Caroline Nelson and Vincent Chou
July 28, 2016 (Mod 2)

Probability Foundations

Law of Total Probability ("mixture rule"):

The Law of Total Probability says the probability of an event is the sum of the probabilities of all possible ways for that event to occur. It relates marginal probabilities with conditional probabilities.

Suppose $B_1, B_2, \ldots, B_N$ form a partition of the sample space, that is they are mutually exclusive and collectively exhaustive.

Mutually exclusivity: $P(B_i, B_j) = 0,\ i \neq j$ (i.e. Both events cannot happen simultaneously)

Collectively exhaustive: $\sum_{i=1}^{N} P(B_i) = 1$ (the sample space includes all possible outcomes)

For a given event A, the Law of Total Probability states:

$$P(A) = \sum_{i=1}^{N} P(A, B_i) = \sum_{i=1}^{N} P(A|B_i)P(B_i)$$

Application: A surveyor is interested in knowing the percentage of people on a college campus that have smoked marijuana. He or she could confront students directly, but due to the possibly incriminating nature of the question, the students' answers may not be truthful.

One solution is to allow them to flip a coin, and the question they answer corresponds to the outcome of the flip. If the outcome is heads, the student will answer question 1 (Does your social security number end in an odd number?), and if it is tails, they answer question 2 (Have you ever smoked marijuana?). The question itself is only known to the coin-flipper; all the researcher knows is the number of people who said "Yes" and the number who said "No", blind to the question that was answered.

Y *= the event that the student answered "Yes"*

Q_1 *= the event that the student answered Question 1*

Q_2 *= the event that the student answered Question 2*

We are interested in the probability that someone has ever smoked marijuana; for that to be the case, they would have had to answer Question 2. This probability is $P(Y|Q_2)$. Using the Law of Total Probability, we know that:

$$P(Y) = P(Y|Q_1) * P(Q_1) + P(Y|Q_2) * P(Q_2)$$

The probability of answering Question 1 ($P(Q_1)$) or answering Question 2 ($P(Q_2)$) are both 0.5 (the coin flip), and the probability of answering "Yes" given Question 1 ($P(Y|Q_1)$) is also 0.5 (the probability that a social security number ends in an odd number).

Let's say 35% of people answered "Yes". How do we figure out how many said "Yes" to Question 2? Filling in what we know, we get:

$$0.35 = 0.5 * 0.5 + 0.5 * P(Y|Q_2)$$

Solving for $P(Y|Q_2)$ we find that $P(Y|Q_2) = 0.20.$
So, 20% of the people asked Q_2 responded "Yes".

Bayes' Rule:

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

$P(A|B)$ is the posterior probability, $P(A)$ is the prior probability, $P(A|B)$ is the likelihood, and $P(B)$ is the marginal probability of B.

<u>Application:</u> Suppose there are 1024 regular quarters and 1 double-headed quarter in a jar (1025 quarters total). Someone has reached in and pulled out a coin (without looking at it). They flip the coin 10 times, and it is heads every time. What is the probability that they are flipping the double-headed quarter?

T : Holding the double-headed quarter

D : Data (10 flips, all heads)

We are looking for $P(T|D)$. Let's apply Bayes' Rule:

$$P(T|D) = \frac{P(T)*P(D|T)}{P(D)}$$

We know $P(T) = \frac{1}{1025}$ (one double-headed quarter in a jar of 1025 quarters) , and we know that $P(D|T) = 1$ (since there is no chance that this data did not happen if the double-headed quarter is being flipped).

Using the Law of Total Probability, we can find $P(D)$, where T' is the compliment of T:

$$P(D) = P(D|T) * P(T) + P(D|T') * P(T')$$
$$= 1 * \frac{1}{1025} + \frac{1}{2^{10}} * (\frac{1024}{1025})$$

$$= \frac{2}{1025}$$

$$\therefore\ P(T|D) = \frac{\frac{1}{1025}(1)}{(\frac{2}{1025})} = \frac{1}{2}$$

Given D, there is a 50% chance that the coin being flipped is the double-headed quarter.

Application: The trial of a man with a positive DNA test.

G: The event that the man is guilty
D: Data (positive DNA test)

We want the probability that given a positive DNA test, the man is guilty ($P(G|D)$).
Once again, Bayes' Rule says:

$$P(G|D) = \frac{P(G)*P(D|G)}{P(D)}$$

$P(G) = \frac{1}{10{,}000{,}000}$, the odds of any random New Yorker being the actual guilty person
$P(D|G) = 1$, assuming that if he is guilty, the DNA test will be positive
$P(D|G') = \frac{1}{1{,}000{,}000}$, the odds of a false positive on the DNA test

The probability that a man is guilty given a positive DNA test is about $\frac{1}{11}$. There is less than a 10% chance that the man is guilty even after a positive DNA test. The positive DNA test does, however, increase the odds that the man is guilty from $\frac{1}{10{,}000{,}000}$ to $\frac{1}{11}$. So while the positive test did not definitively show that the man is guilty, it drastically increased the probability that he is.