1 Conditional Probability

1.1 Bayes' Formula

Bayes' Formula is a powerful – and very famous – application of the conditional probability formula. Often, it can be difficult to calculate a conditional probability $\mathbb{P}(A|B)$ of interest. However, the other way around, $\mathbb{P}(B|A)$, could be easier to find. Bayes' Formula gives us a way of finding $\mathbb{P}(A|B)$, provided that we know $\mathbb{P}(A)$, $\mathbb{P}(B)$, and $\mathbb{P}(B|A)$. Recall that the conditional probability formula is given by

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

We also know that

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

So, solving for $\mathbb{P}(A \cap B)$, we have

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \qquad \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A).$$

Then, setting these terms equal to each other, we have

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

Thus, we get

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

It is often useful to apply the conditional Law of Total Probability in the denominator; that is,

$$\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^C)\mathbb{P}(B|A^C).$$

So, sometimes, Bayes' Rule is stated like so:

$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^C)\mathbb{P}(B|A^C)}}.$$

Now, there is a general version of Bayes' Rule; suppose that $B \subset \Omega$ is an event and $A_1, \ldots, A_n$ partitions the sample space Ω. Then, for each $1 \leq j \leq n$, we have

$$\boxed{\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j)\mathbb{P}(B|A_j)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}}.$$

Here[1],

- The $\mathbb{P}(A_j)$ are called **prior probabilities**.
- The $\mathbb{P}(A_j|B)$ are called **posterior probabilities**.
- The events A_j are called **hypotheses**.
- The event B is called the **evidence**.

Often, we want to see which hypothesis is more likely given the evidence.

[1] In this course, we aren't expected to memorize these.

(Example.) Suppose that a doctor gives a patient a test for cancer. Before the test, all we know is that, on average, 1 in every 1000 women develop this cancer. According to the manufacturer, this test is 99% accurate at detecting the cancer, when it is there. However, there is a 5% chance it will show positive when the cancer is not there (i.e. there is a 1% chance of a false negative, and a 5% chance of a false positive). To make things more clear:

- If she has the cancer, there's a 1% chance the test will incorrectly say negative, and a 99% chance the test will correctly say positive.
- If she does not have the cancer, there's a 5% chance the test will incorrectly say positive.

Now, suppose that the patient tests positive. *With what probability does she actually have the cancer?*

Let C be the event that she has this cancer. Let P be the event that this test is positive. We want to find $\mathbb{P}(C|P)$. We know that

$$\mathbb{P}(C) = 0.001 \qquad \mathbb{P}(P|C) = 0.99 \qquad \mathbb{P}(P|C^C) = 0.05.$$

By Bayes' Rule, we have

$$\mathbb{P}(C|P) = \frac{\mathbb{P}(C)\mathbb{P}(P|C)}{\mathbb{P}(C)\mathbb{P}(P|C) + \mathbb{P}(C^C)\mathbb{P}(P|C^C)} = \frac{0.001(0.99)}{0.001(0.99) + 0.999(0.05)} \approx 1.94\%.$$

Remark: Given that the test is supposed to be 99% accurate and that the patient tested positive, one would think that the patient has the cancer; so, a near 2% probability that the patient has cancer is quite surprising. However, there are a few reasons why this may be the case.

1. Even though the test is fairly accurate when you do have the cancer, when you don't have the cancer there is a decent chance (5%) that it will be incorrect.
2. This cancer is *extremely* rare; it's so rare that there is a much better chance that this test will incorrectly read positive than if you actually have cancer.

Now, without any additional information (i.e. without the test), we could only assume that the patient had this cancer with probability 0.1%. After testing positive, this probability increases by quite a bit, but it is still only 1.94%. This example demonstrates that it is very difficult to design an accurate test for a rare disease; in this example, the probability of a false positive is much more likely than the patient actually having cancer.