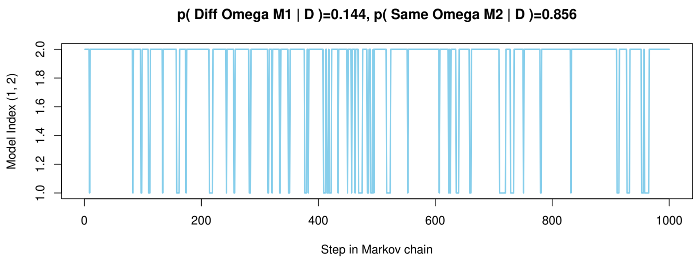
This finding suggest that a better model may consist in one where ω_1 has a separate value, while $\omega_2 = \omega_3 =$



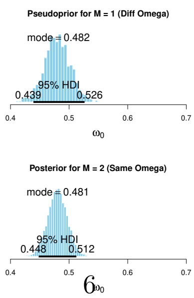
(a) Exercise 12.2 A: posterior estimates of differences between ω_j associated to different conditions under model 1.



(b) Exercise 12.2 A: posterior estimates of ω_i under the model admitting different ω's, and pseudopriors for the model setting $\omega_i = \omega_0$.



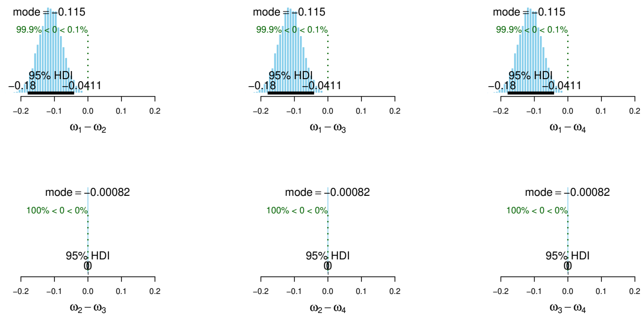
(c) Exercise 12.2 A: trace plot of the model index along steps in the Markov Chain. We see that model 2, the one allowing a single ω_0, is the preferred one.



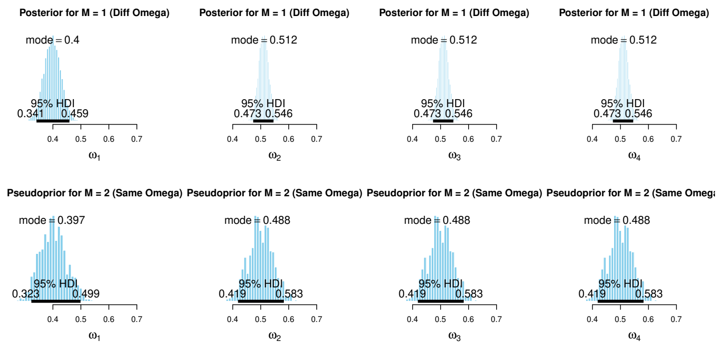
(d) Exercise 12.2 A: Posterior for ω_0 under model 2 and corresponding pseudoprior under model 1.

Figure 4: Exercise 12.2 A.

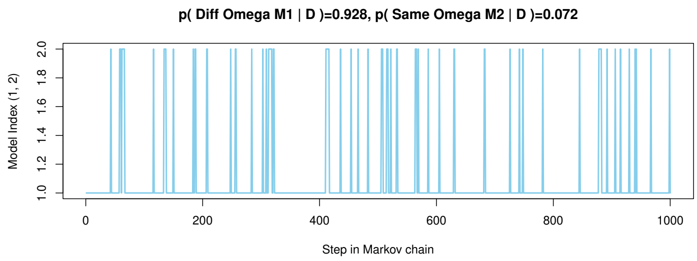
ω_4. Indeed, Figure 5 shows that such model is preferred under the model selection framework. If we inspect the estimated differences between parameter pairs, we see that posterior estimates for $\omega_1 - \omega_j$, $j \in 2, 3, 4$, almost completely lie in the negative semi-axis again. Showing that the assumption of ω_1 having a different value is credible. Interestingly, this plot alone (Figure 5a) is pretty self explanatory, just like in the case of Figure 4a. We may conclude that posterior parameter estimation and inspection is the most meaningful approach in this situation, since it avoids the risk of neglecting important details or getting wrong conclusions, while giving us the power to formulate the same final statements as model selection. However, model comparison may be more useful in the case where we formulate two completely different models, rather than the same one allowing or not for parameters variation.



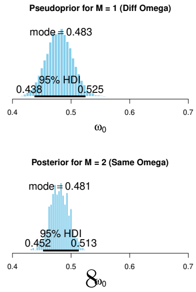
(a) Exercise 12.2 C: posterior estimates of differences between ω_j associated to different conditions under model 1.



(b) Exercise 12.2 C: posterior estimates of ω_i under the model admitting two different ω's, and pseudopriors for the model setting $\omega_i = \omega_0$.



(c) Exercise 12.2 C: trace plot of the model index along steps in the Markov Chain. We see that model 1, the one allowing two ω values, is the preferred one.



(d) Exercise 12.2 C: Posterior for ω_0 under model 2 and corresponding pseudoprior under model 1.

Figure 5: Exercise 12.2 C.