

STAT 238 - Bayesian Statistics

Lab One

Aditya Guntuboyina

27 Jan 2026

1 More on Example 1 from Lecture

In Lectures 1 and 2, we studied the following problem.

Problem 1.1. *Suppose I just tested positive for Covid. Do I really have Covid?*

Let Θ denote the binary parameter which represents whether I truly have Covid or not ($\Theta = 1$ when I have Covid and $\Theta = 0$ when I don't). Let X denote the binary outcome of the Covid test so that $X = 1$ represents the positive test. We need to calculate the probability:

$$\mathbb{P}\{\Theta = 1 \mid \text{test data and background information}\} \quad (1)$$

where test data is simply $X = 1$, and the background information refers to things like “I have been strictly quarantining for the past 3 weeks”, “I do not have symptoms such as fever” etc.

We used the probability model (below B stands for background information):

$$\text{prior: } \mathbb{P}(\Theta = 1 \mid B) = 0.02$$

$$\text{likelihood: } \mathbb{P}(X = 1 \mid \Theta = 1, B) = 0.99 \quad \mathbb{P}(X = 1 \mid \Theta = 0, B) = 0.04.$$

With these probability assignments, we use the Bayes rule to compute (1) as

$$\begin{aligned} \mathbb{P}(\Theta = 1 \mid X = 1, B) &= \frac{\mathbb{P}(X = 1 \mid \Theta = 1, B)\mathbb{P}(\Theta = 1 \mid B)}{\mathbb{P}(X = 1 \mid \Theta = 1, B)\mathbb{P}(\Theta = 1 \mid B) + \mathbb{P}(X = 1 \mid \Theta = 0, B)\mathbb{P}(\Theta = 0 \mid B)} \\ &= \frac{0.99 * 0.02}{0.99 * 0.02 + 0.04 * 0.98} = 0.3356. \end{aligned}$$

This probability is not very high even though the test has very good false positive and false negative rates. This is because the prior probability $\mathbb{P}(\Theta = 1 \mid B)$ is very low (0.02). So, even with the positive test result, it is more likely than not that we are covid-free.

On the other hand, if we pose the problem as hypothesis testing:

$$H_0 : \Theta = 0 \quad \text{versus} \quad H_1 : \Theta = 1$$

and calculate the p -value as

$$\mathbb{P}\{X = 1 \mid H_0\} = \mathbb{P}(X = 1 \mid \Theta = 0) = 0.04,$$

we get a different result (if we use the standard cutoff 0.05 on the p -value). This leads to rejection of the null hypothesis and declaring that I have Covid.

The use of p -values has been linked to serious issues such as lack of reproducibility (see for example the paper "The reproducibility of research and the misinterpretation of p -values" by David Colquhoun). In this context, we can calculate the probability of reproducibility of the positive test as follows. Let X_2 denote the outcome of the second test (and $X_1 = X$ will now denote the outcome of the first test):

$$\begin{aligned}\mathbb{P}(X_2 = 1|X_1 = 1, B) &= \mathbb{P}(X_2 = 1|\Theta = 1, X_1 = 1, B)\mathbb{P}(\Theta = 1|X_1 = 1, B) \\ &+ \mathbb{P}(X_2 = 1|\Theta = 0, X_1 = 1, B)\mathbb{P}(\Theta = 0|X_1 = 1, B).\end{aligned}$$

To calculate the probabilities in the right hand side above, we make the following assignment:

$$\begin{aligned}\mathbb{P}(X_2 = 1|\Theta = 1, X_1 = 1, B) &= \mathbb{P}(X_2 = 1|\Theta = 1) = 0.99 \\ \mathbb{P}(X_2 = 1|\Theta = 0, X_1 = 1, B) &= \mathbb{P}(X_2 = 1|\Theta = 0) = 0.04.\end{aligned}$$

This assumption means that conditional on my Covid status Θ , the two test outcomes X_1 and X_2 are independent. Using this assignment, it is straightforward to calculate the reproducibility probability as follows (note that we already calculated $\mathbb{P}(\Theta = 1 | X_1 = 1, B) = 1 - \mathbb{P}(\Theta = 0 | X_1 = 1, B) = 0.3356$)

$$\mathbb{P}(X_2 = 1|X_1 = 1, B) = 0.99 * 0.3356 + 0.04 * (1 - 0.3356) = 0.35882.$$

Thus the positive test will be reproducible with probability only 35.88% which means that it is more likely to get a negative test the second time.

