

Week 1 Review and Posterior Probability Calculation Example

Duke University

Review Terminology

Before we do the computation, let us review the 3 “probabilities”:

Prior Probability: The probability of the **hypothesis/assumption**. Often the hypothesis/assumption is about some parameters θ and therefore, we will denote this as $\mathbb{P}(\text{hypothesis})$ or $\mathbb{P}(\theta = \text{some value})$.

Likelihood: The probability of the **data/result** we observe given the hypothesis we make. It is denoted as $\mathbb{P}(\text{data} \mid \text{hypothesis})$ or $\mathbb{P}(\text{data} \mid \theta = \text{some value})$.

Posterior Probability: The probability of the **hypothesis/assumption** we update after we have obtained the data. It is denoted as $\mathbb{P}(\text{hypothesis} \mid \text{data})$ or $\mathbb{P}(\theta = \text{some value} \mid \text{data})$.

Example

In the following, we will focus on one example to see how to apply what we have learned in Week 1:

Suppose there are two coins, an unbiased one and a biased one. The unbiased one is equally likely to get a head or a tail, while the biased coin has higher probability, 0.8, to get a head. We randomly choose a coin from the two, flip it 3 times and get 1 head and 2 tails. What is the posterior probability that the coin is biased?

From the problem, since we have no information about the 2 coins, we will assign **equal** prior to each coin. That is

$$\mathbb{P}(\text{biased}) = \mathbb{P}(\text{unbiased}) = \frac{1}{2}.$$

If we use p to denote the parameter representing the probability of getting heads in the 2 coins, we can also write our prior probabilities to be

$$\mathbb{P}(p = 0.8) = \mathbb{P}(p = 0.5) = \frac{1}{2}.$$

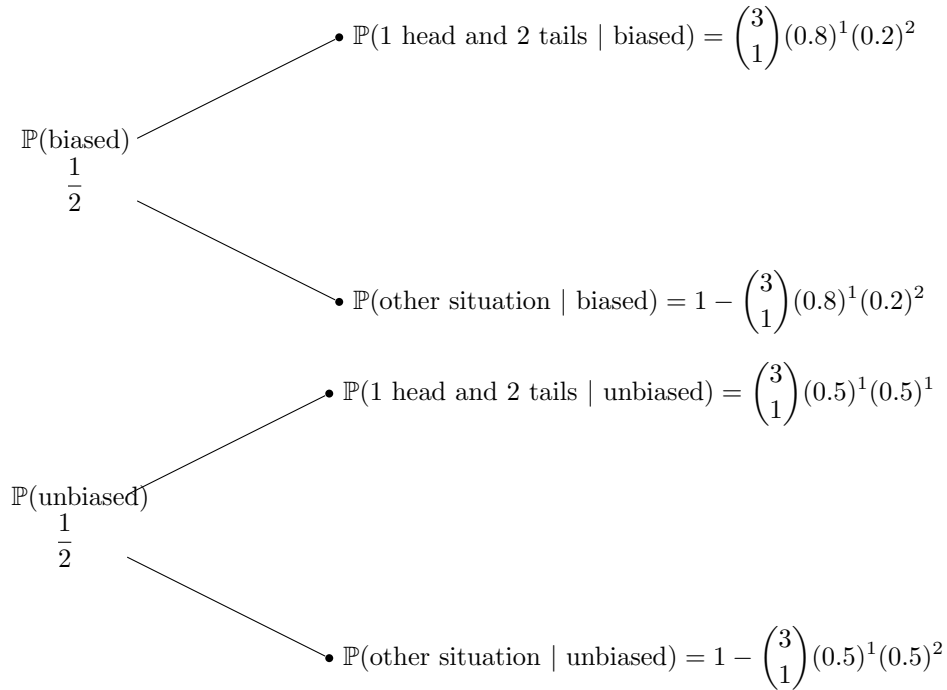
Note: Since we flip the **same** coin 3 times, the likelihood of getting heads each time is the same. So we can either calculate the posterior probability all at once, or **update** the posterior probabilities **sequentially**. However, if we switch coins during the flips, or like in the HIV example that we used ELISA as well as Western Blot for the 3 tests, the parameters involved in the likelihood in each step will change. In this case, we have to update the posterior probability sequentially.

Method 1: View 3 Flips as One Whole Process

We will first show how to calculate the posterior probability all at once:

Probability tree

We can build the probability tree as follows. We will explain how to get the likelihood of each branch in the next section.



