MITOCW | MIT6_041F11_lec05_300k.mp4

OK. So let us start. All right. So today we're starting a new unit in this class. We have covered, so far, the basics of probability theory-- the main concepts and tools, as far as just probabilities are concerned. But if that was all that there is in this subject, the subject would not be rich enough. What makes probability theory a lot more interesting and richer is that we can also talk about random variables, which are ways of assigning numerical results to the outcomes of an experiment.

So we're going to define what random variables are, and then we're going to describe them using so-called probability mass functions. Basically some numerical values are more likely to occur than other numerical values, and we capture this by assigning probabilities to them the usual way. And we represent these in a compact way using the so-called probability mass functions.

We're going to see a couple of examples of random variables, some of which we have already seen but with different terminology. And so far, it's going to be just a couple of definitions and calculations of the type that you already know how to do. But then we're going to introduce the one new, big concept of the day.

So up to here it's going to be mostly an exercise in notation and definitions. But then we got our big concept which is the concept of the expected value of a random variable, which is some kind of average value of the random variable. And then we're going to also talk, very briefly, about close distance of the expectation, which is the concept of the variance of a random variable.

OK. So what is a random variable? It's an assignment of a numerical value to every possible outcome of the experiment. So here's the picture. The sample space is this class, and we've got lots of students in here. This is our sample space, omega. I'm interested in the height of a random student. So I'm going to use a real line where I record height, and let's say this is height in inches.

And the experiment happens, I pick a random student. And I go and measure the height of that random student, and that gives me a specific number. So what's a good number in inches? Let's say 60. OK. Or I pick another student, and that student has a height of 71 inches, and so on.

So this is the experiment. These are the outcomes. These are the numerical values of the random variable that we call height. OK. So mathematically, what are we dealing with here? We're basically dealing with a function from the sample space into the real numbers. That function takes as argument, an outcome of the experiment, that is a typical student, and produces the value of that function, which is the height of that particular student.

So we think of an abstract object that we denote by a capital H, which is the random variable called height. And that random variable is essentially this particular function that we talked about here. OK. So there's a distinction that we're making here-- H is height in the abstract. It's the function. These numbers here are particular numerical values that this function takes when you choose one particular outcome of the experiment.

Now, when you have a single probability experiment, you can have multiple random variables. So perhaps, instead of just height, I'm also interested in the weight of a typical student. And so when the experiment happens, I pick that random student-- this is the height of the student. But that student would also have a weight, and I could record it here. And similarly, every student is going to have their own particular weight.

So the weight function is a different function from the sample space to the real numbers, and it's a different random variable. So the point I'm making here is that a single probabilistic experiment may involve several interesting random variables. I may be interested in the height of a random student or the weight of the random student. These are different random variables that could be of interest.

I can also do other things. Suppose I define an object such as H bar, which is 2.58. What does that correspond to? Well, this is the height in centimeters. Now, H bar is a function of H itself, but if you were to draw the picture, the picture would go this way. 60 gets mapped to 150, 71 gets mapped to, oh, that's too hard for me. OK, gets mapped to something, and so on.

So H bar is also a random variable. Why? Once I pick a particular student, that particular outcome determines completely the numerical value of H bar, which is the height of that student but measured in centimeters. What we have here is actually a random variable, which is defined as a function of another random variable. And the point that this example is trying to make is that functions of random variables are also random variables.

The experiment happens, the experiment determines a numerical value for this object. And once you have the numerical value for this object, that determines also the numerical value for that object. So given an outcome, the numerical value of this particular object is determined. So H bar is itself a function from the sample space, from outcomes to numerical values. And that makes it a random variable according to the formal definition that we have here.

So the formal definition is that the random variable is not random, it's not a variable, it's just a function from the sample space to the real numbers. That's the abstract, right way of thinking about them. Now, random variables can be of different types. They can be discrete or continuous.

Suppose that I measure the heights in inches, but I round to the nearest inch. Then the numerical values I'm going to get here would be just integers. So that would make it an integer valued random variable. And this is a discrete random variable.

