

STAT 238 - Bayesian Statistics

Lecture Two

Spring 2026, UC Berkeley

Aditya Guntuboyina

23 Jan 2026

In the last lecture, we started with the following simple example of Bayesian inference.

1 Example 1: Testing and Covid

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

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

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.

Here is an alternative method of reasoning in this problem. We can formulate this as a hypothesis testing problem with

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

The null hypothesis represents not having covid. The p -value in the above testing problem equals:

$$\mathbb{P}\{X = 1|H_0\} = \mathbb{P}(X = 1|\Theta = 0) = 0.04.$$

Usage of the naive standard cutoff 0.05 on the p -value would now lead to rejection of the null hypothesis and declaring that I have Covid. On the other hand, the previous argument (based on probability theory) gave a much higher probability to me not having Covid.

This p -value based method does not even make use of the information given on $\mathbb{P}(\Theta = 1 | B)$ and $\mathbb{P}(X = 1|\Theta = 1, B)$. It only makes use of $\mathbb{P}(X = 1|\Theta = 0)$. Note that what we are after is $\mathbb{P}(\Theta = 0|X = 1)$ while the p -value is $\mathbb{P}\{X = 1 | \Theta = 0\}$. In general, $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ can be quite different. The correct way of relating them is via the Bayes rule. Without using Bayes rule, one cannot argue that $\mathbb{P}(A | B)$ is large or small using largeness or smallness of $\mathbb{P}(B | A)$.

Consider, for example, the case where A represents the event that a person is dead and B represents the event that they were hanged. Here $\mathbb{P}(A | B)$ is quite close to one while $\mathbb{P}(B | A)$ is quite close to zero.

It is therefore quite problematic that one can say something about $\Theta = 0|X = 1$ from only $\mathbb{P}(X = 1|\Theta = 0)$.

Methods such as testing based on p -values (and putting arbitrary cutoffs on them) are not based on probability. They are examples of frequentist reasoning.

2 Example 2: Spots on a patient

Problem 2.1 (From the book “A tutorial introduction to Bayesian Analysis” by James Stone). *Suppose you are a doctor confronted with a patient who is covered in spots. This patient’s symptoms are consistent with chickenpox but they are also consistent with another, more dangerous, disease, smallpox. How would you decide if they have chickenpox or smallpox?*

Here the unknown parameter is Θ which represents the disease status. Θ can take the three values smallpox, chickenpox, neither of them.

The data is that the patient has spots.

We need to calculate the probability:

$$\mathbb{P}\{\Theta = \text{smallpox} | \text{spots} + B\} \tag{2}$$

where B again represents background information representing other symptoms (e.g., fever) that the patient has. Here is one probability assignment which allows us to calculate this probability:

$$\mathbb{P}\{\text{spots} | \text{smallpox}, B\} = 0.9 \quad \mathbb{P}\{\text{spots} | \text{chickenpox}, B\} = 0.8 \quad \mathbb{P}\{\text{spots} | \text{neither}, B\} = 0 \tag{3}$$

and

$$\mathbb{P}\{\text{smallpox} | B\} = 0.001 \quad \mathbb{P}\{\text{chickenpox} | B\} = 0.1 \quad \mathbb{P}\{\text{neither} | B\} = 0.899. \tag{4}$$

Here “neither” refers to an underlying cause for the patient’s condition that is neither smallpox nor chickenpox.

Using this assignment, the required probability (2) can be calculated via Bayes rule and this leads to

$$\mathbb{P}\{\text{smallpox} \mid \text{spots}, B\} \approx 0.011 \quad \mathbb{P}\{\text{chickenpox} \mid \text{spots}, B\} \approx 0.988 \quad \mathbb{P}\{\text{neither} \mid \text{spots}, B\} = 0.$$

So probability theory with the assignment (3) and (4) says that it is highly likely that the patient has chickenpox (smallpox is basically ruled out because it is extremely rare).

2.1 Alternative Solution: Maximum Likelihood

Here is an alternative way of solving this problem using maximum likelihood estimation. The maximum likelihood estimate in this case is smallpox because smallpox leads to a higher probability (0.9) of the observed data (spots) compared to chickenpox (0.8). Maximum Likelihood (widely used in statistics) is not based on probability theory and also seems to be based on the wrong conditional probabilities $\mathbb{P}\{\text{spots} \mid \text{smallpox}\}$ and $\mathbb{P}\{\text{spots} \mid \text{chickenpox}\}$ while we really should be calculating $\mathbb{P}\{\text{smallpox} \mid \text{spots}\}$ and $\mathbb{P}\{\text{chickenpox} \mid \text{spots}\}$.

For the modeling part in Bayesian applications, for the most part, we shall use standard models based on normal distributions, more generally, exponential families. Here is a simple example to illustrate the use of the normal distribution.

3 Example 3: Inference from measurements

Suppose a scientist makes 6 numerical measurements 26.6, 38.5, 34.4, 34, 31, 23.6 on an unknown real-valued physical quantity θ . On the basis of these measurements, what can be inferred about θ ?

Here is the Bayesian solution to this problem. The first step is modeling where we have to write the likelihood and prior. The likelihood represents the probability of the observed data conditional on parameter values. Here the main parameter is θ . In order to write the probability of the observed data, it is helpful to introduce another parameter σ which represents the scale of the noise inherent in the measurement process.

So our parameter vector is (θ, σ) . We work with the normal likelihood:

$$\text{Likelihood} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right). \quad (5)$$

where $n = 6$ and $x_1 = 26.6, x_2 = 38.5, x_3 = 34.4, x_4 = 34, x_5 = 31, x_6 = 23.6$ denote the observed data points. More formally, you can arrive at this likelihood in the following way. Denote potential measurements by X_1, \dots, X_n . Each actual measurement will have some rounding error so the data point 26.6 should be understood as belonging to the interval $[26.6 - \delta, 26.6 + \delta]$ for some small rounding error δ . So the likelihood is:

$$\begin{aligned} \text{likelihood} &= \mathbb{P}\{\text{observed data} \mid \theta, \sigma\} \\ &= \mathbb{P}\{X_1 \in [x_1 - \delta, x_1 + \delta], \dots, X_n \in [x_n - \delta, x_n + \delta] \mid \theta, \sigma\}. \end{aligned}$$

Assuming δ is small, we can use probability-density approximation to write

$$\text{likelihood} \approx \delta^n f_{X_1, \dots, X_n \mid \theta, \sigma}(x_1, \dots, x_n).$$

We are now assuming that:

$$f_{X_1, \dots, X_n | \theta, \sigma}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right). \quad (6)$$

This leads to the likelihood (5) (note that δ^n is being dropped as it is a constant of proportionality which does not affect any further calculations).

We will complete this example in the next lecture.