28

Bayes' Rule

Forget the Formula

Bayes' rule is a probability benchmark. Once you understand it, it indicates that you have moved on to a new level. Suppose that we have partitioned events into mutually exclusive and exhaustive cases E_1, E_2, ..., E_n. That is, exactly one of E_1, E_2, through, E_n will occur. E_i might be the weather (such as temperature will be between 10 and 30 degrees). Suppose also, that for any E_i we know the probability of X (which might be the event that the Raiders win). That is, we have for each case E_i the conditional probability $P(X|E_i)$. For example we might have the mutually exclusive and exhaustive events E_1: temperature is below 10°; E_2: temperature is between 10° and 30°; E_3: temperature is above 30°. And we may have the conditional probabilities that if the temperature is below 10° then the probability that the Raiders win is 75%; if the temperature is between 10° and 30° then the probability that the Raiders win is 85%; if the temperature is above 30° then the probability that the Raiders win is 95%. Suppose further that the probability that the temperature will be less than 10° is 40%; the probability that the temperature will be between 10° and 30° is 30%; and the probability that the temperature be greater than 30° is also 30%. Bayes' rule simply turns the information around to answer the questions: given that the Raiders win, what is the probability the temperature was below 10°, or between 10° and 30°, or greater than 30°. That is, we were given $P(X|E_i)$ and $P(E_i)$; what is $P(E_i|X)$?

There is a simple formula expressing Bayes' rule. However, many students find the formula intimidating. The fact is that we can learn to easily use Bayes' rule without recourse to the formula, simply by application of probability trees. First, I will give Bayes' rule (Bayes' rule and Bayes' formula are the same thing). Those of you who intend to pursue mathematical and engineering type careers, should be able to deal with the formula; it is really not too horrible. As to the proof of the formula, it is merely an application of the definition of conditional probabilities and the law of total probability.

Let E_1 through E_n be mutually exclusive and exhaustive events with known probabilities. Suppose that $P(X|E_i)$ is known for every i. Then, for a given k:

$$P(E_k|X) = \frac{P(X|E_k)\cdot P(E_k)}{\sum P(X|E_i)\cdot P(E_i)}$$

The summation is over all i's (i = 1 to n).

Formula 1 **Bayes' Rule**

☐ **Exercise 1** (This is an optional exercise for those people who want to deal with the formula directly.) Prove Bayes' rule.

Urn Problems

Urn problems are an excellent illustration of Bayes' rule. They are hated by students almost everywhere, but that is largely because they do not know the solution technique that I am going to show you right here.

Suppose that I have two urns, cleverly named *Urn I* and *Urn II*. Suppose that Urn I has 2 red marbles and 2 blue marbles, and Urn II has 1 red and 3 blue marbles. We flip a coin to select an urn. Having selected an urn we select a marble without looking in the urn. It so happens that the marble we chose is red. Our question is: what is the probability that we chose Urn I?

This is a typical Bayesian question, because we have mutually exclusive and exhaustive events (choosing Urn I or Urn II); we have clear conditional probabilities, such as P(choosing red| having chosen Urn I) and we want to reverse the conditional probability: what is P(having chosen Urn I|having chosen red)? The principal tool to solving this problem is to draw a probability graph such as **Figure 1**.

