NOTE ON BAYES'S FORMULA

I want to apologize; the way I wrote the law of total probability in class is useless, but I realized I should have written it a different way. What I wrote is not incorrect; it is a correct formula, but not very useful because it computes $P(A)$ in terms of $P(A)$.

The important thing that you should really understand is that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{so,} \qquad P(A \cap B) = P(A|B)P(B)$$

Similarly,

$$P(B|A) = \frac{P(\cap B)}{P(A)} \qquad \text{so,} \qquad P(A \cap B) = P(B|A)P(A)$$

So

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \tag{0.1}$$

This makes sense: the probability that both A and B happen is the probability that A happens times the probability that B also happens if A happened. But equivalently, the probability that A and B both happen is the probability that B happens times the probability that A happens given that B happened.

Now, the law of total probability was

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i), \tag{0.2}$$

where $B_1, \ldots, B_n$ were disjoint subsets whose union was the entire sample space. By equation (0.1) above, for any B_i,

$$P(A \cap B_i) = P(A|B_i)P(B_i) = P(B_i|A)P(A) \tag{0.3}$$

In class (in the first lecture), I equated expression (0.2) from the law of total probability with

$$P(A) = \sum_{i=1}^{n} P(B_i|A)P(A) \tag{0.4}$$

This is not wrong; we can totally do that, but it's not that helpful (as I was saying in class). It's not very helpful to try to compute the probability of A in terms of the probability of A.

But, because of equation (0.3), we can instead equate what we got in equation (0.2) with

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \qquad \text{(law of total probability)} \tag{0.5}$$

This is the form of the total law of probability which we want to use! This is more useful: it says we can compute the probability of an event as the sum of the products of the probabilities of the mutually exclusive events we have split up our sample space into and the conditional probabilities of A depending on these.

Now, let's write Bayes's formula in terms of this (so if we want to know which event B_k happened given that event A happened, we'd compute):

$$P(B_k|A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \qquad \text{(Bayes's formula)} \tag{0.6}$$

Note that here the numerator $P(A \cap B_k)$ is also equal to $P(B_k|A)P(A)$, but if we wrote it like that, the P(A) would cancel with the P(A) in the denominator and we'd get a tautology: $P(B_k|A) = P(B_k|A)$

We will see this in action in a problem that we understand, namely the Monty Hall problem.