Or maybe I have a scale for measuring height which is infinitely precise and records your height to an infinite number of digits of precision. In that case, your height would be just a general real number. So we would have a random variable that takes values in the entire set of real numbers. Well, I guess not really negative numbers, but the set of non-negative numbers. And that would be a continuous random variable. It takes values in a continuous set.

So we will be talking about both discrete and continuous random variables. The first thing we will do will be to devote a few lectures on discrete random variables, because discrete is always easier. And then we're going to repeat everything in the continuous setting. So discrete is easier, and it's the right place to understand all the concepts, even those who may appear to be elementary. And then you will be set to understand what's going on when we go to the continuous case.

So in the continuous case, you get all the complications of calculus and some extra math that comes in there. So it's important to have been down all the concepts very well in the easy, discrete case so that you don't have conceptual hurdles when you move on to the continuous case.

Now, one important remark that may seem trivial but it's actually very important so that you don't get tangled up between different types of concepts-- there's a fundamental distinction between the random variable itself, and the numerical values that it takes. Abstractly speaking, or mathematically speaking, a random variable, x, or H in this example, is a function.

OK. Maybe if you like programming the words "procedure" or "sub-routine" might be better. So what's the sub-routine height? Given a student, I take that student, force them on the scale and measure them. That's the sub-routine that measures heights. It's really a function that takes students as input and produces numbers as output.

The sub-routine we denoted by capital H. That's the random variable. But once you plug in a particular student into that sub-routine, you end up getting a particular number. This is the numerical output of that sub-routine or the numerical value of that function. And that numerical value is an element of the real numbers.

So the numerical value is a real number, where this capital X is a function from omega to the real numbers. So they are very different types of objects. And the way that we keep track of what we're talking about at any given time is by using capital letters for random variables and lower case letters for numbers.

OK. So now once we have a random variable at hand, that random variable takes on different numerical values. And we want to describe to say something about the relative likelihoods of the different numerical values that the random variable can take.

So here's our sample space, and here's the real line. And there's a bunch of outcomes that gave rise to one particular numerical value. There's another numerical value that arises if we have this outcome. There's another numerical value that arises if we have this outcome. So our sample space is here. The real numbers are here. And what we want to do is to ask the question, how likely is that particular numerical value to occur?

So what we're essentially asking is, how likely is it that we obtain an outcome that leads to that particular numerical value? We calculate that overall probability of that numerical value and we represent that probability using a bar so that we end up generating a bar graph. So that could be a possible bar graph associated with this picture. The size of this bar is the total probability that our random variable took on this numerical value, which is just the sum of the probabilities of the different outcomes that led to that numerical value.

So the thing that we're plotting here, the bar graph-- we give a name to it. It's a function, which we denote by lowercase b, capital X. The capital X indicates which random variable we're talking about. And it's a function of little x, which is the range of values that our random variable is taking.

So in mathematical notation, the value of the PMF at some particular number, little x, is the probability that our random variable takes on the numerical value, little x. And if you want to be precise about what this means, it's the overall probability of all outcomes for which the random variable ends up taking that value, little x.

So this is the overall probability of all omegas that lead to that particular numerical value, x, of interest. So what do we know about PMFs? Since there are probabilities, all these entries in the bar graph have to be non-negative. Also, if you exhaust all the possible values of little x's, you will have exhausted all the possible outcomes here. Because every outcome leads to some particular x.

So the sum of these probabilities should be equal to one. This is the second relation here. So this relation tell us that some little x is going to happen. They happen with different probabilities, but when you consider all the possible little x's together, one of those little x's is going to be realized. Probabilities need to add to one.

OK. So let's get our first example of a non-trivial bar graph. Consider the experiment where I start with a coin and I start flipping it over and over. And I do this until I obtain heads for the first time. So what are possible outcomes of this experiment?