Calculate likelihoods

The crucial step in Bayesian statistics is to understand how to calculate the likelihood of each situation. In this example, we do not emphasize the order of getting 1 head and 2 tails. Therefore, we should view this process as a **binomial process** (each trial is a Bernoulli trial). Therefore, suppose we assume the probability of getting a head for a general coin is p , then the probability (=likelihood) of getting 1 head and 2 tails given this p is:

$$\mathbb{P}(1 \text{ head and 2 tails; given } p) = \binom{3}{1}p^1(1-p)^2.$$

Here, this p , the probability for a general coin to get a head, is called the **parameter** of the **likelihood**, that is, the **parameter** of the **binomial probability mass function**.

For the 2 coins we have in this example, the likelihoods are

$$\mathbb{P}(1 \text{ head and 2 tails} \mid \text{biased}) = \binom{3}{1}(0.8)^1(0.2)^2 \quad \text{since } p = 0.8 \text{ for the biased coin,}$$

$$\mathbb{P}(1 \text{ head and 2 tails} \mid \text{unbiased}) = \binom{3}{1}(0.5)^1(0.5)^2 \quad \text{since } p = 0.5 \text{ for the unbiased coin.}$$

Calculate posterior probabilities

With the prior we assign (again, you have freedom to assign priors that you think are reasonable.) and the likelihoods we just calculated, we can turn to calculate the **posterior probability** using Bayes' Rule. Since the question only asks for the

posterior probability for the coin to be biased, we omit the calculation for the coin to be unbiased.

$$\begin{aligned} & \mathbb{P}(\text{biased} \mid 1 \text{ head and } 2 \text{ tails}) \\ &= \frac{\mathbb{P}(1 \text{ head and } 2 \text{ tails} \mid \text{biased})\mathbb{P}(\text{biased})}{\mathbb{P}(1 \text{ head and } 2 \text{ tails} \mid \text{biased})\mathbb{P}(\text{biased}) + \mathbb{P}(1 \text{ head and } 2 \text{ tails} \mid \text{unbiased})\mathbb{P}(\text{unbiased})} \\ &= \frac{\binom{3}{1}(0.8)^1(0.2)^2 \times \frac{1}{2}}{\binom{3}{1}(0.8)^1(0.2)^2 \times \frac{1}{2} + \binom{3}{1}(0.5)^1(0.5)^2 \times \frac{1}{2}} = \frac{32}{157} \approx 0.2038217. \end{aligned}$$

Takeaway

As you can see from the Bayes' Rule

$$\mathbb{P}(\text{biased} \mid 1 \text{ head and } 2 \text{ tails}) \propto \mathbb{P}(1 \text{ head and } 2 \text{ tails} \mid \text{biased}) \times \mathbb{P}(\text{biased}).$$

(The denominator is used to normalized the posterior probability so that it is always between 0 and 1.)

We can **summarize** this as

$$\underbrace{\mathbb{P}(\text{hypothesis} \mid \text{data})}_{\text{posterior probability of hypothesis}} \propto \underbrace{\mathbb{P}(\text{data} \mid \text{hypothesis})}_{\text{likelihood of data}} \times \underbrace{\mathbb{P}(\text{hypothesis})}_{\text{prior probability of hypothesis}}.$$

Another takeaway from this process is, since we only care about the situation when we get 1 head and 2 tails, we never even used any likelihoods involving “other situation”. This may reduce our workload once we are proficient with the Bayes' Rule. We do not need to draw the entire probability tree to figure out the posterior probability.

R codes

A simple R code can be

```
> prior <- c(1/2, 1/2)
> likelihood <- c(dbinom(1, 3, 0.8), dbinom(1, 3, 0.5)) # calculate likelihoods for p=0.8 and p=0.5
> posterior <- prior * likelihood / sum(prior * likelihood)
```

The results are

```
> posterior
[1] 0.2038217 0.7961783
```

Method 2: Update posterior probabilities step by step

In this method, we will show how to use updating method to get the same answer as before. Since we need to update our posteriors step by step, we need to know the **results** for every single step. However, the problem does not give us any information about which flip we get the head, and which ones we get tails.

Important Claim: Here I make an argument that the order of each result on each flip will NOT change our final answer. (You may think about why this makes sense intuitively.)