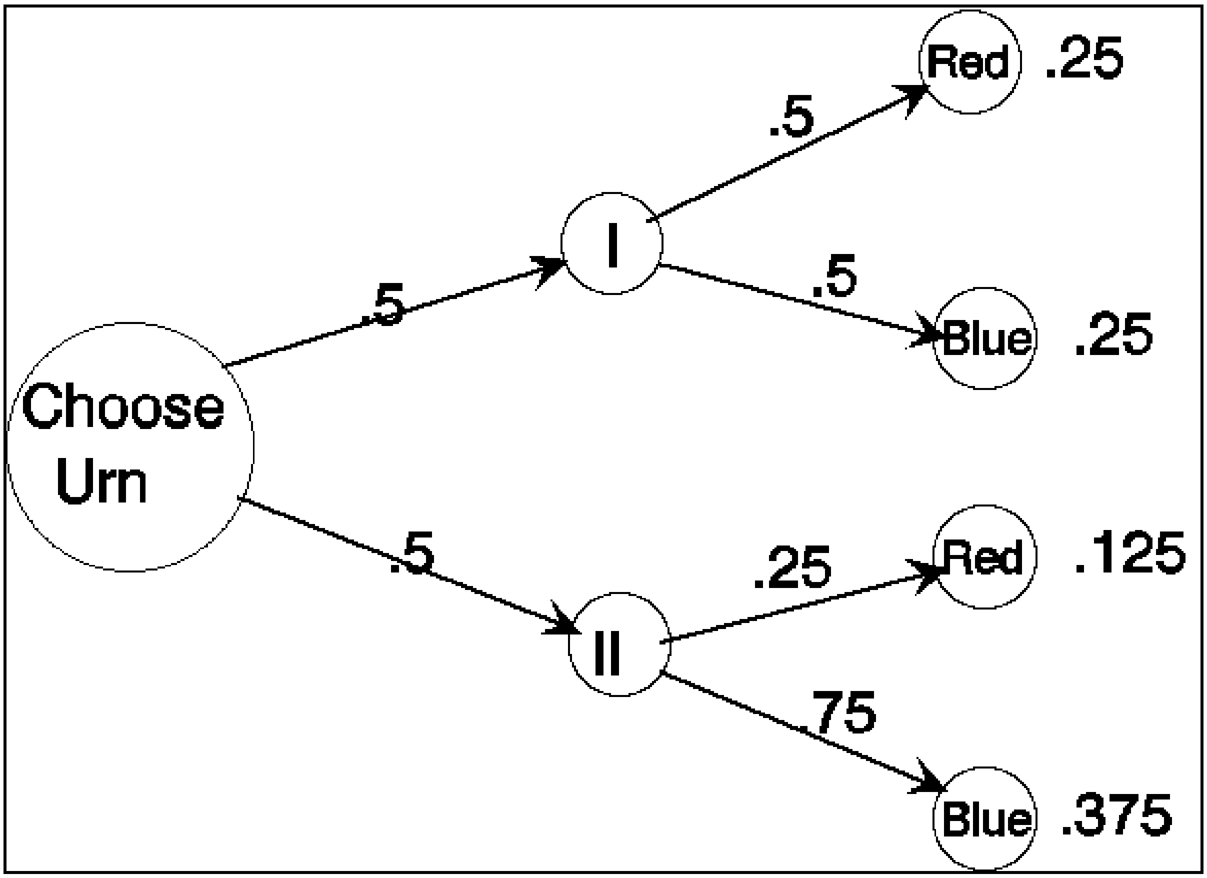

Figure 1 Graph for Urn Problem

We want to find the conditional probability P(having chosen Urn I | having chosen red). From the definition of conditional probability, we want:

$$\frac{\text{P}(\textit{Choosing red } \textbf{and } \textit{Choosing Urn I})}{\text{P}(\textit{Choosing red})}$$

From the graph we can see that P(Choosing red and Choosing Urn I) = .25, and that P(Choosing red) = .25 + .125 = .375. Therefore, P(having chosen Urn I | having chosen red) = .25/.375 = 2/3. That is, the probability that we chose Urn I is two-thirds.

☐ **Exercise 2** Suppose that in the above urn problem we choose the urn by throwing a die. We choose Urn I if the die is a 1, otherwise we choose Urn II. If someone else were to know that we chose a Red marble, but they did not know which urn we chose, what is the best estimate that could be made of whether we chose Urn I?

☐ **Exercise 3** Urn I has 1 red marble and 9 blue marbles. Urn II has 4 red marbles and 1 blue marble. The urn is chosen by a flip of a coin. If a red marble is selected at random from whichever urn was chosen, what is the probability that it was Urn II that was chosen?

☐ **Exercise 4** Urn I has 1 red marble and 5 blue marbles. Urn II has 4 blue and no red marbles. The urn is selected by the throw of a die. If the die is a 1 or a 2, Urn I is chosen. Given that a blue marble was chosen, what is the probability that it is Urn II that was chosen?

The Raiders Example Solved

This Section is Philosophical and Gives all the Good Jargon

I will repeat the Raiders' problem: The probability that the temperature falls below 10° is 40% and in that case the probability that the Raiders win is .75. The probability that the temperature is between 10° and 30° is 30% in which case the probability that the Raiders win is 85%. The probability that the temperature is greater than 30° is 30% in which case the probability that the Raiders win is 95%. Given all of these facts, suppose that we happen to hear later that the Raiders won, but we know nothing else. What, we might ask, is the probability that the temperature was below 10°? The *a priori* probability that it is below 10° is 40%. We want the *a posteriori* probability that the temperature was 10°. *A priori* means roughly *prior*, and *a posteriori* means roughly *later* (go to a dictionary for the real meaning and don't tell anyone where you got these pseudo-definitions). The whole point of Bayes' theorem is that we are going to **reevaluate a probability given information**. The problem is summarized in **Figure 2**.