One possible outcome is that I obtain heads at the first toss, and then I stop. In this case, my random variable takes the value 1. Or it's possible that I obtain tails and then heads. How many tosses did it take until heads appeared? This would be x equals to 2. Or more generally, I might obtain tails for k minus 1 times, and then obtain heads at the k-th time, in which case, our random variable takes the value, little k.

So that's the experiment. So capital X is a well defined random variable. It's the number of tosses it takes until I see heads for the first time. These are the possible outcomes. These are elements of our sample space. And these are the values of X depending on the outcome. Clearly X is a function of the outcome. You tell me the outcome, I'm going to tell you what X is.

So what we want to do now is to calculate the PMF of X. So Px of k is, by definition, the probability that our random variable takes the value k. For the random variable to take the value of k, the first head appears at toss number k. The only way that this event can happen is if we obtain this sequence of events.

T's the first k minus 1 times, tails, and heads at the k-th flip. So this event, that the random variable is equal to k, is the same as this event, k minus 1 tails followed by 1 head. What's the probability of that event?

We're assuming that the coin tosses are independent. So to find the probability of this event, we need to multiply the probability of tails, times the probability of tails, times the probability of tails. We multiply k minus one times, times the probability of heads, which puts an extra p at the end. And this is the formula for the so-called geometric PMF.

And why do we call it geometric? Because if you go and plot the bar graph of this random variable, X, we start at 1 with a certain number, which is p. And then at 2 we get p(1-p). At 3 we're going to get something smaller, it's p times (1-p)-squared. And the bars keep going down at the rate of geometric progression. Each bar is smaller than the previous bar, because each time we get an extra factor of 1-p involved.

So the shape of this PMF is the graph of a geometric sequence. For that reason, we say that it's the geometric PMF, and we call X also a geometric random variable. So the number of coin tosses until the first head is a geometric random variable.

So this was an example of how to compute the PMF of a random variable. This was an easy example, because this event could be realized in one and only one way. So to find the probability of this, we just needed to find the probability of this particular outcome. More generally, there's going to be many outcomes that can lead to the same numerical value. And we need to keep track of all of them.

For example, in this picture, if I want to find this value of the PMF, I need to add up the probabilities of all the outcomes that leads to that value. So the general procedure is exactly what this picture suggests. To find this probability, you go and identify which outcomes lead to this numerical value, and add their probabilities.

So let's do a simple example. I take a tetrahedral die. I toss it twice. And there's lots of random variables that you can associate with the same experiment. So the outcome of the first throw, we can call it F. That's a random variable because it's determined once you tell me what happens in the experiment.

The outcome of the second throw is another random variable. The minimum of the two throws is also a random variable. Once I do the experiment, this random variable takes on a specific numerical value. So suppose I do the experiment and I get a 2 and a 3. So this random variable is going to take the numerical value of 2. This is going to take the numerical value of 3. This is going to take the numerical value of 2.

And now suppose that I want to calculate the PMF of this random variable. What I will need to do is to calculate Px(0), Px(1), Px(2), Px(3), and so on. Let's not do the entire calculation then, let's just calculate one of the entries of the PMF.

So Px(2)-- that's the probability that the minimum of the two throws gives us a 2. And this can happen in many ways. There are five ways that it can happen. Those are all of the outcomes for which the smallest of the two is equal to 2. That's five outcomes assuming that the tetrahedral die is fair and the tosses are independent. Each one of these outcomes has probability of 1/16. There's five of them, so we get an answer, 5/16.

Conceptually, this is just the procedure that you use to calculate PMFs the way that you construct this particular bar graph. You consider all the possible values of your random variable, and for each one of those random variables you find the probability that the random variable takes on that value by adding the probabilities of all the possible outcomes that leads to that particular numerical value.

So let's do another, more interesting one. So let's revisit the coin tossing problem from last time. Let us fix a number n, and we decide to flip a coin n consecutive times. Each time the coin tosses are independent. And each one of the tosses will have a probability, p, of obtaining heads.

