for any A and B. So if the chance of rain on any given day is 0.5, the chance of rain on two consecutive days is not 0.25, but probably a bit higher.

1.3 The cookie problem

We'll get to Bayes's theorem soon, but I want to motivate it with an example called the cookie problem.[1] Suppose there are two bowls of cookies. Bowl 1 contains 30 vanilla cookies and 10 chocolate cookies. Bowl 2 contains 20 of each.

Now suppose you choose one of the bowls at random and, without looking, select a cookie at random. The cookie is vanilla. What is the probability that it came from Bowl 1?

This is a conditional probability; we want $\mathrm{p}(\text{Bowl 1}|\text{vanilla})$, but it is not obvious how to compute it. If I asked a different question—the probability of a vanilla cookie given Bowl 1—it would be easy:

$$\mathrm{p}(\text{vanilla}|\text{Bowl 1}) = 3/4$$

Sadly, $\mathrm{p}(A|B)$ is *not* the same as $\mathrm{p}(B|A)$, but there is a way to get from one to the other: Bayes's theorem.

1.4 Bayes's theorem

At this point we have everything we need to derive Bayes's theorem. We'll start with the observation that conjunction is commutative; that is

$$\mathrm{p}(A \text{ and } B) = \mathrm{p}(B \text{ and } A)$$

for any events A and B.

Next, we write the probability of a conjunction:

$$\mathrm{p}(A \text{ and } B) = \mathrm{p}(A)\,\mathrm{p}(B|A)$$

Since we have not said anything about what A and B mean, they are interchangeable. Interchanging them yields

$$\mathrm{p}(B \text{ and } A) = \mathrm{p}(B)\,\mathrm{p}(A|B)$$

[1]Based on an example from `http://en.wikipedia.org/wiki/Bayes'_theorem` that is no longer there.

That's all we need. Pulling those pieces together, we get

$$\mathrm{p}(B)\,\mathrm{p}(A|B) = \mathrm{p}(A)\,\mathrm{p}(B|A)$$

Which means there are two ways to compute the conjunction. If you have $\mathrm{p}(A)$, you multiply by the conditional probability $\mathrm{p}(B|A)$. Or you can do it the other way around; if you know $\mathrm{p}(B)$, you multiply by $\mathrm{p}(A|B)$. Either way you should get the same thing.

Finally we can divide through by $\mathrm{p}(B)$:

$$\mathrm{p}(A|B) = \frac{\mathrm{p}(A)\,\mathrm{p}(B|A)}{\mathrm{p}(B)}$$

And that's Bayes's theorem! It might not look like much, but it turns out to be surprisingly powerful.

For example, we can use it to solve the cookie problem. I'll write B_1 for the hypothesis that the cookie came from Bowl 1 and V for the vanilla cookie. Plugging in Bayes's theorem we get

$$\mathrm{p}(B_1|V) = \frac{\mathrm{p}(B_1)\,\mathrm{p}(V|B_1)}{\mathrm{p}(V)}$$

The term on the left is what we want: the probability of Bowl 1, given that we chose a vanilla cookie. The terms on the right are:

- $\mathrm{p}(B_1)$: This is the probability that we chose Bowl 1, unconditioned by what kind of cookie we got. Since the problem says we chose a bowl at random, we can assume $\mathrm{p}(B_1) = 1/2$.
- $\mathrm{p}(V|B_1)$: This is the probability of getting a vanilla cookie from Bowl 1, which is $3/4$.
- $\mathrm{p}(V)$: This is the probability of drawing a vanilla cookie from either bowl. Since we had an equal chance of choosing either bowl and the bowls contain the same number of cookies, we had the same chance of choosing any cookie. Between the two bowls there are 50 vanilla and 30 chocolate cookies, so $\mathrm{p}(V) = 5/8$.

Putting it together, we have

$$\mathrm{p}(B_1|V) = \frac{(1/2)\,(3/4)}{5/8}$$

which reduces to 3/5. So the vanilla cookie is evidence in favor of the hypothesis that we chose Bowl 1, because vanilla cookies are more likely to come from Bowl 1.

This example demonstrates one use of Bayes's theorem: it provides a strategy to get from $\mathrm{p}(B|A)$ to $\mathrm{p}(A|B)$. This strategy is useful in cases, like the cookie problem, where it is easier to compute the terms on the right side of Bayes's theorem than the term on the left.

1.5 The diachronic interpretation

There is another way to think of Bayes's theorem: it gives us a way to update the probability of a hypothesis, H, in light of some body of data, D.

This way of thinking about Bayes's theorem is called the **diachronic interpretation**. "Diachronic" means that something is happening over time; in this case the probability of the hypotheses changes, over time, as we see new data.

Rewriting Bayes's theorem with H and D yields:

$$\mathrm{p}(H|D) = \frac{\mathrm{p}(H)\,\mathrm{p}(D|H)}{\mathrm{p}(D)}$$

In this interpretation, each term has a name:

- $\mathrm{p}(H)$ is the probability of the hypothesis before we see the data, called the prior probability, or just **prior**.
- $\mathrm{p}(H|D)$ is what we want to compute, the probability of the hypothesis after we see the data, called the **posterior**.
- $\mathrm{p}(D|H)$ is the probability of the data under the hypothesis, called the **likelihood**.
- $\mathrm{p}(D)$ is the probability of the data under any hypothesis, called the **normalizing constant**.

Sometimes we can compute the prior based on background information. For example, the cookie problem specifies that we choose a bowl at random with equal probability.

In other cases the prior is subjective; that is, reasonable people might disagree, either because they use different background information or because they interpret the same information differently.

The likelihood is usually the easiest part to compute. In the cookie problem, if we know which bowl the cookie came from, we find the probability of a vanilla cookie by counting.

The normalizing constant can be tricky. It is supposed to be the probability of seeing the data under any hypothesis at all, but in the most general case it is hard to nail down what that means.

Most often we simplify things by specifying a set of hypotheses that are

Mutually exclusive: At most one hypothesis in the set can be true, and

Collectively exhaustive: There are no other possibilities; at least one of the hypotheses has to be true.

I use the word **suite** for a set of hypotheses that has these properties.

In the cookie problem, there are only two hypotheses—the cookie came from Bowl 1 or Bowl 2—and they are mutually exclusive and collectively exhaustive.

In that case we can compute $\mathrm{p}(D)$ using the law of total probability, which says that if there are two exclusive ways that something might happen, you can add up the probabilities like this:

$$\mathrm{p}(D) = \mathrm{p}(B_1)\,\mathrm{p}(D|B_1) + \mathrm{p}(B_2)\,\mathrm{p}(D|B_2)$$

Plugging in the values from the cookie problem, we have

$$\mathrm{p}(D) = (1/2)\,(3/4) + (1/2)\,(1/2) = 5/8$$

which is what we computed earlier by mentally combining the two bowls.

1.6 The M&M problem

M&M's are small candy-coated chocolates that come in a variety of colors. Mars, Inc., which makes M&M's, changes the mixture of colors from time to time.