

Conditional probability is an absolutely basic idea that we use all the time. It's the probability that some event occurs, given certain information about it. For example, an insurance company wants to know, what's the probability that you'll live for the next 10 years given your medical history? Or a typical investor wants to know, what's the probability that this stock is going to rise given its stock price gyrations for the past month?

There are people who actually think you could do that, the chartists. That, not knowing anything about the nature of the company, or the business that the stock is part of, that just by watching the price gyration you can make a better guess on what the stock will do tomorrow than you could otherwise.

Another good example is for a system engineer. What's the probability that the system is going to overload, given the recent history of the rate at which requests have been coming in? And finally, as a joke that I like to think about, is, what's the probability that you're a cat owner, given that you're sitting in the cat section of the Angell Memorial Veterinary Hospital?

OK. So let's look concretely at a very simple example of conditional probability that's meant to be illustrative, where we look at a die, and rolling a fair die.

OK. Now if I'm thinking about an ordinary fair die, I've got 6 outcomes that are equally likely. The outcomes are one, two, three, four, five, six. And if I ask, what's the probability that in one roll, I roll a one? Well it's going to be the number of outcomes involving my rolling a one, divided by the total number of outcomes. It's one-sixth. The probability of any given face of a six-sided fair die is one-sixth.

But suppose I give you some additional information. Knowledge about the roll can change the judgment of probabilities. Suppose that I tell you that I rolled an odd number. And now I want to know what's the probability that I rolled a one? And the answer will now be that, given that it's an odd number, the only possibilities are one, three, and five. And so the probability has changed to one-third. That should be a straightforward enough idea.

One way to understand conditional probability is as a kind of experiment, where first you try to roll an odd number and then you do you decide what final roll you're going to make. Let's look at that tree, if we were describing it that way.

So the first branch of the tree that we'll use to build a probability space is to say, OK, among the six possible

outcomes, what are the chances that we rolled an odd number? Well, it's 50/50, because there are three of each. So there's one-half chance that yes, you rolled an odd number. In which case those are the possible outcomes. Or a one-half chance that no, you didn't roll an odd number, and the possible outcomes then are two, four, and six.

Now, once you're here with one, three, and five, let's ask whether you rolled a one. The probability that you did roll a one, we've already agreed is one-third. It's equally likely to be any one of those three outcomes. Which means that it's two-thirds that you wind up rolling either a three or a five. And likewise, here, the probability if you didn't roll an odd number, that is, you rolled an even number, the probability that next you'll roll a one is zero. Or that you won't roll a one is probability one.

So this is a kind of standard way that we have of trying to build up a set of probabilities for outcomes. And if we look at this tree, well, first of all, we can use it to assign some probabilities. Because the probability of your rolling a one is one-sixth, as it should be. It's one-half times one-third, which is the usual way we would calculate the probability of this outcome.

By the way, we could calculate the probability of the outcome being three or five. It would be one-half times two-thirds or one-third. And finally, the probability of rolling an even number would just be one-half. One-half times one.

Now, what's going on here? Well, if you look at this number, one-third, it is what we said was the probability of a one, given that you rolled an odd number. So that's where this label came from. Likewise, this number, two-thirds, is the probability that you didn't roll a one, given that you rolled an odd number. And finally, this number is the probability that you didn't roll a one, given that you rolled an even number. And it's certain.

Let's do another example to get this idea across. Let's go back to Monty Hall, which we've seen before. Remember how we have the probability labels on these branches, which we figured out.

So if we look at this number, one-third, what is it? Well this is where the prize is at Location One and the contestant has picked Door One. And that one-third, we figured out that once the prize is at Door One, in fact, wherever the prize is, the probability that the contestant will pick One is one-third. This number, one-third, is the probability the contestant will pick One given that the prize is at Door One. Yeah?

Here's another third. This is, similarly, the probability that the contestant will pick Door Two given that the prize is at Door Three. It's symmetric to this one. But here's something a little bit different. Here's one-half. This is the probability that Door Three will be opened by Carol given that the prize is at One-- that's that branch-- and the contestant picked One.

