

Week 3 Summary

Duke University

Summary

Since this week contains a lot of information, I will first go over the big picture of what we need to know in this week. Although we have presented many examples, including hypothesis tests, credible intervals, point estimates for parameters and prior choice discussions under different circumstances, we do not require students to master every single piece of the calculations. In fact, you will see in Week 3 Lab that you can obtain most of the results from the function `bayes_inference`. We instead expect students to understand the main concepts, how to interpret the results from Bayes inference, and how different priors may influence the results.

If your goal is merely to pass the quizzes, this summary would be very likely to help you survive. However, if you would like to understand some more discussions about the videos, I suggest you keep on reading the file **Lecture Video Explanations**.

- **Loss Functions:** Videos **Losses and Decision Making** and **Working with Loss Functions** talk about the best estimators correspond to the 3 loss functions, linear, squared and 0/1 losses. I am not going to repeat them here. However, students are required to know how to find the best estimator according to the given distribution of the variable and the given loss function.
- **Expected Loss:** In the video **Minimizing Expected Loss for Hypothesis Testing**, we provide one method to compare 2 hypotheses: to calculate the 2 expected losses under each hypothesis to decide which hypothesis we should take. Here, the expected loss $\mathbb{E}(L(d))$ is in fact the **conditional** expected loss $\mathbb{E}(L(d) \mid \text{data})$, which is conditioned on the **given data we have observed**.

Before we go into the detail of the calculation, let us be clear of the meaning of $\mathbb{P}(H_1 \mid \text{data})$. Here, please understand this term as **given data, the probability that hypothesis H_1 will happen**.

We summarize the calculation of expected loss. Let us denote a decision d and its corresponding loss function $L(d)$. Suppose d **supports** H_1 . Then the definition of $L(d)$ can be written as:

$$L(d) = \begin{cases} 0, & H_1 \text{ happens, meaning } d \text{ is right} \\ w_1, & H_2 \text{ happens, meaning } d \text{ is wrong.} \end{cases}$$

Then according to the definition of expected value, $\mathbb{E}(f(X)) = \sum_{\text{all values of } X} f(\text{the value of } X) \times \mathbb{P}(X \text{ takes this value})$, we can calculate $\mathbb{E}(L(d) \mid \text{data})$ as follows:

$$\mathbb{E}(L(d) \mid \text{data}) = L(H_1 \text{ happens}) \times \mathbb{P}(H_1 \text{ does happen} \mid \text{data}) + L(H_2 \text{ happens}) \times \mathbb{P}(H_2 \text{ does happen} \mid \text{data})$$

Let us come back to the example in the video. We have two competing hypothesis:

$$H_1 : \text{ Patient does not have HIV, } \quad \text{vs} \quad H_2 : \text{ Patient does have HIV}$$

If we let d_1 to be

$$d_1 : \quad \text{Decide that patient does not have HIV, meaning } d_1 \text{ supports } H_1$$

When we say d_1 is right, it is equivalent to, H_1 happens. When d_1 is wrong, means H_2 happens. Therefore, we have

$$E(L(d_1)) = (0) \times (0.88) + (1000) \times (0.12) = (L(H_1 \text{ happens}))(\mathbb{P}(H_1 \mid \text{data})) + (L(H_2 \text{ happens}))(\mathbb{P}(H_2 \mid \text{data}))$$

- **Bayes Factor:** We introduce this concept in the video **Posterior Probabilities of Hypothesis and Bayes Factors**. Here we need to bring in the **odd** concept. Just like we usually say, there is a “50-50 chance” that the coin toss will give a head, **odd** is just the ratio of the probabilities between 2 events.

Prior Odd: Ratio of the prior probabilities under 2 hypotheses. If we are interested in the case that H_1 over H_2 , we calculate the odd as follow:

$$O[H_1 : H_2] = \frac{\mathbb{P}(H_1)}{\mathbb{P}(H_2)}$$

Posterior Odd: Likewise, we can define the posterior odd H_1 over H_2 as the ratio between the posterior probabilities under the 2 hypotheses.

$$PO[H_1 : H_2] = \frac{\mathbb{P}(H_1 | \text{data})}{\mathbb{P}(H_2 | \text{data})}$$