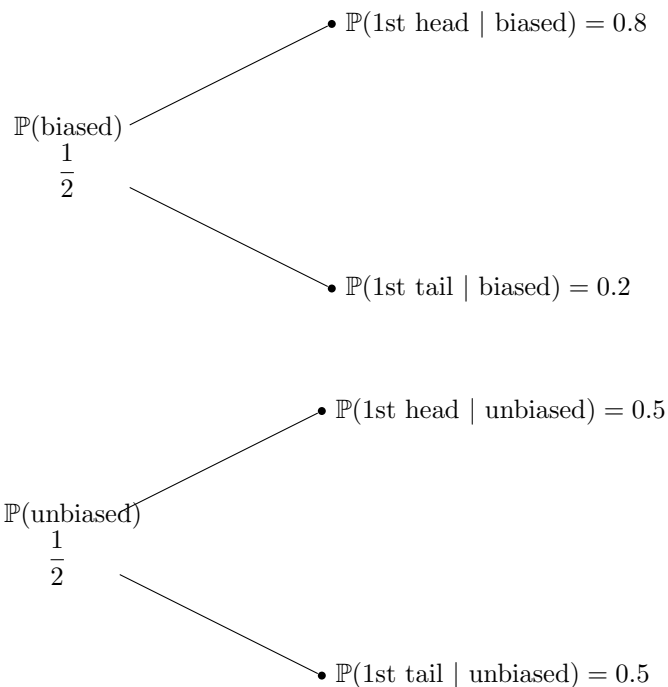
Since the order of 3 coin flip results will not change our final posterior probability, we assume the head and tails show up in the following order:

1st flip: head 2nd flip: tail 3rd flip: tail.

Here I encourage you to try other orders and see how the posterior probabilities change during the updating process.

Probability trees

Under 1st flip:



We update the posterior probabilities after the 1st flip using Bayes' Rule:

$$\begin{aligned} & \mathbb{P}(\text{biased}) \\ &= \frac{\mathbb{P}(\text{1st head} \mid \text{biased})\mathbb{P}(\text{biased})}{\mathbb{P}(\text{1st head} \mid \text{biased})\mathbb{P}(\text{biased}) + \mathbb{P}(\text{1st head} \mid \text{unbiased})\mathbb{P}(\text{unbiased})} \\ &= \frac{(0.8) \times \frac{1}{2}}{(0.8) \times \frac{1}{2} + (0.5) \times \frac{1}{2}} = \frac{8}{13} \approx 0.6153846. \end{aligned}$$

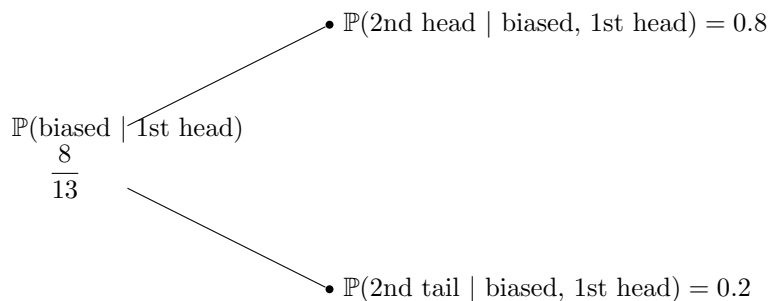
Therefore,

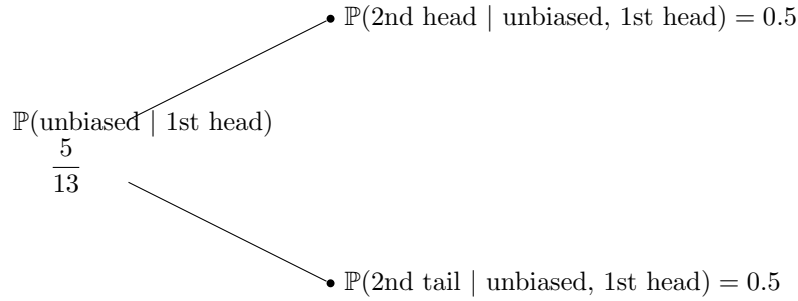
$$\mathbb{P}(\text{unbiased} \mid \text{1st head}) = 1 - \mathbb{P}(\text{biased} \mid \text{1st head}) = 1 - 0.6153846 = 0.3846154.$$

As we can see, with just 1 head, we tend to believe the coin is biased (with probability 0.6153846) instead of unbiased. But will this always be the case? We will see when we update the posterior probabilities after we have more information.