And when the prize is at One and the contestant picks One, Carol, we said in our model, is equally likely to open the two possible doors that have goats that she's able to open.

And so that's one-half, is this conditional probability. The probability that she'll open Door Three given that we're in this location in the tree. Given that the prize is at One and pick is at One.

So the point is simply that we were reasoning about conditional probability in the very way we began defining the tree model that we were using to define probability spaces in the first place. We were implicitly using conditional probabilities to label the probabilities that left each vertex of the tree.

And in fact, formally speaking, what we were using was the product rule. Which is that, the probability that an A event occurs and a B event occurs is simply the probability that the A event-- that's the first branch of the tree-- times the probability of B given A. That's the fundamental rule of conditional probabilities. That's the product rule, and it's something to be memorized.

In fact, this product rule is not a corollary. It's really the definition of conditional probability. So all of the previous discussion was motivation of the following definition.

If A and B are events in a probability space, the probability of B, given A, is defined to be the probability that A and B occur-- that is, A intersection B-- relative to the probability of A.

So that's the formal definition. So this formal definition justifies the product rule by definition. Because you just multiply both sides by the probability of A. And you get probability of A times the probability of B, given A is the probability of the intersection. Notice that implicit in this definition is, the probability of A better not be zero. So you can't condition on an event that has zero probability. Probability of B given A is only defined if probability of A is positive.

OK. If you have a tree that's of depth three, then you need a product rule for three consecutive choices. And it

generalizes in a straightforward way.

Namely, the probability of A and B and C-- the first branch is A and the second branch is B and the third branch is C-- is the probability that you do A on the first branch, times the probability that you do B on the second branch, given that you did A on the first branch, and times the probability that you do C on the third branch, given that you did A on the first and B on the second.

And this product rule for three could, in fact, be proved simply by substitution, using the product rule for two twice. And of course it generalizes to any finite number of sets.

Another useful way to think about probability, that may be a more intuitive than the idea of choosing to whether or not to roll odd and then choosing whether or not to roll a one, usually what you think of is, you rolled the dice and then you're giving me some information about what that roll was. I don't think about the odds of rolling odd or not. I just tell you it's odd, and now tell me what is the probability that, among those odd outcomes, it was one.

So a way to formalize that is, you could think of conditioning on an event A as defining a new probability function on the sample space. Once you're given that A occurred, I can now think that all the probabilities of the sample of the outcomes have changed.

So I'll define a new probability measure relative to A, where all the outcomes that are not in A are going to be assigned probability zero. Because they can't happen given that A occurred. And all of the probabilities of outcomes of points in A, they get their probability raised in proportion to A. Because now A is going to be whole probability space.

Let's be a little bit more formal about that, to be precise. We're going to define a new probability function. Probability sub A, on the same sample space, where the probability of an outcome is zero if the outcome is not in A, and its old probability relativized to the probability of A, if ω is in A.

So that's the definition of the probability with respect to A of ω . It's a new probability measure on the same sample space. So to verify that this new thing is a probability space, you have to verify that the sum of the outcome probabilities is one. And that's a little exercise that I would encourage you to stop now and work out a piece of paper. Because it's trivial, but it's worth checking that you follow the definitions.

The claim is simply that this new measure, probability sub A--

[NO SOUND]

--will satisfy all of the rules of probability. Because it is a probability measure.

So for example, I have the difference rule restated for conditional probabilities. Given that probability sub A is a probability measure, it satisfies the difference rule. Which means when I translate it into a conditional probability statement, I get that the probability of B minus C, given A, is equal to the probability of B, given A, minus the probability of B intersection C, given A.

It's exactly the same as the standard difference rule, except that I have it made everything conditioned on A. And so we automatically get all of these rules for conditional probability that we had holding for probability. Which will be helpful. We won't have to think about proving them again.