PROFESSOR: Certain kinds of random variables keep coming up, so let's look at two basic examples now, namely uniform random variables and binomial random variables. Let's begin with uniform, because we've seen those already. So a uniform random variable means that all the values that it takes, it takes with equal probability.

So the threshold variable Z took all the values from 0 to 6 inclusive, each with probability 1/7. So it was a basic example of a uniform variable. And other examples that come up, if D is the outcome of a fair die-- dies are six-sided. Dice are six-sided. So the probability that it comes up 1 or 2 or 6 is 1/6 each.

Another game is the four-digit lottery number where it's supposed to be the case that the four digits are each chosen at random, which means that the possibilities range from four 0's up through four 9's for 10,000 numbers. And they're supposed to be all equally likely. So the probability that the lottery winds up with 00 is the same as that it ends up with 1 is the same that it ends up with four 9's. It's 1/10,000. So that's another uniform random variable.

Let's prove a little lemma that will be of use later. It's just some practice with uniformity. Suppose that I have R1, R2, R3 are three random variables. They're mutually independent.

And R1 is uniform. I don't really care about the other two. I do care technically that they are only taking the values. They only take values that R1 can take as well. So I haven't said that on this slide, but that's what we're assuming.

And then I claim is that each of the pairs, the probability that R1 equals R2-- the event that R1 is equal to R2 is independent of the event that R2 is equal to R3, which is independent of the event that R1 is equal to R3. Now, these events overlap. There's an R1 here and an R1 there and there's an R2 here and an R2 there. So even though the R1, R2, R3 are mutually independent, it's not completely clear. In fact, it isn't really clear that these events are mutually independent.

But in fact, they're not mutually independent. In fact, they're pairwise independent. They're obviously not three-way independent-- that is, mutually independent-- because if I know that R1 is equal to R2 and I know that R2 is equal to R3, it follows that R1 is equal to R3.

So given these two, the probability of this changes dramatically to certainty. So this is the

useful lemma, which is that if I have the three variables and I look at the three possible pairs of values that might be equal that whether any two of them are equal is independent of each other.

Now, let me give a handwaving argument. There's a more rigorous argument based on total probability that appears as a problem in the text. But the intuitive ideas, let's look at the case that R1 is the uniform variable, and R1 is independent of R2 and R3. So certainly, that implies that R1 is independent of the event that R2 is equal to R3, because R1 isn't mutually independent, both R1 and R2. Doesn't matter what they do, so it's independent of this event that R2 is equal to R3.

Now, because R1 is uniform, it has probability p of equaling every possible value that it can take. And since R2 and R3 only take a value that R1 could take, the probability that R1 hits the value that R2 and R3 happens to have is still p. That's the informal argument.

So in other words, the claim is that the probability that R1 is equal to R2 given that R2 is equal to R3 is simply the probability that R1 happens to hit R2, whatever values R2 has. This equation says that R1 equals R2 is independent of R2, R3. And in fact, in both cases, it's the same probability that R1 is equal to any given value, the probability of R being uniform has of equaling each of its possible values.

You can think about that, see if it's persuasive. It's an OK argument, but I was bothered by it. I found that it took me-- I wasn't happy with it until I sat down and really worked it out formally to justify this somewhat handwavy proof of the lemma.

All right. Let's turn from uniform random variables to binomial random variables. They're probably the most important single example of random variable that comes up all the time. So the simplest definition of a binomial random variable is the one that you get by flipping n mutually independent coins. Now, they have an order, so you can tell them apart. Or again, you can say that you flip one coin n times, but each of the flips is independent of all the others.

Now, there's two parameters here, an n and a p, because we don't assume that the flips are fair. So there's one parameter is how many flips there are. The other parameter is the probability of a head, which might be biased that heads are more likely or less likely than tails. The fair case would be when p was 1/2.

So for example, if n is 5 and p is 2/3, what's the probability that we consecutively flip head,

head, tail, tail, head? Well, because they're independent, the probability of this is simply the product of the probability that I flip a head on the first toss, which is probability of H, which is p; probability of H on the second toss; probability of T on the third; T on the fourth; T on the fifth.

So I can replace each of those by 2/3 is the probability of a head. 2/3, 1/3. 1 minus 2/3 is the probability of a tail. 1/3, 2/3. And I discover that the probability of HHTTH is 2/3 cubed and 1/3 squared.

Or abstracting the probability of a sequence of n tosses in which there are i heads and the rest are tails, n minus i tails, is simply the probability of a head raised to the i-th power times the probability of a tail, namely 1 minus p, raised to the n minus i-th power. Given any particular sequence of H's and T's of length n, this is the probability that's assigned to that sequence. So all sequences with the same number of H's have the same probability. But of course, with different numbers of H's they have different probabilities.

Well, what's the probability that you actually toss i heads and n minus i tails in the n tosses? That's going to be equal to the number of possible sequences that have this property of i heads and n minus i tails. Well, the number of such sequences is simply choose the i places for the n heads out of-- choose the i places for the heads out of the n tosses. So it's going to be n choose i. So we've just figured out that the probability of tossing i heads and n minus i tails is simply n choose i times p to the i, 1 minus p to the n minus i.

In short, the probability that the number of heads is i is equal to this number. And this is the probability that's associated with whether the binomial variable with parameters n and p is equal to i is n choose i, p to the i, 1 minus p to the n minus i. This is a pretty basic formula. If you can't memorize it, then make sure it's written on any crib sheet you take to an exam.

So the probability density function, it abstracts out some properties of random variables. Basically, it just tells you what's the probability that the random variable takes a given value for every possible value. So the probability density function, PDF of R, is a function on the real values. And it tells you for each a what's the probability that R is equal to a.

So what we've just said is that the probability density function of the binomial random variable characterized by parameters n and p at i is n choose i, p to the i, 1 minus p to the n minus i, where we're assuming that i is an integer from 0 to n. If I look at the probability density function for a uniform variable, then it's constant.

The probability density function on any possible value v that the uniform variable can take is the same. This applies for v in the range of U. So in fact, you could say exactly what it is. It's simply 1 over the size of the range of U, if U is uniform.

A closely related function that describes a lot about the behavior of a random variable is the cumulative distribution function. It's simply the probability that R is less than or equal to a. So it's a function on the real numbers, from reals to reals, where CDF R of a is the probability that R is less than or equal to a. Clearly given the PDF, you can get the CDF. And given the CDF, you can get the PDF. But it's convenient to have both around.

Now the key observation about these is that once we've abstracted out to the PDF and the CDF, we don't have to think about the sample space anymore. They do not depend on the sample space. All they're telling you is the probability that the random variable takes a given value, which is in some ways, the most important data about a random variable. You need to fall back on something more general than the PDF or the CDF when you start having dependent random variables, and you need to know how the probability that R takes a value changes, given that s has some property or takes some other value.

But if you're just looking at the random variable alone, essentially everything you need to know about it is given by its density or distribution functions. And you don't have to worry about the sample space. And this has the advantage that both the uniform distributions and binomial distributions come up [AUDIO OUT]

--and it means that all of these different random variables, based on different sample spaces, are going to share a whole lot of properties. Everything that I derive based on what the PDF is is going to apply to all of them. That's why this abstraction of a random variable in terms of a probability density function is so valuable and key. But remember, the definition of a random variable is not that it is a probability density function, rather it's a function from the sample space to values.