Under 2nd flip:

Now we will use the previous posterior probabilities as our new prior probabilities to plug in the probability tree:





To update the posterior probabilities after the 2nd flip, we have

$$\begin{aligned}
 & \mathbb{P}(\text{biased} \mid \text{1st head, 2nd tail}) \\
 &= \frac{\mathbb{P}(\text{2nd tail} \mid \text{biased, 1st head})\mathbb{P}(\text{biased} \mid \text{1st head})}{\mathbb{P}(\text{2nd tail} \mid \text{biased, 1st head})\mathbb{P}(\text{biased} \mid \text{1st head}) + \mathbb{P}(\text{2nd tail} \mid \text{unbiased, 1st head})\mathbb{P}(\text{unbiased} \mid \text{1st head})} \\
 &= \frac{(0.2) \times \frac{8}{13}}{(0.2) \times \frac{8}{13} + (0.5) \times \frac{5}{13}} = \frac{16}{41} \approx 0.3902439,
 \end{aligned}$$

$$\mathbb{P}(\text{unbiased} \mid \text{1st head, 2nd tail}) = 1 - \mathbb{P}(\text{biased} \mid \text{1st head, 2nd tail}) = 1 - 0.3902439 = 0.6097561.$$

As we can see, after the 2nd flip, the posterior probability to believe that the coin is biased towards head drop dramatically to about 0.39. This is how the new data helps us to correct our previous “belief”.

Note: If the above calculation seems complicated to you, since we compute the probabilities conditioning on 2 flips, you may also simplify the notation to the following, assuming you have already used the posterior probability after the 1st flip to start with:

$$\begin{aligned}
 & \mathbb{P}(\text{biased} \mid \text{2nd tail}) \\
 &= \frac{\mathbb{P}(\text{2nd tail} \mid \text{biased})\mathbb{P}(\text{biased})}{\mathbb{P}(\text{2nd tail} \mid \text{biased})\mathbb{P}(\text{biased}) + \mathbb{P}(\text{2nd tail} \mid \text{unbiased})\mathbb{P}(\text{unbiased})} \\
 &= \frac{(0.2) \times \frac{8}{13}}{(0.2) \times \frac{8}{13} + (0.5) \times \frac{5}{13}} = \frac{16}{41} \approx 0.3902439.
 \end{aligned}$$

Under 3rd flip:

Now you may be more familiar with the process of updating. We just show directly the calculation process to update the

posterior probabilities after the 3rd flip (I use H to represent “head” and T to represent “tail”):

$$\begin{aligned} & \mathbb{P}(\text{biased} \mid \text{1st H, 2nd T, 3rd T}) \\ &= \frac{\mathbb{P}(\text{3rd T} \mid \text{biased, 1st H, 2nd T})\mathbb{P}(\text{biased} \mid \text{1st H, 2nd T})}{\mathbb{P}(\text{3rd T} \mid \text{biased, 1st H, 2nd T})\mathbb{P}(\text{biased} \mid \text{1st H, 2nd T}) + \mathbb{P}(\text{3rd T} \mid \text{unbiased, 1st H, 2nd T})\mathbb{P}(\text{unbiased} \mid \text{1st H, 2nd T})} \\ &= \frac{(0.2) \times \frac{16}{41}}{(0.2) \times \frac{16}{41} + (0.5) \times \frac{25}{41}} = \frac{32}{157} \approx 0.2038217. \end{aligned}$$

This is exactly the same result as what we got using Method 1!

R Codes

We can use the following R codes to record all the posterior probabilities we have found during each step.

```
> prior <- c(0.5, 0.5)
> # Initialize likelihood as a dataframe for both head and tail
> likelihood <- data.frame(H = c(0.8, 0.5), T= c(0.2, 0.5))
> # Initialize an empty dataframe to store all the posterior probabilities
> postProb <- data.frame("biased" = double(), "unbiased" = double())
```

Then we calculate the posterior probability and update it as our new **prior**. You may rewrite the codes as a **for** loop in R.

```
> # After 1st flip we get H:
> posterior <- prior * likelihood$H / sum(prior * likelihood$H)
> # Store this into the dataframe 'postProb' and update prior
> postProb[1, ] <- posterior
> prior <- posterior

> # Update the 2nd and 3rd flip:
> posterior <- prior * likelihood$T / sum(prior * likelihood$T)
> postProb[2, ] <- posterior
> prior <- posterior
> posterior <- prior * likelihood$T / sum(prior * likelihood$T)
> postProb[3, ] <- posterior
```

We can print all the posterior probabilities generated after every step:

```
> postProb
      biased  unbiased
1 0.6153846 0.3846154
2 0.3902439 0.6097561
3 0.2038217 0.7961783
```