PROFESSOR: So let's look now at the mathematical foundations of probability theory, the basic definitions of which we've just been doing examples up until now. So a key concept is a probability space. And that's what we're going to talk about in this segment.

So the abstract setting of a probability space is the first thing you start off with is the set of outcomes, which is what we saw we were doing with the tree models in the previous videos. And the condition that we require is that there should be a countable set of outcomes.

So there's something called the sample space. And the sample space is designed to model all the possible things that can happen as the result of your random experiment, all the possible outcomes. And we require that there be a countable number.

Now, the examples that we've seen so far have only had a finite number. But we will shortly see a bunch of examples where we really need an infinite number. But only a countable infinite number. That's part of the definition of a probability space-- the set of outcomes.

The next thing is a probability function whose task is to assign probabilities to the outcomes. So the condition is that the probability function, Pr, gives every element in S, every outcome, is going to be assigned a probability of between 0 and 1 inclusive. So every outcome gets a probability between 0 and 1.

But the constraint on the probability function is that if I sum up the probabilities of all the outcomes-- omega is an outcome in the sample space S-- and I take the sum of all of those probabilities of omega, they have to sum to 1. That's the crucial condition that defines a probability function on a sample space. And the two together are what are called a probability space. A sample space with a probability function is a probability space.

So the purpose of the tree model that we were using is really to figure out which probability space to use. And the mathematics doesn't really start until you have the probability space. Up until that, it's the modeling part that's very important mathematically. But you can't say that the model is right or wrong. It's a model, and its rightness or wrongness is a matter of judgment and comparison to how it stacks up against reality and things that we care about.

When we're using the tree model, it's the leaves of the tree that correspond to the outcomes. And the outcome probabilities, which are crucial for having a probability space, we got by

reasoning about the probabilities to assign to each possible branch of the tree as you worked your way from root to leaf.

So the other key concept that we saw already is the idea of an event. An event, formally, is nothing but a subset of the sample space. An event is some set of outcomes. Presumably, the event is an event that you're interested in, like winning.

And the definition of the probability of an event is simply the sum of the probabilities of all the outcomes in the event. And we were using this already for both Monty Hall and for the poker hands. But this is the official general definition-- that once we have a probability function that assigns probabilities to outcomes, then we can use that to define the probability of an event. This is the definition of the probability of an event-- simply the sum of the outcome probabilities.

And as an immediate corollary of this definition, what we get is something that's central to probability theory. It's called the sum rule. And it says that if you have a bunch of events that are pairwise disjoint-- so there's no outcome in common to A0 or A1 or A1 or A2 and so on-- then the probability of the union of the A's, the probability that one of these events occurs, one or more of these events occurs, is simply the sum of the individual probabilities.

And that is a rule that we'll be using all the time. It's very convenient for computing things. If you just break them up into separate cases, then you can handle the separate cases-- each A0, A1-- separately, and then add up the probabilities.

And in some approaches to probability, more general ones, this is actually taken as an axiom. It's the axiom that defines a probability space, but where you start with an assignment of probabilities to events. But in the discrete case, we don't have to worry about that. It's a corollary of the way we define the probability.

And that, of course, is a crucial rule-- sometimes called the countable sum rule. But we're just going to call it the sum rule. Expressed in concise notation, it's the probability of the union of the Ai's, as i ranges over the non-negative integers, is simply the sum of the individual probabilities of those events.

Now, why it's called discrete probability that we're studying is because we have a countable sample space. And as we saw, that discrete combinatorics is the combinatorics of countable and even finite sets, really. The crucial reason why we're sticking to discrete probability is that

allows us to work with sums instead of integrals.

If you start allowing continuous intervals of time and the probability, say, of throwing a dart and it landing at a given interval on the line and a whole bunch of other situations where it's natural to want to use continuous probabilities, you're forced into defining a probability in terms of integrals, because every outcome has probability 0. And the theoretical basis of it is considerably more complicated. And we don't need it for, in fact, virtually any purposes that come up in computer science. And so we will, happily, not have to study integral calculus or measure theory, really, and just get by with sums.

So let's quickly point out some rules that are now corollaries. They're really derived rules of probability theory that follow as a consequence of the countable sum rule. And the first one is the difference rule. The probability of A minus B is simply equal to the probability of A minus the probability of A intersection B.

Now, notice how much this looks like the difference rule for cardinalities-- that the cardinality of the finite set A minus B is simply the cardinality of A minus the cardinality of A intersection B. And indeed, the proof of this is just like the proof of cardinality. It follows directly from the sum rule for probabilities, which corresponds, of course, to the sum rule for cardinalities.

Namely, by the sum rule for probabilities, what we know is that A is equal set, theoretically, to A intersection B plus A minus b. That is, A breaks up into the points that it has in common with B and the points that it doesn't have in common with B. Since those are disjoint, you can add them. So the probability of A is equal to the probability of A intersection B plus probability of A minus B. And so I just transpose the A minus B to the left hand side. And I get the difference rule, which is a rule that's worth remembering.

Similarly, we have inclusion-exclusion. If A and B are not disjoint, then the probability of A union B is equal to the probability of A plus the probability of B minus the probability of the intersection. And the proof is, in fact, exactly like the corresponding rule for cardinalities of finite sets. And of course, it generalizes to more sets. This is an example of the inclusion-exclusion for three sets in terms of probability.

Another useful, it turns out, consequence is that the probability that A or B happens is guaranteed to be less than or equal to the probability that A happens plus the probability of B happens. And this follows as a trivial consequence of the inclusion-exclusion rule for two sets, because the probability of A union B is equal to this plus this minus some probability, namely,

the probability of the intersection.

So you're taking away something non-negative from these two in order to equal that. In particular, then, this must be less than or equal to that.

And the closely related phenomenon is [? basi-- ?] [AUDIO OUT] The probability that A or B happens is greater than or equal to the probability that A happens.

And finally, we can generalize that to a countable collection of sets. If I have a bunch of events-- A0, A1, and so on-- then the probability that at least one of them occurs, the probability of the union of the Ai's, is less than or equal to the sum of their probabilities.

This is, again, another kind of obvious rule, not hard to prove. We're not going to bother proving it, because it really is obvious. But we will get some mileage out of it later on.

So to summarize, the key concept here is a probability space. It consists of a countable set of outcomes, the sample space, and a probability function that assigns values between 0 and 1 to every outcome such that the sum of the probabilities is 1. And when we're using our tree model and so on, our objective is to construct one of these things. Usually, the hard part will be verifying that, in fact, the way we've assigned probabilities has the property that the sum of them is equal to 1.