Conditional Probability and Bayes Theorem

1 The Monty Hall Paradox

We start with a puzzle – the famous Monty Hall paradox. Suppose you are a contestant in game show. The host shows you three closed doors. One door leads to a brand new car, and the other two doors each lead to a lemon. If you are able to find the door leading to the car, you get the car! You are asked to pick select of the doors, but not open it. You go ahead and open one of the doors. At this point, the host opens one of the other two doors that reveals a lemon, and asks: "Do you want to stay with your choice, or switch?" What should you do? Does it matter?

A priori, without any information each of the three doors is equally likely to be the one leading to the car, so at first glance it seems that it does not matter whether you stay or you switch. Turns out it does! Let us first understand this using basic counting; we will then explain this in the context of conditional probability.

Suppose the three doors are labeled A, B, and C. Without loss of generality, we may assume that you selected A as your first choice. There are 3 equally likely possibilities: the car is behind door A, the car is behind door B, or the car is being door C. Let us consider each of these in turn.

If the car is behind A, then the host can open any one of the other two doors since both lead to lemons. And in this case, your correct response (in hindsight) is to stay.

If the car is behind B, then the host will open door C. In this case, your correct response is to switch.

Similarly, if the car is behind C, then the host will open door B, and your correct response is to switch.

So in 2 out of the 3 equally likely cases, you should switch. This means you are more likely to win the car if you switch.

2 Conditional Probability

The *conditional probability of an event E_1 given event E_2* is the probability that event E_1 occurs, given that event E_2 occurs. It is denoted by $\Pr[E_1|E_2]$ and can be defined as:

$$\Pr[E_1|E_2] = \frac{\Pr[E_1 \cap E_2]}{\Pr[E_2]}.$$

As an example, consider the roll of two fair dice. A priori, the probability that the roll of the first dice yields a 5 is 1/6. But if we are told that the sum of the two dice is 9, then the probability of obtaining a 5 in the first roll, given this new information changes.

$$\Pr[\text{the first roll is a 5} \mid \text{the sum is 9}] = \frac{\Pr[\text{the first roll is a 5 AND the sum is 9}]}{\Pr[\text{the sum is 9}]} = \frac{\frac{1}{36}}{\frac{4}{36}} = \frac{1}{4}.$$

3 Bayes Theorem

Bayes Theorem relates one conditional probability (e.g., the probability of a hypothesis H given an observation E) with its inverse (the probability of an observation given a hypothesis). Bayes Theorem is used heavily in statistics, analysis of data sets, machine learning, information retrieval, and several diverse applications in science and engineering.

To understand Bayes Theorem, it is crucial to note that $\Pr[H|E]$ is quite different from $\Pr[E|H]$. Bayes Theorem states that

$$\Pr[H|E] = \frac{\Pr[E|H] \cdot \Pr[H]}{\Pr[E]}.$$

The theorem directly follows from the definition of conditional probability.

$$\Pr[H|E] = \frac{\Pr[H \cap E]}{\Pr[E]} = \frac{\Pr[E|H] \cdot \Pr[H]}{\Pr[E]}.$$

The above equation can also be easily seen using Venn diagrams, as covered in class.

3.1 Explaining Monty Hall paradox using Bayes Theorem

Let us understand the Monty Hall paradox using Bayes Theorem. When the host asks you whether you would like to switch, the calculation you should do is to determine the probability that you will win the prize given the information provided to you.

As before, suppose without loss of generality that the door you select is labeled A. Also, let us label the door that the host opens as B. Now, consider the following calculations.

$$\Pr[\text{prize behind A} \mid \text{host opens B}] = \frac{\Pr[\text{host opens B} \mid \text{prize behind A}]\,\Pr[\text{prize behind A}]}{\Pr[\text{host opens B}]} = \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

$$\Pr[\text{prize behind B} \mid \text{host opens B}] = \frac{\Pr[\text{host opens B} \mid \text{prize behind B}]\,\Pr[\text{prize behind B}]}{\Pr[\text{host opens B}]} = 0.$$

$$\Pr[\text{prize behind C} \mid \text{host opens B}] = \frac{\Pr[\text{host opens B} \mid \text{prize behind C}]\,\Pr[\text{prize behind C}]}{\Pr[\text{host opens B}]} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Thus, you would be better off switching to one of the other two unopened doors.

3.2 Another application of Bayes Theorem

Here is an example that illustrates the use of Bayes Theorem in biostatistics. Suppose the a prior probability of you being infected with the H1N1 virus is 10^{-5}. Further suppose that a blood test is 99% accurate and you test positive. How likely is it that you actually have the virus? Let us do the calculations.

$$
\begin{aligned}
\Pr[\text{virus}] &= 0.00001 \\
\Pr[\text{no virus}] &= 0.99999 \\
\Pr[\text{positive test} \mid \text{virus}] &= 0.99 \\
\Pr[\text{positive test} \mid \text{no virus}] &= 0.01 \\
\Pr[\text{virus} \mid \text{positive test}] &= \frac{\Pr[\text{positive test} \mid \text{virus}] \cdot \Pr[\text{virus}]}{\Pr[\text{positive test}]} \\
&= \frac{0.99 \cdot 0.00001}{\Pr[\text{positive test}]} \\
&= \frac{0.0000099}{\Pr[\text{positive test}]} \\
\Pr[\text{no virus} \mid \text{positive test}] &= \frac{\Pr[\text{positive test} \mid \text{no virus}] \cdot \Pr[\text{no virus}]}{\Pr[\text{positive test}]} \\
&= \frac{0.01 \cdot 0.99999}{\Pr[\text{positive test}]} \\
&= \frac{0.0099999}{\Pr[\text{positive test}]}
\end{aligned}
$$

Thus, even after testing positive, you are *1000 times more likely* not to have the virus than have it.