ECE 3530/CS 3130: Engineering Probability & Stats
Lecture Notes Spring 2015

Lecture #8

1 Bayes Law

Bayes Law is simply a re-writing of the conditional probability definition. Given any two events A and B,

$$P[B|A] = \frac{P[B \cap A]}{P[A]} \tag{1}$$

The first line can be rewritten as $P[B \cap A] = P[B|A]\,P[A]$. But also note that we could have used the definition of conditional probability to write:

$$P[A|B] = \frac{P[B \cap A]}{P[B]}$$

So we can also write the joint probability as $P[B \cap A] = P[A|B]\,P[B]$. Thus $P[B|A]\,P[A] = P[A|B]\,P[B]$. The simplest definition of Bayes' Law is to take this and divide both sides by $P[A]$ to get:

$$P[B|A] = \frac{P[A|B]\,P[B]}{P[A]} \tag{2}$$

In short, Bayes' Law is a way to convert $P[A|B]$, to the other way around, $P[B|A]$, using the probabilities of the two events A and B.

Often, we won't be given $P[A]$. Instead, we'll have to find it ourselves using the partition, $\{B, B^C\}$, and the law of total probability:

$$P[A] = P[A \cap B] + P\left[A \cap B^C\right] = P[A|B]\,P[B] + P\left[A|B^C\right]P\left[B^C\right]$$

In this case, Bayes' Law becomes:

$$P[B|A] = \frac{P[A|B]\,P[B]}{P[A|B]\,P[B] + P[A|B^C]\,P[B^C]} \tag{3}$$

The Walpole book writes Bayes' Law in its most drawn-out form, *i.e.*, for the case where the partition has k sets $B_1, B_2, \ldots, B_k$, instead of just the two sets B and B^C. In this case, the conditional probability of the rth set, B_r, is

$$P[B_r|A] = \frac{P[A|B_r]\,P[B_r]}{P[A|B_1]\,P[B_1] + P[A|B_2]\,P[B_2] + \cdots + P[A|B_k]\,P[B_k]} \tag{4}$$

There's no majic, in Bayes' Law. It is just a rewriting of the conditional probability definition, additionally using the law of total probability.

1.1 Examples

Example: Neural Impulse Actuation
A embedded sensor is used to monitor a neuron in a human brain. We monitor the sensor reading for 100 ms and see if there was a spike within the period. If the person thinks of flexing his knee, we will see a spike with probability 0.9. If the person is not thinking of flexing his knee (due to background noise), we will see a spike with probability 0.01. For the average person, the probability of thinking of flexing a knee is 0.001 within a given period.

1. What is the probability that we will measure a spike? Answer: Let S be the event that a spike is measured, and its complement as NS. Let the event that the person is thinking about flexing be event T, and its complement is event NT.

$$\begin{aligned} P[S] &= P[S|T]\,P[T] + P[S|NT]\,P[NT] \\ &= 0.9 * 0.001 + 0.01 * 0.999 = 0.0009 + 0.00999 = 0.01089 \end{aligned}$$

2. What is the probability that the person wants to flex his knee, given that a spike was measured? Answer:

$$\begin{aligned} P[T|S] &= \frac{P[T \cap S]}{P[S]} = \frac{P[S|T]\,P[T]}{P[S]} \\ &= \frac{0.9 * 0.001}{0.01089} = 0.0826 \end{aligned}$$

3. What is the probability that the person wants to flex his knee, given that no spike was measured? Answer:

$$\begin{aligned} P[T|NS] &= \frac{P[T \cap NS]}{P[NS]} = \frac{P[NS|T]\,P[T]}{P[S]} \\ &= \frac{0.1 * 0.001}{1 - 0.01089} \approx 1.01 \cdot 10^{-4} \end{aligned}$$

It wouldn't be a good idea to create a system that sends a signal to flex his knee, given that a spike was measured. Some other system design should be considered.

Def'n: *a priori*
prior to observation

For example, prior to observation, we know $P[T] = 0.001$, and $P[NT] = 1 - 0.001 = 0.999$.

Def'n: *a posteriori*
after observation

For example, after observation, we know $P[T|S] = 0.0826$, and $P[NT|NS] = 1 - 0.0001 = 0.9999$.

We did another example on the first day of class: refer to the *Rare Condition Tests* example.

Another one from the Mlodinow reading:

Example: Mlodinow, p117
What is the probability that an asymptomatic woman between 40 and 50 years of age who has a positive mammogram actually has cancer? You know that 7 percent of mammograms show cancer when there is none, and the actual incidence for this population is 0.8 percent, and the false-negative rate is 10 percent.

Note: False positive: Test is positive, but it is a "false alarm", because the person does not have the disease. False negative: The test did not discover the disease, but the person actually has it.

Solution: Let M be the event that a mammogram is positive. Let C be the event that the person has cancer. Here is the provided information:

1. $P\left[M|C^C\right] = 0.07$. The "false positive rate".
2. $P\left[C\right] = 0.008$. The prior information.
3. $P\left[M^C|C\right] = 0.1$.

Applying Bayes' Law:

$$P\left[C|M\right] = \frac{P\left[M|C\right]P\left[C\right]}{P\left[M|C\right]P\left[C\right] + P\left[M|C^C\right]P\left[C^C\right]}$$

From 3., $P\left[M|C\right] = 1 - P\left[M^C|C\right] = 1 - 0.1 = 0.9$. From 2., $P\left[C^C\right] = 0.992$. Thus

$$P\left[C|M\right] = \frac{(0.9)(0.008)}{(0.9)(0.008) + (0.07)(0.992)} = 0.094.$$

Note that it is ALWAYS possible to design a test to have a 0% false positive rate, or to have a 0% false negative rate.