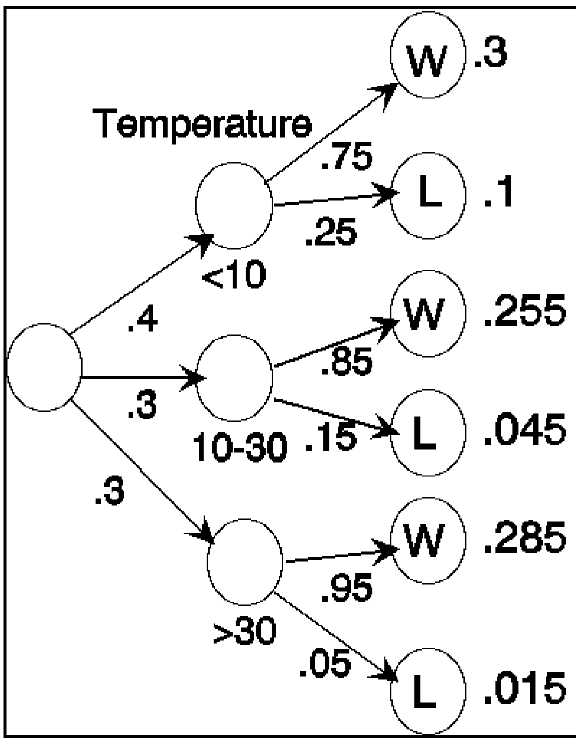

Figure 2 The Probability Tree for Raiders

The probabilities to the right of the leaves are the products of the probabilities of the branches leading to the leaves. The left most branch is the (a priori) unconditional probability of the temperature. The next probability is the probability of winning or losing given the temperature. The product of the two is the joint probability of winning or losing and the temperature being in the given range. (This follows from the law of multiplication.) We have that the joint probability of the temperature being below 10° **and** the raiders winning is .3 as opposed to our input that the probability that the raiders win **given** the temperature is below 10° is .75. To derive the unconditional probability that the Raiders win, we use the law of total probability and add all the disjoint cases.

P(Win) = P(Win **and** temp < 10°) + P(Win **and** 10° < temp < 30°)

+ P(Win **and** temp > 30°) = .3 + .255 + .285 = .84.

We want to find the conditional probability: P(Temperature < 10°| Raiders win). By the definition of conditional probability we have:

$$P(temp < 10° \mid Raiders\ win) = \frac{P((temp < 10°)\ \textbf{and}\ Raiders\ win)}{P(Raiders\ win)}$$

This is .3/.84 = .357. The a priori estimate that the temperature be less than 10° was .4. The a posteriori probability that the temperature be less than 10° is a little lower. The Raiders were likely to win anyway, so the fact that they did win was not very informative.

The Probability of Having Dreaded College Administrator Disease

College administrator disease is one of the most dreaded diseases ever known to man, even though its incidence away from college campuses is one in a million. The victim is marked by lethargic movement, slow thought, the inability to blink both eyes simultaneously, and by the sort of too-high salary that ruins mens' souls (and womens' souls too). The only known test for this disease is an intravenous shot of 100 proof bourbon. If the patient immediately lights up and starts writing a memo then the test is considered positive. The test is quite reliable. It is known that if the patient does not have the disease, that the probability of a positive test is only 4%. If the patient does have the disease, the probability of a negative test is 1%.

Suppose that we do a routine test of a person who is not associated with a college. Let us suppose that the test comes back positive. Our question, given the positive test, is what is the probability that the person has CA disease?

First of all, notice that this question, like all Bayesian questions, is one of re-evaluating a probability under the presence of new information (the test result). We have an *a priori* probability of 1/1,000,000 that the patient has CA disease. We want to solve for the *a posteriori* probability of the disease.[1] The normal estimate most people make is that the probability of having the disease must be about 96% since the probability of a *false negative* is only 4%. But notice, that 4% is a conditional probability. We want the different conditional probability that the person has the disease given a positive test result. As before, we can solve this problem by recourse to a probability tree, **Figure 3**.