Let's consider the random variable, which is the total number of heads that have been obtained. Well, that's something that we dealt with last time. We know the probabilities for different numbers of heads, but we're just going to do the same now using today's notation.

So let's, for concreteness, n equal to 4. Px is the PMF of that random variable, X. Px(2) is meant to be, by definition, it's the probability that a random variable takes the value of 2. So this is the probability that we have, exactly two heads in our four tosses.

The event of exactly two heads can happen in multiple ways. And here I've written down the different ways that it can happen. It turns out that there's exactly six ways that it can happen. And each one of these ways, luckily enough, has the same probability-- p-squared times (1-p)-squared. So that gives us the value for the PMF evaluated at 2.

So here we just counted explicitly that we have six possible ways that this can happen, and this gave rise to this factor of 6. But this factor of 6 turns out to be the same as this 4 choose 2. If you remember definition from last time, 4 choose 2 is 4 factorial divided by 2 factorial, divided by 2 factorial, which is indeed equal to 6. And this is the more general formula that you would be using.

In general, if you have n tosses and you're interested in the probability of obtaining k heads, the probability of that event is given by this formula. So that's the formula that we derived last time. Except that last time we didn't use this notation. We just said the probability of k heads is equal to this. Today we introduce the extra notation.

And also having that notation, we may be tempted to also plot a bar graph for the Px. In this case, for the coin tossing problem. And if you plot that bar graph as a function of k when n is a fairly large number, what you will end up obtaining is a bar graph that has a shape of something like this.

So certain values of k are more likely than others, and the more likely values are somewhere in the middle of the range. And extreme values-- too few heads or too many heads, are unlikely. Now, the miraculous thing is that it turns out that this curve gets a pretty definite shape, like a so-called bell curve, when n is big.

This is a very deep and central fact from probability theory that we will get to in a couple of months. For now, it just could be a curious observation. If you go into MATLAB and put this formula in and ask MATLAB to plot it for you, you're going to get an interesting shape of this form. And later on we will have to sort of understand where this is coming from and whether there's a nice, simple formula for the asymptotic form that we get.

All right. So, so far I've said essentially nothing new, just a little bit of notation and this little conceptual thing that you have to think of random variables as functions in the sample space. So now it's time to introduce something new. This is the big concept of the day. In some sense it's an easy concept.

But it's the most central, most important concept that we have to deal with random variables. It's the concept of the expected value of a random variable. So the expected value is meant to be, let's speak loosely, something like an average, where you interpret probabilities as something like frequencies.

So you play a certain game and your rewards are going to be-- use my standard numbers-- your rewards are going to be one dollar with probability 1/6. It's going to be 2 dollars with probability 1/2, and four dollars with probability 1/3. So this is a plot of the PMF of some random variable. If you play that game and you get so many dollars with this probability, and so on, how much do you expect to get on the average if you play the game a zillion times?

Well, you can think as follows-- one sixth of the time I'm going to get one dollar. One half of the time that outcome is going to happen and I'm going to get two dollars. And one third of the time the other outcome happens, and I'm going to get four dollars. And you evaluate that number and it turns out to be 2.5. OK. So that's a reasonable way of calculating the average payoff if you think of these probabilities as the frequencies with which you obtain the different payoffs.

And loosely speaking, it doesn't hurt to think of probabilities as frequencies when you try to make sense of various things. So what did we do here? We took the probabilities of the different outcomes, of the different numerical values, and multiplied them with the corresponding numerical value.

Similarly here, we have a probability and the corresponding numerical value and we added up over all x's. So that's what we did. It looks like an interesting quantity to deal with. So we're going to give a name to it, and we're going to call it the expected value of a random variable. So this formula just captures the calculation that we did.

How do we interpret the expected value? So the one interpretation is the one that I used in this example. You can think of it as the average that you get over a large number of repetitions of an experiment where you interpret the probabilities as the frequencies with which the different numerical values can happen.