- 0% false positive rate: Make the test always return negative.
- 0% false negative rate: Make the test always return positive.

Both are meaningless tests. But they point out, if you design a test, don't simply ask for one rate to describe the accuracy of the test. You need BOTH the false negative and false positive (or their complements) to evaluate the effectiveness of a test.

Example: Boot Sequence
A computer at boot time must pass (P) three tests for a certain OS; failure (F) of any test stops the boot sequence. A certain type of computer passes the first test 90% of the time, passes the second test 80% of the time, and passes the third test 70% of the time. Given that the boot sequence stopped, what is the probability that it failed the first test?

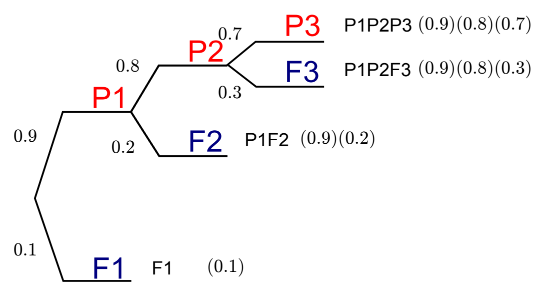

Figure 1: A computer at boot time must pass (P) three tests for a certain OS; failure (F) of any test stops the boot sequence.

Solution: See Figure 1. Let's call $P1, P2, P3$ the events that the computer passes the 1st, 2nd, and 3rd tests, respectively; $F1, F2, F3$ are that it failed those tests. Let BS be the event that the boot sequence stopped. Then the question is asking for $P[F1|BS]$. Using Bayes Law,

$$P[F1|BS] = \frac{P[F1 \cap BS]}{P[BS]} = \frac{P[F1]}{P[F1] + P[P1 \cap F2] + P[P1 \cap P2 \cap F3]}$$

The numerator has $P[F1 \cap BS] = P[F1]$ because $F1$ is a subset of BS. That is, BS is the union of three disjoint events: $F1$, $P1 \cap F2$, and $P1 \cap P2 \cap F3$. Plugging in from Figure 1,

$$P[F1|BS] = \frac{0.1}{0.1 + (0.9)(0.2) + (0.9)(0.8)(0.3)} = 0.202.$$

Note the answer is not the *a priori* probability of failing the first test, which is 10%.

1.2 Bayesian Philosophy

The Bayesian philosophy is that one can learn more about wether or not something is true from observation, even noisy observation. Thus it relates to the classic philosophical question, "How do we know what we know?". However, for hundreds of years, it was sidelined by mathematicians and scientists. Why? Consider this application of Bayes' Law:

- Let B be the event that "voltage is equal to current times resistance", and
- Let $A_1, A_2, \ldots$ be some imperfect binary measurements which are affected by the truth of this statement.

To understand the philosophical beef, From our existing understanding, assume we can easily come up with $P[A_1|B]$ and $P\left[A_1|B^C\right]$. We can use Bayes' Law to come up with the $P[B|A_1]$, the probability that the statement is true, given A_1; but ONLY if we have the probability $P[B]$. That is, we need to know beforehand the probability that our statement is true. In real life, this statement is not random – it is either true or not true. However, before you know whether it is true or not, Bayes would argue, you have to quantify how firmly you believe that B is true or not. Keep track of $P[B]$, and then $P[B|A_1]$ when the first experiment is done, in order to quantify how sure you are about B. And if more measurement events $A_2, A_3, \ldots$ are recorded, you can keep track of $P[B|A_1 \cap A_2]$ and then $P[B|A_1 \cap A_2 \cap A_3]$, and so on. The probability of B given all of the observations you've taken then gives you a quantitative way of tracking your belief in the truth of statement B. However, the only way to do this is, at first before any observations are recorded, to assume some $P[B]$. Bayes suggested setting it to 1/2 when no information is available. Mathematicians and scientists of his time thought that was unjustified and thus the whole idea of using Bayes' Law was nonsense. As a result, Bayesian statistics was sidelined, and almost unheard of until the 20th century.

However, there are many engineering applications in which Bayes' Law is perfectly matched. For example, Bayes' Law is fundamental math used by the British to decode German radio messages, as depicted in the movie on Alan Turing called emphThe Imitation Game (2014). There was a large set of possible codes, that is, settings for the German coding machine. It was infeasible to try each possible code to decode a message. However, by observing characteristics of the coded message, one could say that certain sets of codes had higher likelihood. The machines designed to break the code thus applied Bayes Law, using the set of all codes as the partition $B_1, \ldots, B_k$, and the characteristics observed from the messages as the observations $A_1, A_2, \ldots$. By listening to many messages, eventually, a few posterior probabilities $P[B_r|A_1, A_2, \ldots]$ would become very high, and the rest relatively low; these would be tested by decoding the message with them. When a code could successfully decode a message, the code was broken. Bayes' Law was perfect for this problem because, before observing any message, it was perfectly acceptable to assign equal prior probabilities to each possible code.

This historical and philosophical perspective is discussed both in the Mlodinow reading (posted on Canvas), and more extensively in the book, *The Theory That Would Not Die*, by S. B. McGrayne, Yale University Press, 2012.