10.1 Model Comparison as Hierarchical Modeling

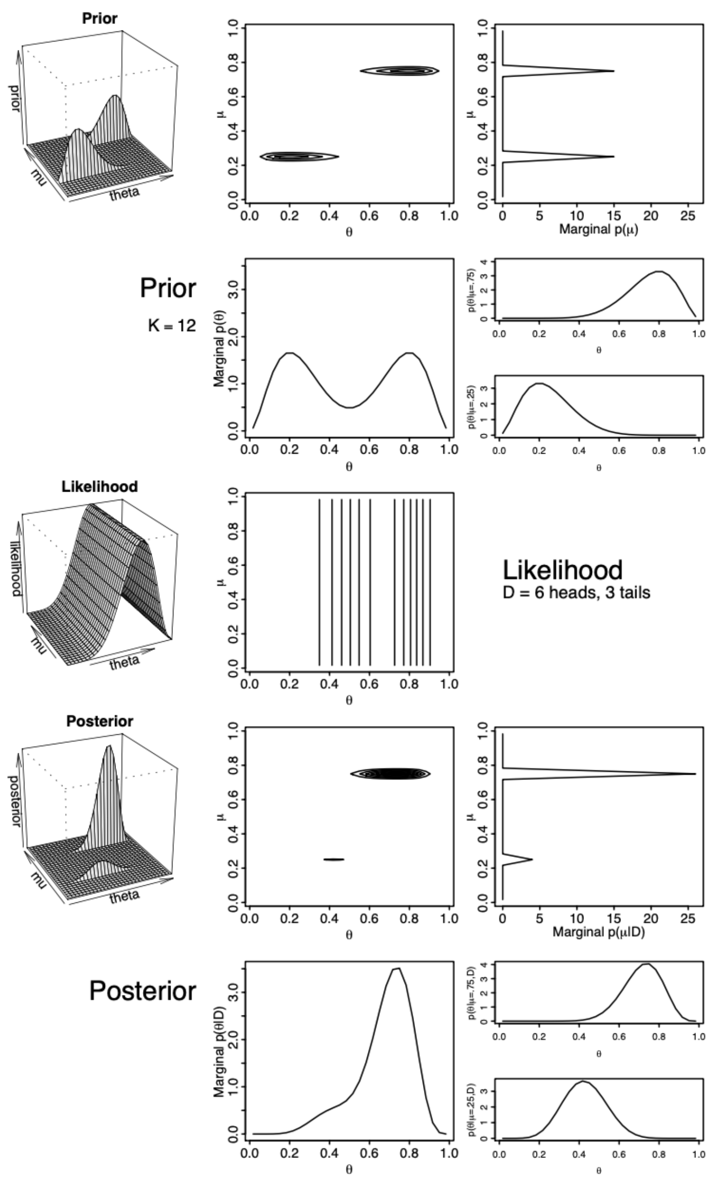
Zongyi Liu

2023-06-22

10.1 Model Comparison as Hierarchical Modeling

Consider again the simple hierarchical dependency in which a single coin with bias θ depends on a hyperparameter μ. The dependency and Bayesian updating for this were plotted in the previous chapter.

In that previous scenario, the hyperparameter μ was assumed to have any possible value in the interval [0,1]. The plot below shows a case in which the hyperparameter μ is instead allowed only two values: $\mu = 0.25$ and $\mu = 0.75$. We can think of this as being a prior with two ridges, one ridge over each of the possible values.



The dependence of θ on μ is determined as before by equation 9.2, in this case with $K = 12$. Therefore, when $\mu = .25$, the prior distribution over θ is $beta(\theta, 3, 9)$, and when $\mu = .75$, the prior distribution over θ is $beta(\theta, 9, 3)$.

In this example we flip a coin $N = 9$ times, and observe $z = 6$ heads. The likelihood is shown in the middle row of Figure 10.1, and the posterior is obtained, as always, by multiplying the likelihood and the prior, point-by-point across the parameter space. The posterior shows two ridges, like the prior, because anywhere the prior is zero, the posterior must also be zero.

The posterior also reveals how much we believe in each candidate value of μ, as can be seen in the graph of $p(\mu|D)$ in the right panel of the fourth row of Figure 10.1.

This comparison of beliefs in two values of μ is tantamount to comparison of beliefs in two models at the level of θ, namely, the model $beta(\theta|3, 9)$ versus the model $beta(\theta|9, 3)$.

When summing the posterior distribution over θ, we are effectively computing, at each value of $\mu = \mu^*$, the evidence for the model $\mu = \mu^* : p(D|\mu) = \int d\theta p(D|\theta)p(\theta|\mu)$.