There's another interpretation that's a little more visual and that's kind of insightful, if you remember your freshman physics, this kind of formula gives you the center of gravity of an object of this kind. If you take that picture literally and think of this as a mass of one sixth sitting here, and the mass of one half sitting here, and one third sitting there, and you ask me what's the center of gravity of that structure. This is the formula that gives you the center of gravity.

Now what's the center of gravity? It's the place where if you put your pen right underneath, that diagram will stay in place and will not fall on one side and will not fall on the other side. So in this thing, by picture, since the 4 is a little more to the right and a little heavier, the center of gravity should be somewhere around here. And that's what for the math gave us. It turns out to be two and a half.

Once you have this interpretation about centers of gravity, sometimes you can calculate expectations pretty fast. So here's our new random variable. It's the uniform random variable in which each one of the numerical values is equally likely. Here there's a total of n plus 1 possible numerical values. So each one of them has probability 1 over (n + 1).

Let's calculate the expected value of this random variable. We can take the formula literally and consider all possible outcomes, or all possible numerical values, and weigh them by their corresponding probability, and do this calculation and obtain an answer. But I gave you the intuition of centers of gravity. Can you use that intuition to guess the answer?

What's the center of gravity infrastructure of this kind? We have symmetry. So it should be in the middle. And what's the middle? It's the average of the two end points. So without having to do the algebra, you know that's the answer is going to be n over 2.

So this is a moral that you should keep whenever you have PMF, which is symmetric around a certain point. That certain point is going to be the expected value associated with this particular PMF. OK. So having defined the expected value, what is there that's left for us to do?

Well, we want to investigate how it behaves, what kind of properties does it have, and also how do you calculate expected values of complicated random variables. So the first complication that we're going to start with is the case where we deal with a function of a random variable.

OK. So let me redraw this same picture as before. We have omega. This is our sample space. This is the real line. And we have a random variable that gives rise to various values for X. So the random variable is capital X, and every outcome leads to a particular numerical value x for our random variable X. So capital X is really the function that maps these points into the real line.

And then I consider a function of this random variable, call it capital Y, and it's a function of my previous random variable. And this new random variable Y takes numerical values that are completely determined once I know the numerical value of capital X. And perhaps you get a diagram of this kind.

So X is a random variable. Once you have an outcome, this determines the value of x. Y is also a random variable. Once you have the outcome, that determines the value of y. Y is completely determined once you know X. We have a formula for how to calculate the expected value of X.

Suppose that you're interested in calculating the expected value of Y. How would you go about it? OK. The only thing you have in your hands is the definition, so you could start by just using the definition. And what does this entail? It entails for every particular value of y, collect all the outcomes that leads to that value of y. Find their probability. Do the same here. For that value, collect those outcomes. Find their probability and weight by y.

So this formula does the addition over this line. We consider the different outcomes and add things up. There's an alternative way of doing the same accounting where instead of doing the addition over those numbers, we do the addition up here. We consider the different possible values of x, and we think as follows-- for each possible value of x, that value is going to occur with this probability. And if that value has occurred, this is how much I'm getting, the g of x.

So I'm considering the probability of this outcome. And in that case, y takes this value. Then I'm considering the probabilities of this outcome. And in that case, g of x takes again that value. Then I consider this particular x, it happens with this much probability, and in that case, g of x takes that value, and similarly here.

We end up doing exactly the same arithmetic, it's only a question whether we bundle things together. That is, if we calculate the probability of this, then we're bundling these two cases together. Whereas if we do the addition up here, we do a separate calculation-- this probability times this number, and then this probability times that number.

So it's just a simple rearrangement of the way that we do the calculations, but it does make a big difference in practice if you actually want to calculate expectations. So the second procedure that I mentioned, where you do the addition by running over the x-axis corresponds to this formula. Consider all possibilities for x and when that x happens, how much money are you getting? That gives you the average money that you are getting.

All right. So I kind of hand waved and argued that it's just a different way of accounting, of course one needs to prove this formula. And fortunately it can be proved. You're going to see that in recitation. Most people, once they're a little comfortable with the concepts of probability, actually believe that this is true by definition. In fact it's not true by definition. It's called the law of the unconscious statistician. It's something that you always do, but it's something that does require justification.