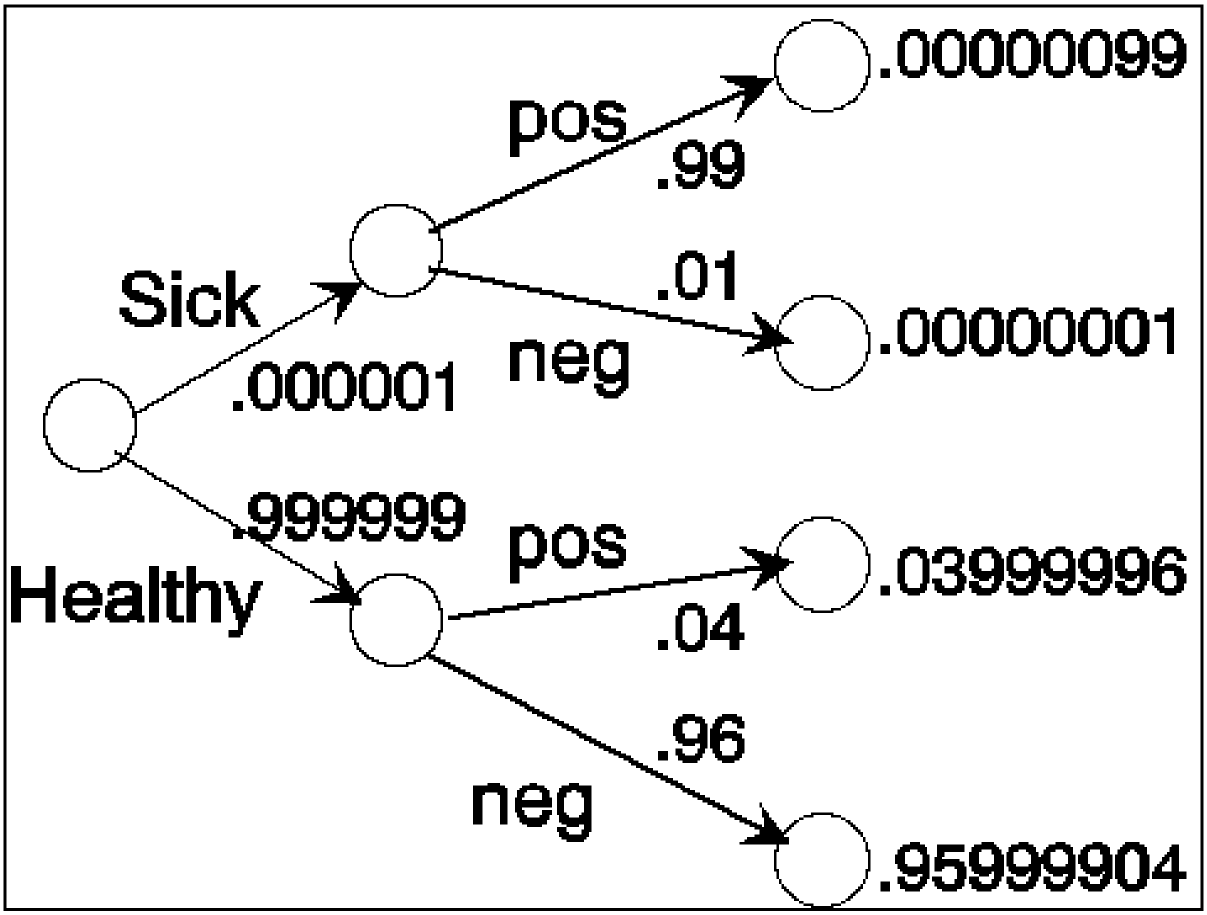

Figure 3 The Probability Tree for College Administrator Disease

We want to solve for the probability of a patient having the disease given the that the patient tested positive. By the definition of conditional probability, we are asking for:

$$\frac{\text{P}(\textit{having the disease}\ \textbf{and}\ \textit{testing positive})}{\text{P}(\textit{testing positive})}$$

. From the graph, we can see that the

[1]This jargon goes over very big in philosophy classes and can also be used at parties.

P(having the disease **and** testing positive) = .00000099. Similarly, we can see that P(testing positive) = .00000099 + .03999996 = .04000095. Hence, P(Having CA disease| Having tested positive) = .0000247. That is, the a priori probability of having CA disease is one-in-a-million. Having a positive test, the a posteriori probability becomes roughly one in forty-thousand. A simple way of seeing that the answer is correct is to do a thought experiment. Consider, that we test one-million (non-college) people for CA disease. We would expect one person to actually have the disease (and to test positive). We would expect roughly 4% to test positive (since the probability of a false positive is 4%). That is, we should have about 40,000 positive tests. Hence, only about one person in 40,000 who tests positive should have the disease. Another way of understanding the above result is to consider the case of an eradicated disease. Even if the test that scores positive for the disease is highly reliable, since the disease is eradicated, one's probability of having the disease is zero.

1. We want $P(E_k|X)$. By definition this is $P(E_k{\cdot}X)/P(X)$.
   The numerator is $P(X|E_k){\cdot}P(E_k)$ and can be seen by rearranging the formula for conditional probability. The denominator uses the same technique to rewrite $P(X) = \sum P(X{\cdot}E_i)$; this is just the law of total probability.

2. P(chose Urn I | chose Red) = P(chose Urn I and chose Red)/P(chose Red) = (1/12)/((1/12) + (5/24) = 2/7

3. P(Urn II | Red) = P(Urn II and Red)/P(Red) = .4/(.4 + .05) = 8/9

4. P(Urn II | blue) = P(Urn II and blue)/P(blue) = (2/3)/((2/3) + (5/18)) = 12/17