2 Law of Total Probability and Bayes Rule

In Bayesian statistics, the rules of probability are used mostly for the following:

1. to compute the marginal distribution of X based on knowledge of the conditional distribution of X given $\Theta = \theta$ (i.e., the likelihood) as well as the marginal distribution of Θ (i.e., the prior)
2. to compute the conditional distribution of Θ given $X = x$ (i.e., the posterior) based on the same knowledge of the conditional distribution of X given $\Theta = \theta$ (i.e., the likelihood) as well as the marginal distribution of Θ (i.e., the prior).

The formula for the first item above is sometimes called the Law of Total Probability (LTP), while the formula for the second item is called the Bayes Rule. The precise formulae differ according to whether X and Θ are discrete or continuous. It is natural to consider the following four separate cases.

2.1 X and Θ are both discrete

The LTP is

$$\mathbb{P}\{X = x\} = \sum_{\theta} \mathbb{P}\{X = x|\Theta = \theta\}\mathbb{P}\{\Theta = \theta\}$$

and the Bayes rule is

$$\mathbb{P}\{\Theta = \theta|X = x\} = \frac{\mathbb{P}\{X = x|\Theta = \theta\}\mathbb{P}\{\Theta = \theta\}}{\mathbb{P}\{X = x\}} = \frac{\mathbb{P}\{X = x|\Theta = \theta\}\mathbb{P}\{\Theta = \theta\}}{\sum_{\theta} \mathbb{P}\{X = x|\Theta = \theta\}\mathbb{P}\{\Theta = \theta\}}.$$

2.2 X and Θ are both continuous

Here LTP is

$$f_X(x) = \int f_{X|\Theta=\theta}(x) f_\Theta(\theta) d\theta$$

and Bayes rule is

$$f_{\Theta|X=x}(\theta) = \frac{f_{X|\Theta=\theta}(x) f_\Theta(\theta)}{f_X(x)} = \frac{f_{X|\Theta=\theta}(x) f_\Theta(\theta)}{\int f_{X|\Theta=\theta}(x) f_\Theta(\theta) dx}.$$

2.3 X is discrete while Θ is continuous

LTP is

$$\mathbb{P}\{X = x\} = \int \mathbb{P}\{X = x|\Theta = \theta\} f_\Theta(\theta) d\theta$$

and Bayes rule is

$$f_{\Theta|X=x}(\theta) = \frac{\mathbb{P}\{X = x|\Theta = \theta\} f_\Theta(\theta)}{\mathbb{P}\{X = x\}} = \frac{\mathbb{P}\{X = x|\Theta = \theta\} f_\Theta(\theta)}{\int \mathbb{P}\{X = x|\Theta = \theta\} f_\Theta(\theta) d\theta}.$$

2.4 X is continuous while Θ is discrete

LTP is

$$f_X(x) = \sum_{\theta} f_{X|\Theta=\theta}(x) \mathbb{P}\{\Theta = \theta\}$$

and Bayes rule is

$$\mathbb{P}\{\Theta = \theta|X = x\} = \frac{f_{X|\Theta=\theta}(x) \mathbb{P}\{\Theta = \theta\}}{f_X(x)} = \frac{f_{X|\Theta=\theta}(x) \mathbb{P}\{\Theta = \theta\}}{\sum_{\theta} f_{X|\Theta=\theta}(x) \mathbb{P}\{\Theta = \theta\}} \quad (2)$$

Next is a simple application for the case where Θ is discrete and X is continuous.

3 A Simple Model Selection Application

Suppose Θ has the $Ber(0.5)$ distribution i.e.,

$$\mathbb{P}\{\Theta = 0\} = \mathbb{P}\{\Theta = 1\} = 0.5.$$

Next assume that X_1, \dots, X_n have the following distributions conditional on $\Theta = \theta$:

$$X_1, \dots, X_n | \Theta = 0 \stackrel{\text{i.i.d.}}{\sim} f_0 \text{ where } f_0(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

and

$$X_1, \dots, X_n | \Theta = 1 \stackrel{\text{i.i.d.}}{\sim} f_1 \text{ where } f_1(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-|x| \sqrt{\frac{2}{\pi}}\right).$$

f_0 is the standard normal density and f_1 is a Laplace (double-exponential) density. Both densities have the same maximal value of $1/\sqrt{2\pi}$. Based on the information given, calculate the conditional distribution of Θ given $X_1 = x_1, X_2 = x_2, \dots, X_6 = x_6$ (i.e., $n = 6$) where

$$x_1 = -0.55, x_2 = -1.11, x_3 = 1.23, x_4 = 0.29, x_5 = 1.56, x_6 = -1.64. \quad (3)$$