All right. So this gives us basically a shortcut for calculating expected values of functions of a random variable without having to find the PMF of that function. We can work with the PMF of the original function. All right. So we're going to use this property over and over.

Before we start using it, one general word of caution-- the average of a function of a random variable, in general, is not the same as the function of the average. So these two operations of taking averages and taking functions do not commute. What this inequality tells you is that, in general, you can not reason on the average.

So we're going to see instances where this property is not true. You're going to see lots of them. Let me just throw it here that it's something that's not true in general, but we will be interested in the exceptions where a relation like this is true. But these will be the exceptions. So in general, expectations are average, something like averages. But the function of an average is not the same as the average of the function.

OK. So now let's go to properties of expectations. Suppose that alpha is a real number, and I ask you, what's the expected value of that real number? So for example, if I write down this expression-- expected value of 2. What is this?

Well, we defined random variables and we defined expectations of random variables. So for this to make syntactic sense, this thing inside here should be a random variable. Is 2 -- the number 2 --- is it a random variable? In some sense, yes. It's the random variable that takes, always, the value of 2.

So suppose that you have some experiment and that experiment always outputs 2 whenever it happens. Then you can say, yes, it's a random experiment but it always gives me 2. The value of the random variable is always 2 no matter what. It's kind of a degenerate random variable that doesn't have any real randomness in it, but it's still useful to think of it as a special case.

So it corresponds to a function from the sample space to the real line that takes only one value. No matter what the outcome is, it always gives me a 2. OK. If you have a random variable that always gives you a 2, what is the expected value going to be? The only entry that shows up in this summation is that number 2. The probability of a 2 is equal to 1, and the value of that random variable is equal to 2. So it's the number itself. So the average value in an experiment that always gives you 2's is 2.

All right. So that's nice and simple. Now let's go to our experiment where age was your height in inches. And I know your height in inches, but I'm interested in your height measured in centimeters. How is that going to be related to your height in inches?

Well, if you take your height in inches and convert it to centimeters, I have another random variable, which is always, no matter what, two and a half times bigger than the random variable I started with. If you take some quantity and always multiplied by two and a half what happens to the average of that quantity? It also gets multiplied by two and a half. So you get a relation like this, which says that the average height of a student measured in centimeters is two and a half times the average height of a student measured in inches.

So that makes perfect intuitive sense. If you generalize it, it gives us this relation, that if you have a number, you can pull it outside the expectation and you get the right result. So this is a case where you can reason on the average. If you take a number, such as height, and multiply it by a certain number, you can reason on the average. I multiply the numbers by two, the averages will go up by two.

So this is an exception to this cautionary statement that I had up there. How do we prove that this fact is true? Well, we can use the expected value rule here, which tells us that the expected value of alpha X, this is our g of X, essentially, is going to be the sum over all x's of my function, g of X, times the probability of the x's. In our particular case, g of X is alpha times X. And we have those probabilities. And the alpha goes outside the summation. So we get alpha, sum over x's, x Px of x, which is alpha times the expected value of X.

So that's how you prove this relation formally using this rule up here. And the next formula that I have here also gets proved the same way. What does this formula tell you? If I take everybody's height in centimeters-- we already multiplied by alpha-- and the gods give everyone a bonus of ten extra centimeters. What's going to happen to the average height of the class? Well, it will just go up by an extra ten centimeters.

So this expectation is going to be giving you the bonus of beta just adds a beta to the average height in centimeters, which we also know to be alpha times the expected value of X, plus beta. So this is a linearity property of expectations. If you take a linear function of a single random variable, the expected value of that linear function is the linear function of the expected value. So this is our big exception to this cautionary note, that we have equal if g is linear.

OK. All right. So let's get to the last concept of the day. What kind of functions of random variables may be of interest? One possibility might be the average value of X-squared. Why is it interesting? Well, why not.

