

**PROFESSOR:** The simplest bound that a random variable differs by much from its expectation is due to a guy named Markov, a Russian probability theorist. And this is Markov's bound that we're going to talk about.

Let's illustrate it with a memorable example of IQ. In the MIT context, it may be a radical idea. But IQ was this thing that was invented for intelligence quotient in the late 19th century, I believe. Might have been early 20th.

It was meant as an effort to break the mold at Harvard of hiring the children of wealthy alumni. And the idea was to have merit-based admissions. And it was going to be some objective measure that did not depend on social class of the ability that people had.

And Harvard was going to admit students based on merit and their intelligence quotient. So the original design of the intelligent quotient by a bunch of psychologists was that the average was supposed to be 100 over the whole population, which, of course, is-- around here, there just aren't very many people with an IQ of just 100.

Anyway, let's ask this extreme question. Yes, around the elite universities, there are a lot of people with IQs much higher than 100. But what fraction of the population could possibly have an IQ as high as 300?

Now, I'm not sure that an IQ of as high as 300 has ever been recorded. But we're talking logically here. Is it possible for a lot of people to have an IQ of greater than or equal to 100?

And the answer is no. You can't possibly have more than 1/3 of the population have an IQ of 300, because if more than 1/3 had an IQ of 300, then that third alone would contribute 1/3 of 300 to the average, which would be greater than 100. So you can't have more than 1/3 of the population have an IQ of triple the expected value of the IQ. So that's the basic bound.

So we can restate it this way. The probability that a randomly chosen person has an IQ greater than or equal to 100 we can say is absolutely less than or equal to the expected value of IQ, namely 100 divided by 300. And just parameterizing it, if we ask, what's the probability that the IQ is greater than or equal to some amount  $x$ , it's less than or equal to  $100/x$  by exactly that reasoning.

And this is basically Markov's bound, except there's one implicit fact that we're using in

deriving the previous identity, or inequality, that IQ is bounded by 100. Our logic was that you can't have more than population  $x$  with an IQ of more than  $100x$ , because that would contribute  $x$  times  $100/x$ , or more than 100 to the average. And the average is only 100.

That's only a problem if there are no negative terms, negative IQs, to offset the excess contribution of the fraction of the population that has this high IQ. But we're implicitly using the fact that IQ is never negative. IQ runs from zero up to unlimited amount. But it's never negative.

And that means that that contribution from the  $1/3$  of the population that has an IQ of over 300 can't be offset by negative values. It's there, and it messes up the average. It forces the average up. So we were using the fact that IQ is always non-negative.

And by this very same reasoning, I'm not going to belabor you with a more formal proof. There's a trivial one in the text. It's easy. The theorem, Markov's bound, says that if  $R$  is non-negative, then the probability that  $R$  is greater than or equal to  $x$  is less than or equal to the expectation of  $R$  divided by  $x$ .

And this holds for any  $x$  greater than 0. Of course, it's silly to state if this bound is greater than or equal to 1. It's not an interesting bound, since probability is never greater than or equal to 1. So we might as well just restrict ourselves to  $x$ 's that are greater than the expectation of  $R$ , because those are the only  $x$ 's that are going to give us a nontrivial bound that's less than 1.

Again, if  $R$  is non-negative, then the probability that  $R$  exceeds an amount  $x$  is less than or equal to the expectation of  $R$  over  $x$ . And that's the Markov bound. If you restate it in terms of deviation from the mean, you could formulate it this way-- the probability that  $R$  is greater than or equal to a constant times its mean--  $\mu$  is an abbreviation for the expectation of  $R$ -- is less than or equal to  $1/c$ .

So now, we can understand that as a bound on the deviation from the mean-- above the mean, in this case-- that  $R$ , as the factor of the expectation increases, the probability decreases proportionally. So the probability that  $R$  is greater equal to 3 times the expected amount is less than or equal to  $1/3$ , which was what we saw with the IQ example.

So look, this Markov bound, in general, is very weak. As I said, I don't think there's ever been an IQ recorded that was as high as 300. And in almost all the examples that you come across, there'll be other information that allows you to deduce tighter bounds on the probability that a

random variable is significantly bigger than its expectation.

But if you don't have any information about the random variable, other than that it's non-negative, then as a matter of fact, Markov bound is tight. You can't possibly reach a stronger contribution, because there are non-negative random variables where the probability that they are greater than or equal to a given amount  $x$  is, in fact, equal to their expectation divided by  $x$ . So the Markov bound is weak in application, but it's the strongest condition you can make on the very limited hypotheses that it makes about properties of the random variable.

And it's also pretty obvious, I hope from this example that we've talked about, but the amazing thing is how useful it is. We will get mileage out of it by using it in clever ways. So let's talk about the first clever way.

And suppose that we're thinking about IQ is greater than or equal to 100. But I bring into the story another fact that we haven't mentioned before, which is, let's suppose that as a matter of fact, IQs of less than 50 just don't occur. I think they might actually, but there's a certain point where you just are not functioning at all. And it's not clear that it makes sense to ever talk about somebody who's in a coma as having an IQ. Maybe they have an IQ of 0.

But let's assume that pragmatically IQ is never less than or equal to 50. Now, if I tell you that I know that IQ is greater than or equal to 50, then I can actually get a better bound out of Markov. Because now, knowing that IQ is greater than or equal to 50, IQ minus 50 becomes a non-negative random variable, which I couldn't be sure it was before, because IQ might have gone below 50.

Now that I know that it's always above 50, IQ minus 50 is non-negative. And Markov's bound will apply to IQ minus 50. And applying it to IQ minus 50 will give you a better bound.

Because now, looking at the probability that the IQ is greater than or equal to 100, of course, that's the same as saying that IQ minus 50 is greater than or equal to 50, which 50 - the average expected value of IQ minus 50 is 100 minus 50. So we're asking whether this is non-negative random variable is greater than or equal to 50. And the answer is, that's less than or equal to its expectation over 50, which is  $1/5$ ,  $50/250$ . And that's a tighter bound than the  $1/3$  we had previously.

This is a general phenomenon that you get and that helps you get slightly stronger bounds out of Markov's bound, namely if you have a non-negative variable, you get a better bound on it

by shifting it so that its mean is 0. As a matter of fact, even if it goes negative, if you shift it up, if you can force it to become above 0 as a minimum, then you can apply Markov's bound to it.