Here is the statistical context for this question. We observe data x_1, \dots, x_n with $n = 6$. We want to use one of the models f_0 or f_1 for this data. The random variable Θ is used to describe the choice of the model. We want to treat both the models on an equal footing so we assumed that Θ has the uniform prior distribution on $\{0, 1\}$.

To calculate the conditional distribution of Θ given the data, we use the formula (2) because Θ is discrete and the data X_1, \dots, X_n are continuous. This gives

$$\begin{aligned}
& \mathbb{P}\{\Theta = 0 \mid X_1 = x_1, \dots, X_n = x_n\} \\
&= \frac{f_{X_1, \dots, X_n \mid \Theta=0}(x_1, \dots, x_n) \mathbb{P}\{\Theta = 0\}}{f_{X_1, \dots, X_n \mid \Theta=0}(x_1, \dots, x_n) \mathbb{P}\{\Theta = 0\} + f_{X_1, \dots, X_n \mid \Theta=1}(x_1, \dots, x_n) \mathbb{P}\{\Theta = 1\}} \\
&= \frac{f_{X_1, \dots, X_n \mid \Theta=0}(x_1, \dots, x_n) \times \frac{1}{2}}{f_{X_1, \dots, X_n \mid \Theta=0}(x_1, \dots, x_n) \times \frac{1}{2} + f_{X_1, \dots, X_n \mid \Theta=1}(x_1, \dots, x_n) \times \frac{1}{2}} \\
&= \frac{f_{X_1, \dots, X_n \mid \Theta=0}(x_1, \dots, x_n)}{f_{X_1, \dots, X_n \mid \Theta=0}(x_1, \dots, x_n) + f_{X_1, \dots, X_n \mid \Theta=1}(x_1, \dots, x_n)} \\
&= \frac{f_0(x_1) f_0(x_2) \dots f_0(x_n)}{f_0(x_1) f_0(x_2) \dots f_0(x_n) + f_1(x_1) f_1(x_2) \dots f_1(x_n)} \\
&= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right)}{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right) + \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\sqrt{\frac{2}{\pi}} \sum_{i=1}^n |x_i|\right)} \\
&= \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right)}{\exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right) + \exp\left(-\sqrt{\frac{2}{\pi}} \sum_{i=1}^n |x_i|\right)}.
\end{aligned}$$

Similarly

$$\mathbb{P}\{\Theta = 1 \mid X_1 = x_1, \dots, X_n = x_n\} = \frac{\exp\left(-\sqrt{\frac{2}{\pi}} \sum_{i=1}^n |x_i|\right)}{\exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right) + \exp\left(-\sqrt{\frac{2}{\pi}} \sum_{i=1}^n |x_i|\right)}.$$

Plugging in the above formula the data values given in (3) for x_1, \dots, x_6 , we obtain

$$\mathbb{P}\{\Theta = 0 \mid X_1 = x_1, \dots, X_6 = x_6\} = 0.72 \quad \text{and} \quad \mathbb{P}\{\Theta = 1 \mid X_1 = x_1, \dots, X_6 = x_6\} = 0.28$$

Thus, conditioning on the data, we have a 72% probability for the normal model compared to 28% probability for the Laplace model. So we would prefer to use the normal distribution here.

Now suppose that we add in an additional observation $x_7 = 5$. It can be checked that

$$\mathbb{P}\{\Theta = 0 \mid X_1 = x_1, \dots, X_7 = x_7\} = 0.001 \quad \text{and} \quad \mathbb{P}\{\Theta = 1 \mid X_1 = x_1, \dots, X_7 = x_7\} = 0.999$$

Now there is overwhelming preference for the Laplace model. This is because $x_7 = 5$ is an outlying observation to which the Laplace model gives much higher probability compared to the Normal model owing to heavy tails of the Laplace density.