It's the simplest function that you can think of. So if you want to calculate the expected value of X-squared, you would use this general rule for how you can calculate expected values of functions of random variables. You consider all the possible x's. For each x, you see what's the probability that it occurs. And if that x occurs, you consider and see how big x-squared is.

Now, the more interesting quantity, a more interesting expectation that you can calculate has to do not with x-squared, but with the distance of x from the mean and then squared. So let's try to parse what we've got up here. Let's look just at the quantity inside here. What kind of quantity is it?

It's a random variable. Why? X is random, the random variable, expected value of X is a number. Subtract a number from a random variable, you get another random variable. Take a random variable and square it, you get another random variable. So the thing inside here is a legitimate random variable. What kind of random variable is it?

So suppose that we have our experiment and we have different x's that can happen. And the mean of X in this picture might be somewhere around here. I do the experiment. I obtain some numerical value of x. Let's say I obtain this numerical value. I look at the distance from the mean, which is this length, and I take the square of that.

Each time that I do the experiment, I go and record my distance from the mean and square it. So I give more emphasis to big distances. And then I take the average over all possible outcomes, all possible numerical values. So I'm trying to compute the average squared distance from the mean.

This corresponds to this formula here. So the picture that I drew corresponds to that. For every possible numerical value of x, that numerical value corresponds to a certain distance from the mean squared, and I weight according to how likely is that particular value of x to arise. So this measures the average squared distance from the mean.

Now, because of that expected value rule, of course, this thing is the same as that expectation. It's the average value of the random variable, which is the squared distance from the mean. With this probability, the random variable takes on this numerical value, and the squared distance from the mean ends up taking that particular numerical value.

OK. So why is the variance interesting? It tells us how far away from the mean we expect to be on the average. Well, actually we're not counting distances from the mean, it's distances squared. So it gives more emphasis to the kind of outliers in here. But it's a measure of how spread out the distribution is.

A big variance means that those bars go far to the left and to the right, typically. Where as a small variance would mean that all those bars are tightly concentrated around the mean value. It's the average squared deviation. Small variance means that we generally have small deviations. Large variances mean that we generally have large deviations.

Now as a practical matter, when you want to calculate the variance, there's a handy formula which I'm not proving but you will see it in recitation. It's just two lines of algebra. And it allows us to calculate it in a somewhat simpler way. We need to calculate the expected value of the random variable and the expected value of the squares of the random variable, and these two are going to give us the variance.

So to summarize what we did up here, the variance, by definition, is given by this formula. It's the expected value of the squared deviation. But we have the equivalent formula, which comes from application of the expected value rule, to the function g of X, equals to x minus the (expected value of X)-squared.

OK. So this is the definition. This comes from the expected value rule. What are some properties of the variance? Of course variances are always non-negative. Why is it always non-negative? Well, you look at the definition and your just adding up non-negative things. We're adding squared deviations. So when you add non-negative things, you get something non-negative.

The next question is, how do things scale if you take a linear function of a random variable? Let's think about the effects of beta. If I take a random variable and add the constant to it, how does this affect the amount of spread that we have? It doesn't affect-- whatever the spread of this thing is, if I add the constant beta, it just moves this diagram here, but the spread doesn't grow or get reduced.

The thing is that when I'm adding a constant to a random variable, all the x's that are going to appear are further to the right, but the expected value also moves to the right. And since we're only interested in distances from the mean, these distances do not get affected. x gets increased by something. The mean gets increased by that same something. The difference stays the same. So adding a constant to a random variable doesn't do anything to it's variance.

But if I multiply a random variable by a constant alpha, what is that going to do to its variance? Because we have a square here, when I multiply my random variable by a constant, this x gets multiplied by a constant, the mean gets multiplied by a constant, the square gets multiplied by the square of that constant. And because of that reason, we get this square of alpha showing up here. So that's how variances transform under linear transformations. You multiply your random variable by constant, the variance goes up by the square of that same constant. OK. That's it for today. See you on Wednesday.