

# STAT 238 - Bayesian Statistics

## Lecture One

Spring 2026, UC Berkeley

Aditya Guntuboyina

21 Jan 2026

### 1 What is Bayesian Statistics?

Bayesian Statistics = Probability Theory

This is the central message of the book “Probability Theory – the logic of Science” by E. T. Jaynes which is arguably the most important book on Bayesian Statistics.

### 2 How does Bayesian Statistics work?

Probability works in the following way. We are interested in knowing whether a certain proposition is true, and we do not have access to full information that would allow us to conclusively determine whether the proposition is true or not. Probability theory allows us to determine a number between 0 and 1 representing how likely it is that the proposition is true based on the available information. This is achieved by the following two steps:

1. **Step One:** The available information that we either possess or that we assume for the sake of argument is converted into **numerical assignments** for the probabilities of certain basic or elementary propositions. This step is often referred to as the **modeling** step.
2. **Step Two:** Based on the probability model, we calculate probabilities of the propositions of interest using the **rules of probability**.

In the context of Bayesian statistics, the unknown proposition is usually written in terms of a variable  $\Theta$  e.g.,  $\Theta \in A$  for some set  $A$ . The goal is to calculate the probability:

$$\mathbb{P}\{\Theta \in A \mid \text{observed data and other background information}\}. \quad (1)$$

The modeling step involves specification of:

$$\mathbb{P}\{\text{observed data} \mid \Theta = \theta \text{ and background information}\} \quad (2)$$

as well as

$$\mathbb{P}\{\Theta = \theta \mid \text{background information}\} \text{ or } f_{\Theta|\text{background information}}(\theta) \quad (3)$$

It is okay to simply think of all these as just probabilities. However, it is common to use the following terminology. (3) is called the prior model (or prior distribution), (2) is called the likelihood, and (1) is called posterior probability.

### 3 Rules of Probability

Probabilities are assigned to propositions (also known as events). Every probability is conditional on some information (this could be available information or some information that we assume for the sake of argument). We shall denote the probability of a proposition  $A$  conditioned on some information  $I$  by  $\mathbb{P}(A | I)$ . When the information  $I$  is clear from context, we sometimes omit it and write the probability  $\mathbb{P}(A | I)$  as simply  $\mathbb{P}(A)$ . Even when we do this, it should always be kept in mind that probabilities are always conditioned on some information.

1. The probability of a proposition always lies between 0 and 1. The probability of an impossible proposition is 0 and the probability of a certain proposition is 1.
2. **Product Rule:**  $\mathbb{P}(A \cap B | I) = \mathbb{P}(A | I)\mathbb{P}(B | A, I) = \mathbb{P}(B | I)\mathbb{P}(A | B, I)$ . Here  $A \cap B$  is the proposition: “both  $A$  and  $B$  are true”. Also  $\mathbb{P}(A | B, I)$  is the probability of  $A$  conditioned on the truth of the proposition  $B$  as well as the information  $I$ . A direct consequence of the product rule is:

$$\mathbb{P}(B | A, I) = \frac{\mathbb{P}(A | B, I)\mathbb{P}(B | I)}{\mathbb{P}(A | I)}.$$

The above formula is known as the Bayes rule.

3. **Sum Rule:** If a proposition  $A$  is broken down into disjoint propositions  $A_1, A_2, \dots$ , then

$$\mathbb{P}(A | I) = \sum_i \mathbb{P}(A_i | I)$$

Disjoint here means that no two of  $A_i$ s can happen simultaneously. For the sum rule, we need

$$A \text{ happens} \iff \text{exactly one of } A_i \text{ happens.}$$

We shall see some justification for these rules later. All other rules of probability follow as a consequence of these rules. For example, the Bayes rule states that:

$$\mathbb{P}(B | A, I) = \frac{\mathbb{P}(A | B, I)\mathbb{P}(B | I)}{\mathbb{P}(A | B, I)\mathbb{P}(B | I) + \mathbb{P}(A | B^c, I)\mathbb{P}(B^c | I)}$$

### 4 Example 1: Testing and Covid

**Problem 4.1.** Suppose I just tested positive for Covid. Do I really have Covid?

Let  $\Theta$  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:

$$\mathbb{P}\{\Theta = 1 | \text{test data and background information}\} \tag{4}$$

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.

In order to calculate the posterior probability (4), we need to introduce probability assumptions. Consider the following model (below  $B$  stands for background information)

$$\mathbb{P}(\Theta = 1 | B) = 0.02 \quad \mathbb{P}(X = 1 | \Theta = 1, B) = 0.99 \quad \mathbb{P}(X = 1 | \Theta = 0, B) = 0.04. \tag{5}$$

$\mathbb{P}(\Theta = 1 \mid B)$  represents the probability of Covid based on background information alone (this is the prior). The fact that it is low (0.02) is meaningful when I know that I have been largely isolating myself for the past few weeks.

$\mathbb{P}(X = 1 \mid \Theta = 1, B) = 0.99$  (true positive rate) and  $\mathbb{P}(X = 1 \mid \Theta = 0, B) = 0.04$  (false positive rate) represent the likelihood.

With these probability assignments, we use the Bayes rule to compute (4) as

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

Thus, under the assumed probability model, there is a 33.56% chance that I am truly covid positive given the positive test. Note that 0.3356 (33.56%) is not very high even though the test has very good false positive and false negative rates. This is because  $\mathbb{P}(\Theta = 1 \mid B)$  (which can be interpreted as probability of having Covid without taking into the account the test result) is very low (0.02).