4.2 Applied to Models and Data

Zongyi Liu

2023-06-09

4.2 Applied to Models and Data

One of the key application that makes Bayes' rule so useful is when the row and column variables are data values and model parameter values, respectively

Generally, a model specifies

$$p(data\ values|parameters\ values\ and\ model\ structure)$$

Given the data:

$$p(parameters\ values\ and\ model\ structure|data\ values)$$

When we have observed some data, we use Bayes' rule to determine our beliefs across competing parameter values in a model, and to determine our beliefs across competing models.

After applying Bayes' rule to data and model parameter, we can get a new table:

Data	**Model Parameter**			**Marginal**
		θ value		
		$\vdots$		
D value	$\cdots$	$p(D,\theta)$ $= p(D\|\theta)p(\theta)$ $= p(\theta\|D)p(D)$	$\cdots$	$p(D)$
		$\vdots$		
Marginal:		$p(\theta)$		

The prior probability of the parameter values is the marginal distribution, $p(\theta)$, which appears in the lower margin of the table.

Then we can introduce some notations:

$$\underbrace{p(\theta|D)}_{\text{posterior}} = \underbrace{p(D|\theta)}_{\text{likelihood}}\ \underbrace{p(\theta)}_{\text{prior}} \Big/ \underbrace{p(D)}_{\text{evidence}}$$

where the evidence is

$$p(D) = \int d\theta\, p(D|\theta)p(\theta).$$

We can also add the conditional term into the equation:

$$\underbrace{p(\theta|D,M)}_{\text{posterior}} = \underbrace{p(D|\theta,M)}_{\text{likelihood}}\,\underbrace{p(\theta|M)}_{\text{prior}} \,/\, \underbrace{p(D|M)}_{\text{evidence}}$$

where the evidence is

$$p(D|M) = \int d\theta\, p(D|\theta,M)p(\theta|M).$$

By Bayes' rule and take the ratio of those equations, we can get:

$$\frac{p(M1|D)}{p(M2|D)} = \underbrace{\frac{p(D|M1)}{p(D|M2)}}_{\text{Bayes factor}}\,\frac{p(M1)}{p(M2)}.$$

This equation says that the posterior beliefs is the ration of the evidences times the ratio of the prior beliefs.

The ratio of the evidences is called the **Bayes factor**.

Also, the quantity $p(D|M)$, which is called the "evidence" in this book, is sometimes instead called the "marginal likelihood" or "prior predictive" by other authors.

$p(M1|D) = p(D|M1)p(M1)/p(D)$, is the "likelihood of the model M1 for the data D".

4.2.1 Data Order In-variance

First we will unpack $p(\theta|D', D)$ and have

$$p(\theta|D',D) = \frac{p(D'|\theta,D)\,p(\theta|D)}{\int d\theta\, p(D'|\theta,D)\,p(\theta|D)}$$

The equation then can be rewritten as

$$p(\theta|D',D) = \frac{p(D'|\theta)\,p(\theta|D)}{\int d\theta\, p(D'|\theta)\,p(\theta|D)}$$

Then we use Bayes' rule again, and we can convert the equation into

$$p(\theta|D',D) = \frac{p(D'|\theta)\,p(D|\theta)\,p(\theta)/p(D)}{\int d\theta\, p(D'|\theta)\,p(D|\theta)\,p(\theta)/p(D)}$$

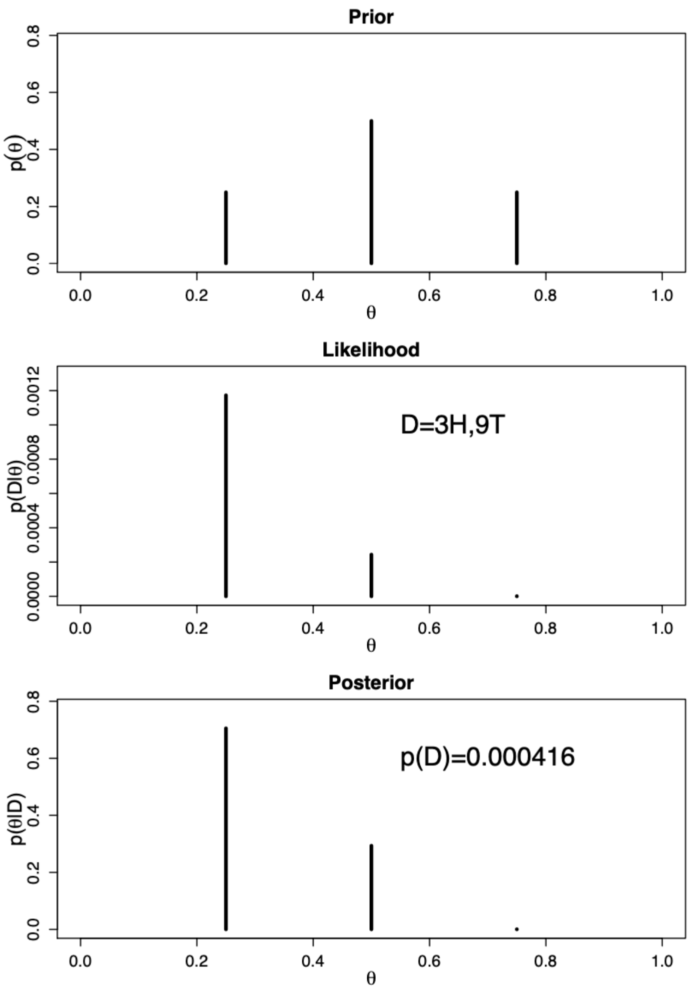
4.2.2 An Example with Coin Flipping

This section we will consider an example when flipping the coin.

First, we specify our prior beliefs. We denote the bias as $\theta = p(H)$, the probability of the coin coming up heads. To keep the example straightforward, suppose that we believe there are only three possible values for the coin's bias: Either the coin is fair, with $\theta = .50$, or the coin is biased with $\theta = .25$ or $\theta = .75$. We believe that the coin is probably fair, but there's some smaller chance it could be biased high or low.

Next, we flip the coin to get some data D and determine the likelihood, $p(D|\theta)$. Suppose we flip the coin 12 times, and it comes up heads 3 times. According to our model of the coin, the probability of coming up heads is θ, and the probability of coming up tails is $1-\theta$.

We can show the results in the plot below



4.2.2.1 $p(D|\theta)$ is not θ

In the examples involving coin flips, it is easy to lose sight of the important fact that $p(D|\theta)$ is different from θ, even though they both are values between 0 and 1 for our current examples. The likelihood $p(D|\theta)$ is a mathematical function of θ. The value of the likelihood function is always a probability. Adding to the confusability is the fact that, in our examples so far, the function that maps θ to $p(D=H|\theta)$ has been the identity function:

$$p(D=H|\theta)=\theta$$