Caroline Nelson and Vincent Chou
July 28, 2016 (Mod 1)

Probability Foundations

Law of Total Probability ("mixture rule"):

The Law of Total Probability is like Kolmogorov's rule 3 in disguise; it says the probability of an event is the sum of all possible ways for that event to occur.

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

Collectively exhaustive: $\sum_{i=1}^{N} P(B_i) = 1$ (this sample space covers all possible results)

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

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

Example: How would you as a researcher get a truthful answer to an embarrassing question?

You're interested in knowing the percentage of people on a college campus that smoke marijuana. You could confront them directly, but due to the possibly embarrassing or incriminating nature of the question, their answers may not be entirely truthful.

So one solution is to give them a coin, and the question they answer depends on whether the flip is heads or tails (the researcher being blind to the coin flip). If it comes up heads, they answer question 1 (Does your social security number end in an odd number?), and if it lands on tails, they answer question 2 (Have you ever smoked marijuana?). The question itself is only known by the coin-flipper; all the researcher knows is the number of people who said "Yes" and the number who said "No".

Y = *"Yes"*

Q_1 = *Question 1*

Q_2 = *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 odds 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 just Q_2? Filling in what we know to our equation 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:

This is Bayes' Rule:

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

Where $P(A|B)$ 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.

You have a prior belief about the likelihood of A ($P(A)$). Then you introduce new information and change your belief ($P(A|B)$).

Example. You have 1024 regular quarters and 1 double-headed quarter in a jar (1025 quarters total). You've reached in and pulled out a coin (without looking at it). You flip the coin 10 times, and it is heads every time. What is the probability that you 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 we have the double-headed quarter).

Using the Law of Total Probability, we can find $P(D)$:

Where T' is the compliment of T:

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

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

So given your D , there is a 50% chance that the coin you are flipping is the double-headed quarter.

Ex. 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:

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

$P(G) = \frac{1}{10000000}$ The odds of any random New Yorker being the actual guilty person

$P(D|G) = 1$ Let's assumed that if he is guilty, the DNA test will be positive

$P(D|G') = \frac{1}{1000000}$ The odds of a false positive on the DNA test.

In the end it works out to be about $\frac{1}{11}$ **.** So 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 $\frac{1}{10000000}$ to $\frac{1}{11}$

So while the test did not definitively show that the man is guilty, it certainly made it much more likely that he is.