Bayes' Theorem ©
Terri Bittner, Ph.D.

Let's start by defining two events *A* and *B*. In this way, we can now refer to probabilities involving these events.

Bayes' Theorem states the following:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (1a)$$

Similarly:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad (1b)$$

That is, the conditional probability that *A* will happen given that we know that *B* has happened (or will happen) is the probability that *both* events happen divided by the probability that event *B* occurs. You can convince yourself of this by thinking about events like the outcomes when you roll a pair of fair dice.

For example, let's let *A* be the event that you roll a 3 with the first die. Let *B* be the event that the sum of the two dice is 7.

What is $P(B|A)$? (What is the probability that you will get a sum of 7 if you know that the first die is a 3? Note that these events are *not* independent even though the outcomes of two separate dice *are* independent.

Now,

$$P(\text{sum of 7}|\text{first die is 3}) = \frac{P(\text{first die is 3 } \textit{and} \text{ sum of 7})}{P(\text{first die is 3})}$$

$$= \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{36} \cdot \frac{6}{1} = \frac{1}{6}$$

Let's think about this for a second. The only way that you can have *both* a sum of 7 and a 3 on the first die is if you roll (3, 4). That is, 3 on the first die and then 4 on the second. The probability of this is $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$, since the two rolls are independent.

The overall answer of $\frac{1}{6}$ makes sense by "brute force" also. What is the probability that you roll a sum of 7 given that you know you rolled a 3 on the first die? Well, it is just the probability that you roll a 4 on the second die. That is the only way you can roll a sum of

7 given the initial condition of a 3 on the first die. This gives us the same answer as with Bayes' Theorem.

Now let's extend it.

Let's use our dice example one more time, but let's define our events differently. Let *A* be the event that you roll a 3 on the first die. Let *B* be the event that you roll a 4 on the second die.

How can we define *P*(rolling a 3 on the first die) given that we are rolling two dice at the same time?

Though it seems overly complex, we can think of it as:

$$P(A) = P(A \text{ and } B) + P(A \text{ and } \textit{not } B)$$

This means that *P*(3 on first die) = *P*(3, 4) + *P*(3, *not* 4)

This may seem stupid at first, but there is a reason for it. Let me write the same thing using our standard mathematical notation:

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) \qquad (2)$$

This is just a fancy way of writing the same thing I wrote above.

Now, let's remember the very first form of Bayes' Theorem that I wrote at the top of the page. That is, recall

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (1a)$$

and

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad (1b)$$

Let's solve equation (1a) for $P(A \cap B)$. We obtain

$$P(A \cap B) = P(A|B) \cdot P(B) \qquad (3)$$

by using simple algebraic substitution.

Now I'm going to substitute equation (3) into equation (2). So, where I see $P(A \cap B)$ or $P(A \cap \overline{B})$, I am going to use the new expression from equation (3).

We obtain:

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$
$$= P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B}) \qquad (4)$$

We have now derived one of the equations from the text (p. 102), and you can use it for the homework problem that we just discussed.

Finally, we will use another substitution to transform equation (1b) into the final complex form of Bayes' Theorem on p. 102 in the text.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})} \qquad (5)$$

Note that since we don't know $P(A|B)$, we had to solve equation (1b) for $P(A \cap B)$ instead of using equation (1a) for this purpose.

While equation (5) looks horribly complicated, it is obtained simply by replacing the numerator of equation (1) by equation (3), and then replacing the denominator of equation (1) by the expression from equation (4). Note that I've replaced *A* with *B* in the denominator's expression from equation (4), but it is exactly the same thing since $P(A \cap B) = P(B \cap A)$.

You use one form or the other of these equations depending on the information you have from the problem. Brute force works for rather simple problems, but it fails to work well when the problems become complex. Define your events very precisely, look at your givens, and it is usually clear which form of Bayes' Theorem you need.