1.7 Bayes Rule

We have previously been looking at how knowledge of some additional information can affect the probability of an event occurring. In this section we will combine our previous tools to form one of the central results in probability, Baye's Rule.

The setup we consider is some event where have a prior belief on what the probability is. For example, we may believe that a certain cancer form of cancer is present in 1% of the population. Therefore we believe the probability a given person would have this cancer is 0.01. Suppose a random person tests positive for this cancer, where the test has some prescribed error probabilities. Then Bayes Rule tells us the probability that this person actually has the cancer given that they positive. Therefore we have updated our belief that this person is positive from 0.01, to some larger value given by Bayes Rule. Essentially Bayes Rule allows us to update our beliefs when given some new information. As terminology you may see from outside this course, your prior belief in an event is called the *prior probability*, while the updated belief given by Bayes Rule is the *posterior probability*. We note that we can tackle these kinds of problems already with the tools we have developed in previous sections, however Bayes' Rule packages this process into one formula.

Lemma 1.7.1. Suppose A and B are events in a sample space Ω. Then we have that:

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

Furthermore we have:

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

We think of A as some prior event, in the example above that would be the event that a random person has the form cancer. We then think of B as some new information, for example B is the event that the person tested positive. Then the event $\mathbb{P}(A|B)$, precisely tells us the probability this person has the disease, given that they tested positive. The second formula given looks slightly more intimidating, but if you look closely all we have done is apply the law of total probability to $\mathbb{P}(B)$ using the event A. In practice, the second formula tends to be more useful.

Example 1.7.1. Suppose a certain form of flu is incident in 1% of the population. A test for this flu is designed, and the manufactures claim the following: If the person has the flu, then the test will detect it successfully 99% of the time. If the person does not have the flu, then the test will incorrectly show a positive result 10% of the time. A random person takes one of these tests. The test shows a positive result, what is the probability that they have the flu?

The challenge with these style of questions is the sheer amount of information that you need to work with. Again, it is always best to identify which events are best to look at. We are interested in the probability that the person has the disease, given that they tested positive. So let F be the event that the person has the flu, while let T be the event that person tested positive. Therefore we need to find $\mathbb{P}(F|T)$. Now we express the information given in terms of the events F and T. As flu is incident in 1% of the population, the prior probability a given person has the flu is 0.01, therefore $\mathbb{P}(F) = 0.01$. Next if the person has the flu, they test positive 99% of the percent of the time, therefore the probability they test positive, given that they have the flu, $\mathbb{P}(T|F) = 0.99$. Similarly if they do not have the flu, denoted F^c, the probability the test positive is 0.1. Hence we have that $(T|F^c) = 0.1$. By recalling that $\mathbb{P}(F^c) = 1 - \mathbb{P}(F)$, we hence have all the terms needed to apply the second form of Bayes Theorem:

$$\mathbb{P}(F|T) = \frac{\mathbb{P}(T|F)\mathbb{P}(F)}{\mathbb{P}(T|F)\mathbb{P}(F) + \mathbb{P}(T|F^c)\mathbb{P}(F^c)} = \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.1 \times 0.99} = 0.091$$

We remark that on a first glance this may seem quite paradoxical. The test itself is 99% accurate, and has a fairly low false positive rate of 10%, however if you test positive, then you will only have the disease 9.1% of the time. Essentially when you test positive, one of two things occur, you are a true positive or a false positive. However as the incidence of the disease is so low in the population (only 1 in 100), a much larger proportion of the positive tests come from the false positives. In other words, the false positive rate dominates the incidence of the disease in the population. Thus when we look at the proportion of people who tested positive, around 9 in 10 of them are actually false positives.

We consider one further example along a similar vein. Usually with these examples you want to identify events first. Then work term-by-term through Bayes Theorem

Example 1.7.2. Suppose a learning algorithm is attempting to identify whether pictures are human faces. When running on training examples, if the input picture was a human then the algorithm would correctly identify it 80% of the time, when the input was non-human then the machine would incorrectly classify it as a human 30% of the time. A random photo from a training set consisting of an equal proportion of human and non-humans photos is input into the algorithm. The algorithm identifies the photo as a human, what is the probability that the photo is indeed a human?

Again we have two events occurring here, firstly let H be the event that photo fed to the algorithm was a human. While let P be the event that algorithm identifies its input as a human. Therefore we are looking to find $\mathbb{P}(H|P)$. We apply Bayes Theorem hence:

$$\mathbb{P}(H|P) = \frac{\mathbb{P}(P|H)\mathbb{P}(H)}{\mathbb{P}(P|H)\mathbb{P}(H) + \mathbb{P}(P|H^c)\mathbb{P}(H^c)}$$

We now work term by term using the information given. We have that $\mathbb{P}(P|H) = 0.8$, as there is an 80% chance of the algorithm identifying the input as human, given that the input is human. As there are an equal proportion of human and non-human photos, and we choose one at random, we have that $\mathbb{P}(H) = \mathbb{P}(H^c) = 0.5$. Finally if the input is not a human, then the algorithm will incorrectly identify it as one 30% of the time, therefore $\mathbb{P}(P|H^c) = 0.3$. Plugging these values in we have:

$$\mathbb{P}(H|P) = \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.3 \times 0.5} = 0.727.$$