We will now work with a geometric random variable and put to use our understanding of conditional PMFs and conditional expectations. Remember that a geometric random variable corresponds to the number of independent coin tosses until the first head occurs. And here p is a parameter that describes the coin. It is the probability of heads at each coin toss.

We have already seen the formula for the geometric PMF and the corresponding plot. We will now add one very important property which is usually called Memorylessness. Ultimately, this property has to do with the fact that independent coin tosses do not have any memory. Past coin tosses do not affect future coin tosses.

So consider a coin-tossing experiment with independent tosses and let X be the number of tosses until the first heads. And X is a geometric with parameter p. Suppose that you show up a little after the experiment has started. And you're told that there was so far just one coin toss. And that this coin toss resulted in tails.

Now you have to take over and carry out the remaining tosses until heads are observed. What should your model be about the future? Well, you will be making independent coin tosses until the first heads.

So the number of such tosses will be a random variable, which is geometric with parameter p. So this duration-- as far as you are concerned-- is geometric with parameter p. Therefore, the number of remaining coin tosses starting from here-- given that the first toss was tails-- has the same geometric distribution as the original random variable X.

This is the Memorylessness property. Now, since X is the total number of coin tosses and since your coin tosses were all of them except for the first one, the random variable that you are concerned with is X minus 1. And so the geometric distribution that you are seeing here is the conditional distribution of X minus 1 given that the first toss resulted in tails, which is the same as the event that X is strictly larger than 1.

So the statement that we have been making is the following in more mathematical language-- that conditioned on X being larger than 1, the random variable X minus 1, which is the remaining number of coin tosses, has a geometric distribution with parameter p.

Let us now give a more precise, mathematical argument. But first, for a special case. Let's us look at the conditional probabilities for the random variable X minus 1. And calculate, for example, the conditional probability that X minus 1 is equal to 3, given that X is larger than 1. Which is the same as saying that the first toss resulted in tails.

Now, the first toss resulted in tails. This is the probability that you will need three more tosses until you observe heads. Needing three more tosses until you observe heads is the event that you had tails in the second toss, tails in the third toss, and heads in the fourth toss.

And all that is conditioned on the first toss having resulted in tails. However, the different coin tosses are independent. So the conditional probabilities, given the event that the first toss was tails should be the same as the unconditional probabilities. The first toss does not change our beliefs about the probabilities associated with the remaining tosses.

Now, this unconditional probability is easy to calculate. It is 1 minus p squared-- because we have two tails in a row-- times p. Now, we observe that this quantity here is the probability that a geometric random variable takes the value of three. Here what have we calculated? We have calculated the PMF of the random variable X minus 1 in a conditional universe where X is larger than 1.

And we evaluated it for a value of 3. The probability that our random variable X minus 1 takes the value of 3. So what we have shown is that this conditional PMF is the same as the unconditional PMF. Now, there is nothing special about the number 3. You can generalize this argument and establish that the conditional probability of X minus 1 given that X is strictly larger than one, for any particular k, is the same as the corresponding probability for the random variable X, which is given by the geometric PMF.

Finally, there is nothing special about the value of 1 that we're using here. In fact, we can generalize and argue as follows-- suppose that I tell you that X is strictly larger than n. That is, the first n tosses resulted in tails. Once more, these past tosses were wasted but have no effect on the future. So the conditional PMF of the remaining number of tosses should be, again, the same.

Therefore, the statement we're making is that this geometric PMF will also be the PMF of X minus n, given that X is strictly larger than n, and this will be true no matter what argument we plug-in into the PMF.

We will now exploit the Memorylessness property of the geometric PMF and use it together with the total expectation theorem to calculate the mean or expectation of the geometric PMF. If we wanted to calculate the expected value of the geometric using the definition of the expectation, we would have to calculate this infinite sum here, which is quite difficult.

Instead, we're going to use a certain trick. The trick is the following-- to break down the expected value calculation into two different scenarios. Under one scenario we obtain heads in the first toss. And in that case the random variable X-- the number of tosses until the first heads-- is equal to 1. And this scenario occurs with probability p.

And we have another scenario with probability 1 minus p where we obtain tails in the first toss. And in that case, our random variable is strictly larger than 1. Now, the expected value of X consists of two pieces.

We have a first toss no matter what. And then we have the number of remaining tosses, which is X minus 1. So this is true by linearity of expectations. Now, the expected value of X minus 1 consists of two pieces using the total expectation theorem.

The probability of the first scenario times the expected value of X minus 1 given that X is equal to 1, plus 1 minus p-- the probability of the second scenario-- times the expected value of X minus 1 given that X is bigger than 1.

Now, this term here is 0. Why? If I tell you that X is equal to 1, then you're certain that's X minus 1 is equal to 0. So this term gives a 0 contribution.

How about the next term? We have a 1 minus p here times this expected value. Now this random variable, conditioned on this event, has the same distribution as an ordinary, unconditioned geometric random variable.

So this expectation here must be the same as the expectation of an ordinary, unconditioned, geometric random variable. And this gives us an equality. Both sides involve the expected value of X. But we can solve this equation for the expected value. And we obtain the end result that the expected value is 1 over p.

By the way, this answer makes intuitive sense. If p is small, this means that the odds of seeing heads is

small. Then in that case, we need to wait longer and longer until we see heads for the first time. Setting aside the specific form of the answer that we found, what we have just done actually illustrates that fairly difficult calculations can become very simple if one breaks down a model or a problem in a clever way. This is going to be a recurring theme throughout this class.