

**PROFESSOR:** Our topic is deviation from the mean, meaning the probability that a random variable returns a value that differs significantly from its mean. Now, the Markov bound gave you a course bound on the probability that  $R$  was overly large using very little information about  $R$ . Not surprisingly, if you know a little bit more about the distribution of  $R$ , simply that it's not negative, you can state tighter bounds.

And this was noticed by a mathematician named Chebyshev. And he has a bound called the Chebyshev bound. Now, it's interesting that the Markov bound, even though it's very weak and seems not very useful, the Chebyshev bound, which generally gives you a significantly stronger, invaluable stronger bound on the probability that a random variable differs much from its mean is actually a trivial corollary of Markov theorem. So that's just a very simple ingenious way to use Markov's bound to derive Chebyshev bound. And let's look at how.

So we're interested in the probability that a random variable  $R$  differs from its mean by an amount  $x$ . The distance between  $R$  and its mean, the absolute value of  $R$  minus  $\mu$ , is greater than or equal to  $x$ . We're trying to get a grip on that probability as a function of  $x$ .

Now, the point is that the event that the distance between  $R$  and its mean is greater than or equal to  $x$ , another way to say that is to square both sides of this inequality. It says that the event that  $R$  minus  $\mu$  squared is greater or equal to  $x$  squared happens. These two events are just different ways of saying the same set. So therefore, their probabilities are equal trivially.

Now, what's nice about this is, of course, that  $R$  minus  $\mu$  squared is a non-negative random variable to which Markov's theorem applies. The square of a real number is always going to be non-negative. So let's just apply Markov's theorem to this new random variable,  $R$  minus  $\mu$  squared. And what does Markov's bound tell us about this probability, that the square variable is greater than or equal to an amount  $x$  squared.

Well, just plug in Markov. And it tells you that this probability that the square variable, that it's as big as  $x$  squared, is simply the expectation of that squared variable divided by  $x$  squared. This is just applying Markov's bound to this variable,  $R$  minus  $\mu$  squared.

Now, this numerator is a weird thing to stare at, expectation of  $R$  minus  $\mu$  squared, and may not seem very memorable. But you should remember, because it's so important that it has

name all its own. It's called the variance of  $R$ . And this is an extra bit of information about the shape of the distribution of  $R$  that turns out to allow you to state much more powerful theorems in general about the probability that  $R$  deviates from its mean by a given amount.

So we could just restate the Chebyshev bound. Just replacing that expectation formula in terms of its name, variance of  $R$ , this is what the Chebyshev bound says. The probability that the distance between  $R$  and its mean is greater than or equal to  $x$  is the variance of  $R$  divided by  $x$  squared, where variance of  $R$  is the expectation of the square of  $R$  minus  $\mu$ .

Now, the very important technical aspect of the Chebyshev bound is that we're getting an inverse square reduction in the probability. Remember, with Markov, the denominator was behaving linearly. And here, it behaves quite quadratically. So these bounds get smaller, much more rapidly as we ask about the probability of differing by a larger amount.

The variance of  $R$ , maybe in a way that will help you remember it is to remember another name that it has. It's called the mean square error. If you think of  $R$  minus  $\mu$  as the error that  $R$  is making in how much it differs from what it ought to be, and we square it, and then we take the average, so we're taking the mean of the squared errors.

And here, we're back to restating Markov bound in terms of the variance. The variance has one difficulty with it. And that leads us to want to look at another object, which is just the square root of the variance, called the standard deviation. So you wonder why-- I mean, if you understand variance, what's the point of taking the square root and working with that?

And the answer is simply that if you think of  $R$  as a random variable whose values have some dimension, like seconds or dollars, then the variance of  $R$  is the expectation of a square variable of  $R$  minus  $\mu$  squared, which means its units are second squared or dollar squared or whatever. And the variance of  $R$  itself is a squared value, which is not reflecting the magnitude of the distance that you expect-- of the kind of errors that you expect  $R$  to make, the distance that you expect part  $R$  to be from its mean.

So we can get the units of this quantity back into matching the units of  $R$  and also get a number that's closer to the kind of variance that you'd expect to observe by just taking the square root. And it's called the standard deviation of  $R$ . If it helps you any, the standard deviation is also called the root mean square error. And you might have heard that phrase.

It comes up all the time in discussions of experimental error. So again, we're taking the error--

means the distance between the random variable and its mean. We're squaring it. We're taking the expectation of that squared error. And then we're taking the square root of it. It's the standard deviation.

So going back to understand what the standard deviation means intuitively in terms of a familiar shaped distribution function for a random variable  $R$ , suppose that  $R$  is a random variable that has this fairly standard kind of bell curved shape or Gaussian shape, that it's got one hump. It's unimodal. And it kind of trails off with some moderate rate, as you get further and further away from the mean.

Well, the mean of a distribution that's shaped like this, it's symmetric around that high point, that's going to be the mean by symmetry. It's equally likely to be-- well, the values average out to this middle value. A standard deviation for a curve like this is going to be an interval that you can interpret as an interval around the mean. And the probability that you're within that interval is fairly high for standard distributions.

Now, we'll see that the Chebyshev bound is not going to tell us much about for arbitrary unknown distribution. But in general, for the typical distributions, you expect to find that the standard deviation tells you that that's where you're most likely to be when you take a random value of the variable.

So let's return to the Chebyshev bound, as we've stated it. And I'm just replacing here, I'm restating the Chebyshev bound, just replacing the variance of  $R$  in the numerator by the square of its square root, by  $\sigma^2$ . It's a useful way to restate it.

Because by restating it this way, it motivates another reformulation of the Chebyshev bound as we reformulated the Markov bound previously in terms of a multiple of something. I'm going to replace  $x$  by a constant times the standard deviation. So I'm going to see the probability that the error is greater than or equal to a constant times the standard deviation.

And this term is going to simplify. Once  $x$  is a constant times the standard deviation, the standard deviations are going to cancel out. And I'm just going to wind up with  $1$  over  $x$  squared. So let's just do that. And there's the formula-- the probability that the distance of  $R$  from its mean is greater than or equal to a multiple  $c$  of its standard deviation is less than or equal to  $1$  over  $c$  squared. So it's getting much more rapidly smaller as  $c$  grows.

Let's look at what that means for just some numbers, to make the thing a little bit more real.

What this assertion is telling us is that  $R$  is probably not going to return a value that's a significant multiple of its standard deviation. For example, what does this formula tell us about the probability that  $R$  is going to be greater than or equal to one standard deviation away from its mean?

Well, it actually tells us nothing. That's the case in which it's no good. Because  $c$  is 1, it's just telling us that the probability is at most 1, which we always know, because probabilities are at most 1.

But if I ask, what's the probability that the error of  $R$  is greater than or equal to twice the standard deviation, then this theorem is telling me something nontrivial. It's telling me that the probability that it's twice the deviation is 1 over 2 squared or  $1/4$ . An arbitrary random variable with standard deviation  $\sigma$  is going to exceed twice-- the error is going to exceed twice the standard deviation at most  $1/4$  of the time, three times at most  $1/9$  of the time, four times at most the  $1/16$  of the time.

So the qualitative message to take away is that, for any random variable whatsoever, as long as it has a standard deviation  $\sigma$ , then you can say some definite things about the probability that the random variable is going to take a value that differs by a large multiple of the standard deviation from its mean. That probability is going to be small and get smaller and rapidly smaller as the multiple of the standard deviation continues.