Using **conditional probability** that we introduced in Week 1, we derive

$$\begin{aligned} \text{posterior odd} = PO[H_1 : H_2] &= \frac{\mathbb{P}(H_1 | \text{data})}{\mathbb{P}(H_2 | \text{data})} = \frac{\frac{\mathbb{P}(H_1 \text{ and data})}{\mathbb{P}(\text{data})}}{\frac{\mathbb{P}(H_2 \text{ and data})}{\mathbb{P}(\text{data})}} = \frac{\mathbb{P}(H_1 \text{ and data})}{\mathbb{P}(H_2 \text{ and data})} \\ &= \frac{\mathbb{P}(\text{data} | H_1)\mathbb{P}(H_1)}{\mathbb{P}(\text{data} | H_2)\mathbb{P}(H_2)} = \underbrace{\frac{\mathbb{P}(\text{data} | H_1)}{\mathbb{P}(\text{data} | H_2)}}_{\text{Bayes factor}} \times \underbrace{\frac{\mathbb{P}(H_1)}{\mathbb{P}(H_2)}}_{\text{prior odd}} \end{aligned}$$

To summarize, we can calculate Bayes factor of H_1 over H_2 using:

$$\text{Bayes factor} = BF[H_1 : H_2] = \frac{\text{posterior odd}}{\text{prior odd}} = \frac{PO[H_1 : H_2]}{O[H_1 : H_2]}$$

or

$$\text{Bayes factor} = BF[H_1 : H_2] = \frac{\mathbb{P}(\text{data} | H_1)}{\mathbb{P}(\text{data} | H_2)}$$

The 1st method tells us how Bayes factor connects prior odd and posterior odd. The 2nd method tells us that Bayes factor only depends on the likelihood of data under the 2 hypotheses. Therefore, when it is not easy to assign a good prior for the hypotheses, Bayes factor still gives us information about the **relative comparison** between the 2 hypotheses.

Note: It is obvious that, if the Bayes factor of H_1 over H_2 is $BF[H_1 : H_2]$, then the Bayes factor of H_2 over H_1 is

$$BF[H_2 : H_1] = \frac{1}{BF[H_1 : H_2]}$$

Note: $BF[H_1 : H_2]$ is interpreted as the evidence to support H_1 over H_2 , or the evidence against H_2 . Bayes factor will never give us the **absolute comparison** between the 2 hypotheses. We can only say which hypothesis is favored over another one because there is relatively stronger evidence to support this hypothesis. To interpret Bayes factor, please refer to the **Jeffreys scale** or **Kass & Raftery scale** in the video.

Note: About Bayes factor when the parameter θ in the likelihood follows a **continuous probability distribution**. In the lecture video, there is a typo about this part. The formula for Bayes factor should be:

$$\text{Bayes factor} = \frac{\mathbb{P}(\text{data} | H_1)}{\mathbb{P}(\text{data} | H_2)} = \frac{\int \mathbb{P}(\text{data}, \theta | H_1) d\theta}{\int \mathbb{P}(\text{data}, \theta | H_2) d\theta} = \frac{\int \mathbb{P}(\text{data} | \theta, H_1) \pi(\theta | H_1) d\theta}{\int \mathbb{P}(\text{data} | \theta, H_2) \pi(\theta | H_2) d\theta}$$

(It should be $\mathbb{P}(\text{data}, \theta | H_1)$ instead of $\mathbb{P}(\text{data} | \theta, H_1)$)

- **Examples of Hypothesis Tests Using Bayes Factors and Prior Choice:** In the rest of the videos, we explore how to use Bayes factors to compare hypotheses, which include:
 - Compare 2 proportions using **Beta-Binomial conjugate family**.
 - Compare 2 paired means using **NormalGamma-Normal conjugate family**. (I will give details of this conjugacy in the Lecture Video Explanations file.)
 - Compare 2 independent means using **NormalGamma-Normal conjugate family**.

In particular, in the last example, we explore the effect of the “**imaginary parameter values**”, the values we use in the prior distribution of the parameters. We see how different values affect the **credible interval**, which will lead to the two paradoxes: Lindley’s paradoxes and the Bartlett’s paradoxes.

The 2 paradoxes suggest the importance of setting appropriate prior when sample size is too large, or when the “**imaginary sample size**” is too large.

- We use heavy-tailed prior as such Cauchy prior to handle the extrasensory perception.
- Suppose we would like to include as little information in the prior as possible, we can consider using the Jeffrey’s prior, which is a limit of some distributions when the “imaginary sample size” goes to infinity.
- To avoid paradoxes, we may also consider the intrinsic prior. The formula for intrinsic prior is too complicated so the video only refers it by names.

Note: Students are required to understand what effects lead to the 2 paradoxes (which is mentioned in the video **Posterior Probability, p -Values and Paradoxes**). However, students do not need to worry about how to do the detailed calculation to obtain the credible intervals and the posterior probability under hypotheses. These will be handled by the `bayes_inference` function in the `statsr` package.

Note: As you may guess, the credible intervals we have obtained in this week are all **equal-tailed credible intervals**.